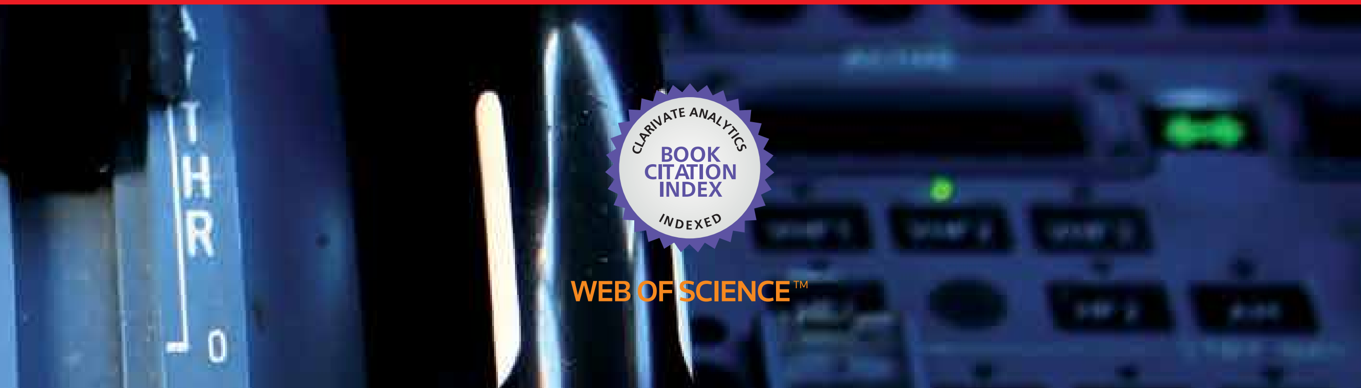
IntechOpen

Recent Advances in Aircraft Technology

Edited by Ramesh K. Agarwal



RECENT ADVANCES IN AIRCRAFT TECHNOLOGY

Edited by **Ramesh K. Agarwal**

Recent Advances in Aircraft Technology
http://dx.doi.org/10.5772/2406
Edited by Ramesh K. Agarwal

Contributors

Josif Boguslavskiy, Andrea LAfflitto, Wassim M. Haddad, Nikolai I Petrov, A Haddad, H Griffiths, Galina Petrova, R.T. Waters, Peter Pong, Subhash Challa, Marco Leo, Luca De Filippis, Giorgio Guglieri, Ahmed Akl, Thierry Henri Gayraud, Pascal Berthou, Juraj Belan, Mohamad Taha, Ahmed AbdElmalek Abdel-Hafez, Curt Kothera, Robert Vocke Iii, Norman Wereley, Edward Bubert, Benjamin K.S. Woods, Matko Orsag, Stjepan Bogdan, Giorgio Cavallini, Roberta Lazzeri, Jozsef Rohacs, Mariusz Wazny, Nicolae Jula, Costin Cepisca, Yu Gu, Francis Barchesky, Haiyang Chao, Marcello Napolitano, Jason Gross, Peter Teunissen, Gabriele Giorgi, Oleksandr Bezvesilniy, Dmytro Vavriv, Melih Cemal Kushan, Sinem Cevik, Yagız Uzunonat, Fehmi Diltemiz, Ramesh K. Agarwal

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Recent Advances in Aircraft Technology
Edited by Ramesh K. Agarwal

p. cm.

ISBN 978-953-51-0150-5

eBook (PDF) ISBN 978-953-51-4345-1

Meet the editor

Professor Ramesh K. Agarwal is the William Palm Professor of Engineering and the director of Aerospace Research and Education Center at Washington University in St. Louis, USA. From 1994 to 2001, he was the Sam Bloomfield Distinguished Professor and Executive Director of the National Institute for Aviation Research at Wichita State University in Kansas. From 1978 to 1994, he worked in various scientific and managerial positions at McDonnell Douglas Research Laboratories in St. Louis; he became the Program Director and McDonnell Douglas Fellow in 1990. Dr. Agarwal received Ph.D in Aeronautical Sciences from Stanford University in 1975, M.S. in Aeronautical Engineering from the University of Minnesota in 1969 and B.S. in Mechanical Engineering from Indian Institute of Technology, Kharagpur, India in 1968. Professor Agarwal has worked in Computational Fluid Dynamics, Rarefied Gas Dynamics and Hypersonic Flows, Flow Control, and more recently in Sustainable Air and Ground Transportation.

Contents

Preface

The book is a compilation of research articles and review articles describing the state of the art and latest advancements in technologies for various areas of aircraft system. The authors contributing to this volume are leading experts in their fields. The book is divided into five sections.

Section one is titled "Aircraft Structure and Advanced Materials". It has five papers, dealing with aircraft structures and advanced materials. In the area of aircraft structures, the topics such as morphing structures and probabilistic approach to fatigue design are covered, while the chapters on advanced materials include the study of advanced materials for jet engines using quantitative metallography, the innovative approaches to gas turbine engine applications and superalloys for aerospace applications.

Section two is titled "Aircraft Control Systems". It contains seven papers dealing with a wide variety of topics. The topics include algorithms for parameter identification of the aircraft dynamics, quadrotor dynamics, graph search algorithms for path planning, GNSS carrier phase-based attitude determination, fuel optimal control problem for UAV formations, measurement and management of uncertainty through data fusion, and subjective factors in flight safety.

Section three is titled "Aircraft Electrical Systems". It has four papers dealing with a wide variety of topics. The topics include a review of power generation system for a more electric aircraft, power electronics for a more electric aircraft, design of an in-flight entertainment system, and methods for reliability analysis of electrical systems of aircrafts.

Section four deals with inspection and maintenance of an aircraft. It has two papers dealing with a number of topics concerning techniques for inspection and maintenance. The first chapter describes the automatic inspection of aircraft components using thermographic and ultrasonic techniques, while the second chapter deals with the analysis of the maintenance process.

The last section of the book contains chapters on miscellaneous topics. One chapter reviews the technologies for sustainable green aviation, another chapter describes the synthetic aperture radar systems for small aircraft, the third chapter describes the avionics design for a fault-tolerant flight control test-bed and the final chapter in this section discusses the lightening strike effects to a radar dome.

Thus the book covers a wide variety of topics related to aircraft technologies in twenty two chapters in a single volume. There is hardly another book that covers such a wide range of topics in a single volume. Therefore, it can serve as a useful source of reference to both researchers and students interested in learning about specific aircraft technologies, as well as obtaining a general overview of the state of the art of many technologies relevant to aircraft systems and their improvement.

Ramesh K. Agarwal
Washington University in St. Louis,
USA

Part 1

Aircraft Structures and Advanced Materials

1

One Dimensional Morphing Structures for Advanced Aircraft

Robert D. Vocke III[1], Curt S. Kothera[2], Benjamin K.S. Woods[1],
Edward A. Bubert[1] and Norman M. Wereley[1]
[1]University of Maryland, College Park, MD
[2]Techno-Sciences, Inc., Beltsville, MD,
USA

1. Introduction

Since the Wright Brothers' first flight, the idea of "morphing" an airplane's characteristics through continuous, rather than discrete, movable aerodynamic surfaces has held the promise of more efficient flight control. While the Wrights used a technique known as wing warping, or twisting the wings to control the roll of the aircraft (Wright and Wright, 1906), any number of possible morphological changes could be undertaken to modify an aircraft's flight path or overall performance. Some notable examples include the Parker Variable Camber Wing used for increased forward speed (Parker, 1920), the impact of a variable dihedral wing on aircraft stability (Munk, 1924), the high speed dash/low speed cruise abilities associated with wings of varying sweep (Buseman, 1935), and the multiple benefits of cruise/dash performance and efficient roll control gained through telescopic wingspan changes (Sarh, 1991; Gevers, 1997; Samuel and Pines, 2007).

While the aforementioned concepts focused on large-scale, manned aircraft, morphing technology is certainly not limited to vehicles of this size. In fact, the development of a new generation of unmanned aerial vehicles (UAVs), combined with advances in actuator and materials technology, has spawned renewed interest in radical morphing configurations capable of matching multiple mission profiles through shape change – this class has come to be referred to as "morphing aircraft" (Barbarino *et al.*, 2011). Gomez and Garcia (2011) presented a comprehensive review of morphing UAVs. Contemporary research is primarily dedicated to various conformal changes, namely, twist, camber, span, and sweep. It has been shown that morphing adjustments in the planform of a wing without hinged surfaces lead to improved roll performance, which can expand the flight envelope of an aircraft (Gern *et al.*, 2002), and more specifically, morphing to increase the span of a wing results in a reduction in induced drag, allowing for increased range or endurance (Bae *et al.*, 2005). The work presented here is intended for just such a one dimensional (1-D) span-morphing application, for example a UAV with span-morphing wingtips depicted in Figure 1. By achieving large deformations in the span dimension over a small section of wing, the wingspan can be altered during flight to optimize aspect ratio for different roles. Furthermore, differential span change between wingtips can generate a roll moment, replacing the use of ailerons on the aircraft (Hetrick *et al.*, 2007). This one dimensional morphing could also be used in the chordwise direction, and is not limited in application to fixed-wing aircraft, as rotorcraft would also benefit from a variable diameter or chord rotor.

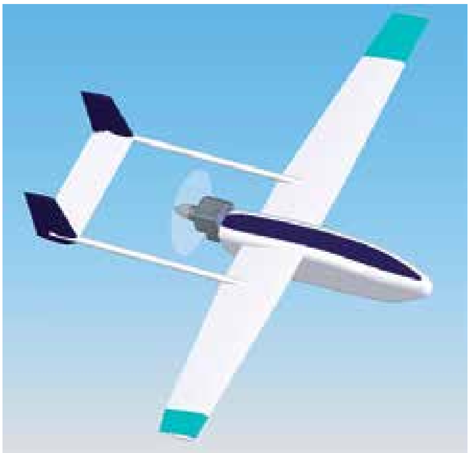


Fig. 1. Illustration of span-morphing UAV showing 1-D morphing wingtips.

A key challenge in developing a one dimensional morphing structure is the development of a useful morphing skin, defined here as a continuous layer of material that would stretch over the morphing structure and mechanism to form a smooth aerodynamic skin surface. For a span-morphing wingtip in particular, the necessity of a high degree of surface area change, large strain capability in the span direction, and little to no strain in the chordwise direction all impose difficult requirements on any proposed morphing skin. The goal of this effort was a 100% increase in both the span and area of a morphing wingtip, or "morphing cell."

Reviews of contemporary morphing skin technology (Thill *et al.*, 2008; Wereley and Gandhi, 2010) yield three major areas of research being pursued: compliant structures, shape memory polymers, and anisotropic elastomeric skins. Compliant structures, such as the FlexSys Inc. Mission Adaptive Compliant Wing (MACW), rely on a highly tailored internal structure and a conventional skin material to allow small amounts of trailing edge camber change (Perkins *et al.*, 2004). Due to the large geometrical changes required for a span-morphing wingtip as envisioned here, metal or resin-matrix-composite skin materials are unsuitable because they are simply unable to achieve the desired goal of 100% increases in morphing cell span and area.

Shape memory polymer (SMP) skin materials are relatively new and have recently received attention for morphing aircraft concepts. They may at first glance seem highly suited to a span-morphing wingtip: shape memory polymers made by Cornerstone Research Group exhibit an order of magnitude change in modulus and up to 200% strain capability when heated past a transition temperature, yet return to their original modulus upon cooling. There have been attempts to capitalize on the capabilities of SMP skins, such as Lockheed Martin's Z-wing morphing UAV concept (Bye and McClure, 2007) and a reconfigurable segmented variable stiffness skin composed of rigid disks and shape memory polymer proposed by McKnight *et al.* (2010). However, electrical heating of the SMP skin to reach transition temperature proved difficult to implement in the wind tunnel test article and the

SMP skin was abandoned as a high-risk option. Additionally, the state-of-the-art of SMP technology does not appear to be well-suited for dynamic control morphing objectives.

With maximum strains above 100%, low stiffness, and a lower degree of risk due to their passive operation, elastomeric materials are ideal candidates for a morphing skin. Isotropic elastomer morphing skins have been successfully implemented on the MFX-1 UAV (Flanagan *et al.*, 2007). This UAV employed a mechanized sliding spar wing structure capable of altering the sweep, wing area, and aspect ratio during flight. Sheets of silicone elastomer connect rigid leading and trailing edge spars, forming the upper and lower surfaces of the wing. The elastomer skin is reinforced against out-of-plane loads by ribbons stretched taught immediately underneath the skin, which proved effective for wind tunnel testing and flight testing. Morphing sandwich structures capable of high global strains have also been investigated (Joo *et al.*, 2009; Bubert *et. al.*, 2010; Olympio *et al.*, 2010). However, suitable improvements over these structures, such as anisotropic fiber reinforcement and a better developed substructure for out-of-plane reinforcement, are desired for a fully functional morphing skin.

The present research therefore focuses on the development of a passive anisotropic elastomer composite skin with potential for use in a 1-D span-morphing UAV wingtip. The skin should be capable of sustaining 100% active strain with negligible major axis Poisson's ratio effects, giving a 100% change in surface area, and should also be able to withstand typical aerodynamic loads, assumed to range up to 200 psf (9.58 kPa) for a maneuvering flight surface, with minimal out-of-plane deflection. The following will describe the process of designing, building, and testing a morphing skin with these goals in mind, and will compare the performance of the final article to the initial design objectives.

2. Conceptual development

The primary challenge in developing a morphing skin suitable as an aerodynamic surface is balancing the competing goals of low in-plane actuation requirements and high out-of-plane stiffness. In order to make the skin viable, actuation requirements must be low enough that a reasonable actuation system within the aircraft can stretch the skin to the desired shape and hold it for the required morphing duration. At the same time, the skin must withstand typical aerodynamic loads without deforming excessively (e.g., rippling or bowing), which would result in degradation to the aerodynamic characteristics of the airfoil surface.

To achieve these design goals, a soft, thin silicone elastomer sheet with highly anisotropic carbon fiber reinforcement, called an elastomeric matrix composite (EMC), would be oriented such that the fiber-dominated direction runs chordwise at the wingtip, and the matrix-dominated direction runs spanwise (Figure 2a). Reinforcing carbon fibers controlling the major axis Poisson's ratio of the sheet would limit the EMC to 1-D spanwise shape change (Figure 2b). For a given skin stiffness, actuation requirements will increase in proportion to the skin thickness, t_s, while out-of-plane stiffness will be proportional to t_s^3 by the second moment of the area. To alleviate these competing factors, a flexible substructure is desired (Figure 2c) that would be capable of handling out-of-plane loads without greatly adding to the in-plane stiffness. This allows a thinner skin which, in turn, reduces actuation requirements. The combined EMC sheet and substructure form a continuous span-morphing skin.

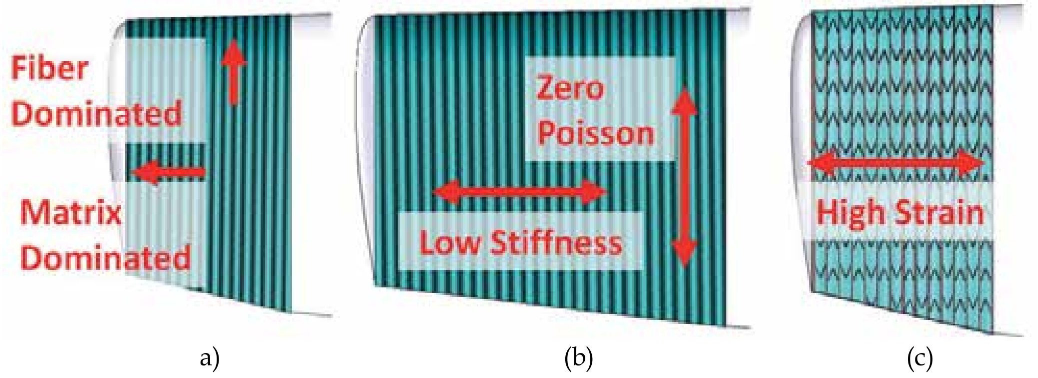


Fig. 2. Design concept as a span morphing wingtip. (Bubert *et al.*, 2010)

To motivate the goal of low in-plane stiffness for this research, the skin prototype was designed to be actuated by a span-morphing pneumatic artificial muscle (PAM) scissor mechanism described separately by Wereley and Kothera (2007). The PAM scissor mechanism shown in Figure 3 was designed to transform contraction of the PAM actuator into extensile force necessary in a span-morphing wing. Based upon the maximum performance of the PAM and the kinematics of the scissor frame, the maximum force output of the actuation system was predicted and a skin stiffness goal was determined such that 100% active strain could be achieved, with the skin simplified as having linear stiffness. A margin of 15% was added to the 100% strain goal to account for anticipated losses due to friction or manufacturing shortcomings in the skin or actuation system.

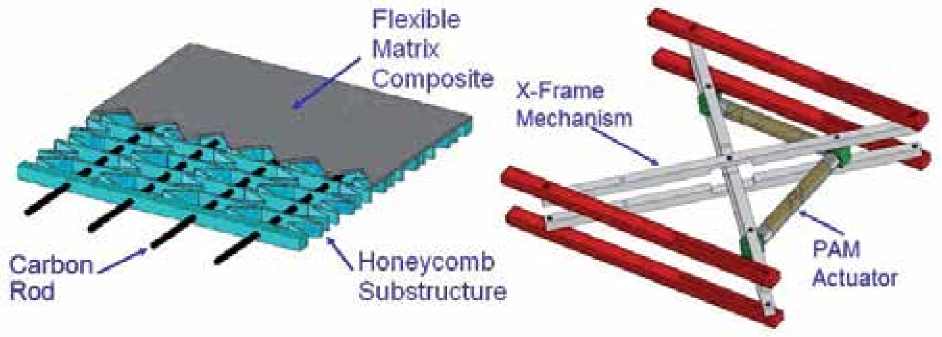


Fig. 3. Morphing skin demonstrator including PAM actuation system.

In addition, minimal out-of-plane deflection of the skin surface under aerodynamic loading was desired. No specific out-of-plane deflection goal was set or designed for, but out-of-plane stiffness of the substructure was kept in mind during the design process. Deflection due to distributed loads was included as a final test to ensure that the aerodynamic shape of a UAV wing morphing structure could be maintained during flight.

3. Skin development

The primary phase of the morphing skin development was to fabricate the EMC sheet that would make up the skin or face sheet. A number of design variables were available for tailoring the EMC to the application, including elastomer stiffness, durometer, ease of handling during manufacturing, and the quantity, thickness, and angle of carbon fiber reinforcement.

3.1 Elastomer selection

Initially, a large number of silicone elastomers were tested for viability as matrix material. Desired properties included maximum elongation well over 100%, a low stiffness to minimize actuation forces, moderate durometer to avoid having too soft a skin surface, and good working properties. Workability became a primary challenge to overcome, as two-part elastomers with high viscosities or very short work times would not fully wet out the carbon fiber layers. While over a dozen candidate elastomer samples were examined, only four were selected for further testing. Table 1 details the silicone elastomers tested as matrix candidates.

Material	Modulus (kPa)	Viscosity (cP)	% Elongation at Break	Comments
DC 3-4207	130	430	100+	difficult to demold
Sylgard-186	410	65,000	100+	too viscous
V-330, CA-45	570	10,000	500	excellent workability
V-330, CA-35	330	10,000	510	excellent workability

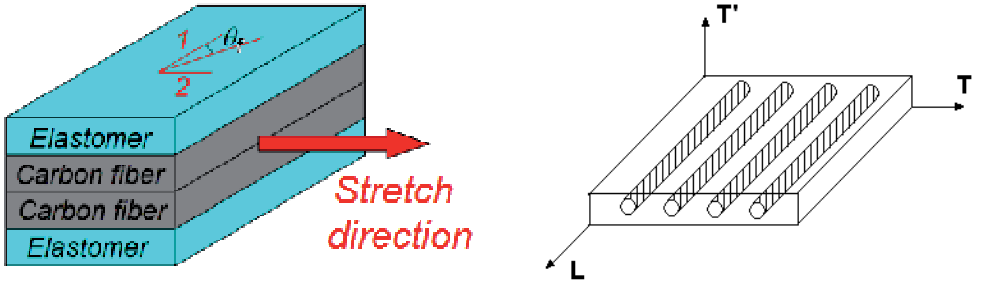
Table 1. Elastomer properties.

The most promising compositions tested were Dow Corning 3-4207 series and the Rhodorsil V-330 series. Both exhibited the desired low stiffness and greater than 100% elongation, but DC 3-4207 suffered from poor working qualities and lower maximum elongation and was not down-selected. Rhodorsil's V-330 series two-part room temperature vulcanization (RTV) silicone elastomer had the desired combination of low viscosity, long working time, and easy demolding to enable effective EMC manufacture, and also demonstrated very high maximum elongation and tear strength. V-330 with CA-35 had the lowest stiffness of the two V-330 elastomers tested. This led to selecting V-330, CA-35 for use in test article fabrication.

3.2 CLPT predictions and validation

Concurrently, using classical laminated plate theory (CLPT), a simple model of the EMC laminate was developed to study the effects of changing composite configuration on performance. The skin lay-up shown in Figure 4a was examined: two silicone elastomer face sheets sandwiching two symmetric unidirectional carbon fiber/elastomer composite laminae. The unidirectional fiber layers are offset by an angle θ_f from the **1**-axis, which corresponds to the chordwise direction. Orienting the fiber-dominated direction along the wing chord controls minor Poisson's ratio effects while retaining low stiffness and high strain capability in the **2**-axis, which corresponds to the spanwise direction.

In order to determine directional properties of the EMC laminate, directional properties of each lamina must first be found. The following micromechanics derivation comes from Agarwal *et al*. (2006). For a unidirectional sheet with the material longitudinal (**L**) and transverse (**T**) axes oriented along the fiber direction as shown in Figure 4b, we assume that



(a) (b)

Fig. 4. (a) EMC lay-up used in CLPT predictions. (b) Unidirectional composite layer showing fiber orientation.

perfect bonding occurs between the fiber and matrix material such that equal strain is experienced by both fiber and matrix in the **L** direction. Based upon these assumptions, the longitudinal elastic modulus is given by the rule of mixtures:

$$E_L = E_f V_f + E_m (1 - V_f) \qquad (1)$$

Here E_L is the longitudinal elastic modulus for the layer, E_f is the fiber elastic modulus, E_m is the matrix elastic modulus, and V_f is the fiber volume fraction. To find the elastic modulus in the transverse direction, it is assumed that stress is uniform through the matrix and fiber. The equation for the transverse modulus, E_T, is:

$$E_T = 1 / (V_f / E_f + (1 - V_f) / E_m) \qquad (2)$$

Calculations based on these micromechanics assumptions supported the intuitive conclusion that thinner EMC skins would have a lower in-plane stiffness modulus in the spanwise direction, E_2. Predictions for the transverse elastic modulus and the minor Poisson's ratio are plotted versus fiber offset angle in Figure 5a and Figure 5b, respectively, as solid lines. In order to provide some validation for the CLPT predictions, three EMC sample coupons were manufactured, consisting of 0.5 mm elastomer face sheets sandwiching two 0.2-0.3 mm composite lamina with a fiber volume fraction of 0.7. Nominal fiber axis offset angles of 0°, 10°, and 20° were used. The measured transverse modulus and minor Poisson's ratio are plotted as circles in their respective figure. As expected, increasing fiber offset angle increases the in-plane stiffness of the EMC, requiring greater actuation forces. Also, it is noteworthy that the inclusion of unidirectional fiber reinforcement at 0° offset angle nearly eliminates minor Poisson's ratio effects as predicted by CLPT theory.

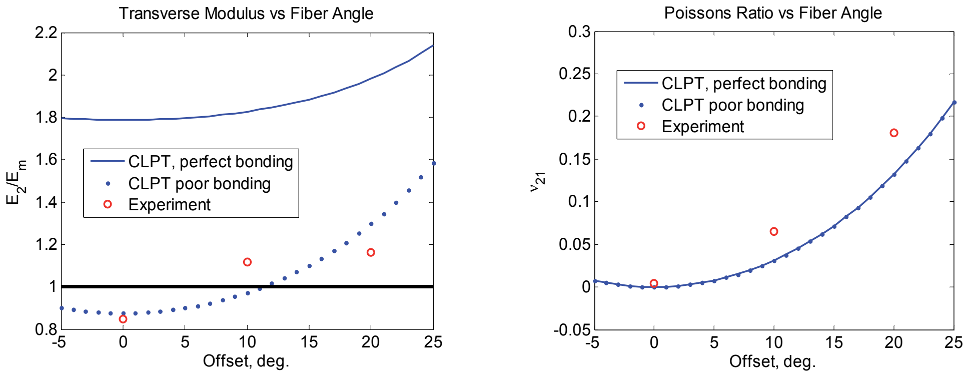
It is of critical importance to note that, according to the assumptions used in deriving the lamina transverse modulus in Eq. (2), the transverse modulus has a lower bound equal to the matrix modulus. This lower bound is shown in Figure 5a as a horizontal black line at $E_2/E_m = 1$. However, the experimental data is close to this lower bound for the 10° and 20° samples, and the modulus is actually below the lower bound for the 0° case. Clearly in this case there is a problem in the micromechanics from which the transverse modulus prediction was derived.

Recall it was assumed that perfect bonding between fiber and matrix occurred, as illustrated in Figure 6a. This implies stress was equally shared between matrix and fiber under transverse loading. Close visual examination of the EMC samples during testing revealed that the fiber/matrix bond was actually very poor, and the matrix pulled away from individual fibers under transverse loading as illustrated in Figure 6b. Thus, the fibers carry no stress in the transverse direction, and the effective cross-sectional area of matrix left to carry transverse force in the lamina is reduced by the fiber volume fraction. For the case of poor transverse bonding exhibited in the fiber laminae, the transverse modulus in Eq. (2) can thus be simplified to:

$$E_{\mathrm{T}} = E_m / (1 - V_f) \tag{3}$$

Using Eq. (3) to calculate transverse modulus for the fiber laminae, new CLPT predictions for EMC non-dimensionalized transverse modulus and minor Poisson's ratio are also plotted in Figure 5a and 5b, respectively. Much better agreement is seen between the analytical and experimental values for E_2/E_{m}. In spite of the poor bond between fiber and matrix material in the EMCs, the fiber stiffness still appears to contribute to the transverse stiffness at higher fiber offset angles. The minor Poisson's ratio is also influenced by the fiber offset angle. The EMC's longitudinal modulus, not shown, also remains high. These findings clearly indicate the fiber continues to contribute to the longitudinal stiffness of the fiber laminae even when bonding between matrix and fiber is poor.

To explain this contribution, it is hypothesized that friction between fiber and matrix help share load between the two materials in the longitudinal direction, while the matrix is free to pull away from the fiber in the transverse direction. This would explain the stiffening effect seen in the transverse modulus at increased offset angles and the controlling effect the fiber appears to have on Poisson's ratio at very low offset angles.



(a) (b)

Fig. 5. Comparison of CLPT predictions with experimental data for three different fiber angles (a) non-dimensionalized transverse elastic modulus E_2/E_m, (b) minor Poisson's ratio ν_{21}.

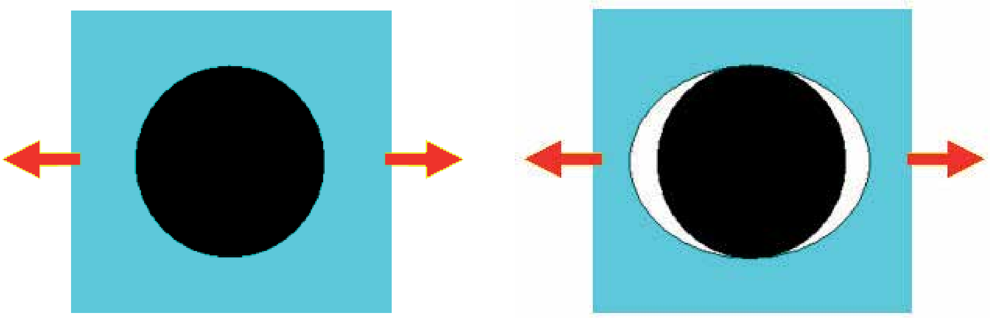


Fig. 6. Fiber/matrix bond (a) assumed perfect bonding and equal transverse stress sharing in CLPT, (b) actual condition with poor fiber/matrix bond and no fiber stress under transverse loading.

Based upon these CLPT results, a fiber offset angle of 0° was selected to minimize transverse stiffness and also to minimize the minor Poisson's ratio. As the analytical and experimental results in Figure 5b indicate, a 0° fiber offset angle can resist chordwise shape change during spanwise morphing. While this conclusion appears obvious, the results demonstrate that with the appropriate correction to micromechanics assumptions in the transverse direction, simple CLPT analysis can be more confidently used to predict EMC directional properties. This simplifies the morphing skin design procedure by allowing in-plane EMC stiffness to be predicted by analytical methods.

3.3 EMC fabrication and testing

A key issue in this study was developing a dependable and repeatable skin manufacturing process. The final manufacturing process involved a multi-step lay-up process, building the skin up through its thickness (Figure 7). First, a sheet of elastomer was cast between two aluminum caul plates using shim stock to enforce the desired thickness. Secondly, unidirectional carbon fiber was applied to the cured elastomer sheet, with particular attention paid to the alignment of the fibers to ensure that they maintained their uniform spacing and unidirectional orientation (or angular displacement, depending on the sample). Enough additional liquid elastomer was then spread on top of the carbon to wet out all of the fibers. An aluminum caul plate was placed on top of the lay-up, compressing the carbon/elastomer layer while the elastomer cured. The third and final step in the skin lay-up process was to build the skin up to its final thickness. The bottom sheet of skin with attached carbon fiber was laid out on a caul plate. As in the first step, shim stock was used to enforce the desired thickness (now the full thickness of the skin) and liquid elastomer was poured over the existing sheet. A caul plate was then placed on top of this uncured elastomer and left for at least 4 hours. Once cured, the completed skin was removed from the plates, trimmed of excess material, and inspected for flaws. A successfully manufactured skin had a consistent cross-section and no air bubbles or visible flaws.

Several EMC sheets were originally manufactured in an effort to experimentally test the effect of fiber thickness and orientation on in-plane and out-of-plane characteristics and to attempt to optimize both. Table 2 describes the nominal dimensions and fiber angle values

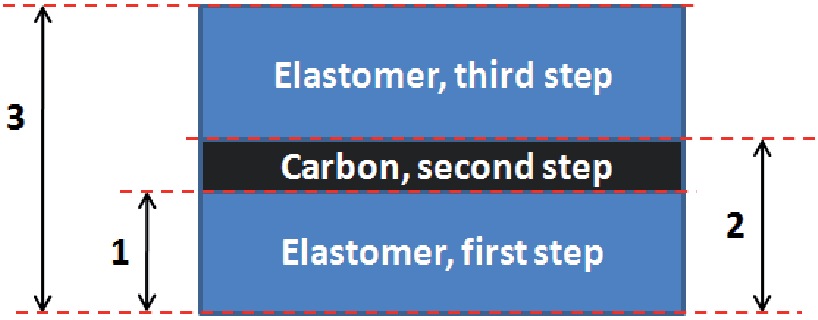
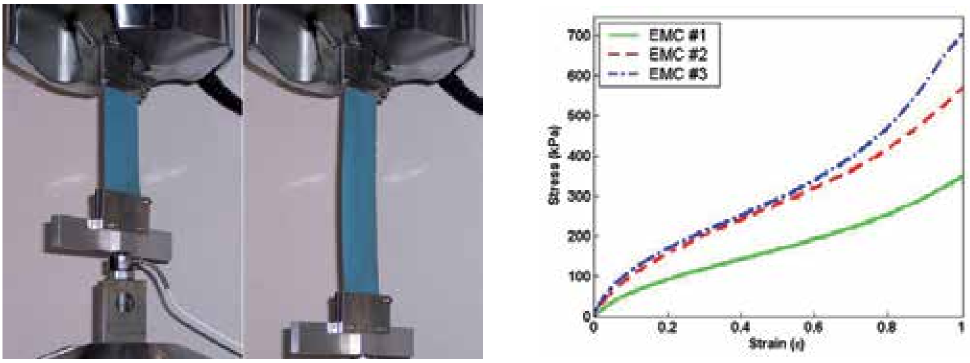


Fig. 7. Progression of skin manufacturing process.

for the three EMC samples. EMC #3 was not intended to be used in the final morphing skin demonstrator, but instead was an academic exercise intended to increase out-of-plane stiffness at the expense of in-plane stiffness.

	Sheet thickness (mm)	Fiber orientation (deg)	Fiber layer thickness (mm)	Total Thickness (mm)
EMC #1	0.5	0	0.4	1.4
EMC #2	0.5	0	0.7	1.7
EMC #3	0.5	+15, 0, -15 deg	0.8	1.8

Table 2. Summary of EMC sample properties.



(a) (b)

Fig. 8. In-plane skin testing, (a) EMC sample taken to 100% strain; (b) data from EMC samples.

Sample strips measuring 51 mm x 152 mm were cut from the three EMCs and tested on a Material Test System (MTS) machine. Each sample was strained to 100% of its original length and then returned to its resting position. The test setup is depicted in Figure 8a and data from these tests are presented in Figure 8b. Notice the visibly low Poisson's ratio effects as the EMC is stretched to 100% strain in Figure 8a – there is little measurable reduction in width. It is also important to note that the stress-strain curves measured for each EMC reflect not only the impact of their lay-ups on stiffness, but also improvements in manufacturing ability. Thus, due to improved control of carbon fiber angles and the thickness of elastomer matrix, EMC #3 has roughly the same stiffness as EMC #2, in spite of the larger amount of carbon fiber present and higher fiber angles. EMC #1 exhibited high quality control and linearity of fiber arrangement and has the lowest stiffness of all, regardless of its nominal similarity to EMC #2. Based upon these tests, EMC #1 and EMC #2 were selected for incorporation into integrated test articles. EMC #1 displayed the lowest in-plane stiffness, while EMC #2 had the second lowest stiffness, making them the most attractive candidates for a useful morphing skin.

4. Substructure development

The most challenging aspect of the morphing skin to design was the substructure. Structural requirements necessitated high out-of-plane stiffness to help support the aerodynamic pressure load while still maintaining low in-plane stiffness and high strain capability.

4.1 Honeycomb design

The substructure concept originally evolved from the use of honeycomb core reinforcement in composite structures such as rotor blades. Honeycomb structures are naturally suited for high out-of-plane stiffness, and if properly designed can have tailored in-plane stiffness as well (Gibson and Ashby, 1988). By modification of the arrangement of a cellular structure, the desired shape change properties can be incorporated.

In order to create a honeycomb structure with a Poisson's ratio of zero, a negative Poisson's ratio cellular design presented by Chavez *et al*. (2003), or so-called auxetic structure (Evans *et al*., 1991), was rearranged to resemble a series of v-shaped members connecting parallel rib-like members, as seen in Figure 9. This arrangement gives large strains in one direction with no deflection at all in the other by means of extending or compressing the v-shaped members, which essentially act as spring elements. The chordwise rib members act as ribs in a conventional airplane wing by defining the shape of the EMC face sheet and supporting against out-of-plane loads. The v-shaped members connect the ribs into a single deformable substructure which can then be bonded to the EMC face sheet as a unit, with the v-shaped bending members controlling the rib spacing.

For a standard honeycomb, Gibson and Ashby (1988) describe the in-plane stiffness as a ratio of in-plane modulus to material modulus, given in terms of the geometric properties of the honeycomb cells. By modifying this standard equation, it is possible to describe the in-plane stiffness of a zero-Poisson honeycomb structure with cell geometric properties as illustrated in Figure 10a. Here t is the thickness of the bending (v-shaped) members, ℓ is the length of the bending members, h is the cell height, c is the cell width, and θ is the angle between the rib members and the bending members. Note that in the figure the cell is being stretched vertically and F is the force carried by a bending member under tension. Also note that the depth of the cell, denoted as b, is not represented in Figure 10.

With the geometry of the cell defined, an expression can be found for the honeycomb's equivalent of a stress-strain relationship. For small deflections, the bending member between points **1** and **2** can be considered an Euler-Bernoulli beam as shown in Figure 10b, with the forces causing a second mode deflection similar to a pure moment. From Euler-

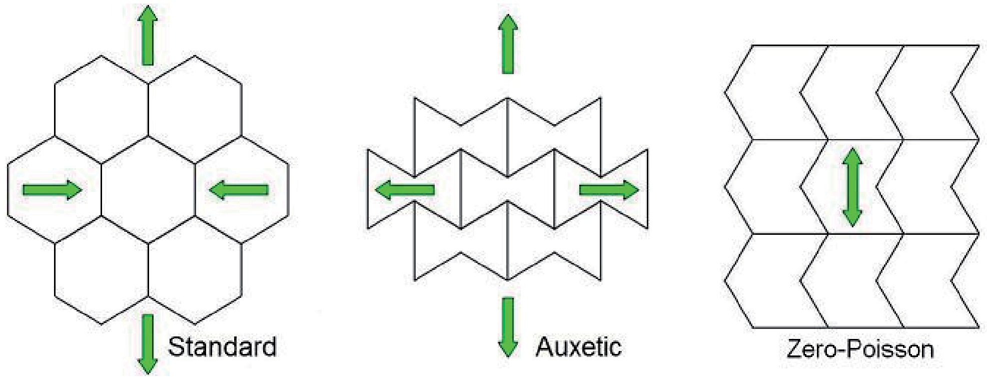
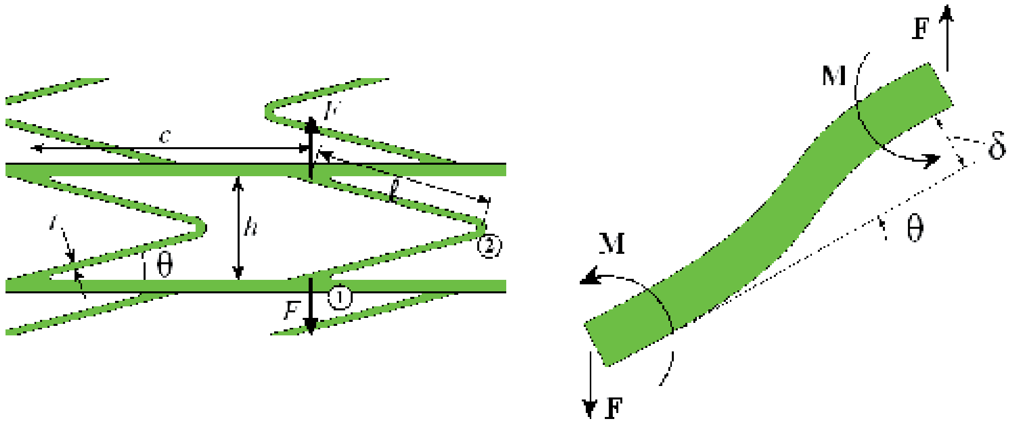


Fig. 9. Comparison of standard, auxetic, and modified zero-Poisson cellular structures showing strain relationships.



(a) (b)

Fig. 10. (a) Geometry of zero-Poisson honeycomb cell, (b) Forces and moments on bending member leg.

Bernoulli theory, the cosine component of the force F will cause a bending deflection δ (Shigley *et al.*, 2004):

$$\delta = \frac{F\cos\theta\ell^3}{12E_0 I} \tag{4}$$

Here E_0 is the Young's Modulus of the honeycomb material and I is the second moment of the area of the bending member; in this case $I = bt^3/12$. In order to determine an effective tensile modulus for the honeycomb substructure, the relationship in Eq. (4) between force and displacement needs to be transformed into an equivalent stress-strain relationship. The equivalent stress through one cell can be found by using the cell width c and honeycomb depth b to establish a reference area, and the global equivalent strain is determined by non-dimensionalizing the v-shaped member's bending deflection 2δ by the cell height h. These equivalent stresses and strains are used to determine a transverse stiffness modulus for the honeycomb, E_2:

$$\sigma_2 = \frac{F}{cb}, \tag{5}$$

$$\varepsilon_2 = \frac{\delta \cos\theta}{h/2}, \tag{6}$$

$$E_2 = \frac{\sigma_2}{\varepsilon_2} \tag{7}$$

Substituting Eqns. (5) through (7) into Eq. (4) and simplifying yields the following expression for the stiffness of the overall honeycomb relative to the material modulus:

$$\frac{E_2}{E_0} = \left(\frac{t}{l}\right)^3 \frac{\sin\theta}{\frac{c}{l}\cos^2\theta} \tag{8}$$

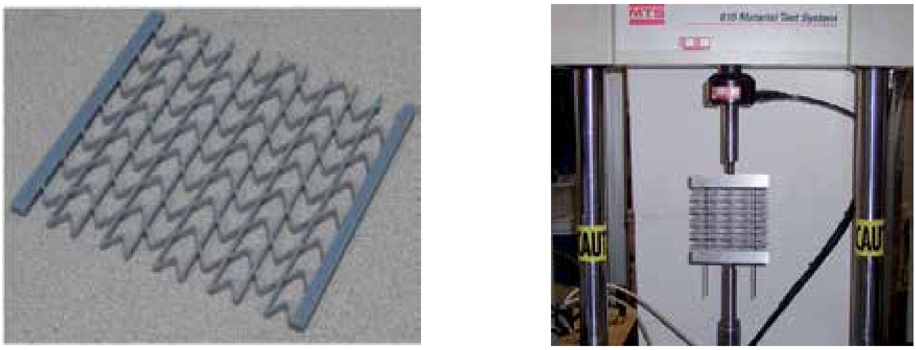
Because this modified Gibson-Ashby model assumes the bending member legs to be beams with low deflection angles and low local strains, Eq. (8) should only be valid for global strains that result in small local deflections. However, it will be shown that due to the nature of the honeycomb design, relatively large global strains are achievable with only small local strains.

With this fairly simple equation, the cell design parameters can easily be varied and their effect on the overall in-plane stiffness of the structure can be studied. For fixed values of t, h, c, and b, the modulus ratio of the structure, E_2/E_0, increases with the angle θ. Noting the definitions in Figure 10a, it can be seen that decreasing θ consequently affects the bending member length l, as the upper and lower ends must meet to form a viable structure. Thus, for a given cell height h, minimum stiffness limitations are introduced into the design from a practicality standpoint in that the bending members must connect to the structure and cannot intersect one another. Lower in-plane stiffness can be achieved by increasing cell width to accommodate lower bending member angles.

In Figure 11a, an example is given of a zero-Poisson substructure designed in a commercial CAD software and produced on a rapid prototyping machine out of a photocure polymer. Using this method, a large number of samples could be fabricated with variations in bending member angle, θ. By testing these structures on an MTS machine (Figure 11b), a comparison could be made between the predicted effect of bending member angle on in-plane stiffness and the actual observed effect.

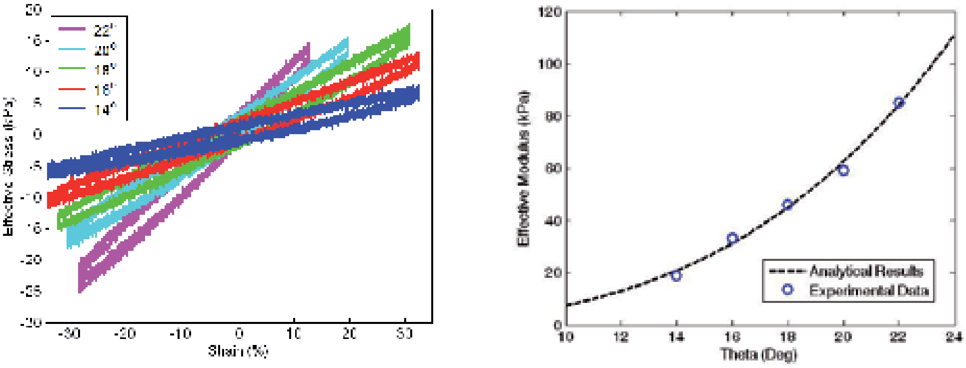
The stress-strain test data from a series of rapid prototyped honeycombs is presented in Figure 12a. Each honeycomb was tested over the intended operating range, starting at a reference length of 67% of resting length (pre-compressed) and extending to 133% of resting length to achieve 100% total length change. To test the validity of the modified Gibson-Ashby model, comparisons of experimental data and analytical predictions were made. The stiffness modulus of each experimentally tested honeycomb was determined by applying a linear least squares regression to the data in Figure 12a. The resulting stiffnesses were then plotted with the analytical predictions from Eq. (8) in Figure 12b.

The strong correlation between the analytical predictions and measured behaviour suggests the assumptions made in the modified Gibson-Ashby equation are accurate over the intended operating range of the honeycomb substructure, and local strains are indeed relatively low. Having low local strain is a benefit as it will increase the fatigue life of the substructure. The low local strains were verified with a finite element analysis that predicted a maximum local strain of 1.5% while undergoing 30% compression globally, a 20:1 ratio. This offers hope that a honeycomb substructure capable of high global strains with a long fatigue life can be designed by minimizing local strain, an area which should be a topic of further research. Further details regarding this structure can be found by consulting Kothera *et al.* (2011).



(a) (b)

Fig. 11. (a) Example of Objet PolyJet rapid-prototyped zero-Poisson honeycomb, (b) morphing substructure on MTS machine.



(a) (b)

Fig. 12. (a) Stress-strain curves of substructures of various interior angles, (b) In-plane substructure stiffness, analytical versus experiment.

To minimize the in-plane stiffness of the substructure, the lowest manufacturable bending member angle, 14°, was selected for integration into complete morphing skin prototypes.

Furthermore, this testing demonstrated the usefulness of the modified Gibson-Ashby equation for future honeycomb substructure design efforts. The in-plane stiffness of zero-Poisson honeycomb structures can be predicted.

5. One dimensional morphing demonstrator

5.1 Carbon fiber stringers

One unfortunate aspect of the zero-Poisson honeycomb described above is the lack of bending stiffness about the in-plane axis perpendicular to the rib members. Another structural element is needed to reinforce the substructure for out-of-plane loads. In order to reinforce the substructure, carbon fiber "stringers" were added perpendicular to the rib members. Simply comprised of carbon fiber rods sliding into holes in the substructure, the stringers reinforce the honeycomb against bending about the transverse axis.

The impact of the stringers on the in-plane stiffness of the combined skin was imperceptible. Fit of the stringer through the holes in the substructure was loose and thus the assembly had low friction. Additionally, the EMC sheet and bending members of the substructure kept the substructure ribs stable and vertical, preventing any binding while sliding along the stringers.

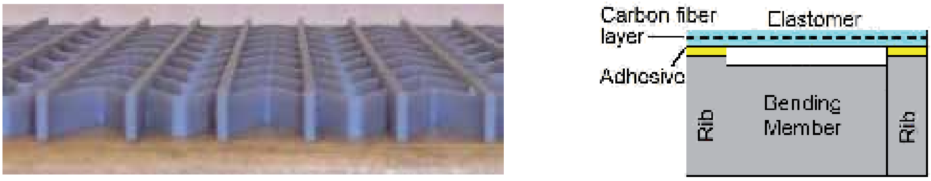
5.2 EMC/substructure adhesive

In order to integrate the EMC face sheets with the honeycomb substructure and carry in-plane loads, a suitable bonding agent was necessary. The desired adhesive was required to bond the silicone EMC to the plastic rapid-prototyped honeycomb sufficiently to withstand the shear forces generated while deforming the structure. In addition, the adhesive also needed to be capable of high strain levels in order to match the local strain of the EMC at the bond site. Loads imposed on the adhesive by distributed loads (such as aerodynamic loads on the upper surface of a wing) were not taken into account in this preliminary study.

Due to the fact that the substructure, and not the EMC itself, would be attached to the actuation mechanism, the adhesive was required to transfer all the force necessary to strain the EMC sheet. Based upon the known stiffness of the EMCs selected for integration into the morphing skin prototype, the adhesive was required to withstand up to 10.5 N/cm of skin width for 100% area change. The adhesive was to bond the EMC along a strip of plastic 2.54 cm deep, so the equivalent shear strength required was 41.4 kPa. A couple silicone-based candidate adhesives were selected for lap shear evaluation, all of which were capable of high levels of strain. Test results indicated that Dow Corning (DC) 700, Industrial Grade Silicone Sealant, a one-part silicone rubber that is resistant to weathering and withstands temperature extremes, was most capable of bonding the EMC skin to the substructure, as it had a safety factor of 2.

5.3 Morphing structure assembly

A 152 mm x 152 mm morphing skin sample was fabricated from EMC #1. A 14° angle honeycomb was used for the substructure, and DC 700 adhesive was used to bond the EMC to the honeycomb substructure. To assist in the attachment, the rib members of the honeycomb core were designed with raised edges on one side, as shown in Figure 13a. This figure shows a side view of the zero-Poisson honeycomb, where it can be seen that the top surface has the ribs extended taller than the bending members. Therefore, the bonding layer can be applied to the raised rib surfaces and pressed onto the EMC without bonding the bending members to the EMC. A sectional side view of a single honeycomb cell, shown in Figure 13b, illustrates conceptually how the bonded morphing skin looks. A thin layer of adhesive is shown between the EMC and the ribs of the honeycomb, but it does not affect the movement of the bending members. The outermost two ribs on the substructure were each 26 mm wide, providing large bonding areas to carry the load of the skin under strain. This left 100 mm of active length capable of undergoing high strain deformation.



(a) (b)

Fig. 13. EMC-structure bonding method – (a) honeycomb core; (b) single cell diagram.

The configuration of the morphing skin design is summarized in Table 3. The assembled morphing skin sample was used to assess in-plane and out-of-plane stiffness before fabricating a final 165 mm x 330 mm full scale test article for combination and evaluation with the PAM actuation system described in Section 2.

EMC	**Honeycomb**	**Adhesive**	**Active Length**
EMC #1, 1.4 mm thick, two CF layers at 0^{o}	14^{o} zero-Poisson rapid prototyped VeroBlue	DC-700	100 mm

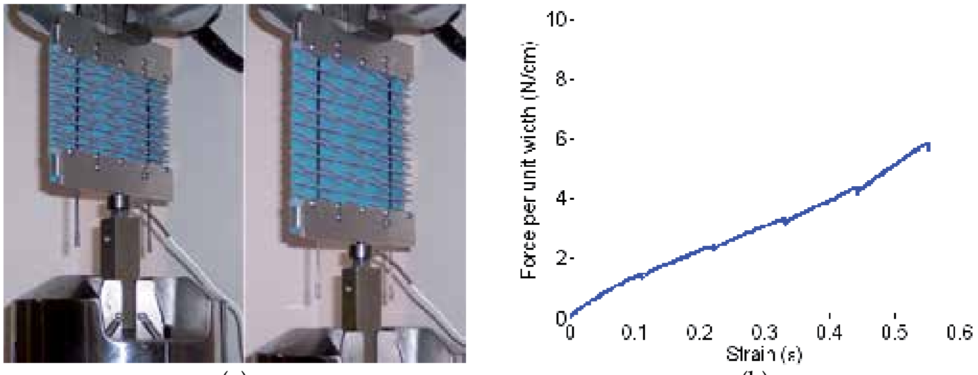
Table 3. Morphing skin configuration.

5.4 In-plane testing

The morphing skin sample was tested on an MTS machine to 50% strain. The level of strain was limited in order to prevent unforeseen damage to the morphing skin before it could be tested for out-of-plane stiffness as well. In Figure 14a, the morphing skin is shown undergoing in-plane testing, with results presented in Figure 14b. Note that the test procedure strained the specimen incrementally to measure quasi-static stiffness, holding the position briefly before starting with the next stage. Relaxation of the EMC sheet is the cause for the dips in force seen in the figure.

Based upon the individually measured stiffnesses of the EMC and substructure components used in the morphing skin and the stiffness of the skin overall, the energy required to strain each structural element can be determined (Figure 15), with the adhesive strain energy found by subtracting the strain energy of the other two components from the total for the morphing skin. The strain energy contribution of each element is broken down in energy per unit width required to strain the sample from 10 cm to 20 cm.

It can be seen that the adhesive had a considerable strain energy requirement, more than double that of the honeycomb substructure. When designing future morphing skins, the energy to strain the adhesive layer must be taken into account to ensure sufficient actuation force is available to meet strain requirements. More careful attention to minimizing the amount of adhesive used to bond the skin and substructure would also likely reduce the in-plane stiffness of the morphing skin by a non-trivial amount.



(a) (b)

Fig. 14. Morphing skin sample in-plane testing – (a) Skin #1 on MTS; (b) Data from morphing skin in-plane testing.

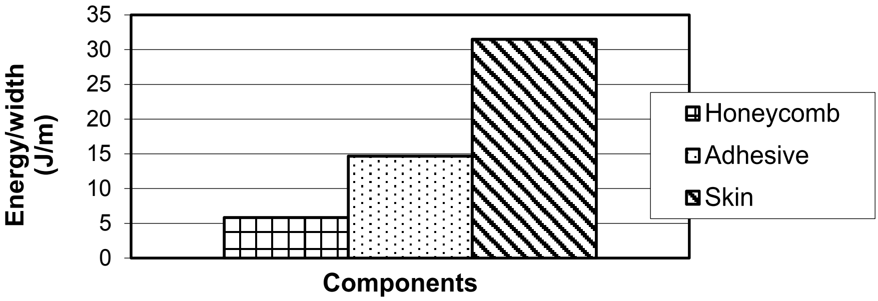
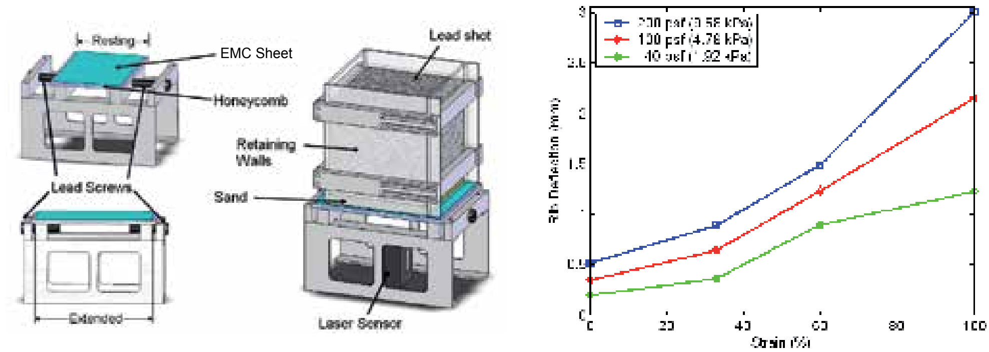


Fig. 15. Contributions to morphing skin strain energy.

5.5 Out-of-plane testing

The final phase of evaluation for morphing skin sample required measuring out-of-plane deflection under distributed loadings, approximating aerodynamic forces. A number of testing protocols were investigated, including ASTM standard D 6416/D 6416M for testing simply supported composite plates subject to a distributed load. This particular test protocol is intended for very stiff composites, not flexible or membrane-like composites. A simpler approach to the problem was adopted wherein acrylic retaining walls were placed above the morphing skin sample into which a distributed load of lead shot and sand could be poured. The final configuration of the out-of-plane deflection testing apparatus can be seen in Figure 16a. A set of lead-screws stretched the morphing skin sample from rest to 100% strain. The acrylic retaining walls could be adjusted to match the active skin area, and were tall enough to contain lead shot equivalent to a distributed load of 200 psf (9.58 kPa). By applying a thin layer of sand directly to the surface of the skin, the weight of the lead shot was distributed relatively evenly over the surface of the EMC. Moreover, as the skin deflected under load, the sand would adjust to conform to the surface and continue to spread the weight of the lead. A single-point laser position sensor was also placed underneath to measure the maximum deflections at the center of the skin, between the rib members.



(a) (b)

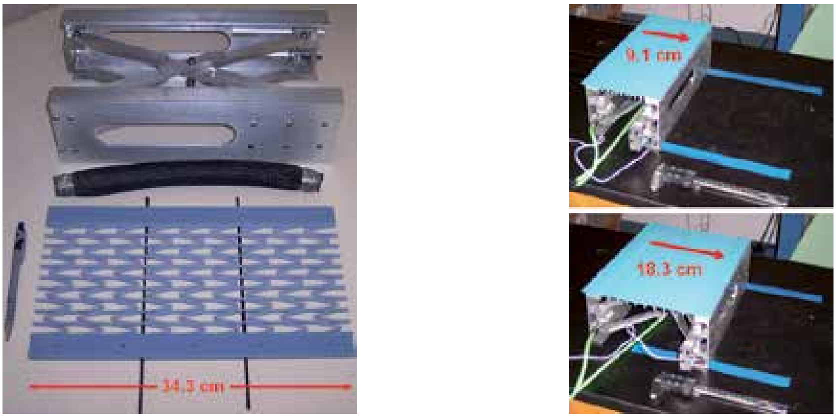
Fig. 16. (a) Out-of-plane deflection test apparatus design. (b) Out-of-plane deflection results as measured on the center rib.

The test procedure for each morphing skin covered the full range of operation, from resting (neutral position) to 100% area change. Lead-screws were used to set the skin to a nominal strain condition between 0% and 100% of the resting length. The laser position sensor shown below the skin in the figure was positioned in the center of a honeycomb cell at the center of the morphing skin, where the greatest deflection is seen. This positioning was achieved using a small two-axis adjustable table. The laser was zeroed on the under-surface of the EMC, and the relative distance to the bottom of an adjacent rib was measured. This established a zero measurement for rib deflection as well. A layer of sand with known weight was poured onto the surface of the EMC, and lead shot sufficient to load the skin to one of the three desired distributed loads was added to the top of the sand. Wing loadings of 40 psf (1.92 kPa), 100 psf (4.79 kPa), and 200 psf (1.92 kPa) were simulated. Once the load had been applied, measurements were taken at the same points on the EMC and the adjacent rib to determine deflection. These measurements were repeated for four different strain conditions (0, 25%, 66%, 100%) and the three different noted distributed loads.

Experimental results from the morphing skins is provided in Figure 16b. It was observed that, relative to the rib deflections, the EMC sheet itself deflected very little (less than 0.25 mm). The results therefore ignore the small EMC deflections and show only the maximum deflection measured on the rib at the midpoint of each morphing skin. Overall, the morphing skin deflections show that as the skin is strained and unsupported length increases, the out-of-plane deflection increases. Naturally, the deflection increases with load as well. Based on observation and on these results, the EMC sheets appeared to carry a greater out-of-plane load than expected, probably due to tension in the skin. EMC deflections between ribs remained low at all loading and strain conditions, while the substructure experienced deflections an order of magnitude greater. Future iterations of morphing skins will require stiffer substructures to withstand out-of-plane loads.

5.6 Full scale integration and evaluation

After proving capable of reaching over 100% strain with largely acceptable out-of-plane performance, the morphing skin sample from the previous subsection was used as the basis for a larger test article. A 34.3 cm x 14 cm morphing skin, nominally identical to the morphing skin sample in configuration, was fabricated and attached to the actuation assembly. The actuation assembly, honeycomb substructure, and completed morphing cell can be seen in Figure 17. Individual components of the system are pictured in Figure 17a, while the assembled morphing skin test article appears in Figure 17b. The active region stretches from 9.1 cm to 18.3 cm with no transverse contraction, thus, producing a 1-D, 100% increase in surface area with zero Poisson's ratio.



(a) (b)

Fig. 17. Integration of morphing cell – (a) actuation and substructure components; (b) complete morphing cell exhibiting 100% area change.

To characterize the static performance of the morphing cell, input pressure to the PAM actuators was increased incrementally and the strain of the active region was recorded at each input pressure, and a load cell in line with one PAM recorded actuator force for comparison to predicted values. This measurement process was repeated three times, recording strain, input pressure, and actuator force at each point. Note that the entire upper surface of the EMC is not the active region: each of the fixed-length ends of the honeycomb was designed and manufactured with 25.4 mm of excess material to allow adequate EMC bonding area and an attachment point to the mechanism. This inactive region can be seen on the top and bottom of the honeycomb shown in Figure 17a. The two extremities of the arrows in Figure 17b also account for the inactive region at both ends of the morphing skin.

The static strain response to input actuator pressure is displayed in Figure 18. Strain is seen to level off with increasing pressure due to a combination of mechanism kinematics and the PAM actuator characteristics, but the system was measured to achieve 100% strain with the PAMs pressurized to slightly over 620 kPa.

The measured system performance matches analytical predictions very closely. The previously mentioned analytical predictions and associated experimental data for the actuation system and skin performance are also repeated in this figure. The morphing cell performance data, while not perfectly linear, approximately matches the slope of the experimental skin stiffness and intersects the actuation system experimental data near 100% extension. Furthermore, although the performance data falls roughly 15% short of original predictions, the morphing skin meets the design goal, validating the analytical design process. Losses were not included in the original system predictions. However, the margin of error included in the original design for friction, increased skin stiffness, and other losses enabled the final morphing cell prototype to achieve 100% strain. It should also be noted that 100% area increases could be achieved repeatedly at 1 Hz using manual actuator pressurization.

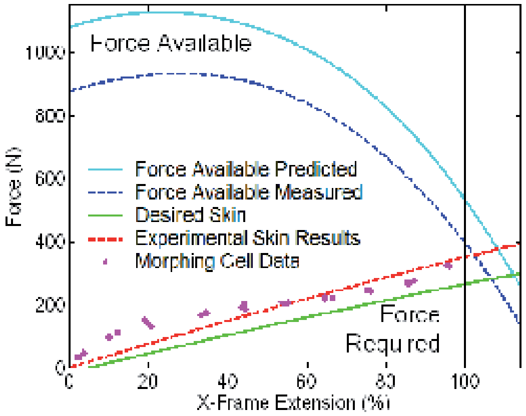


Fig. 18. Morphing cell data comparison with predictions.

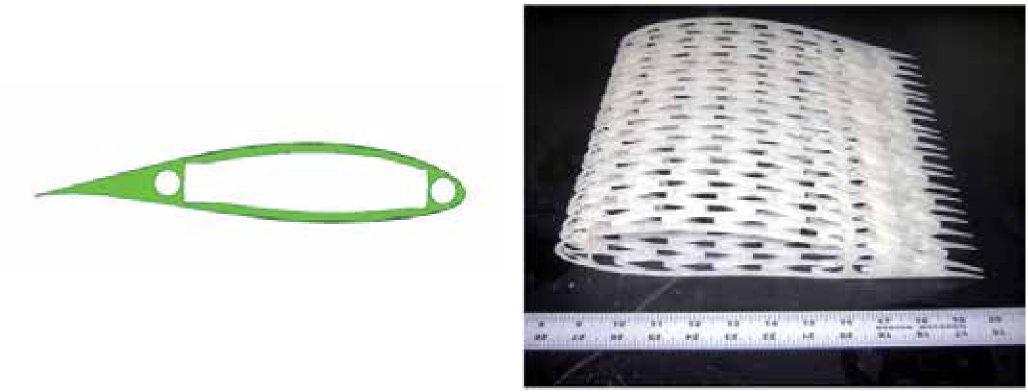
6. Wind tunnel prototype

Building on the success of the 1-D morphing demonstrator, a wind tunnel-ready morphing wing was designed and tested. A key technical issue addressed here was determining the scalability of the skin and substructure manufacturing processes for use on a real UAV. Thus, the prototype airfoil system was designed such that future integration with a candidate UAV is feasible, and experimentally evaluated as a wind tunnel prototype. Nominal design parameters for the prototype are a 30.5 cm chord wing section capable of 100% span extension over a 61.0 cm active morphing section with less than 2.54 mm of out-of-plane deflection between ribs due to dynamic pressures consistent with a 130 kph maximum speed.

6.1 Structure development

Initially, the planar core design was extruded and cut into the form of a NACA 63_3-618 airfoil with a chord of approximately 30.5 cm and span of 91.4 cm. A segment of the resulting morphing airfoil core appears in Figure 19a. While this morphing structure is capable of achieving greater than 100% length change itself, it has insufficient spanwise bending and torsional stiffness and so does not constitute a viable wing structure. The structure was therefore augmented with continuous sliding spars. Additionally, the center of the wing structure was hollowed out to potentially accommodate an actuation system for the span extension.

The final form of the morphing airfoil core is shown in Figure 19b. This figure shows a shell-like section mostly around the center of the airfoil, where an actuator could be located. Both the leading and trailing edges feature circular cut-outs to accommodate the carbon fiber spars, and near the trailing edge is a solid thickness airfoil shape for more rigidity where the airfoil is thinnest. The spars were sized using simple Euler-Bernoulli beam approximations and a desired tip deflection of less than 6.4 mm at full extension.



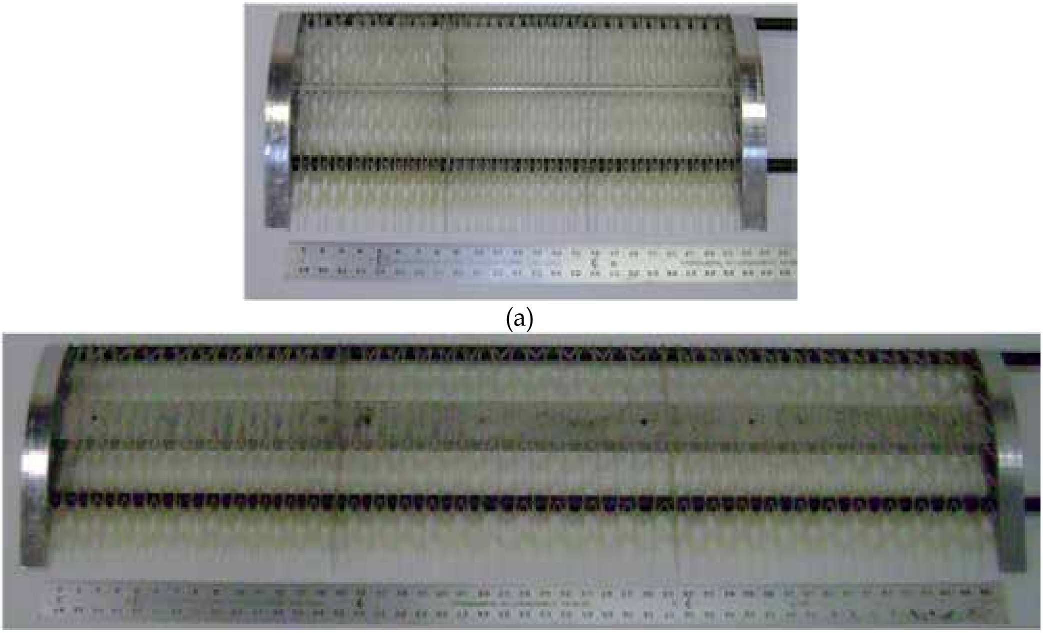
(a) (b)

Fig. 19. (a) Final substructure design, cross-section view (b) Manufactured substructure, side view.

Due to the complex geometry of the morphing core and the desire for rapid part turn around, a stereo lithographic rapid-prototyping machine was again used to manufacture the morphing core sections from an acrylic-based photopolymer. The viability of this approach for flyable aircraft applications would have to be studied, but the material/manufacturing approach was sufficient for this proof-of-concept structural demonstrator. Other fabrication techniques such as investment casting, electrical discharge machining, etc. could be considered when fabricating this structure to meet full scale aircraft requirements. It should also be noted that the prototype will feature three of the core segments shown in Figure 19b. They will be pre-compressed when the EMC skin is bonded to allow for more expansion capability and introduce a nominal amount of tension in the EMC skin.

Figure 20a shows the core sections together between two aluminum end plates, with the leading edge and trailing edge support spars. The end plates were sized to provide a suitably large bonding surface for attaching the skin on the tip and root of the morphing section. In this configuration, the core sections are initially contracted such that the active span length is 61.0 cm. In terms of the aircraft, this contracted state will be considered the neutral, resting state because the EMC skin will not be stretched here and a potential actuation system would not be engaged. Hence, this is the condition in which the skin would be bonded to the core. Also shown in Figure 20b is the same arrangement in the fully extended (100% span increase) state with a span of 122.0 cm. The figure shows that the spacing between each of the rib-like members has nearly doubled from what was shown in the contracted state. This figure helps illustrate the large area morphing potential of this technological development in a way that could not be seen once the skin was attached.

Spanwise bending and torsional stiffness was provided by two 1.91 cm diameter carbon fiber spars. The spars were anchored at the leftmost outboard portion of the wing but were free to slide through the inboard end plates, thus allowing the wing to extend while maintaining structural integrity. The spars were sized in bending to deflect less than 2.54 cm at 100% extension under the maximum expected aerodynamic loads. Note that the spars are also capable of resolving torsional pitching moments, but as the express purpose of the present work was to demonstrate the feasibility of a span morphing wing, these torsional properties were not directly evaluated.



(a)

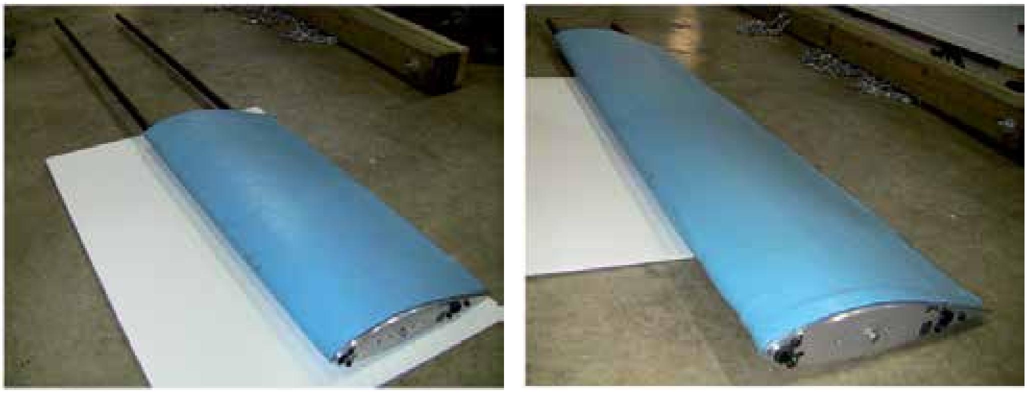
(b)

Fig. 20. Assembled core with spars and end plates – (a) contracted state; (b) extended state.

6.2 Prototype integration

The skin was bonded to the morphing substructure using DC-700. The skin was attached to each rib member, but not to the v-shaped bending members. Particular caution was used when bonding the skin to the end plates, as all of the tensile stress in the skin was resolved through its connection to the end plates.

At the resting condition with no elastic energy stored in the skin, Figure 21 shows the 0% morphing state with a 61.0 cm span. Increasing the span by another 61.0 cm highlights the full potential of this morphing system as the prototype wing section doubles its initial span, which has gone from 61.0 cm to 122 cm to show the 100% morphing capability (Figure 21b). Recall from the design that the wing section chord stays constant during these span extensions, so the morphing percentages indicated (e.g., 100%) are consistent with the increase in wing area. As a fixed point of reference in each of these figures, note that the length of the white poster board underneath the prototype wing section does not change. Note also that this demonstration will use fixed-length internal spreader bars to hold the structure in different morphing lengths. Actuation was achieved by manually stretching the skin/core structure and then attaching the appropriate spreader bar to maintain the stretched distance.



(a) (b)

Fig. 21. Prototype morphing wing demonstration – (a) resting length, 0% morphing; (b) 61.0 cm span extension, 100% morphing.

6.3 Wind tunnel testing

Having shown that the prototype morphing wing section could achieve the goal of 100% span morphing for a total 100% wing area increase, the final test that was performed placed the wing section in a wind tunnel. The purpose of this test was to ensure that the EMC skin and core could maintain a viable airfoil shape at different morphing states under true aerodynamic loading, with minimal out-of-plane deflection between ribs. An open circuit wind tunnel at the University of Maryland with a 50.8 cm tall, 71.1 cm wide test section was used in this test. An overall view of the test section is shown in Figure 22a, with the wing at its extended length, and a close-up view of the test section is shown in Figure 22b looking upstream from the trailing edge.

With only a 50.8 cm tall test section in the wind tunnel, where only this span length of the prototype morphing wing would be placed in the wind flow, while the remaining span and support structure was below the tunnel. This is illustrated in Figure 22a, where the full extension condition (100% morphing) is shown. It should also be noted that while only a 50.8 cm span section of the wing is in the air flow, this is sufficient to determine whether or not the skin and core can maintain a viable airfoil shape in the presence of representative aerodynamic conditions, which was the primary goal of this test. That is, the morphing core motion and skin stretching is consistent and substantially uniform across the span of the prototype, so any characteristics seen in one small section of the wing could similarly be seen or expected elsewhere in the wing, making this 50.8 cm span "sampling" a reasonable measure of system performance.

Both the cruise (105 kph) and maximum (130 kph) rated speeds of the candidate UAV were tested. Three angles-of-attack (0o, 2o, 4o) and three wing span morphing conditions (0%, 50%, 100%) were also included in the test matrix. Tests were performed by first setting the morphing condition of the wing section, then positioning the wing for the desired angle-of-attack (AOA). With these values fixed, the tunnel was turned on and the speed was increased incrementally, stopping at the two noted test speeds while experimental observations were made. Tests were completed at each of the conditions in the table indicated with an x-mark. Note that tests were not performed at two of the angles-of-attack at the 100% morphing condition. This was because the skin began to debond near the trailing edge at one of the end plates. This occurred over a section approximately 7.6 cm in span at the 100% morphing condition, though the majority of the prototype remained intact. After removing the wing section from the wind tunnel and inspecting the debonded corner, it was discovered that very little adhesive was on the skin, core, and end plate. Thus, the likely cause for this particular debonding was inconsistent surface preparation, which can easily be rectified in future refinements. Note that the upper surface of the trailing edge experiences relatively small dynamic pressures compared to the rest of the wing, so that this debonding was most likely unrelated to the wind tunnel test. Rather, it was the result of manufacturing inconsistency.

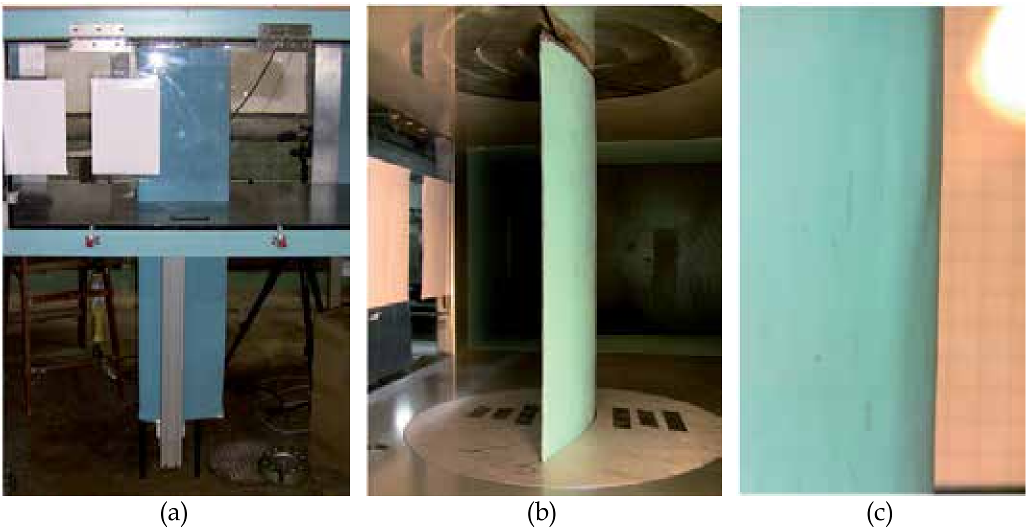


Fig. 22. Wind tunnel test setup – (a) Overall wind tunnel setup at 100% morphing; (b) Wing installed in wind tunnel – from trailing edge; (c) Picture of wing section leading edge at 130 kph, 100% morphing.

During execution of the test matrix, digital photographs (Figure 22c) were taken of the leading edge at each test point to determine the amount of skin deflection (e.g., dimpling) that resulted from the dynamic pressure. The leading edge location was chosen as the point to measure because the pressure is highest at the stagnation point. Pictures were taken perpendicular to the air flow direction and angled from the trailing edge, looking forward on the upper skin surface. Grids were taped to the outside of the transparent wall on the opposite side of the test section to provide reference lengths for processing. The grids form 12.7 mm squares and are located 35.6 cm behind the airfoil in the frame of view, which is also 35.6 cm from the camera lens. These can be seen in Figure 22c. Using image processing, the maximum error in the measurements was determined to be ±7%. This error can be attributed to vibration of the wind tunnel wall, which the camera lens was pressed against, or deviations in the focus of the pictures. In all the data processed, the maximum discernible out-of-plane deflection was approximately 0.51 ± 0.04 mm, which is well within the goal of less than 2.54 mm. In reference to the 30.5 cm chord and 5.49 cm thickness, this deflection accounts for only 0.17% and 0.93%, respectively. Additionally, in observing this experiment, it can be qualitatively stated the morphing wing held its shape remarkably well under all tested conditions. This can be confirmed through visual inspection of the figures, as well.

7. Conclusions

This work explored the development of a continuous one dimensional morphing structure. For an aircraft, continuous morphing wing surfaces have the capability to improve efficiency in multiple flight regimes. However, material limitations and excessive complexity have generally prevented morphing concepts from being practical. Thus, the goal of the present work was to design a simple morphing system capable of being scaled to UAV or full scale aircraft. To this end, a passive 1-D morphing skin was designed, consisting of an elastomer matrix composite (EMC) skin with a zero-Poisson honeycomb substructure intended to support out-of-plane loads. In-plane stiffness was controlled to match the capabilities of an actuator by careful design and testing of each separate skin component. Complete morphing skins were tested for in-plane and out-of-plane performance and integrated with the actuator to validate the design process on a small-scale morphing cell section.

Design goals of 100% global strain and 100% area change were demonstrated on a laboratory prototype using the combined morphing skin and actuation mechanism. The morphing skin strained smoothly and exhibited a very low in-plane Poisson's ratio. Actuation frequencies of roughly 1 Hz were achieved.

This work was then extended to a full morphing UAV-scale wing suitable for testing in a wind tunnel. The system was assembled as designed and demonstrated its ability to increase span by 100% while maintaining a constant chord. Wind tunnel tests were conducted at cruise (105 kph) and maximum speed (130 kph) conditions of a candidate UAV test platform, at 0^o, 2^o, and 4^o angles-of-attack, and at 0%, 50%, and 100% extensions. At each test point, image processing was used to determine the maximum out-of-plane deflection of the skin between ribs. Across all tests, the maximum discernable out-of-plane deflection was little more than 0.5 mm, indicating that a viable aerodynamic surface was maintained throughout the tested conditions.

8. Acknowledgement

This work was sponsored by the Air Force Research Laboratory (AFRL) through a Phase I STTR (contract number FA9550-06-C-0132), and also by a Phase I SBIR project from NASA Langley Research Center (contract number NNX09CF06P).

9. References

Agarwal, B. D., Broutman, L. J., and Chandrashekhara, K. (2006). *Analysis and Performance of Fiber Composites*, John Wiley & Sons, Hoboken.

Bae, J.S., Seigler, T.M. and Inman, D.J. (2005). ''Aerodynamic and Static Aeroelastic Characteristics of a Variable-Span Morphing Wing,'' *Journal of Aircraft*, 42(2): 528-534. doi: 10.2514/1.4397

Barbarino, S., Bilgen, O., Ajaj, R.M., Friswell, M.I., and Inman, D.J. (2011). "A Review of Morphing Aircraft," *Journal of Intelligent Material Systems and Structures,* 22: 823-877. doi:10.1177/1045389X11414084.

Bubert, E.A., Woods, B.K.S., Lee, K., Kothera, C.S., and Wereley, N.M. (2010). "Design and Fabrication of a Passive 1D Morphing Aircraft Skin," *Journal of Intelligent Material Systems and Structures*, 21(17):1699-1717 doi: 10.1177/1045389X10378777

Buseman, A. (1935) "Aerodynamic Lift at Supersonic Speeds," Ae. Techl. 1201, Report No. 2844 (British ARC, February 3, 1937), Bd. 12, Nr. 6: 210-220.

Bye, D.R. and McClure, P.D. (2007). ''Design of a Morphing Vehicle,'' *48th AIAA Structures, Structural Dynamics, and Materials Conference*, 23-26 April, Honolulu, HI, Paper No. AIAA-2007-1728.

Chaves, F. D., Avila, J., and Avila, A. F. (2003). "A morphological study on cellular composites with negative Poisson's ratios,"*44th AIAA Structures, Structural Dynamics, and Materials Conference*, Norfolk, VA, Paper No. AIAA 2003-1951.

Evans, K.E., Nkansah, M.A., Hutchinson, I.J., and Rogers, S.C. (1991). "Molecular network design," *Nature*, 353: 124.

Flanagan, J.S., Strutzenberg, R.C., Myers, R.B., and Rodrian, J.E. (2007). "Development and Flight Testing of a Morphing Aircraft, the NextGen MFX-1,"*AIAA Structures, Structural Dynamics and Materials Conference*, Honolulu, HI. Paper No. AIAA-2007-1707.

Gern, F.H., Inman, D.J., and Kapania, R.K. (2002). "Structural and Aeroelastic Modeling of General Planform Wings with Morphing Airfoils," *AIAA Journal*, 40(4): 628-637. doi: 10.2514/2.1719

Gevers, D.E. (1997). "*Multi-purpose Aircraft*," US Patent No. 5,645,250.

Gibson, L. J. and Ashby, M. F. (1988). *Cellular Solids: Structure and Properties*, Pergamon Press, Oxford.

Gomez, J. C., and Garcia, E. (2011). Morphing unmanned aerial vehicles. *Smart Materials and Structures*, 20(10):103001. doi:10.1088/0964-1726/20/10/103001

Hetrick, J. A., Osborn, R. F., Kota, S., Flick, P. M., and Paul, D. B. (2007). "Flight Testing of Mission Adaptive Compliant Wing,"*48th AIAA Structures, Structural Dynamics, and Materials Conference*, Honolulu, HI, Paper No. AIAA 2007-1709.

Joo, J.J. Reich, G.W. and Westfall , J.T. (2009). "Flexible Skin Development for Morphing Aircraft Applications Via Topology Optimization." *Journal of Intelligent Material Systems and Structures,* 20(16):1969-1985.

Kothera, C.S., Woods, B.K.S., Bubert, E.A., Wereley, N.M., and Chen, P.C. (2011). "Cellular Support Structures for Controlled Actuation of Fluid Contact Surfaces." U.S. Patent 7,931,240. Filed: 16 Feb 2007. Issued: 26 Apr 2011.

McKnight, G. Doty, R. Keefe, Herrera, A.G. and Henry, C. (2010). "Segmented Reinforcement Variable Stiffness Materials for Reconfigurable Surfaces." *Journal of*

Intelligent Material Systems and Structures, 21:1783-1793, doi:10.1177/ 1045389X10386399

Munk, M. M. (1924). "Note on the relative Effect of the Dihedral and the Sweep Back of Airplane Wings," NACA Technical Note 177.

Olympio, K.R., and Gandhi, F. (2010). "Flexible Skins for Morphing Aircraft Using Cellular Honeycomb Cores," *Journal of Intelligent Material Systems and Structures*, 21:1719-1735, doi:10.1177/1045389X09350331

Parker, H.J. (1920). "The Parker Variable Camber Wing," Report #77 Fifth Annual Report, *National Advisory Committee for Aeronautics,* Washington, D.C.

Perkins, D. A., Reed, J. L., and Havens, E. (2004). "Morphing Wing Structures for Loitering Air Vehicles," *45th AIAA Structures, Structural Dynamics & Materials Conference*, Palm Springs, CA, Paper No. AIAA 2004-1888.

Sarh, B., (1991). "*Convertible Fixed Wing Aircraft*," US Patent No. 4,986,493.

Samuel, J.B. and Pines, D.J. (2007). "Design and Testing of a Pneumatic Telescopic Wing for Unmanned Aerial Vehicles," *Journal of Aircraft*, 44(4) DOI: 10.2514/1.22205

Shigley, J., Mishke, C., and Budynas, R. (2004). *Mechanical Engineering Design*, McGraw-Hill, New York.

Thill, C., Etches, J., Bond, I., Potter, K., and Weaver, P. (2008). "Morphing Skins," *The Aeronautical Journal*, 112(1129):117-139.

Wereley, N. and Gandhi, F. (2010). "Flexible Skins for Morphing Aircraft." *Journal of Intelligent Material Systems and Structures*, 21: 1697-1698, doi:10.1177/1045389X10393157.

Wereley, N. M. and Kothera, C. S. (2007). "Morphing Aircraft Using Fluidic Artificial Muscles," *International Conference on Adaptive Structures and Technologies*, Ottawa, ON, Paper ID 171.

Wright, O. and Wright, W. (1906). "Flying-Machine" U.S. Patent 821,393. Filed: 23 Mar 1903. Issued: 22 May, 1906.

A Probabilistic Approach to Fatigue Design of Aerospace Components by Using the Risk Assessment Evaluation

Giorgio Cavallini and Roberta Lazzeri
University of Pisa-Department of Aerospace Engineering
Italy

1. Introduction

Fatigue design of aerospace metallic components is carried out by using two methodologies: damage tolerance and safe-life. At present, regulations mainly recommend the use of the former, which entrusts safety to a suitable inspections plan. Indeed, a crack or a flaw is supposed to have been present in the component since the beginning of its operative life, and it must remain not critical, i.e. it must not cause a catastrophic failure in the life period between two following inspections, (Federal Aviation Administration, 1998; Joint Aviation Authorities, 1994; US Department of Defence, 1998). When a crack is detected, the component is repaired or substituted and the structural integrity is so restored.

If the damage tolerance criterion cannot be applied, the regulations state the safe-life criterion should be used, i.e. components must remain free of crack for their whole operative life and, at their ends, components must be in any case substituted.

Therefore, both methodologies have deterministic bases and a single value (usually the mean value) is associated to each parameter that can influence the fatigue phenomenon, which on the contrary has a deep stochastic behaviour.

To take these items into account and in order to protect against unexpected events, it is necessary to introduce safety factors in the fatigue life design (generally equal to 2 or 3 for damage tolerance and equal to 4 or even more for safe-life). They usually produce heavy or expensive structures and, in the past, they were not always able to protect against catastrophic failures, because the real risk level is in any case unknown. On the one hand, indeed, the inspected structures or the substituted components may be still undamaged, with high costs; on the other hand, highly insidious phenomena, such as Multiple Site Damage and Widespread Fatigue Damage (which are typical of ageing aircrafts) cannot be taken into account very well and in the past they were the causes of some catastrophic accidents.

For these reasons, researchers are hypothesizing the possibility of facing fatigue design in a new way, by using the risk evaluation from a probabilistic point of view. Indeed, the parameters that affect the phenomenon have a statistical behaviour, and this can be described by means of statistical distributions.

In such a way, by using a statistical method, such as the Monte Carlo Method (Besuner, 1987; Hammersley & Handscomb, 1983), all the parameter distributions can be managed and each simulated 'event' can be considered as a possible 'event'. So, the computer simulation of the fatigue life of a big amount of components and the evaluation of the real risk level are possible, making this approach extremely useful.

The Authorities are interested in this approach but, before allowing the use of it as a design criterion, they require impartial evidence, first of all about the reliability of the analytical models used for fatigue simulation and for parameter distribution evaluation.

This paper intends to show a computer code – PISA, Probabilistic Investigation for Safe Aircrafts – and how it can be applied to the fatigue design of typical aerospace components, such as riveted joints, making it possible to integrate damage tolerance and the evaluation of the real risk level connected to the chosen inspection plan, (Cavallini et al., 1997; Cavallini & R. Lazzeri, 2007).

2. The parameters that mainly affect the fatigue phenomenon

A metallic component subjected to repeated loads can fail due to the fatigue phenomenon, (Schijve, 2001). Many research activities on this subject are known from both the theoretical and the experimental points of view and it is well known that fatigue has a random behaviour, with a high number of parameters (mechanical behaviour of the material, loads, geometry, manufacturing technologies, etc.) that can affect it.

Almost all these parameters have a statistical behaviour, but some among them play a more important role compared to the others and must be taken into account with their distributions, while the others can be assumed to be constant.

In detail, we can assume four main parameters as statistically distributed:

- the Initial Fatigue Quality (IFQ), described by using the (Equivalent) Initial Flaw Size, (E)IFS, or the Time To Crack Initiation, TTCI, distribution, (Manning et al., 1987);
- the crack grow rate, CGR (constant C in the Paris law);
- the fracture toughness K_{Ic}, and
- the inspection reliability, i.e. the Probability of crack Detection, PoD.

They are described in the following.

2.1 The Initial Fatigue Quality (IFQ)

Structural components can have, until the end of the manufacturing process, defects due to metallurgical effects, scratches, roughness, inclusions, welding defects, etc.

So, the IFQ can be considered as a property linked to the material and the manufacturing process. Defects can be the starting point for fatigue cracks. For this reason it is extremely important to know their position and size, but, as they are very small, they are very difficult to be measured even if by using very sophisticated inspection methods.

As a consequence, this information can be reached only through an indirect evaluation by means of a 'draw-back' procedure starting from experimental data about detectable cracks.

Therefore, a 'tool' able to characterize the component initial condition is necessary, in order to predict the fatigue life. At present, two approaches are available, Fig. 1:

- the (Equivalent) Initial Flaw Size distribution,
- the Time To Crack Initiation distribution.

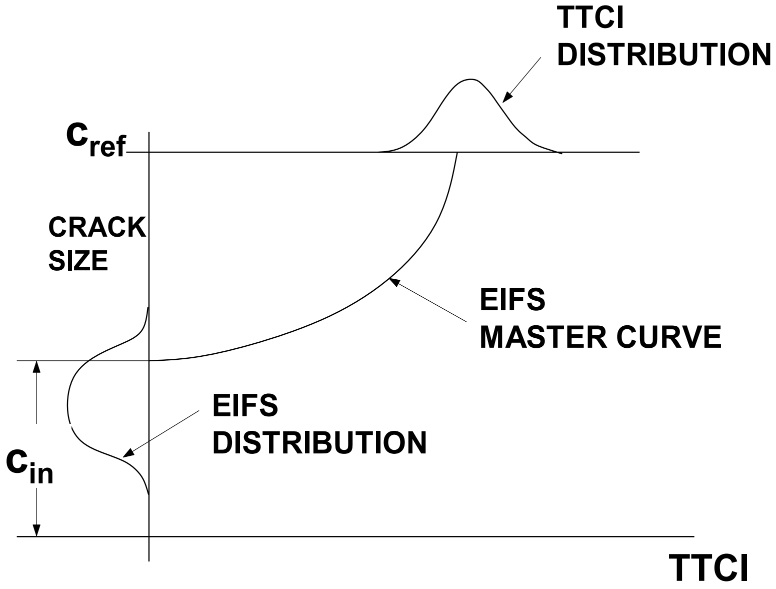


Fig. 1. TTCI and EIFS distributions.

2.1.1 The Time To Crack Initiation distribution (TTCI)

The TTCI model can be used to describe fatigue crack nucleation in metallic components. It can be defined as the time (in cycles, flights or flight hours) necessary for an initial defect to grow to a detectable reference crack, c_{ref}. In such a way, this method does not reveal the crack dimension during the early steps of the component life. The TTCI can be described by using a Weibull, (Manning et al., 1987), or a lognormal distribution (Liao & Komorowski, 2004). It must be noted that the TTCI distribution is not only a material property, as it is connected to the crack growth, and so to the loading fatigue spectrum.

2.1.2 The (Equivalent) Initial Flaw Size distribution (EIFS)

The EIFS is the (fictitious) dimension of a crack at time t=0. The use of the adjective 'equivalent' indicates that the initial flaw is not the actual one and its size is only 'equivalent' to it because it is very difficult to account for the influence of all the relevant parameters. The distribution can be numerically obtained starting from experimental crack size data by using a 'fictitious' backward integration of the crack growth. It is affected by the material properties and the stress distribution: cold-working, rivet interference, … have to be taken into account, too.

The EIFS can be described by using a lognormal or a Weibull (Manning et al., 1987) distribution. In the present paper, a lognormal distribution is assumed.

2.2 The Crack Growth Rate, CGR (C constant in the Paris law)

Different models are available to describe the crack growth law according to Linear Elastic Fracture Mechanics and the distribution of the involved parameters. We assumed to use the simple and effective Paris law $dc/dN=C(\Delta K)^m$. The parameter m is assumed to be constant and all the scatter is consolidated in C, which is assumed to belong to a normal distribution.

2.3 The Fracture Toughness K_{Ic}

Fracture toughness is a very important material property because it identifies a failure criterion (crack instability). Unfortunately, few experimental data are available to characterize its distribution. Anyway, a normal distribution (Hovey et al., 1991), or a lognormal distribution (Johnson, 1983; Schutz, 1980) can be supposed. We assumed the fracture toughness can be described through a lognormal distribution.

2.4 The Probability of crack Detection (PoD)

Non destructive inspections are among the principal items of the damage tolerance methodology. Indeed, during inspection, it is supposed that cracks are detected and the component can be re-qualified for further use. This action depends on many parameters, included the human factor and so it can be described only by using a probabilistic approach. Usually, (Lincoln, 1998), we define the inspection capability as the 90% probability of crack detection with the 95% of confidence.

Many distributions have been proposed for the Probability of Detection, (Tong, 2001; Ratwani, 1996).

In the present work we assumed a three parameters Weibull distribution, [Lewis et al., 1978]:

$$PoD = 1 - e^{-\left[\frac{c-c_{\min}}{\lambda - c_{\min}}\right]^{\beta}} \tag{1}$$

where, $c_{\min}$ is the minimum detectable crack size, c is the actual crack size and λ and β are parameters connected to the chosen inspection method.

3. The Monte Carlo method

A tool is necessary to manage all the parameter distributions together and at the same time. Some reliable approaches are available - FORM (First Order Reliability Method), SORM (Second Order Reliability Method) and many others (Madsen et al., 1986) – but we decided to use the Monte Carlo method because of its simplicity and effectiveness, as it can handle high numbers of different distributions for the stochastic variables to simulate many different deterministic situations. In addition, the Monte Carlo method easily allows the introduction of the repairs, that is a non-continuous change in the crack size.

The Monte Carlo method is based on a very easy assumption: the probability of an event p_f – in the present paper the component failure – is evaluated by using the analytical expression

$$p_f = n / N \tag{2}$$

where N is the total simulation number and n is the number of positive results.

Each simulation reproduces only one deterministic event, in which, for each deterministic or random variable a value is assumed; for the stochastic parameters, the value is randomly obtained from its distribution.

After a high number of trials, the method converges to the solution.

The only disadvantage of this method is that it requires a high number of simulations to have a low probability of the event. As an example, if the required probability is 10^{-6}, with a confidence level of 95%, at least (Grooteman, 2002)

$$N_{trials} > \frac{3}{10^{-6}} = 3x10^6 \tag{3}$$

are necessary.

4. The PISA code and the simulation of the fatigue phenomenon

The PISA code (Cavallini et al., 1997; Cavallini & R. Lazzeri, 2007), developed at the Department of Aerospace Engineering of the University of Pisa, allows the simulation of the whole fatigue life of typical aerospace components, such as simple plane panels, riveted lap-joint panels, Fig. 2, or stiffened panels, subjected to constant amplitude fatigue loading.

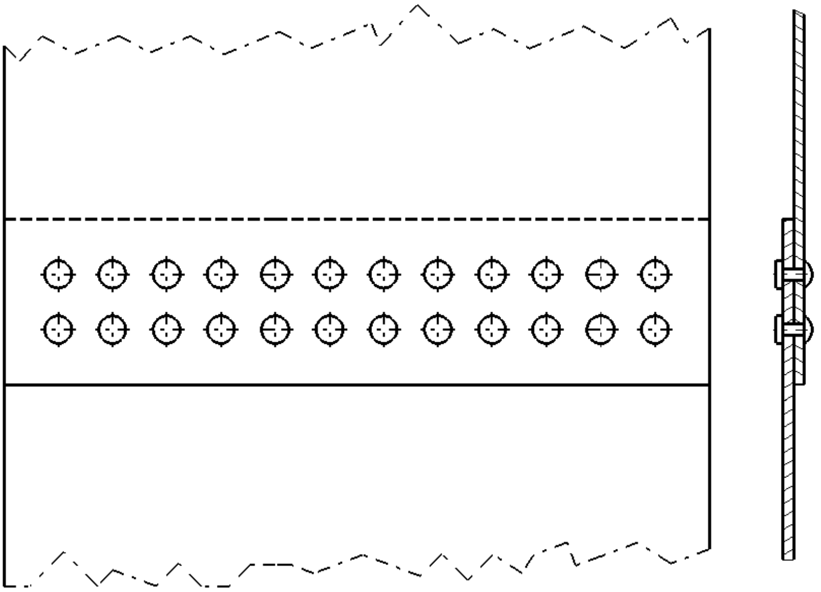


Fig. 2. Riveted lap-joint panel.

As far as the panel is concerned, the following hypotheses can be made:

Geometrical	Plane panel
	Uniform thickness
	Constant rivet pitch
	Rivets with or without countersunk head
	Through cracks
	Cracks on one or both hole sides and orthogonal to the load direction
Physical	Uniform stress
	Plane stress
	Uniform pin load in the same row
	Rivets with extremely high stiffness
	Fretting and corrosion effects are negligible

Table 1. Assumed geometrical and physical hypotheses.

The code can simulate crack nucleation, growth, inspection actions and failures in components subjected to uniform loading. The basic idea is that the damage process can be simulated as the continuous growth of an initial defect due to metallurgical effects and/or the manufacturing process, and/or other parameters.

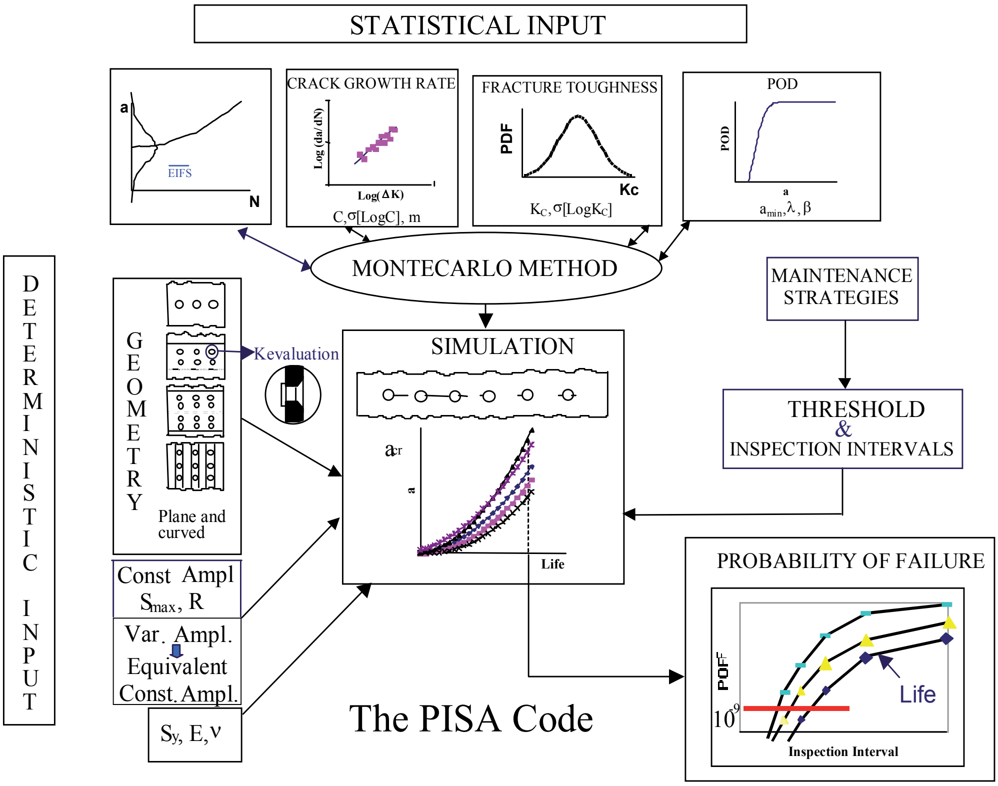


Fig. 3. Structure of the code.

Many deterministic simulations can be run and the risk assessment can be evaluated by using the Monte Carlo method, Fig. 3. In each simulation, the deterministic parameters are assigned by taking off a value from their own distributions.

The first phase – crack nucleation at holes – is simulated by using the EIFS distribution, which can be considered as an indication of the initial fatigue quality. In this context, this approach has to be preferred to the TTCI, as it allows to consider the whole life as the only propagation phase.

The second phase – crack growth – is simulated by using the simple, well-known Paris law $dc/dN=C(\Delta K)^m$

The core for the evaluation of the crack growth is the expression of the stress intensity factor *K* from the beginning of the life to the final failure.

In the stress intensity factor evaluation, suitable corrective factors (Sampath & Broek, 1991; Kuo et al., 1986) have been used to take into account the different boundary conditions, and the load transfer inside the joints has been simulated by using the Broek & Sampath model, (Sampath & Broek, 1991).

In detail, the effect of different boundary conditions can be taken into account by using the composition approach (Kuo et al., 1986).

K is analytically evaluated by means of a corrective coefficient which has been found by splitting the complex geometry into simple problems (open hole, finite width, …), whose solutions are known.

With regard to the open hole

$$K = K^R \cdot CR_1 \cdot CR_2 \cdot CR_n \tag{4}$$

$$K^R = S\sqrt{\pi(c-r)} \tag{5}$$

where *S* is the uniform membrane stress and (*c-r*) the crack length, Fig. 4.

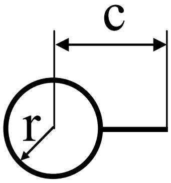


Fig. 4. Crack at an open hole.

As to the rivet effect, the load *P* on the hole has been approximated with a uniform pressure *p* on the hole, (Kuo et al., 1986), Fig. 5.

$$P = t \cdot \int_0^\pi p \cdot sen\theta \cdot r \cdot d\theta = t \cdot p \cdot r \cdot \int_0^\pi sen\theta \cdot d\theta = 2 \cdot p \cdot r \cdot t \tag{6}$$

$$p = \frac{P}{2 \cdot r \cdot t} \tag{7}$$

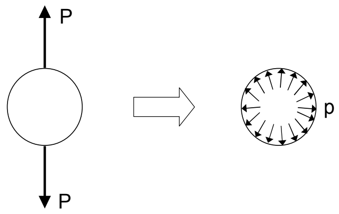


Fig. 5. Approximation of the *P* load with a uniform pressure *p*.

$$K = K^{P} \cdot CP_1 \cdot CP_2 \cdots CP_n \tag{8}$$

$$K^{P} = p\sqrt{\pi\ c} \tag{9}$$

The main corrective factors for open holes and for filled holes taken into account are summarized in Table 2.

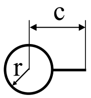
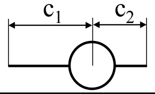
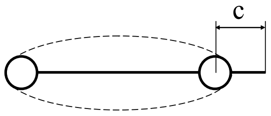
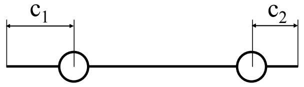
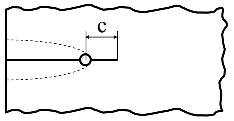
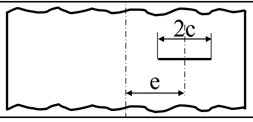
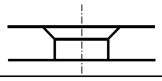
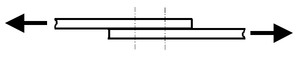
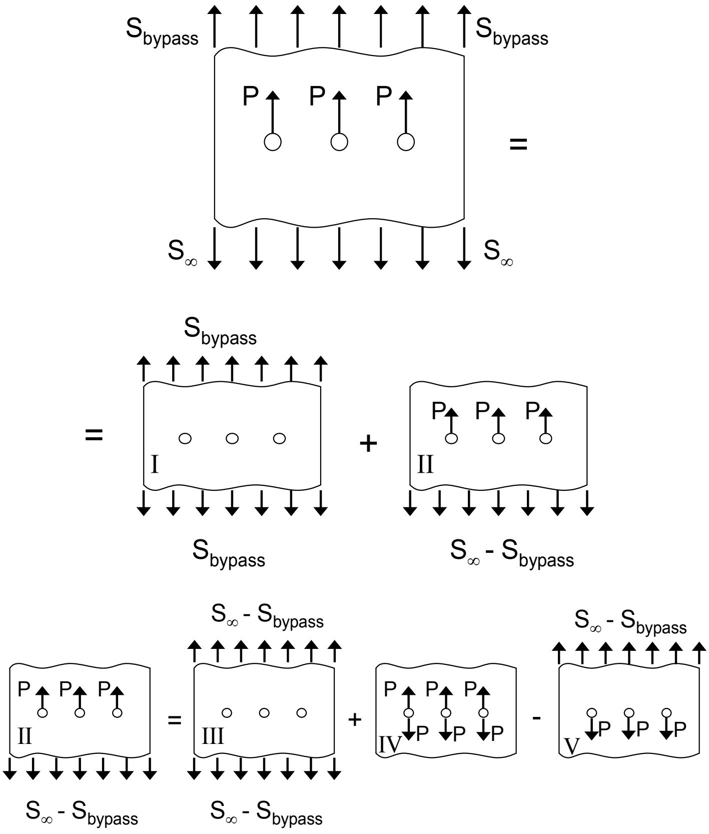
Crack at a hole (Kuo et al., 1986),	
Two cracks at a hole (Kuo et al., 1986),	
Link-up, with one crack (Kuo et al., 1986),	
Link-up, with two cracks (Kuo et al., 1986),	
Edge crack (Kuo et al., 1986),	
Panel finite width (Kuo et al., 1986),	
Countersink (Kuo et al., 1986),	
Secondary bending, (Sampath & Broek, 1991)	

Table 2. The main corrective factors for open holes and for filled holes taken into account in the PISA code.

The Broek & Sampath model joins the solutions related to the open hole and the loaded hole by using the superposition approach, Fig. 6.



$$K = K_{I} + K_{II} = K_{I} + ½(K_{III} + K_{IV})$$

Fig. 6. The Broek & Sampath model.

$$K = \frac{1}{2} \cdot CR \cdot \left(S_{\infty} + S_{bypass}\right) \cdot \sqrt{\pi \cdot c} + \frac{1}{2} \cdot CP \cdot p \cdot \sqrt{\pi \cdot c} \qquad (10)$$

S_{∞} is the membrane uniform stress and S_{bypass} is the bypass stress, i.e. for each row, the stress not transferred by the rivets.

Also rivet interference introduces an additional stress distribution. Its main effect is that only a part of the applied load amplitude S_{max}-S_{min} is effective for crack propagation. This effect can be taken into account by using the Wang model (Wang, 1988) for the evaluation of the lift-off stress S_{o}, corresponding to the separation of the rivet from the hole, and then by introducing in the simulations carried out with the PISA code only the effective amplitude S_{max}-S_{o} for crack propagation.

Once the stress intensity factor is calculated, crack growth simulation can start.

Two collinear cracks are considered as linked according to the Swift criterion, i.e. when their plastic radii r_p - evaluated by using the Irving expression - are tangential.

Inspections at planned intervals are simulated through the PoD distribution, applied at each crack at both hole sides. Though the repair of the hole has the same quality as the pristine panel, the repair itself of the detected crack is not immune from the possibility of having some tiny cracks.

The final failure can happen either for crack instability ($K_{\max}$ higher than the fracture toughness, $K_{\max} \geq K_{Ic}$) or for static failure ($S_{\max}$ higher than the yield stress in the net section evaluated without the plastic zones, $S_{\max} \geq S_{02}$)

5. Experimental activity as a support for the evaluation of the statistical distributions

To support this activity, all the parameter statistical distributions and the coefficients for the used analytical law (for example, the EIFS distribution, *C* and *m* coefficients for the Paris law, etc.) have to be experimentally evaluated. Of course, they cannot be obtained from tests on real components, but we have tested realistic simple specimens and we have demonstrated the applicability of the obtained results to the life evaluation of the actual components.

In this paper the activity carried out to find the distributions of the EIFS and the C constant in Paris law are shown. A similar approach can be used for the definition of the distributions for K_{Ic} and PoD.

5.1 Equivalent Initial Flaw Size distribution evaluation

For the evaluation of the EIFS distribution it was necessary to use a 'fictitious' negative integration (draw-back) which, starting from the experimental crack data at assigned number of cycles, would be able to find the 'equivalent' initial size, i.e. the crack size at N=0.

To support this approach, a wide experimental activity was performed on 29 simple strip lap-joints, Fig. 7, in aluminum alloy 2024-T3 (Cavallini & R. Lazzeri, 2007). They were fatigue tested under a constant amplitude load spectrum with $S_{\max}$=120 MPa and R=0.1. The tests were stopped at a set number of cycles, the specimens were statically broken and the crack dimensions were carefully measured. The tests confirmed an already well known result: all the cracks were found in the most critical location, i.e. in the first row, at the countersunk side.

At present, several numerical codes are available to simulate the growth of a single crack in the long crack range, but few can manage also the short crack range and none is able to carry out a direct negative integration that starting from the experimental crack data can find the initial dimension. For this reason, we decided to use the PISA code itself and the simplified model of a specimen with a through crack at a lap-joint, taking into account the effects of countersink of the hole, membrane stress, by-pass loading and pin load, Fig. 8, (Cavallini & R. Lazzeri, 2007).

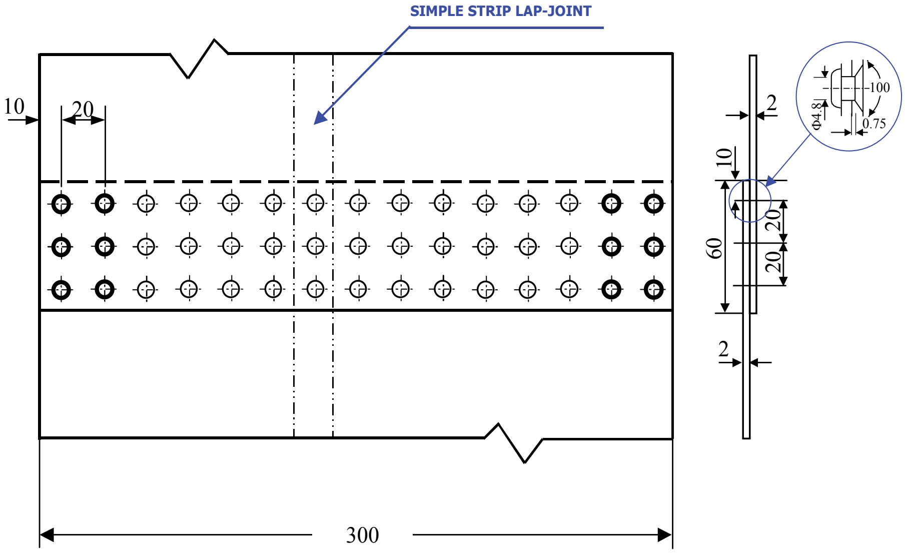


Fig. 7. Specimen geometry, all lengths in mm.

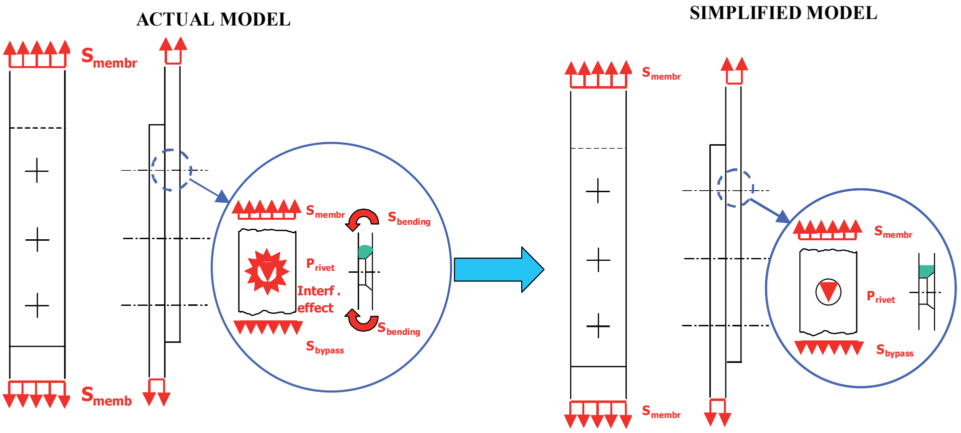


Fig. 8. Simplified model implemented inside the PISA code in order to evaluate the EIFS distribution.

In addition, an iterative positive integration was carried out starting from an initial tentative crack size value and stopping at the same number of cycles of the experimental result. The simulated final crack dimension was compared with the experimental one, and an iterative process was carried out to the required convergence.

In this way, a lognormal distribution for the EIFS was found, with $\mu[\text{Log}_{10}(c_0)]$=-2.88605, and $\sigma[\text{Log}_{10}(c_0)]$=0.28456, Fig. 9.

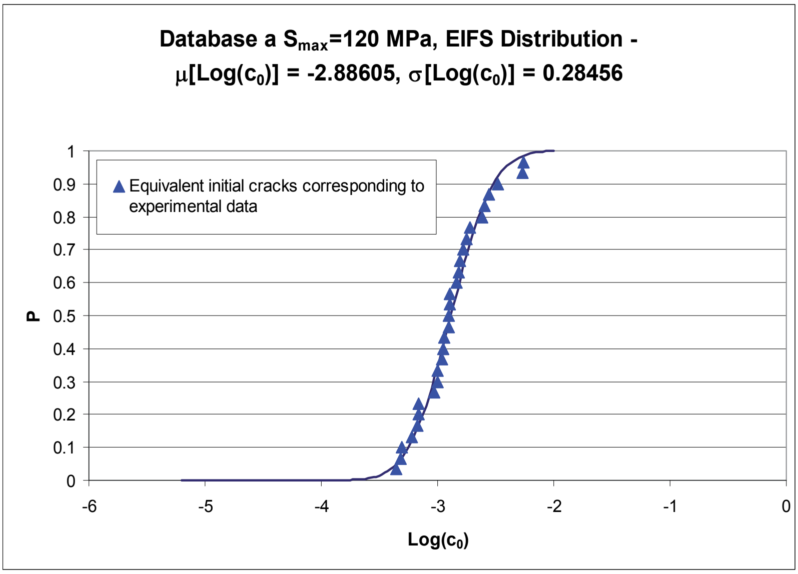


Fig. 9. EIFS distribution obtained by using the draw-back procedure.

5.2 The Crack Growth Rate (CGR) (C constant in the Paris law)

The crack growth law must be experimentally characterized, in order to evaluate the parameters involved in the selected crack growth law (Paris).

For their evaluation, a test campaign on a 2.1 mm thick Center Crack Tension (CCT) specimen with an open hole (4 mm in diameter) in aluminium alloy 2024-T3 was carried out (Imparato & Santini, 1997), Fig. 10.

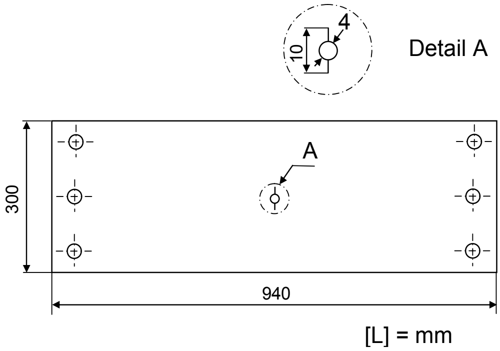


Fig. 10. CCT specimen in Al 2024-T3.

They were pre-cracked and the further crack propagation in the 5 to 40 mm range was investigated by using the Potential Drop Technique.

Tests were carried out under a constant amplitude (C.A.) spectrum, at the same S_{max}=68.7 MPa, with 4 different R=S_{min}/S_{max} values (R=0.1, R=0.25, R=0.4, R=0.55). In Fig. 11 the experimental results for one test are shown.

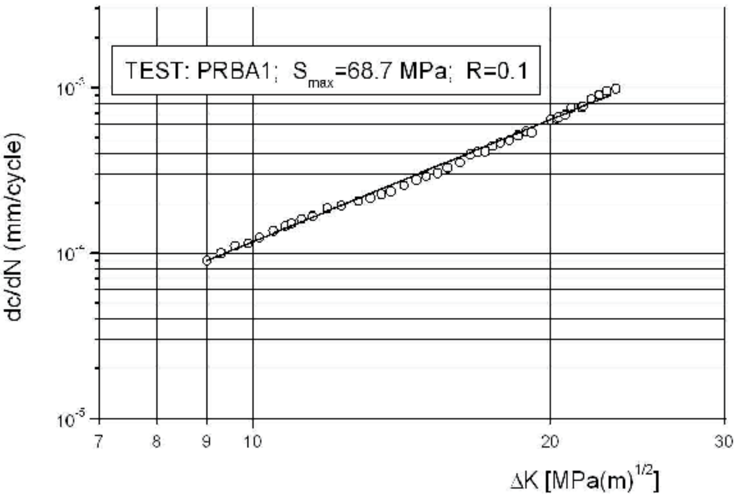


Fig. 11. Result for one CCT coupon test.

The test results were elaborated splitted for the different R values. For the characterization of the linear portion of the curve, we assumed *m* as a deterministic parameter (equal for each test, Fig. 12), and we considered only *C* as normal distributed.

In such a way, for *R*=0.1, we found that *m*=2.555 and *C* is normal distributed with μ[(*C*)]=2.8834x10^{-7}, and σ[(*C*)]=0.036792.

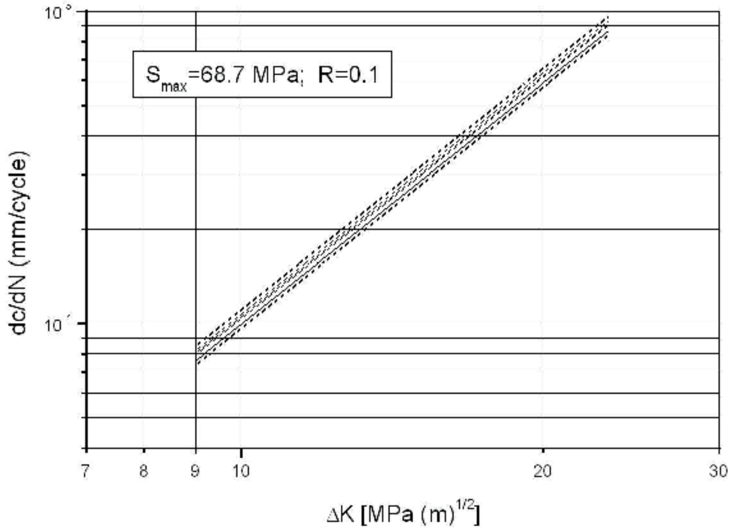


Fig. 12. Elaboration of the results for the tests at the same *R* value.

5.3 The validation of the analytical models implemented inside the PISA code

To validate the approach and to verify the capability of the PISA code to simulate the fatigue behavior of aerospace structural components, further experimental tests were carried out on wide lap-joint panels, in the same aluminum alloy as the simple strips, Fig. 7, loaded under a constant amplitude spectrum ($S_{\max}$ = 120 MPa, R=0.1).

A group of panels was fatigue tested for an assigned number of cycles (1 at 70,000 cycles, 4 at 75,000 cycles, 4 at 80,000 cycles, 3 at 85,000 cycles) and then statically broken to measure the sizes of the nucleated cracks. Also, in these panels the cracks were found only in the most critical row, i.e. the first one, at the countersunk side.

After having statically broken the tested panels, it was not possible, at all the hole sides, to detect a crack and they were considered as run-outs. The run-out effect has been introduced inside the crack size distribution by using the maximum likelihood method (Spindel & Haibach, 1979). Their crack dimensions have been supposed less than 0.1 mm. We supposed that the crack sizes (both with and without the run-outs) belong to two lognormal distributions.

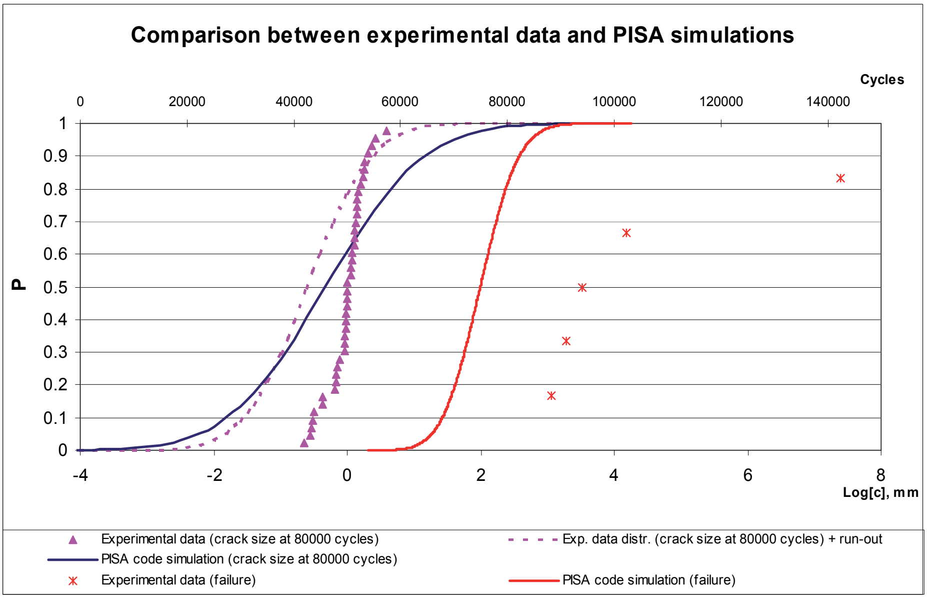


Fig. 13. Comparison between experimental data and PISA simulations (crack dimensions at 80,000 cycles and cycles to failure for the lap-joint panels).

Starting from the EIFS distribution obtained by the simple strip joints and by using the PISA code, the capability of the procedure has been verified by simulating the behavior of the cracks in the most critical row of the lap-joint panels.

The comparison has been made by generating 1000 runs, i.e. by simulating the crack sizes at different number of cycles in the most critical row of 1000 lap-joint panels similar to the tested lap-joint panels (i.e. 15 holes x 2 sides = 30 positions for each panel). In Figure 13 (Cavallini & R. Lazzeri, 2007) the comparison is shown between the predicted crack dimensions and the corresponding experimental results at 80,000 cycles.

The agreement between predictions and experimental results can be seen; in detail, the predicted distribution obtained by using PISA is included between the experimental data distribution with and without the run-outs, Fig. 13.

In addition, 5 further lap-joint panels were fatigue tested till failure. In Fig. 13 the comparison between PISA simulations and the experimental results is also shown. The agreement is good, though the simulation results are a little conservative.

6. Applications of the PISA code

The Pisa code is organized in such a way that all the information about geometry, material characterization, loads, inspection methods, failure criteria are collected in an input file.

The code can be used for the evaluation of the fatigue behaviour of only one component, starting from an initial known crack path, or for the generation of high numbers of deterministic simulations for the probabilistic approach.

6.1 The simulation of a single panel

In Fig. 14, the fatigue behaviour of a very simple panel, with only four open holes (diameter=4 mm), in Al 2024-T3, loaded under a constant amplitude load with S_{max}=100 MPa and R=0.1, is simulated. The initial crack path is extracted from the EIFS distribution, but it would be assigned also as an external input. The assigned life was 100,000 cycles. Inspections were planned every 25,000 cycles. For the probability of detection parameters, we assumed c_{min}=0.65 mm, λ=1.62 mm, β=1.35, (extrapolated from Ratwani, 1996).

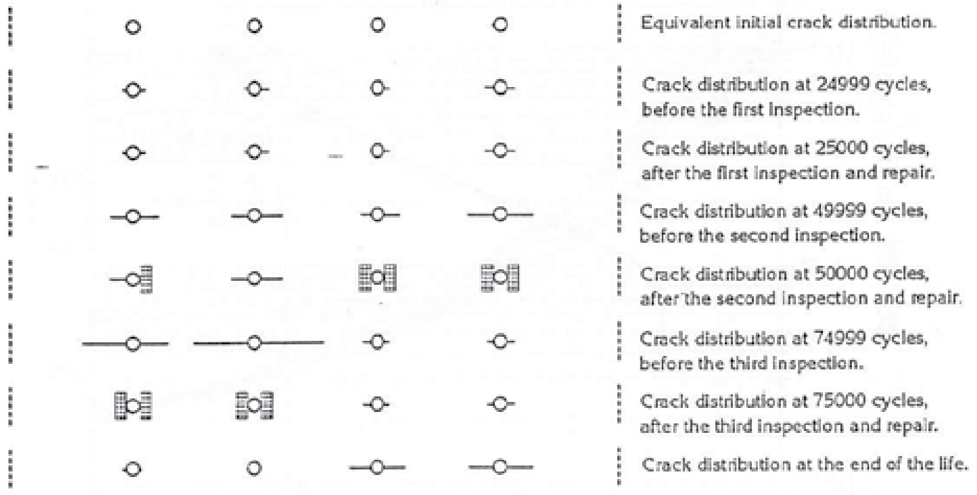

Equivalent initial crack distribution.
Crack distribution at 24999 cycles, before the first Inspection.
Crack distribution at 25000 cycles, after the first inspection and repair.
Crack distribution at 49999 cycles, before the second inspection.
Crack distribution at 50000 cycles, after the second inspection and repair.
Crack distribution at 74999 cycles, before the third inspection.
Crack distribution at 75000 cycles, after the third inspection and repair.
Crack distribution at the end of the life.

Fig. 14. Simulation of the fatigue life of an open hole panel (damage tolerance criterion).

As it can be seen in the Figure, during the first inspection (at 25,000 cycles) cracks were too small and were not detected. They grew till the second inspection (at 50,000 cycles), when five cracks were detected and the panel repaired. The simulation went on till the following inspection, when four more cracks were detected and repaired. With this inspection plan the panel could reach the target life.

6.2 Applications of the PISA code for the probabilistic risk evaluation

Before applying the Pisa code for the evaluation of the probability of failure of an aerospace component, an acceptable risk level must be identified. Indeed, one among the most debated items connected with the application of this methodology is the definition of the 'acceptable' risk level. Usually, 'risk' defines the probability of failure of some components within an assigned period.

Lincoln (Lincoln, 1998), says that for the USAF an acceptable global risk failure is 10^{-7} for flight, even if other authors suggest a safer $r(\mathrm{t}) \leq 10^{-9}$ per hour (Lundberg, 1959).

We fixed $r(\mathrm{t}) \leq 10^{-7}$. To reach a 10^{-7} probability of failure, at least $3\mathrm{x}10^{+7}$ simulations must be run.

Starting from a lap-joint in Al 2024-T3, Fig. 7, loaded at C.A. with $S_{\max}$=120 MPa, R=0.1, our aims were the definition of a 'safe' maintenance plan, the comparison of the effects of the deterministic (safe-life and damage tolerance) and the probabilistic approaches, and the evaluation of their respective advantages and disadvantages, (Cavallini & R. Lazzeri, 2007).

We assumed the following distributions for the stochastic parameters:

- The EIFS fits a lognormal distribution, with $\mu[\mathrm{Log}_{10}(c_0)]$=-2.88605, and $\sigma[\mathrm{Log}_{10}(c_0)]$=0.28456, $[c_0]$ in mm.
- The C parameter in the Paris law is normal distributed, with $\mu[(C)]=2.8834\mathrm{x}10^{-7}$, and $\sigma[(C)]$=0.036792. The corresponding m value is m=2.555.
- The fracture toughness K_{Ic} fits a lognormal distribution, with $\mu[\mathrm{Log}_{10}(K_{\mathrm{Ic}})]$=1.65 (Koolloons, 2002), and COV= $\sigma[\mathrm{Log}_{10}(K_{\mathrm{Ic}})]/\ \mu[\mathrm{Log}_{10}(K_{Ic})]$=0.14, $[K_{\mathrm{Ic}}]=\mathrm{MPa}(m)^{0.5}$ (Schutz, 1980),
- The probability of detection is expressed as (1), with $c_{\min}$=0.65 mm, λ=1.62 mm, β=1.35, (extrapolated from Ratwani, 1996).

At first, we simulated the fatigue behaviour of $3\mathrm{x}10^{+7}$ lap-joint panels without any inspection actions. The number of cycles with probability of failure equal to 10^{-7} is 51,000 cycles, Fig. 15. So, this number of cycles can be fixed for the first inspection (threshold).

The second run was made after having fixed, for each panel, the first inspection at 51,000 cycles. In such a way, we obtained the new probability of failure curve and it was possible to fix the second inspection at 63,000 cycles, that is (63,000 - 51,000) = 12,000 cycles after the first one.

In Fig. 15 the probability of failures corresponding to the different deterministic approaches are also shown.

The safe life criterion requires the component replacement after a portion (as for example ¼) of its mean life. The mean life (probability of failure equal to 50%) corresponds to 74,550 cycles that, divided by four, gives 18,638 cycles. So, for the safe life criterion, after 18,638 cycles the component should be substituted, without any consideration about its real damage condition.

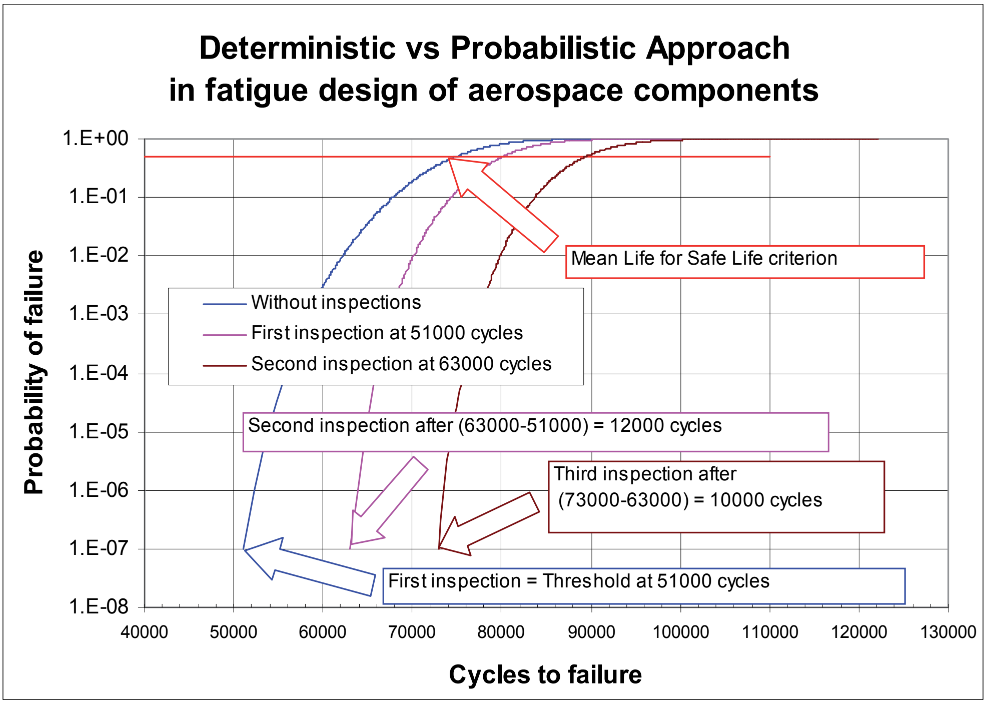


Fig. 15. Example of PISA simulation and maintenance strategy for a lap-joint panel.

The probability of failure at 18,638 cycles is extremely low, so the component could still be used without any loss in safety. In this situation, the component replacement only introduces costs.

As far as the damage tolerance criterion is concerned, the first inspection (threshold) is fixed by evaluating the number of cycles necessary for a crack of assigned dimension (regulations state a 1.27 mm size) to grow till the final failure. The inspections cited below are planned considering the propagation period of a sure visible crack (depending on the selected inspection method, in this case 6.35 mm) till the final failure. A safety factor equal to 2 at the threshold and 3 at the following period are additionally applied.

For this component, by analytical calculation or by using the PISA code itself, it can be found that the period necessary for a crack to grow from 1.27 mm to the final failure is equal to 57,700 cycles, and from 6.35 mm to the final failure is equal to 31,100 cycles.

Therefore, the first inspection will be carried out at 57,700/2=28,850 cycles and the next one after 31,100/3 = 10,367 cycles.

As it can be seen in Fig 15, the corresponding probability of failures is very low, and so the inspection plan, based on the deterministic damage tolerance approach, might be very expensive.

7. Conclusion

In this Chapter, a new possible and useful approach to fatigue design of aerospace metallic components is explained, founded on probabilistic bases, together with a tool – the PISA code - and the experimental test results used for the validation of the tool, and of the approach.

The validation analysis provided good results and therefore the PISA code can be used for the risk assessment analysis and to compare the effect of the deterministic approaches (damage tolerance and safe-life) with those of the probabilistic approach in the fatigue design of a wide lap-joint panel.

The advantages appeared to be very significant:

- the probability of failure can be defined as a design constraint (or goal), for instance 10^{-7}; so the risk level is well defined. In deterministic approaches, this important element is not known and the assumption of conservative values of the inputs can produce uneconomical designs without benefits;
- in each design condition, it is possible to know the "distance" from the critical condition in terms of probability of failure;
- the Multiple Site Damage event can be handled in a logical way because it is one of the possible statistical events. The same problem, faced on deterministic bases, could bring to heavy and/or very expensive solutions;
- components are inspected or withdrawn and substituted only if really necessary, thus avoiding too early inspections or the substitution of intact components.

The comparison between the different approaches, applied to a lap joint panel, shows that a more economic inspection plan can be applied if the probabilistic approach is used, without loss of safety.

Of course, this new methodology can be safely applied only if reliable models for the crack growth are available, and the parameter distributions have been carefully obtained.

8. References

Besuner P.M. (1987). Probabilistic Fracture Mechanics, In: *Probabilistic fracture mechanics and reliability*, Provan Ed., pp. 387-436, Martinus Nijhoff Publ., ISBN 90-247-3334-0, Dordrecht (NL).

Cavallini G., Lanciotti A. & Lazzeri L. (1997). A Probabilistic Approach to Aircraft Structures Risk Assessment, Proceedings of the 19th ICAF Symposium, Edinburgh (UK), June 1997, pp. 421-440.

Cavallini G. & Lazzeri R. (2007). A probabilistic approach to fatigue risk assessment in aerospace components. *Eng. Fracture Mech.*, vol. 74, issue 18, (Dec. 2007), pp. 2964-2970.

Federal Aviation Administration (1998). Federal Aviation Regulations – Part 25. Airworthiness Standards: Transport Category Airplanes, Section 571, Damage-tolerance and fatigue evaluation of structures. Available from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgFAR.nsf/Frameset?OpenPage

Grooteman F. P. (2002). *WP4.4: Structural Reliability Solution Methods – Advanced Stochastic Method*, Admire Document N. ADMIRE-TR-4.4-03-3.1/NLR-CR-2002-544.

Hammersley J. M. & Handscomb D. C. (1983). *Monte Carlo Methods,* Chapman and Hall Publ., ISBN 0-412-15870-1, New York.

Hovey P. W., Berens A. P. & Skinn D. A. (1991). *Risk Analysis for Aging Aircraft Volume 1 – Analysis*, Flight Dynamics Directorate, Wright Laboratory, Wright-Patterson AFB, OH 45433-6553.

Imparato G. & Santini L. (1997). *Prove sperimentali sul comportamento a fatica di strutture con danneggiamento multiplo*, Thesis in Aeronautical Engineering, Department of Aerospace Engineering, University of Pisa.

Joint Aviation Authorities (1994). Joint Airworthiness Requirements, JAR-25, Large Aeroplanes, Section 1, Subpart D, JAR 25.571, Damage-tolerance and fatigue evaluation of structures.

Johnston G. O. (1983). Statistical scatter in fracture toughness and fatigue crack growth rates, In: *Probabilistic fracture Mechanics and Fatigue Methods: Applications for structural design and maintenance*, ASTM STP 798, pp. 42-66, Bloom J.M. & Ekvall J.C., American Society for Testing Materials.

Koolloons M. (2002). Details on Round Tobin Tests, ADMIRE Document, ADMIRE-TR-5.1-04-1.1/NLR.

Kuo, A., Yasgur, D. & Levy, M. (1986). Assessment of damage tolerance requirements and analyses - Task I report., *ICAF Doc. 1583*, AFVAL-TR-86-3003, vol. II, AFVAL Wright-Patterson Air Force Base, Dayton, Ohio.

Lewis W. H., Sproat W.H., Dodd B. D. & Hamilton J. M. (1978). *Reliability of nondestructive inspection-final report*, San Antonio Air Logistic Center, Rep. SA-ALC/MME 76-6-38-1.

Liao M. & Komorowski J. P. (2004). Corrosion risk assessment of aircraft structures. *Journal of ASTM International*, vol. 1, no. 8 (September 2004), pp. 183-198.

Lincoln J.W. (1998). Role of nondestructive inspection airworthiness assurance, *RTO AVT Workshop on Airframe Inspection Reliability under field/depot conditions*, Brussels, Belgium, May 1998.

Lundberg, B. (1959). The Quantitative Statistical Approach to the Aircraft Fatigue Problem, Full-Scale Fatigue Testing of Aircraft Structures, *Proceedings of the 1st ICAF Symposium*, Amsterdam, Netherlands, 1959, Pergamon Press, pp. 393-412 (1961).

Madsen H. O., Krenk S. & Lind N. C. (1986). *Methods of Structural Safety*, Prentice Hall, Inc., ISBN 0-13-579475-7, Englewood Cliffs, NJ.

Manning, S.D., Yang, J.N. & Rudd, J.L. (1987). Durability of Aircraft Structures, In: *Probabilistic Fracture Mechanics and Reliability*, Provan J.W. (ed.), pp. 213-267, Martinus Nijhoff.

Ratwani M. M. (1996). Visual and non-destructive inspection technologies, In: *Aging Combat Aircraft Fleets - Long Term Implications*, AGARD SMP LS-206.

Sampath S. & Broek D. (1991). Estimation of requirements of inspection intervals for panels susceptible to multiple site damage, In: *Structural Integrity of Aging Airplanes*, Atluri S.N., Sampath, S.G. & Tong, P., Editors, , pp. 339-389, Springer-Verlag, Berlin.

Schijve J., (2001). *Fatigue of Structures and Materials*, Kluwer Academic Publishers, ISBN 0-7923-7013-9, Dordrecht, NL.

Schutz W. (1980). Treatment of scatter of fracture toughness data for design purpose, In: *Practical Applications of fracture Mechanics*, AGARD-AG-257, Liebowitz H. (ed).

Spindel J.E. & Haibach E. (1979). The method of maximum likelihood applied to the statistical analysis of fatigue data. *International Journal of Fatigue*, vol. I, no. 2, (April 1979), pp. 81-88.

Tong Y. C. (2001). Literature Review on Aircraft Structural Risk and Reliability Analysis, Department of Defence DSTO, Melbourne. Available from http://dspace.dsto.defence.gov.au/dspace/bitstream/1947/4289/1/DSTO-TR-1110%20PR.pdf

US Department of Defence (1998). Joint Service Specification Guide - Aircraft Structures, JSSG-2006, Available from http://www.everyspec.com/USAF/USAF+(General)/JSSG-2006_10206/.

Wang G. S. (1988). An Elastic-Plastic Solution for a Normally Loaded Center Hole in a finite Circular Body, *Int. J. Press-Ves & Piping*, vol. 33, pp. 269-284.

Study of Advanced Materials for Aircraft Jet Engines Using Quantitative Metallography

Juraj Belan
University of Žilina, Faculty of Mechanical Engineering,
Department of Materials Engineering, Žilina
Slovak Republic

1. Introduction

The aerospace industry is one of the biggest consumers of advanced materials because of its unique combination of mechanical and physical properties and chemical stability. Highly alloyed stainless steel, titanium alloys and nickel based superalloys are mostly used for aerospace applications. High alloyed stainless steel is used for the shafts of aero engine turbines, titanium alloys for compressor blades and finally nickel base superalloys are used for the most stressed parts of the jet engine – the turbine blades. Nickel base superalloys were used in various structural modifications: as cast polycrystalline, a directionally solidified, single crystal and in last year's materials which were produced by powder metallurgy.

So what exactly is a superalloy? Let us have a closer look to its definition. An interesting thing about it is that there is no standard definition of what constitutes a superalloy. The definitions which are provided in the various handbooks and reference books, although somewhat vague, are typically based on the service conditions in which superalloys are utilised. The most concise definition might be that provided by Sims et al. (1987): "...superalloys are alloys based on Group VIII-A base elements developed for elevated-temperature service, which demonstrate combined mechanical strength and surface stability." Superalloys are typically used at service temperatures above 540 C° (1000 F°), and within a wide range of fields and applications, such as components in turbine engines, nuclear reactors, chemical processing equipment and biomedical devices; by volume, its predominant use is in aerospace applications. Superalloys are processed by a wide range of techniques, such as investment casting, forging and forming, and powder metallurgy.

The superalloys are often divided into three classes based on the major alloying constituent: iron-nickel-based, nickel-based and cobalt-based. The iron-nickel-based superalloys are considered to have developed as an extension of stainless steel technology. Superalloys are highly alloyed, and a wide range of alloying elements are used to enhance specific microstructural features (and - therefore - mechanical properties). Superalloys can be further divided into three additional groups based on their primary strengthening mechanism:

- solid-solution strengthened;

- precipitation strengthened;
- oxide dispersion strengthened (ODS) alloys.

Solid-solution strengthening results from lattice distortions caused by solute atoms. These solute atoms produce a strain field which interacts with the strain field associated with the dislocations and acts to impede the dislocation motion. In precipitation strengthened alloys, coherent precipitates resist dislocation motion. At small precipitate sizes, strengthening occurs by the dislocation cutting of the precipitates, while at larger precipitate sizes strengthening occurs through Orowan looping. Oxide dispersion strengthened alloys are produced by mechanical alloying and contain fine incoherent oxide particles which are harder than the matrix phase and which inhibit dislocation motion by Orowan looping (MacSleyne 2008).

Figure 1. provides a representation of the alloy and process development which has occurred since the first superalloys began to appear in the 1940s; the data relates to the materials and processes used in turbine blading, such that the creep performance is a suitable measure for the progress which has been made. Various points emerge from a study of the figure. First, one can see that - for the blading application - cast rather than wrought materials are now preferred since the very best creep performance is then conferred. However, the first aerofoils were produced in wrought form. During this time, alloy development work – which saw the development of the first Nimonic alloys - enabled the performance of blading to be improved considerably; the vacuum introduction casting technologies which were introduced in the 1950s helped with this since the quality and cleanliness of the alloy were dramatically improved.

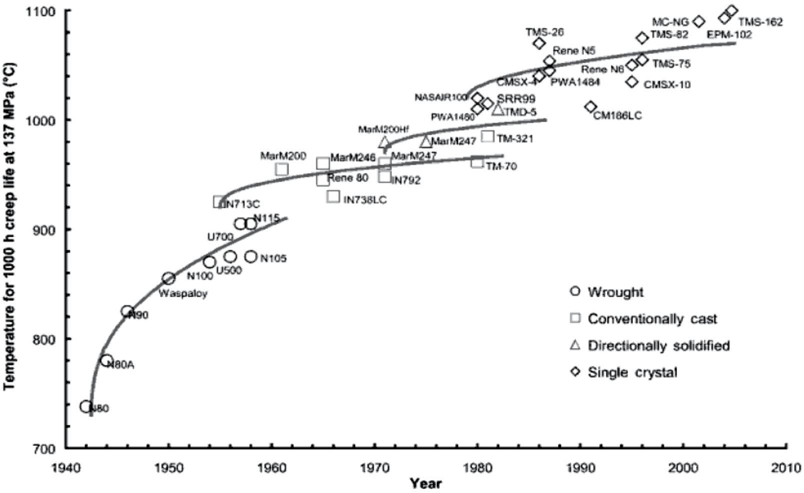


Fig. 1. Evolution of the high–temperature capability of superalloys over a 70 year period, since their emergence in the 1940s (Reed 2006).

Second, the introduction of improved casting methods and - later - the introduction of processing by directional solidification enabled significant improvements to be made; this was due to the columnar microstructures that were produced in which the transverse grain boundaries were absent (see Figure 2.)

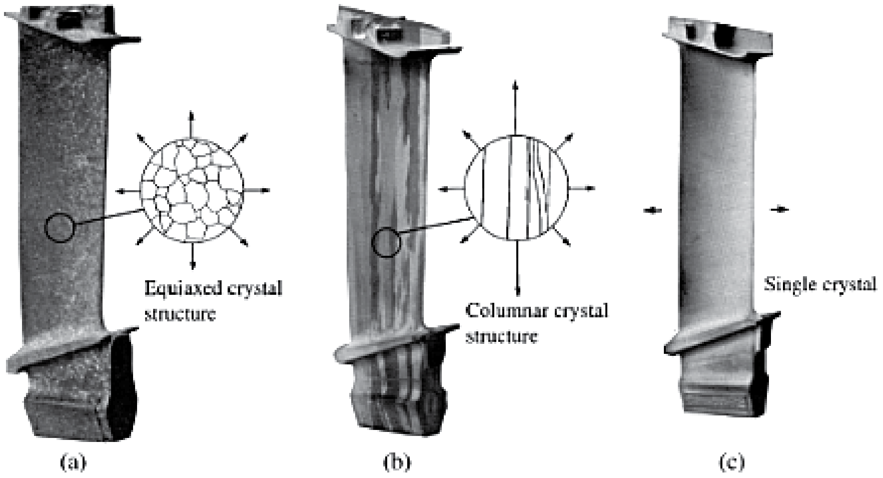


Fig. 2. Turbine blading in the (a) equiaxed-, (b) columnar- and (c) single–crystal forms.

Once this development had occurred, it was quite natural to remove the grain boundaries completely such that monocrystalline (single-crystal) superalloys were produced. This allowed, in turn, the removal of grain boundary strengthening elements such as boron and carbon which had traditionally been added, thereby enabling better heat treatments to reduce microsegregation and induced eutectic content, whilst avoiding incipient melting during heat treatment. The fatigue life is then improved.

Nowadays, single–crystal superalloys are being used in increasing quantities in the gas turbine engine; if the very best creep properties are required, then the turbine engineers turn to them (although it should be recognised that the use of castings in the columnar and equiaxed forms is still practiced in many instances).

In this chapter, a problem of polycrystalline (equiaxed) nickel base superalloy turbine blades - such as the most stressed parts of the aero jet engine - will be discussed.

The structure of polycrystalline Ni–based superalloys - depending on heat–treatment - consists of a solid solution of elements in Ni (γ-phase, an austenitic fcc matrix phase) and inter-metallic strengthening precipitate $Ni_3(Al, Ti)$ (γ'-phase, which is an ordered coherent precipitate phase with a LI2 structure). A schematic showing representative microstructures of both a wrought and a cast nickel-base superalloy is shown in Figure 3. The γ′ precipitates in precipitate strengthened nickel-base superalloys remain coherent up to large precipitate sizes due to the small lattice mismatch between the matrix phase γ and the γ′ precipitates. The γ′ precipitates are usually present in volume fractions in the range of 20-60%, depending on the alloy (Sims et al. 1987), with typical shapes from the spherical at small sizes to cuboid at larger sizes, although more complex dendritic shapes are also observed in some cases (see Figure 4). The alignment of γ′ precipitates along the elastically soft (100) directions is frequently observed. Nickel based superalloys are precipitation hardened, with a typical precipitate size of 0.25-0.5 μm for high temperature applications (Sims et al. 1987).

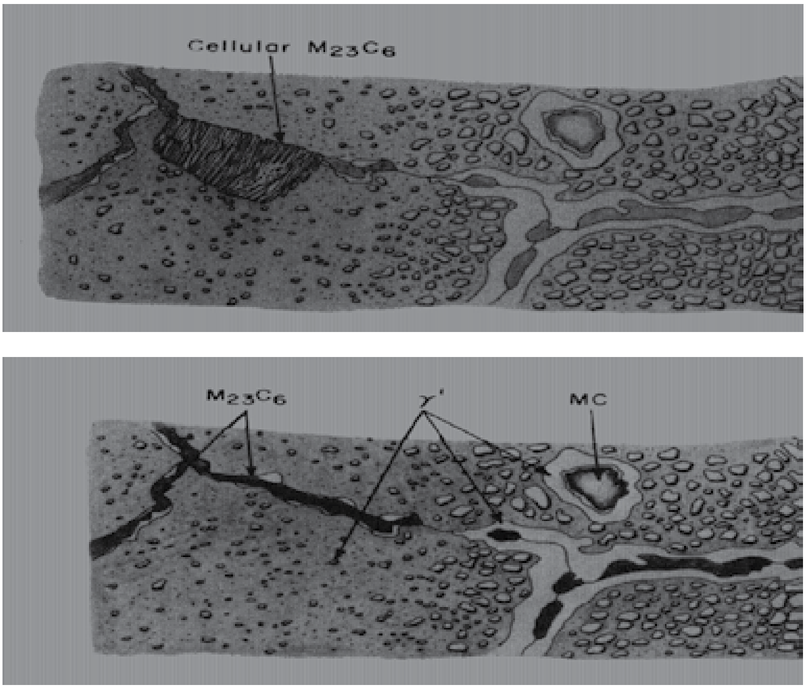


Fig. 3. Structure of a wrought and a cast nickel-base superalloy (M. J. Donachie & S. J. Donachie 2002).

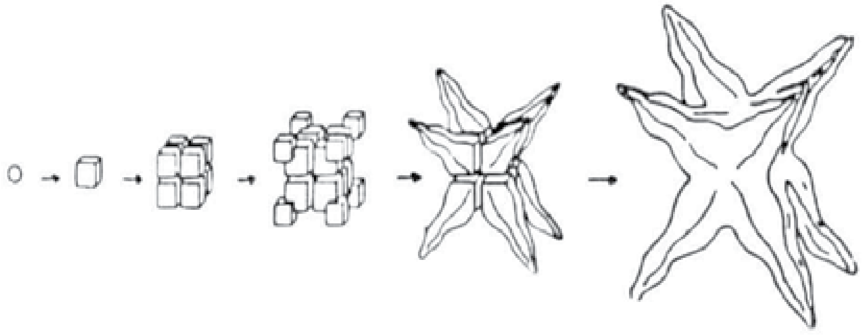


Fig. 4. Schematic showing the evolution of γ' morphology during continuous cooling. Sphere → cube → ogdoadically diced cubes → octodendrite → dendrite (Durand–Charre 1997).

In niobium-strengthened nickel-base superalloys - such as IN-718 - the principal strengthening phase is γ'' (Ni_3Nb), which has a bct ordered DO22 structure. When γ'' precipitates are observed, they form as disk-shaped precipitates on {100} planes with a thickness of 5-9 nm and an average diameter of 60 nm (Durand–Charre 1997).

The next structural components are MC type primary carbides (created by such elements as Cr and Ti) and $M_{23}C_6$ type secondary carbides (created by such elements as Cr, Co, Mo and W). However, except of these structural components, "unwanted" TCP (Topologically

Close-Packed) phases are also presented, such as σ-phase A_xB_y (Cr, Mo)$_x$(Fe, Ni)$_y$, μ–phase A_7B_6 (Co, Fe, Ni)$_7$(Mo, W, Cr)$_6$, Laves phases A_2B (Fe, Cr, Mn, Si)$_2$(Mo, Ti, Nb) and A_3B phases (π Ni_3(AlTa), η Ni_3Ti, δ Ni_3Ta and ε (NiFeCo)$_3$(NbTi)). The shape and size of these structural components have a significant influence on final the mechanical properties of alloys and - mainly - on creep rupture life.

Although alloy-specific heat treatments are generally proprietary, the typical heat treatment of nickel-base superalloys consists of a solution treatment followed by an aging step (precipitation and coarsening). For additional details on alloy-specific heat treatments, see Sims et al. (1987) and M. J. Donachie & S. J. Donachie (2002). Nickel-base superalloys are highly-alloyed: because of the complexity which this adds, many experimental studies use binary or ternary alloys as model alloy systems. The nickel-rich region of the binary Ni-Al alloy system is frequently used as a model alloy system. The Al-Ni phase diagram is shown in Figure 5, and we will use it to consider the typical heat treatments of nickel-base superalloys.

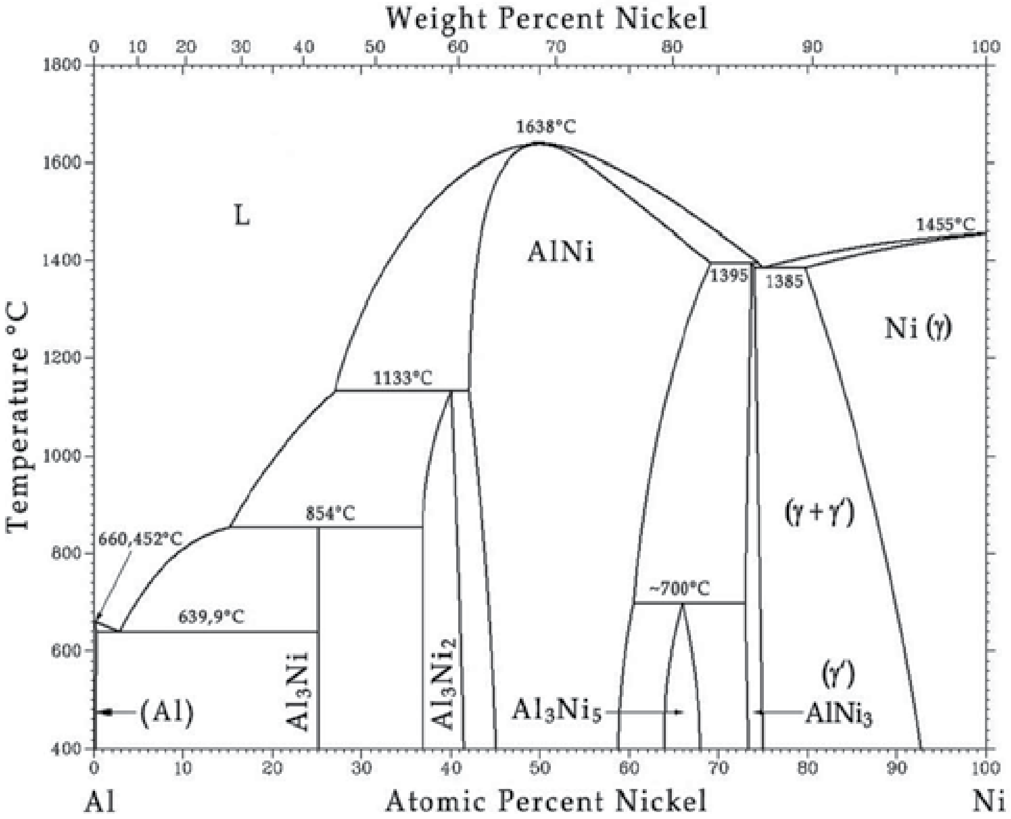


Fig. 5. Al-Ni phase diagram (ASM, 1992).

The solution treatment occurs above the γ′ solvus and is required for the ordered γ′ to go into the solution. The γ′ solvus separates the γ + γ′ and γ regions in Figure 5. This is usually followed by a quench (air, water or oil, depending upon the alloy) to room temperature. The aging occurs at a temperature below the solvus temperature and allows for the homogeneous nucleation, growth and coarsening of γ', followed by air or furnace cooling to room temperature. Although heterogeneous nucleation is observed on grain boundaries and dislocations - for example - nucleation is primarily homogeneous. The temperature and duration of the aging treatment are selected so as to optimise the morphology, alignment and size distribution of γ' precipitates. The resulting microstructure, in addition to its dependence on heat-treatment parameters, is also dependent on the physical properties of the alloy (and their isotropic or anisotropic nature) such as the lattice mismatch, the coherent γ' interface energy, the volume fraction of γ' and the elastic properties of the matrix and precipitate.

Polycrystalline turbine blades typically work within a temperature range from 705°C up to 800°C. As such, they must be protected from heat by various heat-proof layers; for example an alitise layer, MCrAlY coating or TBC (Thermal Barrier Coating). For this reason, dendrite arm-spacing, carbide size and distribution, morphology, the number and value of the γ'-phase and protective layer degradation are very important structural characteristics for the prediction of a blade’s lifetime as well as the aero engine itself. In this chapter, the methods of quantitative metallography (Image Analyzer software NIS – Elements for carbide evaluation, the measurement of secondary dendrite arm-spacing and a coherent testing grid for γ'-phase evaluation) are used for the evaluation of the structural characteristics mentioned above on experimental material – Ni base superalloy ŽS6K.

For instance, a precipitate γ' size greater than 0.8 μm significantly decreases the creep rupture life of superalloys and a carbide size greater than 5 μm is not desirable because of the initiation of fatigue cracks (Cetel, A. D. & Duhl, D. N. 1988).

For this reason, the needs of new methods of the evaluation of non–conventional structure parameters were developed. Quantitative metallography, deep etching and colour-contrast belong to the basic methods. The analysis of quantitative metallography has a statistical nature. The elementary tasks of quantitative metallography are:

- Dendrite arm-spacing evaluation;
- Carbide size and distribution;
- Volume ratio of evaluated gamma prime phase;
- Number ratio of evaluated gamma prime phase;
- Size of evaluated gamma prime phase;
- Protective alitise layer degradation.

The application of quantitative metallography and colour contrast on the ŽS6K Ni–base superalloy are the main objectives discussed in this chapter.

2. Description of experimental methods and experimental material

2.1 Experimental methods

For the evaluation of structural characteristics the following methods of quantitative metallography were used:

- Carbide distribution and average size was evaluated by the software NIS-Elements;
- Secondary dendrite arm-spacing measurement;

- For the number of γ'-phase particles, a coherent testing grid with 9 square shape area probes was used;
- For the volume of γ'-phase particles, a coherent testing grid with 50 dot probes made from backslash crossing was used.

Secondary dendrite arm spacing was evaluated according to Figure 6 and calculated with formula (1). The changing of the distance between the secondary dendrite arms "d" is an important characteristic because of base material, matrix γ, degradation via the equalising of chemical heterogeneity and also grain size growing.

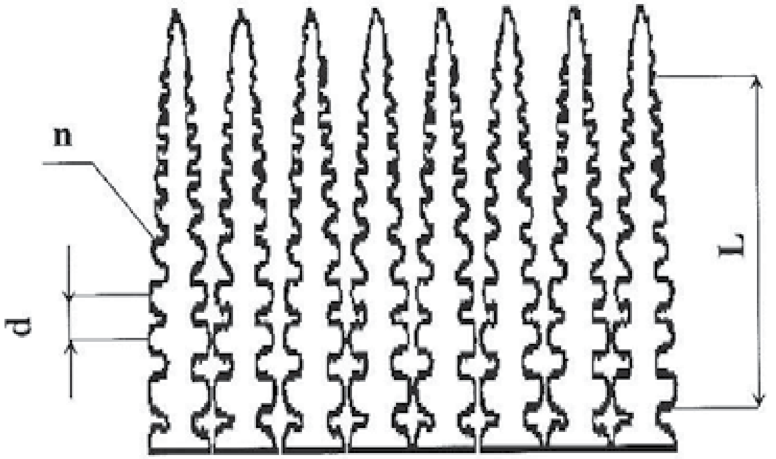
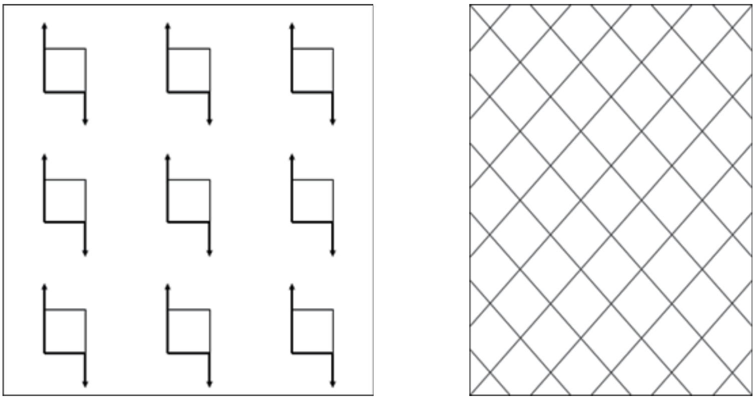


Fig. 6. Scheme for the evaluation of secondary dendrite arm-spacing.

$$d = \frac{L}{n} \cdot \frac{1}{z} \cdot 1000 \quad (\mu m) \tag{1}$$

- where "L" is a selected distance on which secondary arms are calculated (the distance is usually chosen with the same value as used magnification "z" – the reason why this is so is in order to simplify the equation), "n" is the number of secondary dendrite arms and "z" is the magnification used.

For the evaluation of the γ- and γ'-phases the method of coherent testing grid was used, and the number of γ' "N" was evaluated by a grid with 9 square-shaped area probes (Figure 7a) and the volume of γ' "V" was evaluated by grid with 50 dot probes (Figure 7b). Afterwards, measurement of the values was calculated with formulas (2) and (3). For a detailed description of the methods used, see (Skočovský & Vaško 2007, Tillová & Panušková 2008, Tillová et al. 2011). The size of γ' is also important from the point of view of creep rupture life. A precipitate with a size higher than 0.8 µm can be considered to be heavily degraded and as causing decreasing mechanical strength at higher temperatures.



a) number of γ′ particles b) volume of γ′ particles

Fig. 7. Coherent testing grid for γ′ evaluation.

$$N = 1{,}11 \cdot z^2 \cdot x_{str} \cdot 10^{-9} \quad \left(\mu m^{-2}\right) \tag{2}$$

- where "N" is a number of γ′ particles, "z" is the magnification used, "x_{str}" is the medium value of γ′-phase measurements.

$$V = \frac{n_s}{n} \cdot 100 \quad (\%) \tag{3}$$

- where "V" is a volume of γ′ particles, "n_s" is the medium value of γ′-phase measurement and "n" is a number of dot probes (when using a testing grid with 50 dot probes, the equation become more simple: $V=2n_s$).

2.2 Experimental material

The cast Ni–base superalloy ŽS6K was used as an experimental material. Alloy ŽS6K is a former USSR superalloy which was used in the DV–2 jet engine. It is used for turbine rotor blades and whole-cast small-sized rotors with a working temperature of up to 800 ÷ 1050°C. The alloy is made in vacuum furnaces. Parts are made by the method of precise casting. The temperature of the liquid at casting in a vacuum to form is 1500 ÷ 1600°C, depending on the part′s shape and its quantity. The cast ability of this alloy is very good, with only 2 ÷ 2.5% of shrinkage. Blades made of this alloy are also protected against hot corrosion, with a protective heat-proof alitise layer, and so they are able to work at temperatures of up to 750°C for 500 flying hours.

This alloy was evaluated at the starting stage, the stage with normal heat treatment after 600, 1000, 1500 and 2000 hours of regular working (for these evaluations, real ŽS6K turbine blades with a protective alitise layer were used as an experimental material), and different samples made from the same experimental material ŽS6K after annealing at 800 °C/ 10 and 800 °C/15 hours. This was followed by cooling at various rates in water, oil and air. The chemical composition in wt % is presented in Table 1.

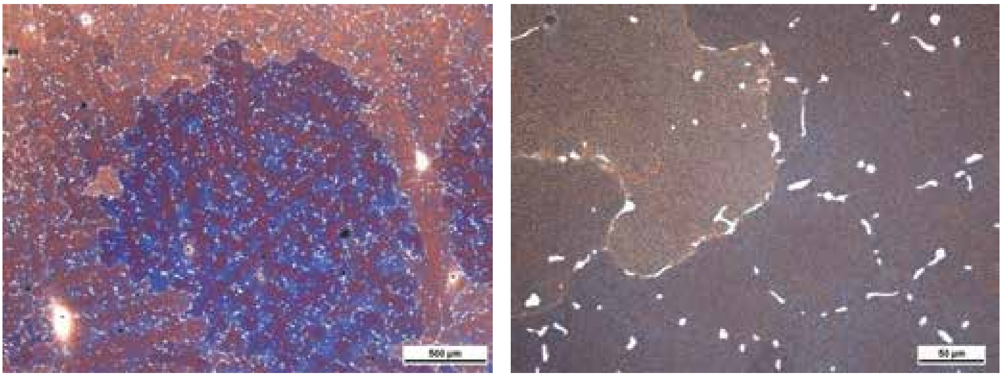
A typical microstructure of the ŽS6K Ni–base superalloy as cast is shown by Figures 8 and 9. The microstructure of the as–cast superalloy consists of significant dendritic segregation

C	Ni	Co	Ti	Cr	Al	W	Mo	Fe	Mn
0.13 ÷0.2	Bal.	4.0 ÷ 5.5	2.5 ÷ 3.2	9.5 ÷ 12	5.0 ÷ 6.0	4.5 ÷ 5.5	3.5 ÷ 4.8	2	0.4

Adulterants			
P	**S**	**Pb**	**Bi**
0.015	0.015	0.001	0.0005

Table 1. Experimental alloy's chemical composition.

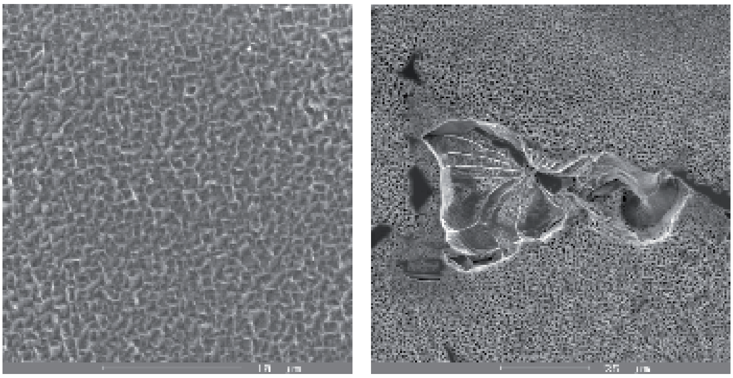
caused by chemical heterogeneity (Fig. 8a) and particles of primary MC and secondary $M_{23}C_6$ carbides (Fig. 8b). Primary carbides MC (where M is (Ti, Mo and W)) are presented as block-shaped particles, mainly inside grains. Secondary carbides are presented by "Chinese" script-shaped particles on grain boundaries.



a) dendritic segregation b) MC and $M_{23}C_6$ carbides

Fig. 8. Microstructure of as–cast Ni–base superalloy ŽS6K, Beraha III.

However, the microstructure also contains a solid solution of elements in the nickel matrix – the so-called γ-phase (Ni (Cr, Co and Fe)) and strengthening-phase, which is a product of artificial age–hardening and has a significant influence on mechanical properties and creep rupture life – so-called γ′-phase (gamma prime, Ni_3 (Al and Ti)), Fig. 9a. Of course, both of these phases - γ (gamma) and γ′ (gamma prime) - create an eutectic γ/γ′, Fig. 9b.



a) matrix and γ′ phases b) γ/γ′ eutectic

Fig. 9. Ni–base superalloy ŽS6K microstructure, as–cast.

3. Experimental results and discussion

3.1 Carbide evaluation

Polycrystalline and columnar grain alloys contain carbon additions to help improve grain–boundary strength and ductility. While the addition of carbon is beneficial to grain boundary ductility, the large carbides that form can adversely affect fatigue life. Both low- and high-cycle fatigue-cracking have been observed to initiate with the large (lengths greater than 0.005 mm) carbides presented in these alloys. When polycrystalline alloys were cast in a single crystal form, it was determined that carbides did not impart any beneficial strengthening effects in the absence of grain boundaries, and thus could be eliminated. Producing essentially carbon–free single crystal alloys led to significant improvements in fatigue life as large carbide colonies were no-longer present to initiate fatigue cracks (Cetel, A. D. & Duhl, D. N. 1988).

The first characteristic were carbide size and its distribution evaluated. Specimens made of the ŽS6K superalloy were compared at the starting stage (non-heat-treated, as-cast) after 800°C/10 hrs and 800°C/15 hrs. The cooling rate depends on the cooling medium; in our case these were air, oil and water. The results for the ratio of carbide particles in the observed area are in Figure 10 and the results on the average carbide size are in Figure 11.

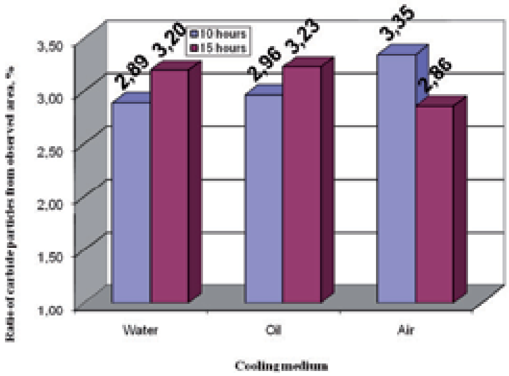


Fig. 10. The ratio of carbide particles from the observed area.

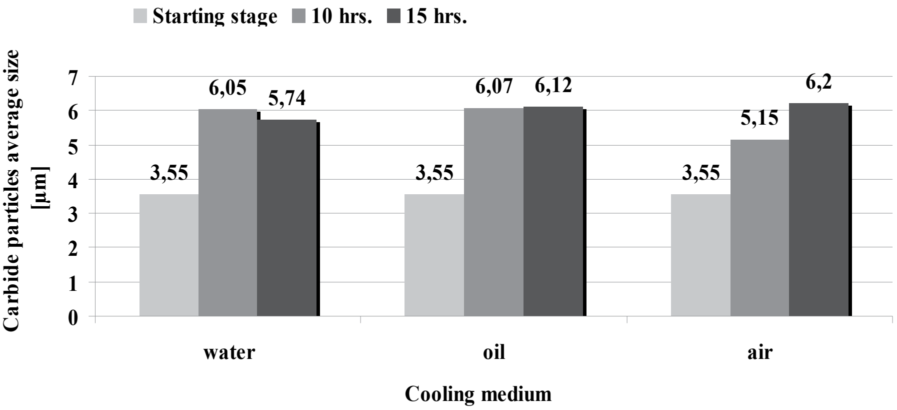


Fig. 11. Average carbide size [μm].

From the relations presented (Figure 11) it is obvious that the holding time on various temperatures for annealing and cooling in selected mediums does not have a significant influence on carbide particle size. More significant, the influence on the ratio of carbide particles has a cooling rate (Figure 10). With increasing speed of cooling and a longer holding time on the annealing temperature, the carbide particles' ratio decreases.

Generally, we can suppose that carbide particles are partially dissolved with the temperature of annealing and elements, which are consider as an carbide creators (in this case mainly Ti) have create a new particles of γ′ phase. This phenomenon has an influence on decreasing the segregated carbide percentage ratio. With an increase of the cooling rate (water, oil), an amount of the γ′-phase decreases and the carbides percentage ratio is higher. At slow cooling and a longer time of holding is higher amount of γ′ segregate and, therefore, the ratio of carbides decreases. It is all happen according to scheme:

$$MC + \gamma \rightarrow M_{23}C_6 + \gamma'$$

The microstructures which are equivalent to these evaluations are in Figures 12 and 13. For carbide evaluation, etching is not necessary. All of the micrographs are non-etched.

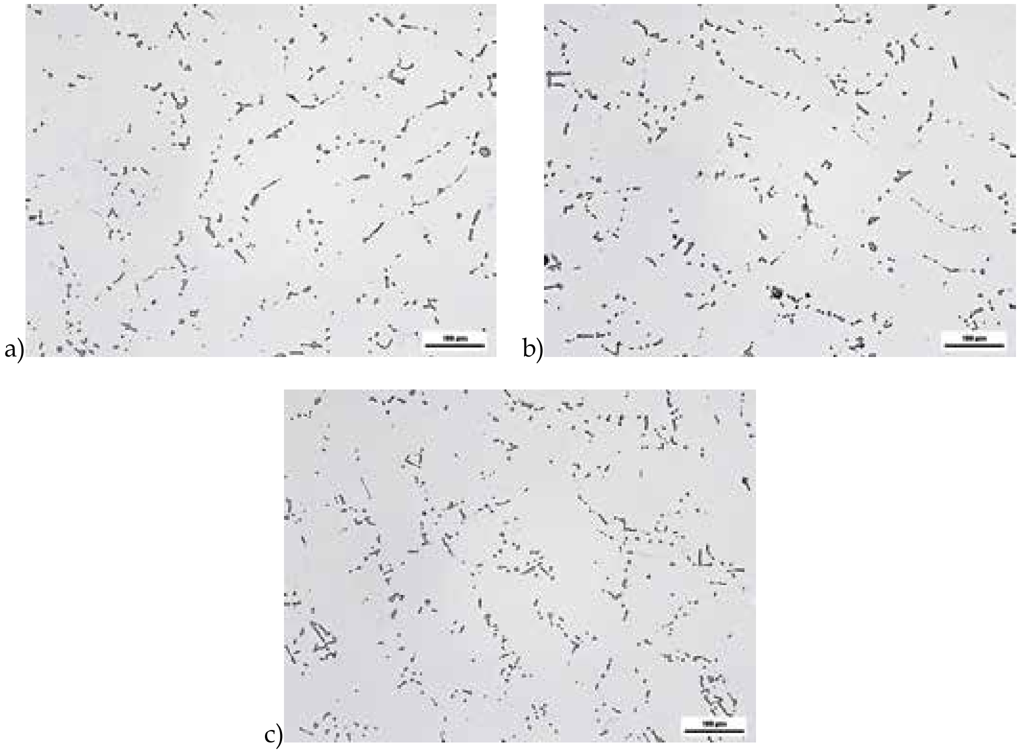


Fig. 12. Microstructure of ŽS6K, carbides ratio after 800°C annealing/10 hrs: a) water cooling; b) oil cooling; c) air cooling.

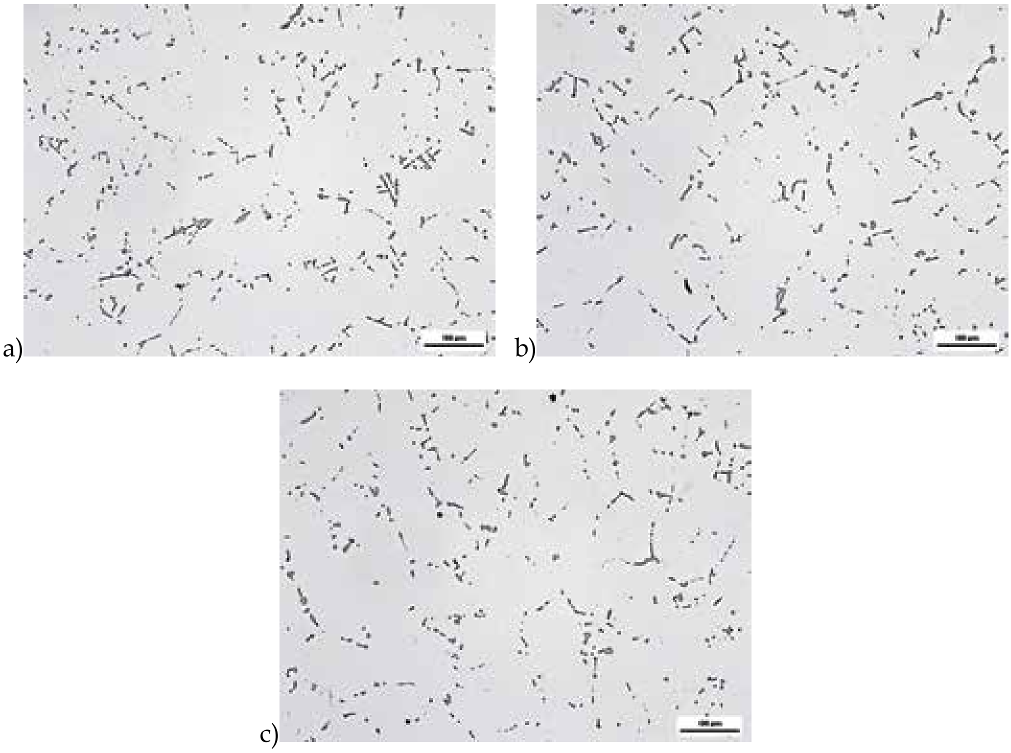


Fig. 13. Microstructure of ŽS6K, carbides ratio after 800°C annealing/15 hrs: a) water cooling; b) oil cooling; c) air cooling.

3.2 Evaluation of secondary dendrite arm-spacing

The second characteristic which is evaluated is dendrite arm-spacing. In this evaluation, two different approaches were taken. For the first evaluation, non-heat treated ŽS6K specimens were used and compared with loading at 800°C/10(15) hrs. The results of these first evaluations can be seen in Table 2 and Figures 14, 15 and 16. The second evaluation was performed on ŽS6K turbine blades used in the DV-2 (LPT – Low Pressure Turbine and HPT – High Pressure Turbine) aero jet engine at the starting stage (basic heat treatment) and after an engine exposition (at real working temperatures) for 600, 1000, 1500 and 2000 hours. Again, the results are in Table 3 and the microstructures are in Figure 17.

Secondary dendrite arm spacing [μm]			
ŽS6K – starting stage		185.19	
Cooling medium			
	Water	**Oil**	**Air**
ŽS6K/10hrs.	126.58	131.58	138.89
ŽS6K/15hrs.	113.64	131.58	156.25

Table 2. Results from secondary dendrite arm-spacing for a non-heat treated ŽS6K alloy

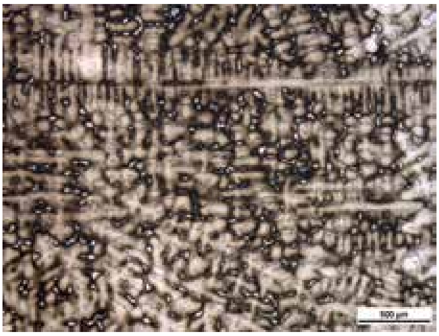


Fig. 14. Dendritic segregation of ŽS6K, starting stage, Marble etchant.

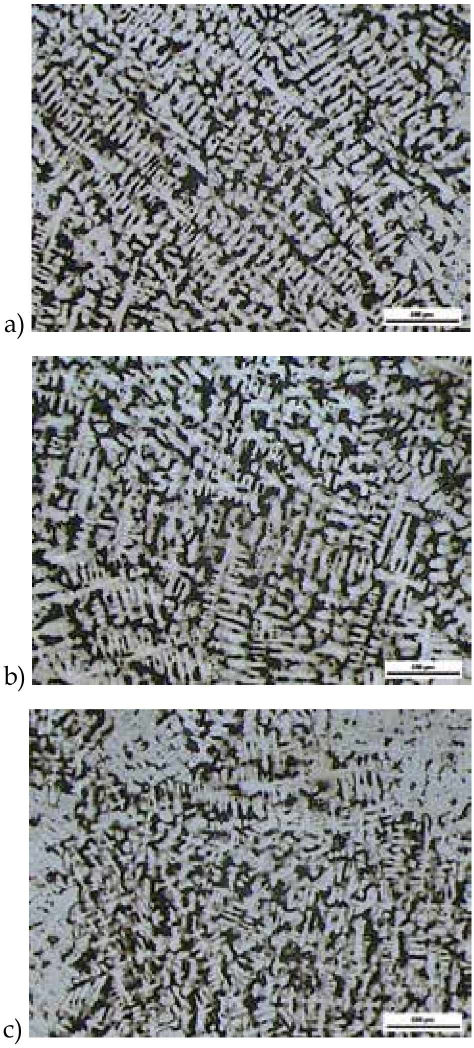
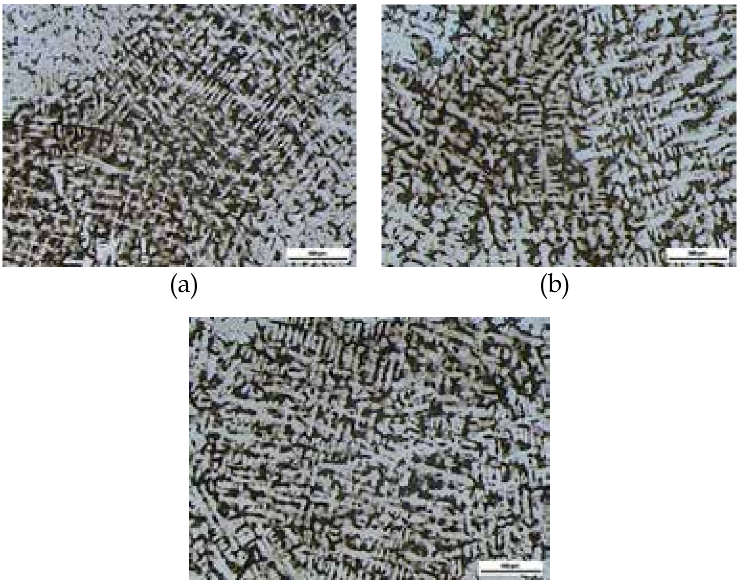


Fig. 15. Dendritic segregation of ŽS6K, 800°C/10 hrs: a) water cooling; b) oil cooling,; c) air cooling, Marble etchant.

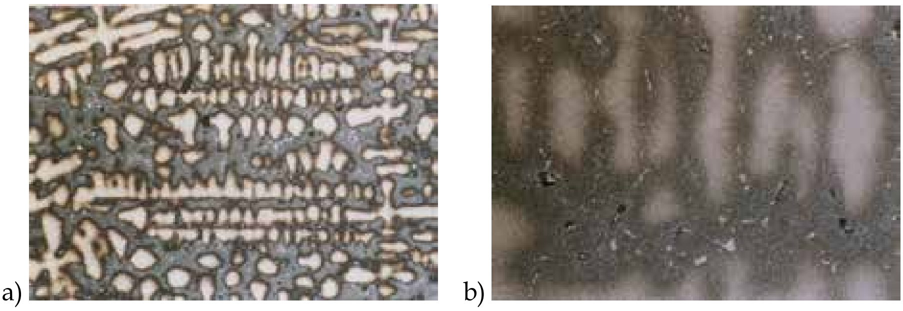


(a) (b) (c)

Fig. 16. Dendritic segregation of ŽS6K, 800°C/15 hrs.: a) water cooling, b) oil cooling, c) air cooling, Marble etchant.

Type of blade	Secondary dendrite arm-spacing [μm]
Blade of 1°LPT – starting stage	24.38
Blade of HPT - 600 hrs.	24.78
Blade of HPT - 1000 hrs.	27.98
Blade of HPT - 1500 hrs.	48.73
Blade of HPT - 2000 hrs.	66.66

Table 3. Results from secondary dendrite arm spacing for real turbine blades, heat-treated ŽS6K alloy.



a) b)

Fig. 17. Dendritic segregation of ŽS6K turbine blades: a) 1°LPT – starting stage; b) HPT – after 1500 hrs of work, Marble etchant.

The cast materials are characterised by dendritic segregation, which is caused by chemical heterogeneity. With the influence of holding at an annealing temperature, chemical heterogeneity decreases. This means that the distance between secondary dendrite arms increases (the dendrites are growing). From the results mentioned above (Table 2), it is clear to see that with a higher cooling rate comes a slowing of the diffusion processes and the dendrite arm-spacing decreases in comparison with the starting stage (Figure 14). All of these changes are also obvious in Figures 15 and 16. The ŽS6K dendrite arm-spacing increases in relation to the annealing time, with an annealing temperature and cooling medium of between 113.64 and 156.25 μm.

The same phenomena can be observed with heat-treated turbine blades after various working times. Of course, the secondary dendrite arm-spacing is smaller, but again it has a tendency to growth. So, this confirms the results from Table 2: that a longer time of exposure has a significant influence on dendrite and grain size.

3.3 Evaluation of γ′ morphology

Since the advanced high-strength nickel–base alloys owe their exceptional high temperature properties to the high volume fraction of the ordered γ′-phase that they contain, it should not be surprising that control of precipitate distribution and morphology can profoundly affect their properties. The post-casting processing of these alloys - especially solution heat treatment - can radically affect microstructure.

The high-strength alloys typically contain about 55 ÷ 75 % of the γ′ precipitates which, in the cast condition, are coarse (0.4 ÷ 1.0 μm) and irregularly-shaped cuboid particles (see Figure 9).

The evaluation of the γ′-phase is also divided into two parts, just as the dendrite evaluation was. Firstly, the γ′-phase was evaluated on the cast stage, and secondly on turbine blades. The characteristics of γ'-phase morphology were also measured using the coherent testing grid methods. As was mentioned above, the number and volume of the γ'-phase have a significant influence on the mechanical properties of this alloy, especially on creep rupture life. The average satisfactory size of the γ'-phase is about 0.35–0.45 μm (Figure 18) and also the carbide size should not exceed a size of 5 μm because of fatigue crack initiation (M. J. Donachie & S. J. Donachie 2002). Another risk in using high temperature loading (or annealing) is the creation of TCP phases - such σ-phase or Laves-phase - within the temperature range of 750 °C–800 °C. The results of first evaluation are in Table 4. The microstructures related to this evaluation are in Figures 19 and 20.

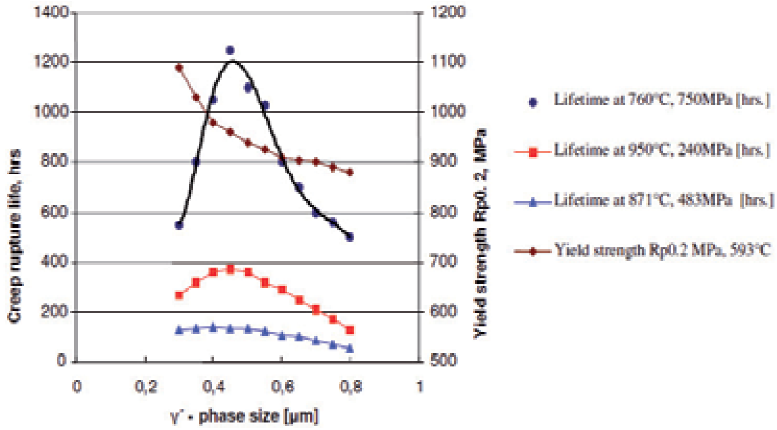


Fig. 18. Influence of γ'-phase size on the lifetime and mechanical properties of Ni superalloy.

Cooling medium	Number of γ' - phase N [μm^{-2}]	Volume of γ' - phase V [%]	Average size of γ' - phase u [μm]
Start. stage	2.47	39.4	0.61
10h water	1.95	56.2	0.54
10h oil	1.60	63	0.63
10h air	1.50	72.4	0.69
15h water	1.90	66.8	0.59
15h oil	1.59	71.8	0.67
15h air	1.49	76.6	0.72

Table 4. Results from γ'-phase evaluation at the cast stage at 800°C/10 (15) hrs.

With exposure for 10 hours at an annealing temperature, the volume of γ'-phase was increased by about 16.8–33% when compared with the starting stage (Figure 19). The significant increase of the γ'-phase was observed at a holding time of 15 hours (Figure 20), and cooling on air, where volume of γ'-phase is 76.6 %.

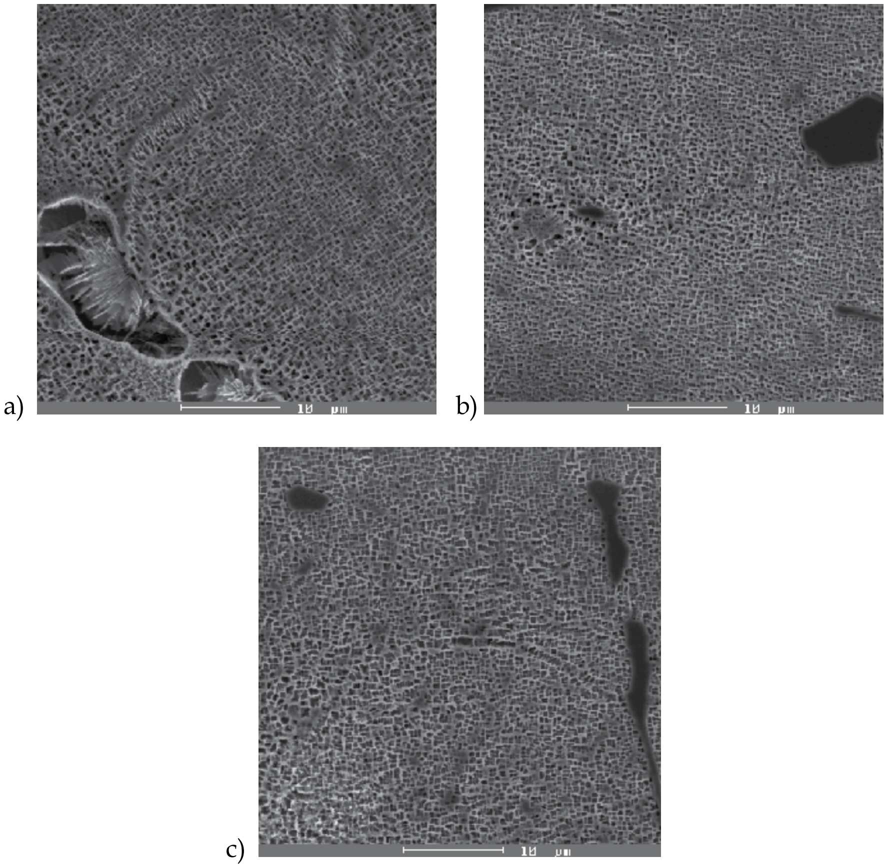


Fig. 19. Morphology of γ'-phase after 800°C/10 hrs: a) air cooling; b) oil cooling; c) water cooling, Marble etchant, SEM.

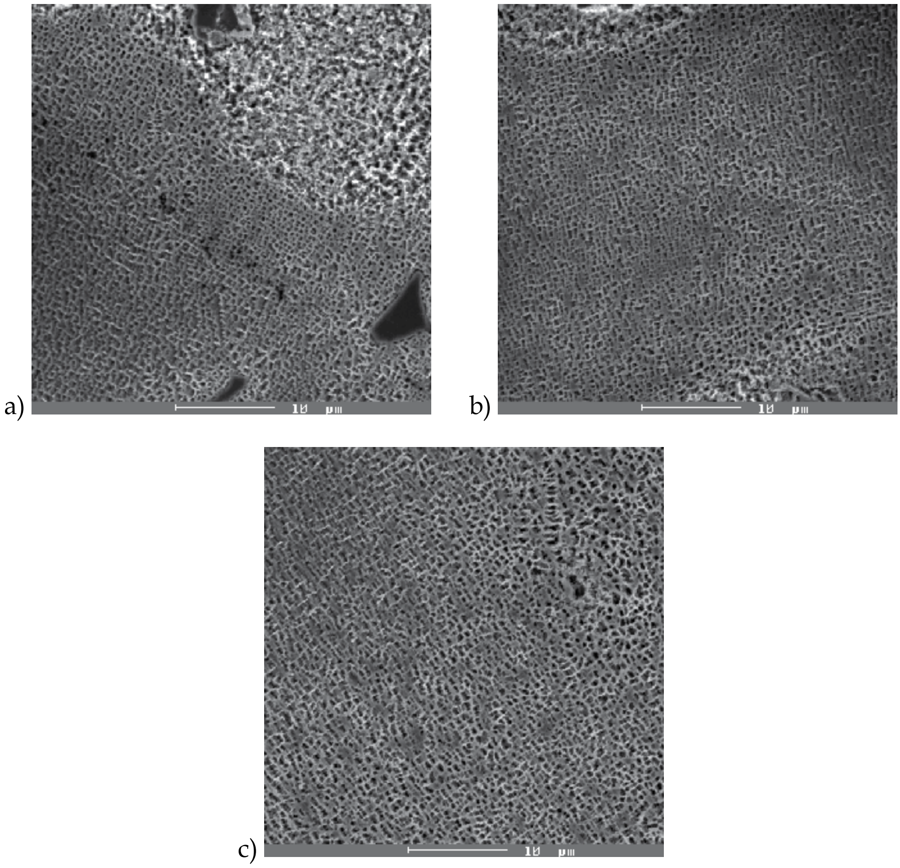


Fig. 20. Morphology of γ′-phase after 800°C/15 hrs: a) air cooling; b) oil cooling; c) water cooling, Marble etchant, SEM.

The highly alloyed nickel–base alloys solidify dendritically and, due to the effects of chemical segregation across the dendrites, a higher concentration of the γ′-phase forms elements such as aluminium and titanium which are more present in the inter-dendritic areas than in the dendrite core. This results in the γ′ solvus (the temperature at which γ′ first precipitates upon cooling) being lower in the core region than the inter-dendritically region.

Varying the cooling rate from the solution heat treatment temperature can significantly affect the γ′ particle size, as rapid rates do not allow sufficient time for the particles to coarsen as they precipitate upon cooling below the γ′ solvus temperature. Increasing the cooling rate of the solution heat treatment temperature from 30 to 120°C/minute results in an average particle size refinement of more than 30% (Figure 21) (Cetel, A. D. & Duhl, D. N. 1988).

By controlling both the solution heat treatment and the cooling rate, both the volume fraction of the fine particles as well as their size can be controlled. Heat treating an alloy close to its γ′ solvus temperature and completely dissolving its coarse γ′ particles can produce consistently high-elevated temperature creep–rupture strength.

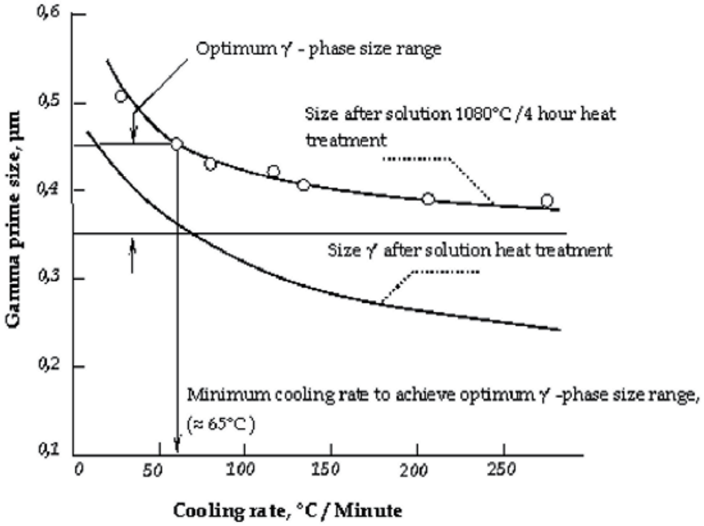


Fig. 21. Optimum γ′ size achieved by rapid cooling of the solution temperature combined with post-solution heat treatment.

Work performed by (Nathal et al. 1987) indicates that the optimum γ′-phase size for an alloy is dependent on the lattice mismatch between the γ- and γ′-phases (Figure 22), which is composition dependent.

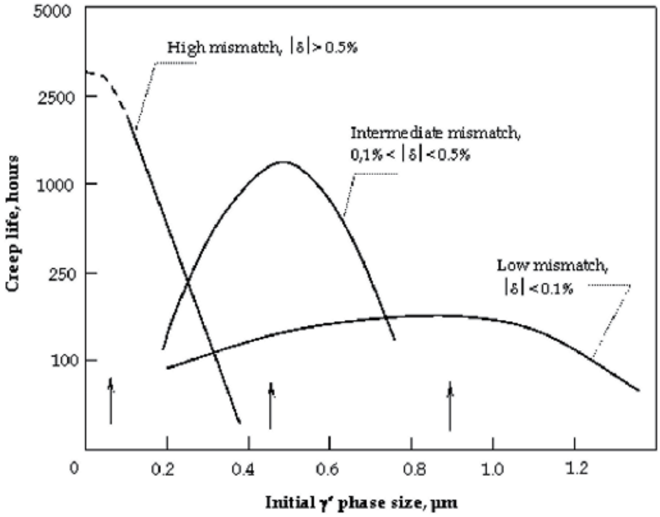


Fig. 22. The optimum γ′-phase size to maximize creep strength is dependent on mismatch between the γ- and γ′-phases (Nathal et al. 1987).

The second evaluation of the γ′-phase was provided on heat-treated turbine blades of a DV-2 aero jet engine after various working times. The results obtained are shown by Table 5. For the evaluation a coherent testing grid was used - the same procedure as in the first evaluation. The microstructures related to this evaluation are shown in Figures 23 and 24.

Time of work [hours]	Number of γ'-phase N [μm^{-2}]	Volume of γ'-phase V [%]	Average size of γ'-phase u [μm^2]
0	0.98542	67.2	0.6819
600	1.1242	67.6	0.60131
1000	1.1004	59	0.53615
1500	0.81938	57.4	0.7005
2000	0.6968	40.6	0.5826

Table 5. Results from the γ'-phase evaluation on heat-treated turbine blades at various working times.

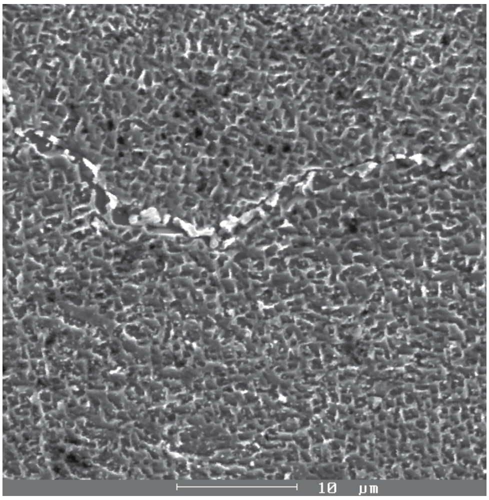
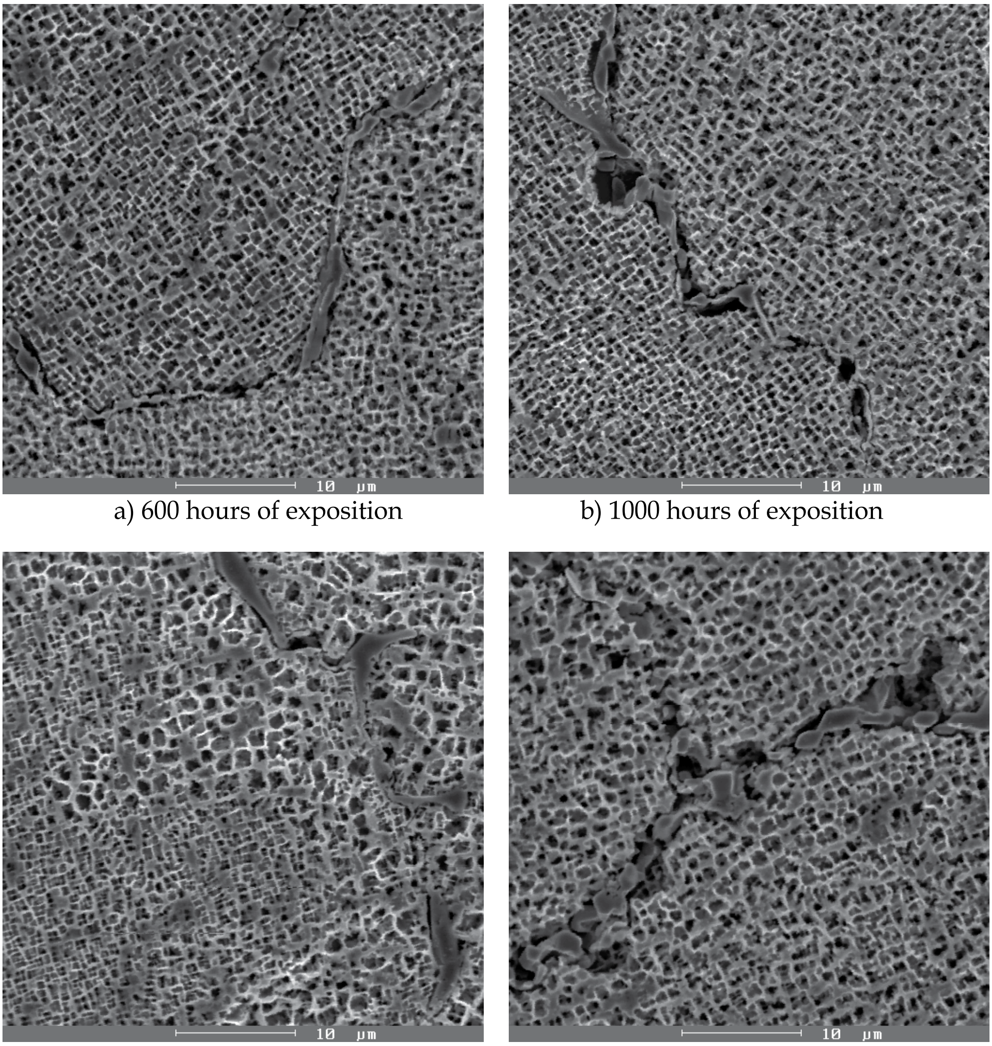


Fig. 23. Morphology of the γ'-phase heat-treated turbine blade, starting stage (0 hours), HCl + H_2O_2 etchant.

The morphology of the γ'-phase at the starting stage is cuboid and distributed equally in the base γ matrix (Figure 23.). With an increase of the hours of work at a temperature of up to 750°C, the γ'-phase morphology changes. The particles of the γ'-phase gradually coarsen (time of work to 1000 hours, Figure 24 a, b), which confirms the results of a number of γ'-phase evaluations "N" (see Table 5). A decrease of this value at a longer duration of work (1500 and 2000 hours) is caused by reprecipitation of new, fine γ'-phase particles in the area between the primal γ'-phase (Figure 24 c, d). From the results in Table 5, it is obvious that the γ'-phase of the ŽS6K alloy coarsen uniformly and increase its volume ratio in the structure after up to 1000 hours of exposition (regular work of a jet engine). However, after longer durations of work (1500 or 2000 hours) there occurs the reprecipitation of new, fine particles of the γ'-phase in the free space of matrix and which has caused structural heterogeneity.

In terms of structure degradation and the prediction of the life time of turbine blades - as well as the jet engine itself – and according to the results in Table 5, after up to 1000 hours of exposition the structure (with "N" = 1.1004, "V" = 59 and average size "u" = 0.53615) is at the "edge" of use because of its mechanical properties, as shown by Figure 18. However, the γ'-phase size is not the only parameter influencing the life time. In addition, the number "N" and volume "V" is important from the point of view of dislocation hardening. When "N" and "V" are smaller, this means that the distances between single particles are greater and that fact causes a decrease of the dislocation hardening effect. On the other hand, $M_{23}C_6$ carbides form a carbide net on the grain boundary which also decreases the creep rupture life by developing brittle grain boundaries. For a comparison of the increasing distance between γ′-phase particles see Figure 25.



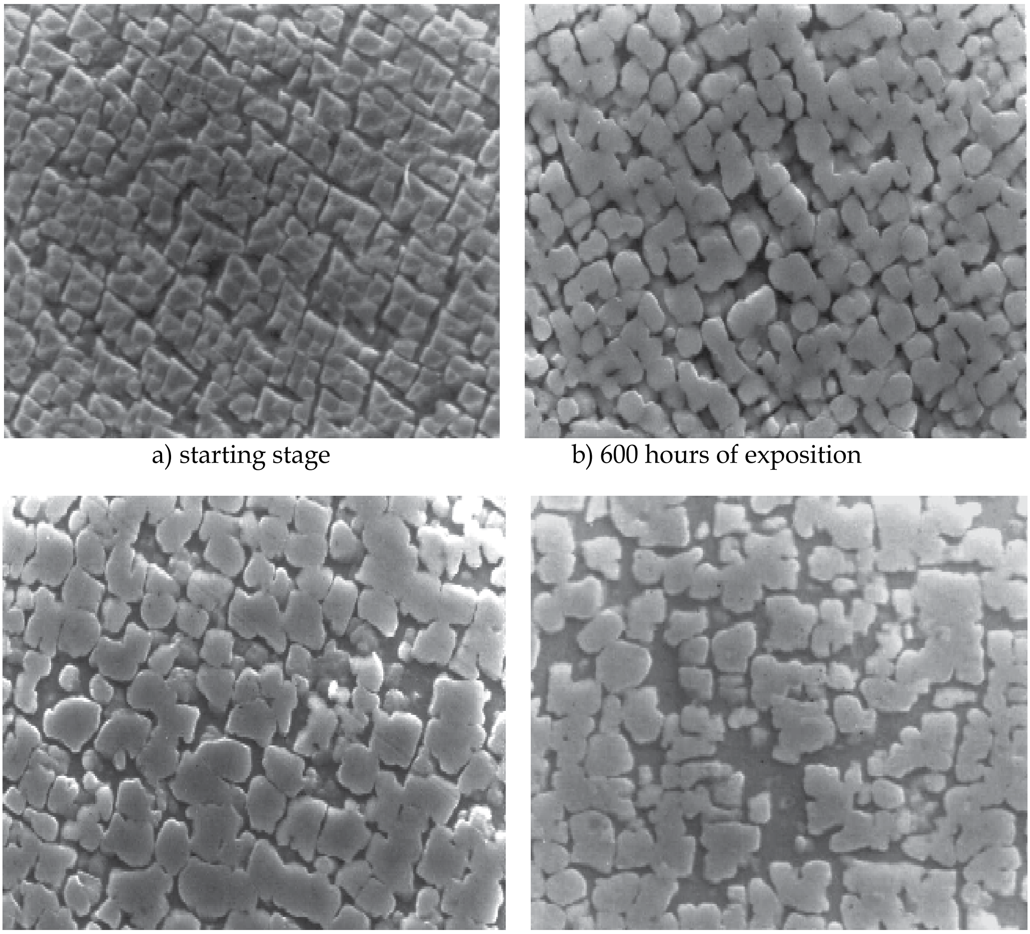
a) 600 hours of exposition

b) 1000 hours of exposition

c) 1500 hours of exposition

d) 2000 hours of exposition

Fig. 24. Morphology of the γ′-phase, heat treated turbine blade made of the ŽS6K alloy, after various exposition, HCl + H_2O_2 etchant.



a) starting stage

b) 600 hours of exposition

c) 1500 hours of exposition

d) 2000 hours of exposition

Fig. 25. Detail of the ŽS6K alloy's γ'-phase showing the increasing distance between γ' particles as an affect of working exposition - at normal working loading of a jet engine - which has a significant influence on the dislocation hardening effect.

3.4 Evaluation of the Al–Si protective layer

To improve the lifetime of turbine blades made from the ŽS6K alloy against a hot corrosion environment, an alitise Al–Si protective layer is used. What is important about this kind of layer is that it does not improve the high temperature properties of the base alloy but only its hot corrosion resistance.

An alitise layer is used for the protection of HPT blades (Figure 26) and only 1° of LPT blades, which means that Al-Si suspension is applied on to the surface of the blades. Silicon is added due to its ability to increase resistance to corrosion in sulphide and sea environments. Generally, an alitise layer AS-2 type is used for corrosion protection of aero gas turbine parts which work at temperatures of up to 950 °C; type AS-1 up to a temperature of 1100 °C (DV–2–I–62: Company standard). The standard procedure of applying Al–Si protective coating is in Table 6.

Heat – treatment	Conditions
Homogenization annealing	In vacuum, temperature 1225 °C, holding 4 hrs, cooling with argon to 900 °C per 10 min.
Alitise AS2	1. Spraying of AS2 layer (AS2 – koloxylin solution 350 ml, Al – powder 112 g, Si – powder 112 g)
	2. Diffusion annealing temperature 1000 °C, 3 hrs, slowly cooled in retort

Table 6. Steps for protective Al–Si coating as applied on to a ŽS6K turbine blade.

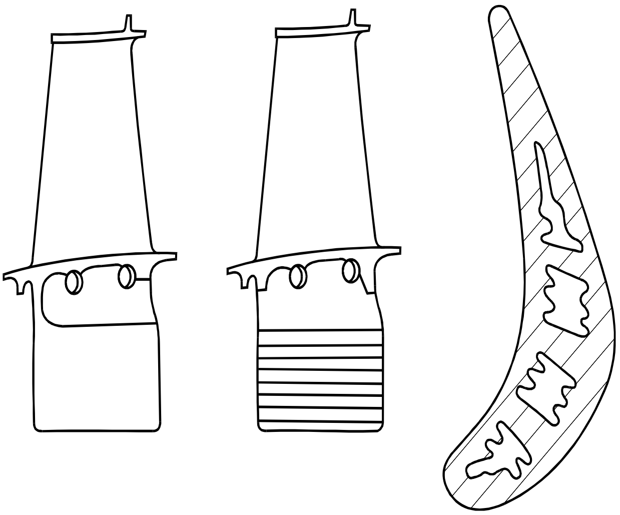


Fig. 26. A high pressure turbine blade of a DV–2 aero jet engine, left-side, right-side and cross-section with cooling chambers.

The Al–Si layer consists of two layers at the starting stage. The upper layer (the aluminium-rich layer) is created by aluminides - a complex compound of Si, Cr and Mo - and by carbides. The lower part of layer (the silicon rich layer) is created mainly by silicon and titanium carbides and the γ matrix. The average thickness of the layer is 0.04 mm. According to the evaluation by metallography, the alitise layer is equally distributed across the whole blade surface at the starting stage (Figure 27 a, b).

In cases of overheating (here, at 1000°C - the normal working temperature is 705°C ÷ 750°C) the alitise layer is significantly degraded (Figure 27 c, d). The layer is non-homogeneous, with a rough surface and in place of the flow edge in the area of the flap pantile is a layer which is evenly broken (Figure 27d). Layer degradation is connected with the diffusion of elementary elements - such as Cr, Ti, Ni and Al - from the base material into the surface area (Table 7.). Where Cr and Ti creates carbides, Ni and Al form fine γ′ particles and Al as itself also creates NiAl (β–phase) and Al_2O_3 oxides on the surface of the layer. With decreasing of the layer's heat resistance, the base material is impoverished, which leads to the growth of γ′ particles and decreasing of its volume.

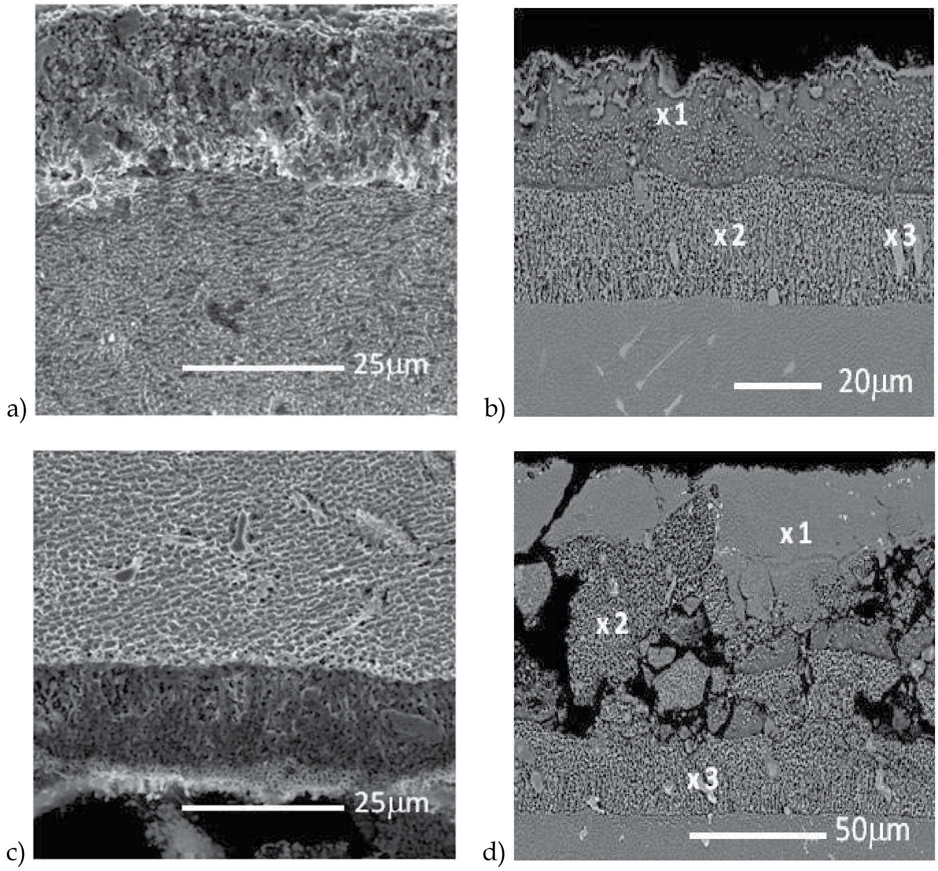


Fig. 27. Alitise layer a, b) starting stage; c, d) after overheating at 1000°C, SEM.

Sample	Marked spots	AlK	SiK	MoK	TiK	CrK	CoK	NiK	Wk
Starting stage, Fig. 27b	1	19.7	2.21	0.23	0.83	6.54	4.14	65.5	0.34
	2	15.01	3.09	2.56	2.28	5.16	4.06	65	2.27
	3	1.57	7.57	17.51	5.82	13.53	3.01	28.45	21.94
Overheating at 1000°C, Fig. 27d	1	9.04	5.94	7.42	0.92	9.25	4.15	52.6	10.6
	2	18.2	3.55	3.56	0.73	6.18	3.81	58.2	5.7
	3	-	9.54	13.5	9.5	11.3	3.25	33.1	19.6

Table 7. Spot analysis of selected particles. The marked spots (in wt%) correspond with Figure 27b, d.

The alitise layer on the blades which have worked at regular conditions is also degraded, which is represented by the changing of the layer thickness and the surface relief. Changes in layer thickness are caused by heterogeneity of the temperature field along the blade and the abrasive and erosive effect of gases and exhaust gases. The level of layer degradation varies, depending on the area of blade. From a metallographic point of view, the highest degradation is in the flap pantile region close to the flow edge, in the case of blades after

1500 and 2000 hours of work (Figure 28 c, d). In region close to the Si sub-layer, needle particles (probably a special form of Cr base carbides) are created which grow depending upon the time of work (compare Figures 27 a, b and 28). These needle particles start to form after 600 hours of loading, which means that after 600 (Figure 28 a, b) hours of work and aero jet engine should to be taken in for overhauling and the old alitise layer replaced by a new one. However, when it comes to the local overheating of the turbine blades, all of the degradation processes are much faster.

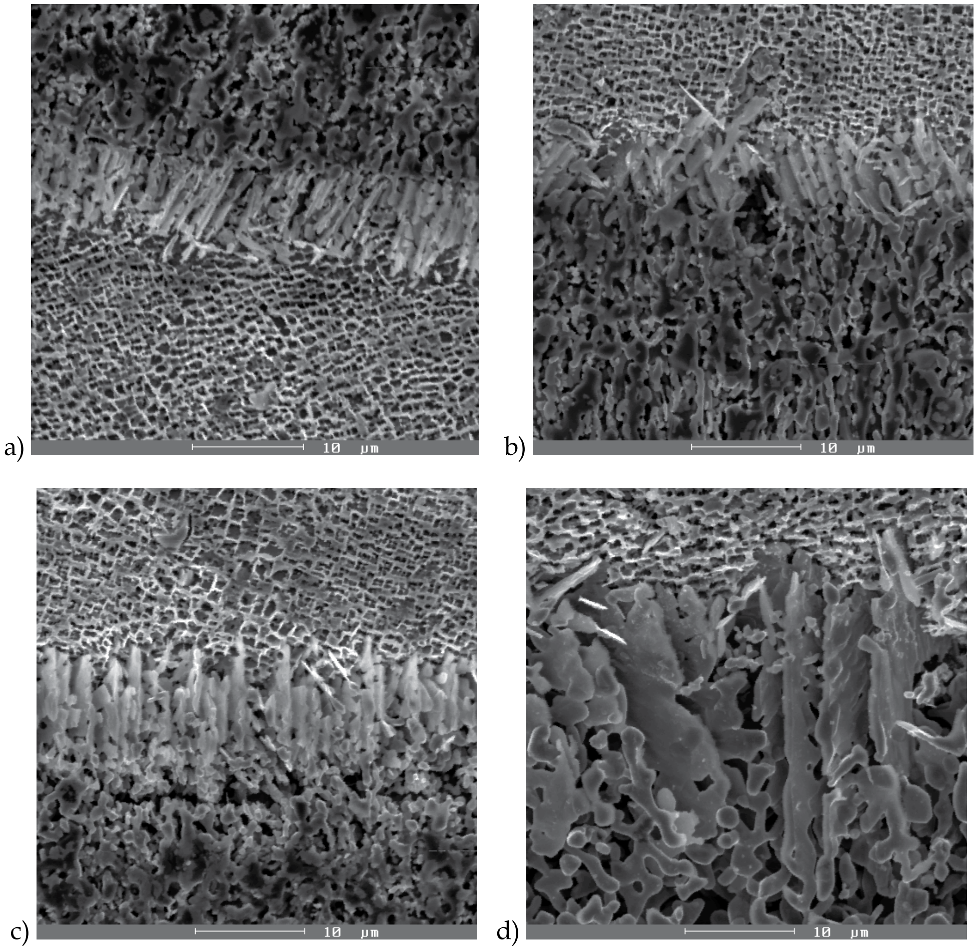


Fig. 28. Creation of needle particles in the region under the Si sub-layer: a) 600 hours; b) 1000 hours; c) 1500 hours; d) 2000 hours of regular work, SEM, Marble etchant.

4. Conclusion

As cast Ni–base ŽS6K superalloy was used as an experimental material, the structural characteristics were evaluated from the starting stage of the sample, after annealing at 800 °C/10 and 800 °C/15 hrs and after various working times in real jet engines with the use of the methods of quantitative metallography. The results are as follows:

- The structure of the samples is characterised by dendritic segregation. In dendritic areas, fine γ'-phase is segregated. In inter dendritic areas, eutectic cells γ/γ' and carbides are segregated.
- The holding time (10–15 hrs.) has a significant influence on the carbide particles' size. The size of the carbides is under a critical level for the initiation of fatigue crack only at the starting stage. An increase in the rate of cooling has a significant effect on the carbide particles' ratio.
- The chemical heterogeneity of the samples with a longer holding time decreases. This is a reason of the fact that there is sufficient time for the diffusion mechanism, which is confirmed by the measurement results of secondary dendrite arm-spacing.
- The volume of the γ'-phase with a longer holding time increases and the γ'-phase size grows. With a higher rate of cooling the γ' particles become finer.
- There was no evidence of the presence of TCP phase even at a high annealing temperature.
- Cooling rate also has an influence on the hardness. At a lower rate of cooling, the internal stresses are relaxed, which causes hardness to increase – a changing of the dislocation structure.

The cooling rates, represented by various cooling mediums, have a significant influence on the diffusion processes which are operating within the structure. These diffusion processes are the main mechanisms for the formation and segregation of carbide particles, the equalising of chemical heterogeneity (represented by dendrite arm-spacing) and segregation of the γ'-phase; they are also responsible for structural degradation of such alloys.

Air - as a cooling medium - provides sufficient time for the realisation of diffusion reactions and it leads to a decrease of chemical heterogeneity, which is presented by an increase of secondary dendrite arm-spacing. Also, this "slow" cooling rate has a positive effect on carbides' segregation and on the morphology, number and volume of the strength precipitate γ' (the precipitate has a greater diameter and its volume increases).

Water is the most intensive cooling medium, which breaks diffusion processes and which leads to an increase of carbide particles in the observed area; the precipitate γ' is smaller, increasing the hardness and at least also increasing the strength.

From a general point of view we can perform cooling in oil, which might be consider as a medium point between cooling in air and cooling in water.

For the turbine blades, which have been worked at normal loading and for various durations (600, 1000, 1500 and 2000 hours), the following results were achieved:

- The medium distance of secondary dendrite arms "d" grows in dependence on the time of work, caused by changes of the grain size of the γ-matrix.
- The gradual dissolving of primary carbides rests and the reprecipitation of secondary carbides on grain boundary. After longer durations of work (1000–2000 hours) it changes its chain morphology onto the carbide net, which has a significant influence on lowering the mechanical properties of the alloy.
- The inter-metallic phase-γ' was evaluated with the methods of quantitative metallography; this evaluation shows gradual morphology changes of the γ'-phase – coarsening, spheroidisation and reprecipitation.

- The alitise layer degradation was expressed by a changing thickness and needle-like Cr carbide segregation at the sub-layer region, which has a negative influence on the layer's lifetime. There is strong recommendation for overhauling after every 500 hours of regular work.

5. Acknowledgment

The authors acknowledge the financial support of the projects VEGA No. 1/0841/11 and No. 1/0460/11; KEGA No. 220-009ŽU-4/2010 and European Union - the Project "*Systematization of advanced technologies and knowledge transfer between industry and universities (ITMS 26110230004)*".

6. References

ASM. (1992). *ASM Handbook Volume 3: Alloy Phase Diagrams* (10th edition), ASM International, ISBN 0–871–70381–5, USA.

Cetel, A. D. & Duhl, D. N. (1988). Microstructure – Property Relationships In Advanced Nickel Base Superalloy Airfoil Castings, *2nd International SAMPE Metals Conference,* pp. 37–48, USA, August 2–4, 1988.

Donachie, M. J. & Donachie, S. J. (2002). *Superalloys – A technical Guide* (2nd edition), ASM International, ISBN 0–87170–749–7, USA.

Durand–Chare, M. (1997). *The Microstructure of Superalloys*, Gordon & Breach Science Publishers, ISBN 90–5699–097–7, Amsterdam, Netherlands.

DV–2–I–62: Company standard, Považské machine industry, Division of Aircraft Engine DV–2, Považská Bystrica, Slovakia, 1989.

MacSleyne, J. P. (2008). Moment invariants for two-dimensional and three-dimensional characterization of the morphology of gamma-prime precipitates in nickel-base superalloys, In: *Doctoral Thesis / Dissertation*, n.d., Available from: <http://www.grin.com/en/doc/263761/moment-invariants-for-two-dimensional-and-three-dimensional-characterization>

Nathal, M. V. (1987) *Met. Trans.*, Vol. 18 A, pp. 1961–1970.

Reed, R. C. (2006). *The superalloys: fundamentals and applications*, Cambridge University Press, ISBN 0–521–85904-2, New York, USA.

Sims, Ch. T., Stoloff, N. S. & Hagel, W. C. (1987). *Superalloys II* (2nd edition), Wiley-Interscience, ISBN 0–471–01147–9, USA.

Skočovský, P. & Vaško, A. (2007). *The quantitative evaluation of cast iron structure* (1st edition), EDIS, ISBN 978-80-8070-748-4, Žilina, Slovak Republic.

Tillová, E. & Panuškova, M. (2008). Effect of Solution Treatment on Intermetallic Phase's Morphology in AlSi9Cu3 Cast Alloy. *Mettalurgija/METABK*, No. 47, pp. 133-137, 1-4, ISSN 0543-5846.

Tillová, E., Chalupová, M., Hurtalová, L., Bonek, M., & Dobrzanski, L. A. (2011). Structural analysis of heat treated automotive cast alloy. *Journal of Achievements in Materials and Manufacturing Engineering/JAMME*, Vol. 47, No. 1, (July 2011), pp. 19-25, ISSN 1734-8412.

4

ALLVAC 718 Plus™ Superalloy for Aircraft Engine Applications

Melih Cemal Kushan[1], Sinem Cevik Uzgur[2],
Yagiz Uzunonat[3] and Fehmi Diltemiz[4]
[1]Eskisehir Osmangazi University
[2]Ondokuz Mayis University
[3]Anadolu University
[4]Air Supply and Maintenance Base
Turkey

1. Introduction

Innovations on the aerospace and aircraft industry have been throwing light upon building to future's engineering architecture at the today's globalization world where technology is the indispensable part of life. On the basis of aviation sector, the improvements of materials used in aircraft gas turbine engines which constitute 50 % of total aircraft weight must protect its actuality continuously. On the other hand utilization of super alloys in aerospace and defense industries can not be ignored because of excellent corrosion and oxidation resistance, high strength and long creep life at elevated temperatures.

Materials that can be used at the homologous temperature of 0.6 Tm and still remain stable to withstand severe mechanical stresses and strains in oxidizing environments are so-called superalloys, usually based on Ni, Fe or Co (Sims et al, 1987). Nickel-based superalloys are the exceptional group of superalloys with superior materials properties. Their excellent properties range from high temperature mechanical strength, toughness to resistance to degradation in oxidizing and corrosive environment. Therefore they are not only used in aerospace and aircraft industry, but also in ship, locomotive, petro- chemistry and nuclear reactor industries.

Inconel 718 is Ni-based, precipitation- hardening superalloy with Nb as a major hardening element, used for high temperature aerospace applications very widely in recent years (Yaman & Kushan, 1998). However, the metastability of the primary strengthening (γ'', gamma double prime) phase is typically unacceptable for applications above about 650°C. As a result, other more costly and difficult to process alloys, like Waspalloy, are used in such applications. Although Waspalloy is strengthened primarily by γ', it is still more susceptible to weld-related cracking than Inconel 718 (Otti et al, 2005). In these circumstances ALLVAC 718 Plus™ come to stage, which is strengthened with uniform cubic FCC inter metallic γ' phase, innovated by ATI ALLVAC Company very recently. In recent years it has been becoming widespread dramatically for using of disc material in aerospace gas turbine engine parts. The most important reason of this is the high yield and ultimate tensile strength and very good corrosion and oxidation resistance of material together with excellent creep resistance at elevated temperatures. Fig. 1 shows that where wrought alloy 718Plus can be used as a disc material for high pressure (HP) compressor as well as for high pressure (HP) turbine discs (Bond & Kennedy, 2005).

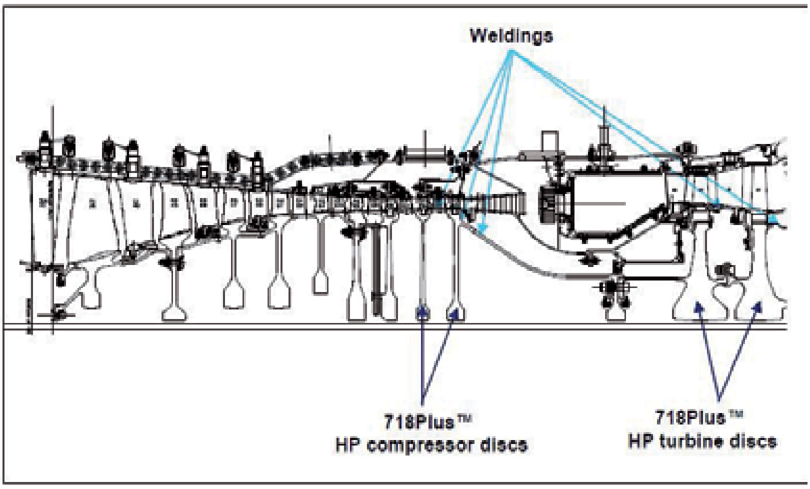


Fig. 1. Potential applications for alloy 718PlusTM in a future high pressure core section (Bond & Kennedy, 2005).

The newly innovated ALLVAC 718 Plus superalloy which is the last version of Inconel 718 has been proceeding in the way to become a material that aerospace and defense industries never replace of any other material with combining its good mechanical properties, easy machinability and low cost.

2. Gas turbine engines

Gas turbine engines, also known as jet engines, power most modern civilian and military aircraft. Fig. 2 shows some sections of this kind of an engine. The inlet (intake) directs outside air into the engine. The compressor (shown in a (a) part of Fig. 2) is situated at the exit of the inlet. In order to produce thrust, it is essential to compress the air before fuel is added. In an axial-flow compressor, the air flows in the direction of the shaft axis through alternate rows of stationary and rotating blades, called stators and rotors, respectively. Modern axial-flow compressors can increase the pressure 24 times in 15 stages, with each set of stators and rotors making up a stage. The compressors in most modern engines are divided into low-pressure and high-pressure sections which run off two different shafts. In the combustor, or burner (shown in a (b) part of Fig. 2), the compressed air is mixed with fuel and burned. Fuel is introduced through an array of spray nozzles that atomize it. An electric igniter is used to begin combustion. The combustor adds heat energy to the air stream and increases its temperature (up to about 1930°C), a process which is accompanied by a slight decrease in pressure (~ 1-2%). For best performances, the combustion temperature should be the maximum obtainable from the complete combustion of the oxygen and the fuel. However, turbine inlet temperatures currently cannot exceed about 1100°C because of material limits. Hence, only part of the compressed air is burned in the combustor; the remainder is used to cool the turbine.

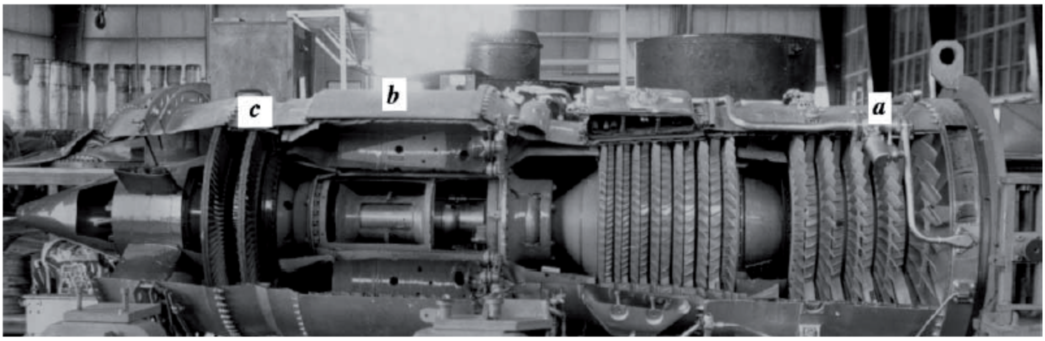


Fig. 2. Some basic sections of a gas turbine engine (Eliaz et al, 2002).

Leaving the combustor, the hot exhaust is passed through the turbine (shown in a (c) part of Fig. 2), in which the gases are partially expanded through alternate stator and rotor rows. Depending on the engine type, the turbine may consist of one or several stages. Like the compressor, the turbine is divided into low-pressure and high-pressure sections (shown in Fig. 3), the latter being closer to the combustor.

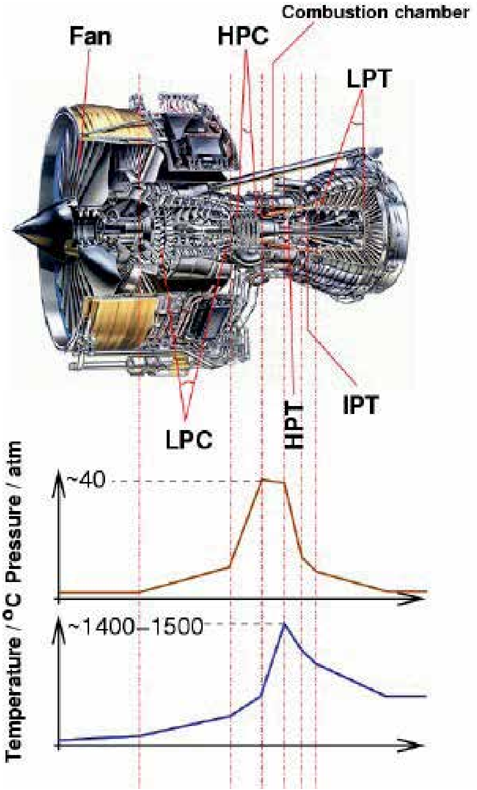


Fig. 3. The temperature and pressure profile in gas turbine (Carlos & Estrada, 2007).

The turbine provides the power to turn the compressor, to which it is connected via a central shaft, as well as the power for the fuel pump, generator, and other accessories. From thermodynamics, the turbine work per mass airflow is equal to the change in the specific enthalpy of the flow from the entrance to the exit of the turbine. This change is related to the temperatures at these points. The temperature at the entrance to the turbine can be as high as 1650°C, considerably above the melting point of the material from which the blades are made.

The gases, leaving the turbine at an intermediate pressure, are finally accelerated through a nozzle to reach the desired high jet-exit velocity. Because the exit velocity is greater than the free stream velocity, thrust is produced. The amount of thrust generated depends on the rate of mass flow through the engine and the leaving jet velocity, according to Newton's Second Law. Thus, the gas is accelerated to the rear, and the engine (as well as the aircraft) is accelerated in the opposite direction according to Newton's Third Law (Eliaz et al, 2002).

Modern gas turbines have the most advanced and sophisticated technology in all aspects; construction materials are not the exception due their extreme operating conditions. The most difficult and challenging point is the one located at the turbine inlet, because, several difficulties associated to it; like extreme temperature, high pressure, high rotational speed, vibration, small circulation area, and so on. These rush characteristics produce effects on the gas turbine components that are shown on the Table 1 (Carlos & Estrada, 2007).

	Oxidation	**Hot corrosion**	**Interdiffusion**	**Thermal Fatigue**
Aircraft	Severe	Moderate	Severe	Severe
Land-based Power Generator	Moderate	Severe	Moderate	Light
Marine Engines	Moderate	Severe	Light	Moderate

Table 1. Severity of the different surface related problems for gas turbine applications (Carlos & Estrada, 2007).

In order to overcome those barriers, gas turbine components are made using advanced materials and modern alloys (superalloys) that contains up to ten significant alloying elements.

3. Superalloys

These alloys have been developed for high-temperature service and include iron-, cobalt- and nickel based materials, although nowadays they are principally nickel based. These materials are widely used in aircraft and power-generation turbines, rocket engines, and other challenging environments, including nuclear power and chemical processing plants. The aero gas turbine was the impetus for the development of superalloys in the early 1940s, when conventional materials available at that time were insufficient for the demanding environment of the turbine. Therefore it can be said that "The development of superalloys made the modern gas turbine possible".

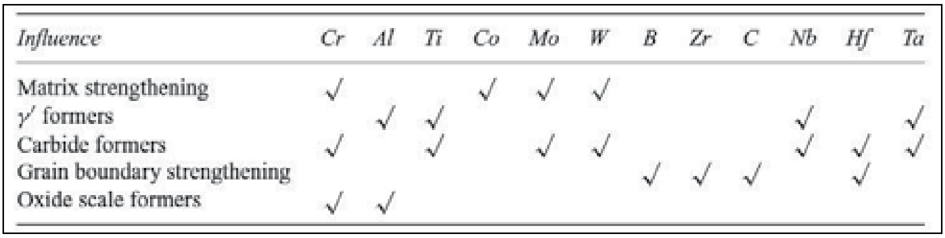
A major application of superalloys is in turbine materials, jet engines, both disc and blades. Initial disc alloys were *Inco 718* and *Inco 901* produced by conventional casting ingot, forged billet and forged disc route. These alloys were developed from austenitic steels, which are still used in industrial turbines, but were later replaced by *Waspaloy* and *Astroloy* as stress and temperature requirements increased. These alloys were turbine blade alloys with a suitably modified heat-treatment for discs. However, blade material is designed for creep, whereas disc material requires tensile strength coupled with low cycle fatigue life to cope with the stress changes in the flight cycle. To meet these requirements *Waspaloy* was thermomechanically processed (TMP) to give a fine-grain size and a 40% increase in tensile strength over the corresponding blade material, but at the expense of creep life. Similar improvements for discs have been produced in *Inco 901* by TMP. More highly-alloyed nickel-based discs suffer from excessive ingot segregation which makes grain size control difficult. Further development led to alloys produced by powder processing by gas atomization of a molten stream of metal in an inert argon atmosphere and consolidating the resultant powder by HIPing to near-net shape. Such products are limited in stress application because of inclusions in the powder and, hence, to realize the maximum advantage of this process it is necessary to produce 'superclean' material by electron beam or plasma melting.

Improvements in turbine materials were initially developed by increasing the volume fraction of γ´ in changing *Nimonic 80A* up to *Nimonic 115*. Unfortunately increasing the (Ti +Al) content lowers the melting point, thereby narrowing the forging range makes processing more difficult. Improved high-temperature oxidation and hot corrosion performance has led to the introduction of aluminide and overlay coatings and subsequently the development of *IN 738* and *IN 939* with much improved hot-corrosion resistance.

Further improvements in superalloys have depended on alternative manufacturing routes, particularly using modern casting technology like Vacuum casting (Smallman & Bishop, 1999).

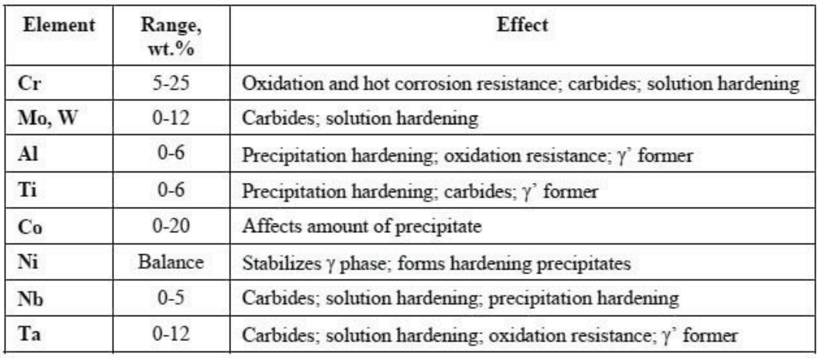
In these alloys γ' (Ni_3Al) and γ'' (Ni_3Nb) are the principal strengtheners by chemical and coherency strain hardening. The ordered γ'-Ni_3Al phase is an equilibrium second phase in both the binary Ni–Al and Ni–Cr–Al systems and a metastable phase in the Ni–Ti and Ni–Cr–Ti systems, with close matching of the γ' and the FCC matrix. The two phases have very similar lattice parameters and the coherency confers a very low coarsening rate on the precipitate so that the alloy overages extremely slowly even at 0.7*T*m. In alloys containing Nb, a metastable Ni_3Nb phase occurs but, although ordered and coherent, it is less stable than γ' at high temperatures (Smallman & Ngan, 2007).

In high-temperature service, the properties of the grain boundaries are as important as the strengthening by γ' within the grains. Grain boundary strengthening is produced mainly by precipitation of chromium and refractory metal carbides; small additions of Zr and B improve the morphology and stability of these carbides. Optimum properties are developed by multistage heat treatment; the intermediate stages produce the desired grain boundary microstructure of carbide particles enveloped in a film of γ' and the other stages produce two size ranges of γ' for the best combination of strength at both intermediate and high temperatures (Smallman & Ngan, 2007). Table 2 indicates the effect of the different alloying elements and Table 3 indicates the common ranges of main alloying additions and their effects on superalloy properties.



Influence	*Cr*	*Al*	*Ti*	*Co*	*Mo*	*W*	*B*	*Zr*	*C*	*Nb*	*Hf*	*Ta*
Matrix strengthening	√			√	√	√						
γ' formers		√	√							√		√
Carbide formers	√		√		√	√				√	√	√
Grain boundary strengthening							√	√	√		√	
Oxide scale formers	√	√										

Table 2. The effect of the different alloying elements (Smallman & Ngan, 2007).



Element	Range, wt.%	Effect
Cr	5-25	Oxidation and hot corrosion resistance; carbides; solution hardening
Mo, W	0-12	Carbides; solution hardening
Al	0-6	Precipitation hardening; oxidation resistance; γ' former
Ti	0-6	Precipitation hardening; carbides; γ' former
Co	0-20	Affects amount of precipitate
Ni	Balance	Stabilizes γ phase; forms hardening precipitates
Nb	0-5	Carbides; solution hardening; precipitation hardening
Ta	0-12	Carbides; solution hardening; oxidation resistance; γ' former

Table 3. Common Ranges of Main Alloying Additions and Their Effects on Superalloys.

The wide range of applications for superalloys has expanded many other areas since they were developed and now includes aircraft and land-based gas turbines, rocket engines, chemical, and petroleum plants. The performance of an industrial gas turbine engines depends strongly on service conditions and the environment in which it operates.

3.1 Iron-nickel-based superalloys

Iron-nickel base superalloys evolved from austenitic stainless steels and are based on the principle of combining both solid-solution hardening and precipitate forming elements. As a class, the iron nickel superalloys have useful strengths to approximately 650°C (1200°C). The austenitic matrix based on nickel and iron, with at least 25 wt % Ni needed to stabilize the FCC phase. Other alloying elements, such as Chromium partition primarily to austenite to provide solid-solution hardening. Most alloys contain 25 to 45 wt % Nickel. Chromium in the range of 15 to 28 wt% is added for oxidation resistance at elevated temperature, while 1 to 6 wt% Mo provides solid solution strengthening. The main elements that facilitate precipitation hardening are titanium, aluminum and niobium.

The strengthening precipitates are primarily γ' (Ni_3Al), η (Ni_3Ti), and γ″ (Ni_3Nb). Elements that partition to grain boundaries, such as Boron and Zirconium, suppress grain boundary creep, resulting in significant increases in rupture life. Boron in quantities of 0.003 to 0.03 wt% and, less frequently, small additions of zirconium are added to improve stress-rupture properties and hot workability. Zirconium also forms the MC carbide ZrC. Another MC carbide (NbC) is found in alloys that contain niobium such as Inconel 706 and Inconel 718. Vanadium is also added in small quantities to iron-nickel superalloys to improve both notch ductility at service temperatures and hot workability. Based on their composition and strengthening mechanisms, there are several groupings of iron-nickel superalloys (Campbell, 2008).

The most common precipitate is γ', typified by A-286, V-57 or Incoloy 901. Some alloys, typified by Inconel (IN)- 718, which precipitate γ″, were formerly classed as iron-nickel-base superalloys but now are considered to be nickel-base.

The most common type of iron-nickel-base superalloys is INCONEL 718 which is a precipitation- hardening alloy used for high-temperature applications. In particular, the reputation of wrought Inconel 718 for being relatively easy to weld is generally attributed to the sluggish precipitation kinetics of the tetragonal γ″ strengthening phase. Inconel 718 is a relatively recent alloy as its industrial use started in 1965. It is a precipitation hardenable alloy, containing significant amounts of Fe, Nb and Mo. Minor contents of Al and Ti are also present. Inconel 718 combines good corrosion and high mechanical properties with and excellent weldability. It is employed in gas turbines, rocket engines, turbine blades, and in extrusion dies and containers.

Ni and Cr contribute to the corrosion resistance of this material. They crystallize as a γ phase (face centred cubic). Nb is added to form hardening precipitates γ″ (a metastable inter metallic compound Ni_3Nb, centred tetragonal crystal). Ti and Al are added to precipitate in the form of intermetallic γ' ($Ni_3(Ti,Al)$, simple cubic crystal). They have a lower hardening effect than particles. C is also added to precipitate in the form of MC carbides (M = Ti or Nb). In this case the C content must be low enough to allow Nb and Ti precipitation in the form of γ' and γ″ particles. Mo is also frequent in Inconel 718 in order to increase the mechanical resistance by solid solution hardening. Finally, a β phase (intermetallic Ni_3Nb), (sometimes called δ phase) can also appear. It is an equilibrium particle with orthorhombic structure. All theses particles can precipitate along the grain boundaries of the γ matrix increasing the intergranular flow resistance of the present alloy. A typical precipitation time temperature (PTT) diagram for this alloy is shown in Fig. 4 (Thomas et al, 2006).

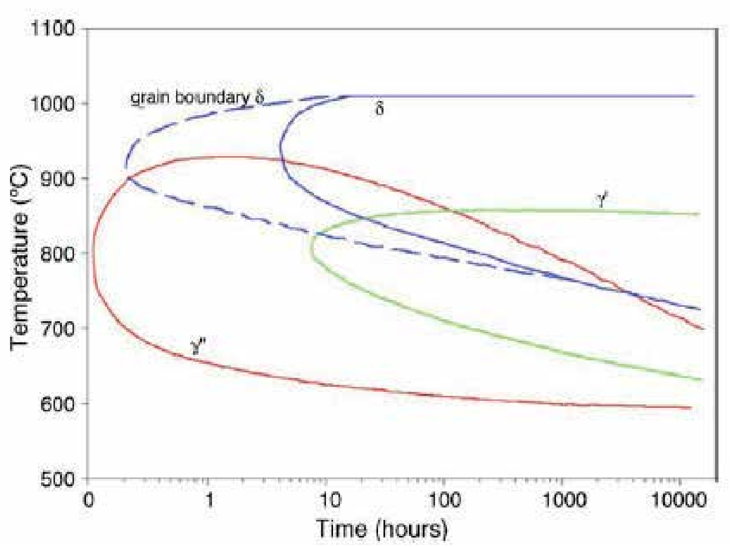


Fig. 4. PTT diagram of different phases in Inconel 718 (Thomas et al, 2006).

3.2 Nickel-based Superalloys

Nickel-based superalloys are an unusual class of metallic materials with an exceptional combination of high temperature strength, toughness, and resistance to degradation in corrosive or oxidizing environments. The nickel-based alloys show a wider range of application than any other class of alloys.

The austenitic stainless steels were developed and utilized early in the 1900s, whereas the development of the nickel-based alloys did not begin until about 1930. In aerospace applications nickel-based superalloys are used widely as components of jet engine turbines. Therefore important position of super-alloys in this area is manifested by the fact that they represent at present more than 50 % of mass of advanced aircraft engines. Extensive use of super-alloys in turbines, supported by the fact that thermo-dynamic efficiency of turbines increases with increasing temperatures at the turbine inlet, became partial reason of the effort aimed at increasing of the maximum service temperature of high-alloyed alloys (Jonsta et al, 2007). Therefore in gas turbine applications alloys with good stability and very low crack-growth rates that are readily inspectable by nondestructive means are desired. Fuel efficiency and emissions are also key commercial and environmental drivers impacting turbine-engine materials. To meet these demands, modern nickel-based alloys offer an efficient compromise between performance and economics. The chemistries of several common and advanced nickel-based superalloys are listed in Table 4 and the parts of gas turbine engine in which Nickel-based superalloys (marked red) commonly used are shown in Fig. 5.

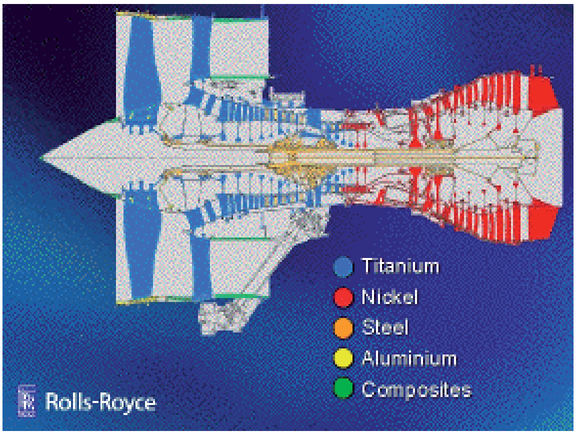
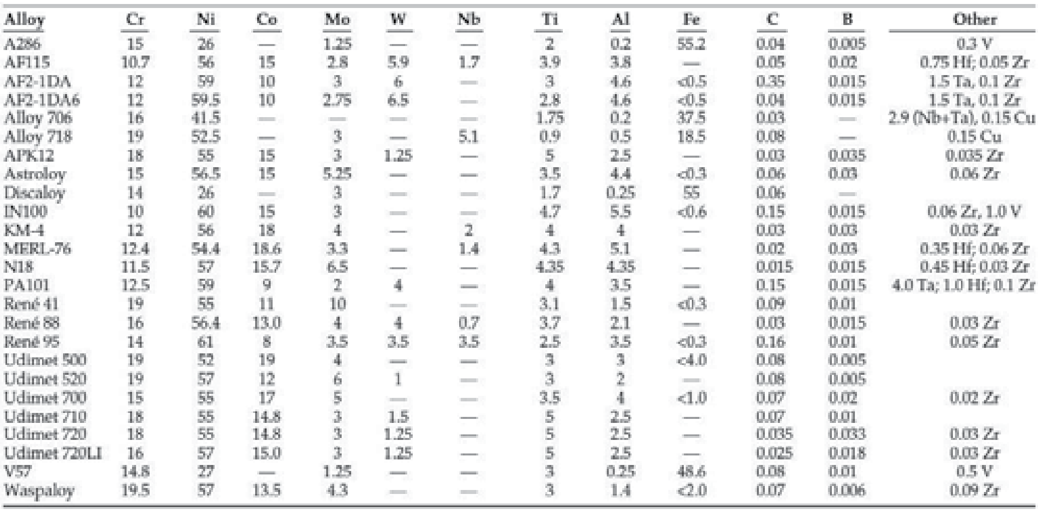


Fig. 5. Commonly used materials in gas turbine engine components.

In the environmental series nickel is nobler than iron but more active than copper. Reducing environments, such as dilute sulfuric acid, find nickel more corrosion resistant than iron but not as resistant as copper or nickel-copper alloys. The nickel molybdenum alloys are more corrosion resistant to reducing environment than nickel or nickel- copper alloys (Philip & Schweitzer, 2003). Nickel-based superalloys are extremely prone to weld cracking.

High-temperature strength of Ni-base superalloys depends mainly, on the volume fraction and morphology of γ´ precipitates. Several basic factors contribute to the magnitude of hardening of the alloy (Sajjadi & Zebarjad, 2006).

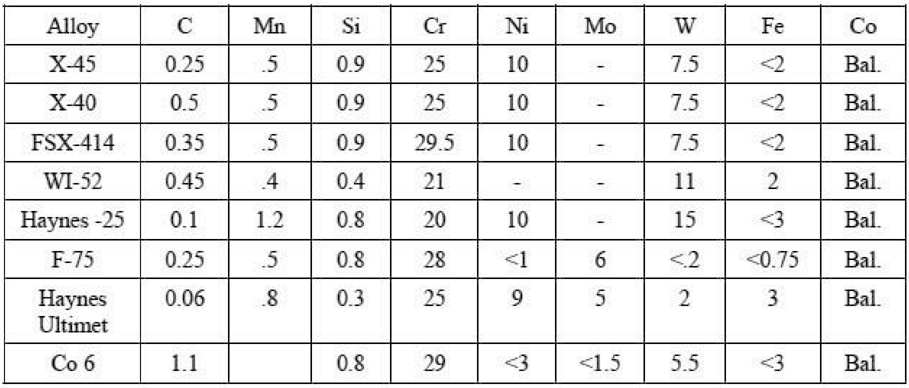


Alloy	Cr	Ni	Co	Mo	W	Nb	Ti	Al	Fe	C	B	Other
A286	15	26	—	1.25	—	—	2	0.2	55.2	0.04	0.005	0.3 V
AF115	10.7	56	15	2.8	5.9	1.7	3.9	3.8	—	0.05	0.02	0.75 Hf; 0.05 Zr
AF2-1DA	12	59	10	3	6	—	3	4.6	<0.5	0.35	0.015	1.5 Ta, 0.1 Zr
AF2-1DA6	12	59.5	10	2.75	6.5	—	2.8	4.6	<0.5	0.04	0.015	1.5 Ta, 0.1 Zr
Alloy 706	16	41.5	—	—	—	—	1.75	0.2	37.5	0.03	—	2.9 (Nb+Ta), 0.15 Cu
Alloy 718	19	52.5	—	3	—	5.1	0.9	0.5	18.5	0.08	—	0.15 Cu
APK12	18	55	15	3	1.25	—	5	2.5	—	0.03	0.035	0.035 Zr
Astroloy	15	56.5	15	5.25	—	—	3.5	4.4	<0.3	0.06	0.03	0.06 Zr
Discaloy	14	26	—	3	—	—	1.7	0.25	55	0.06	—	
IN100	10	60	15	3	—	—	4.7	5.5	<0.6	0.15	0.015	0.06 Zr, 1.0 V
KM-4	12	56	18	4	—	2	4	4	—	0.03	0.03	0.03 Zr
MERL-76	12.4	54.4	18.6	3.3	—	1.4	4.3	5.1	—	0.02	0.03	0.35 Hf; 0.06 Zr
N18	11.5	57	15.7	6.5	—	—	4.35	4.35	—	0.015	0.015	0.45 Hf; 0.03 Zr
PA101	12.5	59	9	2	4	—	4	3.5	—	0.15	0.015	4.0 Ta; 1.0 Hf; 0.1 Zr
René 41	19	55	11	10	—	—	3.1	1.5	<0.3	0.09	0.01	
René 88	16	56.4	13.0	4	4	0.7	3.7	2.1	—	0.03	0.015	0.03 Zr
René 95	14	61	8	3.5	3.5	3.5	2.5	3.5	<0.3	0.16	0.01	0.05 Zr
Udimet 500	19	52	19	4	—	—	3	3	<4.0	0.08	0.005	
Udimet 520	19	57	12	6	1	—	3	2	—	0.08	0.005	
Udimet 700	15	55	17	5	—	—	3.5	4	<1.0	0.07	0.02	0.02 Zr
Udimet 710	18	55	14.8	3	1.5	—	5	2.5	—	0.07	0.01	
Udimet 720	18	55	14.8	3	1.25	—	5	2.5	—	0.035	0.033	0.03 Zr
Udimet 720LI	16	57	15.0	3	1.25	—	5	2.5	—	0.025	0.018	0.03 Zr
V57	14.8	27	—	1.25	—	—	3	0.25	48.6	0.08	0.01	0.5 V
Waspaloy	19.5	57	13.5	4.3	—	—	3	1.4	<2.0	0.07	0.006	0.09 Zr

Table 4. The chemical compositions of several superalloys (wt.%) (Furrer & Fecht, 1999).

3.3 Cobalt-based Superalloys

The cobalt-based superalloys (Table 5) are not as strong as nickel-based superalloys, but they retain their strength up to higher temperatures. They derive their strength largely from a distribution of refractory metal carbides (combinations of carbon and metals such as Mo and W), which tend to collect at grain boundaries (Fig. 6). This network of carbides strengthens grain boundaries and alloy becomes stable nearly up to the melting point. In addition to refractory metals and metal carbides, cobalt superalloys generally contain high levels of Cr to make them more resistant to corrosion that normally takes place in the presence of hot exhaust gases. The Cr atoms react with oxygen atoms to form a protective layer of Cr_2O_3 which protects the alloy from corrosive gases. Being not as hard as nickel-based superalloys, cobalt superalloys are not so sensitive to cracking under thermal shocks as other superalloys. Co-based superalloys are therefore more suitable for parts that need to be worked or welded, such as those in the intricate structures of the combustion chamber (Jovanović et al).



Alloy	C	Mn	Si	Cr	Ni	Mo	W	Fe	Co
X-45	0.25	.5	0.9	25	10	-	7.5	<2	Bal.
X-40	0.5	.5	0.9	25	10	-	7.5	<2	Bal.
FSX-414	0.35	.5	0.9	29.5	10	-	7.5	<2	Bal.
WI-52	0.45	.4	0.4	21	-	-	11	2	Bal.
Haynes -25	0.1	1.2	0.8	20	10	-	15	<3	Bal.
F-75	0.25	.5	0.8	28	<1	6	<.2	<0.75	Bal.
Haynes Ultimet	0.06	.8	0.3	25	9	5	2	3	Bal.
Co 6	1.1		0.8	29	<3	<1.5	5.5	<3	Bal.

Table 5. The chemical composition of some Cobalt-based superalloys (Jovanović et al).

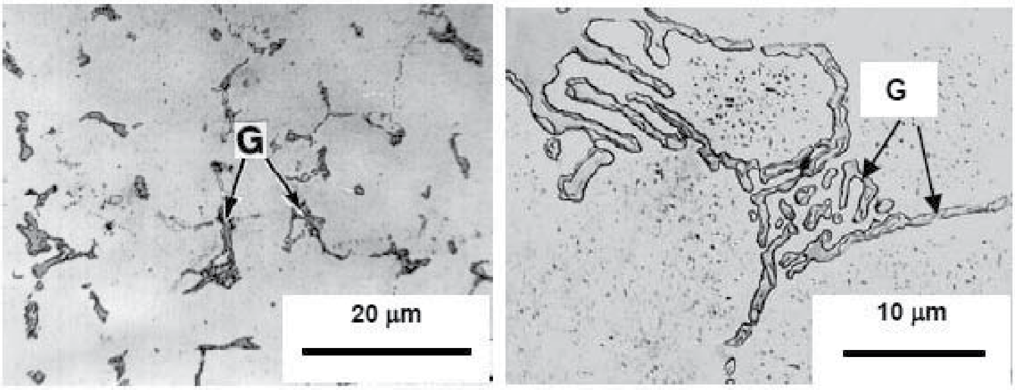


Fig. 6. Optical micrograph of Haynes-25. G mainly M_6C carbides (Jovanović et al).

4. ALLVAC 718 plus™

Inconel 718 is a nickel base superalloy that is used extensively in aerospace applications because of its unique high temperature mechanical properties. Since it was invented by Eiselstein, it has been used as a material of construction for aero-engine and land based turbine components. The reasons for alloy 718's popularity include excellent strength, good hot and cold workability, the best weldability of any of the superalloys and last, but not least, moderate cost. However, the application of the alloy has been limited to a temperature below 650 ◦C, as its properties deteriorate rapidly on exposure above this temperature due to the instability of the main strengthening phase of the alloy, γ″ (Idowu & Ojo, 2007). With prolonged exposure at this temperature or higher, γ″ rapidly overages and transforms to the equilibrium δ phase with an accompanying loss of strength and especially creep life (Kennedy, 2005).

Other wrought, commercial superalloys exist which have significantly greater temperature capability such as Waspaloy and René 41. These alloys are typically γ′ hardened and are significantly more difficult to fabricate and weld. Because of this and because of their intrinsic raw material content, these alloys are significantly more expensive than alloy 718. There have been numerous attempts to develop an affordable, workable 718-type alloy with increased temperature capability. After a number of years of systematic work, including both computer modeling and experimental melting trials, ATI Allvac has developed a new alloy, Allvac® 718Plus™, which offers a full 55°C temperature advantage over alloy 718. The alloy maintains many of the desirable features of alloy 718, including good workability, weldability and moderate cost (Kennedy, 2005).

ATI Allvac has extensively investigated the 718Plus alloy billet properties, both as an internal program and as part of the Metals Affordability Initiative program entitled "Low-Cost, High Temperature Structural Material" for turbine engine ring-rolling applications. The objective of all these programs is to develop an alloy with the following characteristics:

- 55°C temperature advantage based on the Larson-Miller, time-temperature parameter
- Improved thermal stability; equal to Waspaloy at 704°C
- Good weldability; at least intermediate to 718 and Waspaloy alloys

- Minimal cost increase; intermediate to 718 and Waspaloy alloys
- Good workability; better than Waspaloy alloy

The use of 718Plus alloy in elevated temperature applications is of interest for military systems. In particular, the manufacturing difficulties associated with alloys such as Waspaloy provide a need for a material with similar component capabilities, but with better producibility. Initial characterization shows that the alloy exhibits many similarities to Alloy 718, including good workability, weldability and intermediate temperature strength capability (Bergstrom & Bayhan, 2005).

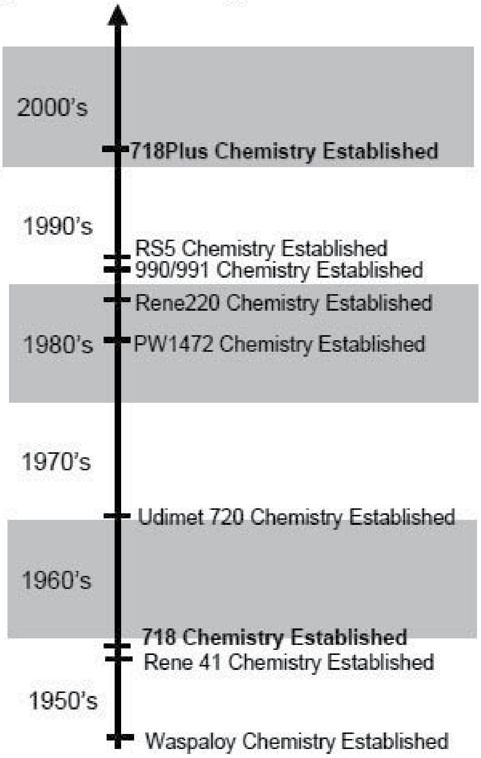
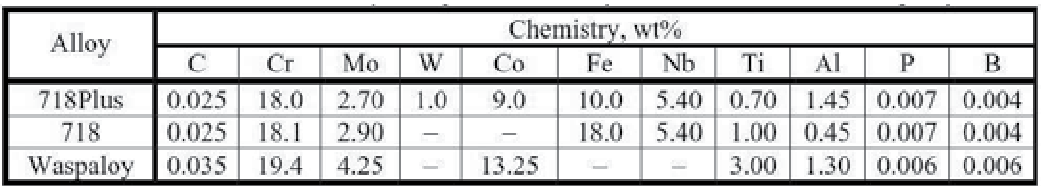

2000's
718Plus Chemistry Established
1990's
RS5 Chemistry Established
990/991 Chemistry Established
Rene220 Chemistry Established
1980's
PW1472 Chemistry Established
1970's
Udimet 720 Chemistry Established
1960's
718 Chemistry Established
Rene 41 Chemistry Established
1950's
Waspaloy Chemistry Established

Fig. 7. Developments leading up to alloy 718 and subsequent efforts to improve capability over 718 (Otti et al, 2005).

Since the advent of the first superalloys over 60 years ago, alloy developers have worked to promote strength and high temperature stability while balancing processability. Processing constraints for many alloy systems preclude their general use for cast and wrought forging applications. Instead these compositions are used in the cast form, or are producible only using powder metallurgy. The development and introduction of alloy 718 in the late 1950's offered a significant breakthrough in malleability and weldability relative to other high strength alloys available at that time including Waspaloy and René 41 which are primarily gamma prime strengthened. Since the introduction of alloy 718 a significant number of alloys have been examined, including cast as well as wrought alloys, with the primary intent to maintain or improve properties and provide increased thermal stability while maintaining favorable processability. Some of the alloys developed subsequently are shown along with 718, Waspaloy, and René 41 in the development timeline of Fig. 7. A key requirement beyond strength, toughness, fatigue, creep, crack growth resistance, and processability which has also driven composition development is weldability (Otti et al, 2005).

4.1 Chemistry

There are lots of wrought alloys in use for gas turbine engine parts, such as Waspalloy, which have high temperature capability. But they are typically much more difficult to manufacture and fabricate into finished parts and also significantly more expensive than alloy 718 (Bond & Kennedy, 2005). Therefore when ALLVAC 718 plus is compared to Inconel 718 this newly modified super alloy has the higher content of Al+ Ti, the higher ratio of Al/ Ti and the addition of W and Co instead of Fe. As a result it provides increased temperature capacity up to 55°C and impressive thermal stability. Therefore it closes the gap between Inconel 718 and Waspalloy, as combining the good processability and weldability of Inconel 718 with the temperature capability of Waspalloy. (Schreiber et al, 2006). The chemical compositions of the ALLVAC 718 plus with Inconel 718 and Waspalloy are given in Table 6.



Alloy	Chemistry, wt%										
	C	Cr	Mo	W	Co	Fe	Nb	Ti	Al	P	B
718Plus	0.025	18.0	2.70	1.0	9.0	10.0	5.40	0.70	1.45	0.007	0.004
718	0.025	18.1	2.90	–	–	18.0	5.40	1.00	0.45	0.007	0.004
Waspaloy	0.035	19.4	4.25	–	13.25	–	–	3.00	1.30	0.006	0.006

Table 6. Nominal chemistry comparison of the ALLVAC 718 plus, Inconel 718 and Waspalloy (Cao, 2005).

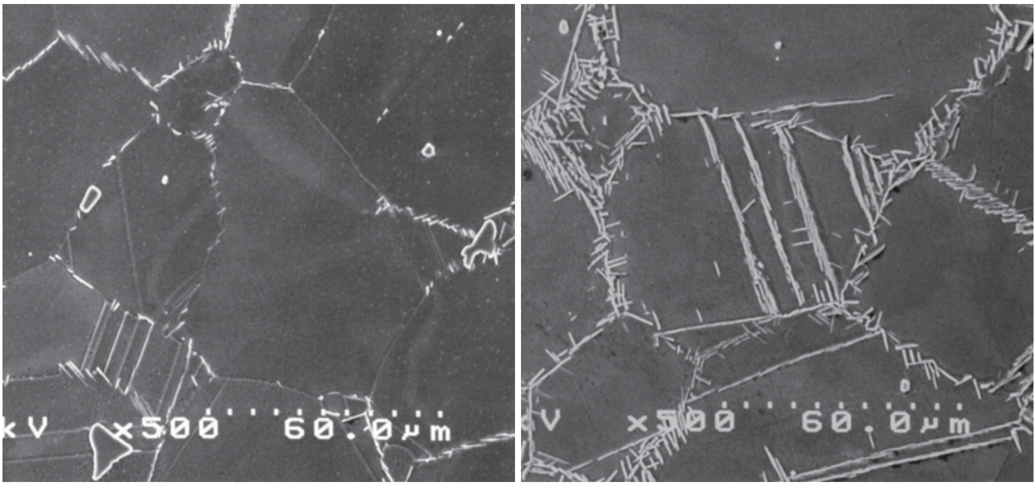
Alloy 718Plus has a much larger content of γ′ and γ ″ than alloy 718 and a smaller amount of δ phase. Solvus temperatures for γ′ and γ″ are also higher in alloy 718Plus. All of these points likely contribute to improved high temperature properties. One of the major differences between alloy 718 and Waspaloy is the speed of the precipitation reaction. The γ″ precipitation in alloy 718 is very sluggish and accounts in part for the good weldability and processing characteristics of the alloy.

In 718- type alloys primarily Fe, Co, Mo and W are the matrix elements. The effects of alloying elements on microstructure, mechanical properties, thermal stability and processing characteristics of alloy are important factors. Niobium is one of the major hardening elements and the other two is Al and Ti. The change in Al/ Ti ratio and the increase in Al+ Ti content converts the alloy into a predominantly γ′ strengthening alloy and it gives the alloy an improved thermal stability. Furthermore the modification on the content of Al and Ti develop the optimum mechanical properties of the alloy. Another factor on the improvement of the mechanical properties and thermal stability is the addition of Co up to about 9 wt%. Still further improvement occurs with Fe content of 10 wt%, 2.8 wt% Mo and 1 wt% W (Cao & Kennedy, 2004). Very small additions of P and B further increases stress rupture and creep resistance.

4.2 Strengthening mechanisms

As mentioned before, the primary strengthening phase is γ′ with a volume fraction ranging from 19.7-23.2 %, depending on the quantity of δ phase. Gamma prime strengthened alloys like Waspaloy and René 41 have much greater stability at higher temperature than γ″ strengthened alloys like 718 since γ″ grows rapidly and partially decomposes to equilibrium δ phase at temperatures in the 650–760°C range. Studies of the γ′ phase in 718Plus alloy show it to be high in Nb and Al, which is very different from the γ′ present in Waspaloy and René 41 and may account for its unique precipitation behavior and strengthening effects.

Like most superalloys there is a strong relationship between processing, structure and properties for alloy 718Plus. Optimum mechanical properties are achieved with a microstructure which has a small amount of rod shaped δ particles on the grain boundaries like that shown in Fig. 8 (a). Excessively high forging temperatures or high solution heat treating temperatures will result in structures with little or no δ phase precipitates that are prone to notch stress rupture failure. It is reported that no notch problems have been experienced using the 954°C solution temperature, probably because some δ phase can be precipitated at this temperature. However, excessively long heating times and possibly large amounts of stored, strain energy can result in large amounts of δ phase appearing on grain boundaries, twin lines and intragranularly, Fig. 8 (b). Such structures can lead to lower than expected tensile and rupture strength (Kennedy, 2005).



(a) (b)

Fig. 8. SEM Micrographs of Alloy 718Plus™ with (a) Preferred δ Phase Morphology and (b) Excessive δ Phase (Kennedy, 2005).

Alloy 718Plus does contain δ phase which is beneficial for conferring stress rupture notch ductility and controlling microstructure during thermo-mechanical processing. However, the volume fraction of the delta phase is considerably less than is found in alloy 718 and tends to be more stable with a much slower growth rate at elevated temperatures. Some γ″ may also be present in 718Plus alloy but in a much lower quantity, less than 7% (Jeniski & Kennedy, 2006).

When Inconel 718 is compared with the ALLVAC 718 plus it is reported that the size of strengthening phases increases in both alloys after long time thermal exposure (Fig. 9), but more significantly in alloy 718. In alloy 718, the average size of γ″+γ′ grows from about 15 nm at as heat-treated condition to almost 100 nm after 500 hrs long time aging at 760°C as indicated in Fig. 10 and the main strengthening phase γ″ grows to about 200nm in estimation (see Fig. 11). However, in alloy 718Plus, the main strengthening phase γ′ coarsens slowly and the average size of γ′ is still about 70 nm as indicated in Figure 9. These important quantitative phase analyses results convince us that alloy 718Plus has a superior stable microstructure in comparison with alloy 718 (Xie et al, 2005).

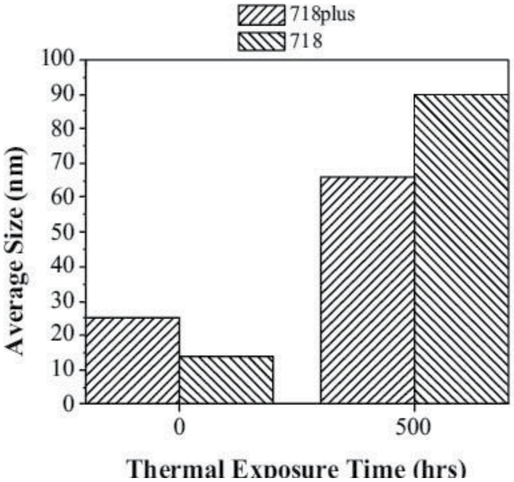


Fig. 9. The size of strengthening phases in Alloy 718 and Alloy 718 plus (Xie et al, 2005).

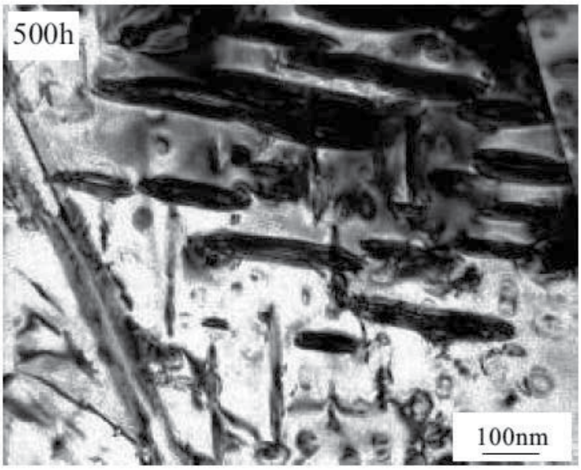


Fig. 10. The coarsening of strengthening phases γ´´ and γ´ in alloy 718 after 760°C (Xie et al, 2005).

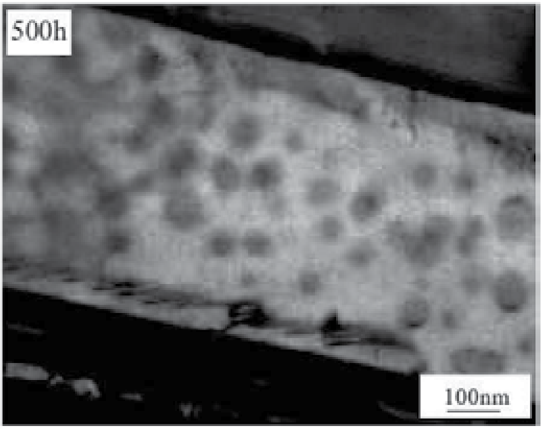


Fig. 11. The coarsening of strengthening phase γ´ in alloy 718 plus after 760°C Aging (Xie et al, 2005).

4.3 Microstructure

The microstructure of 718 plus in as received hot-rolled condition consists of FCC austenitic matrix with an average grain size of 50 µm. Fig. 12 shows the optical micrograph of the alloy. It can be seen that precipitates with round-to-blocky morphology are randomly dispersed within the microstructure (Vishwakarma et al, 2007).

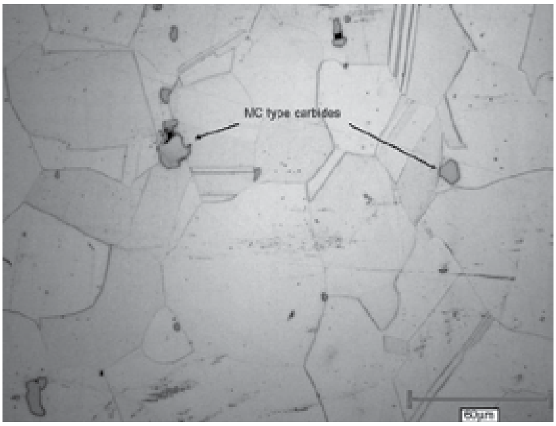


Fig. 12. Optical microstructure of as received 718 plus alloy (Vishwakarma et al, 2007).

SEM/EDS analysis of the precipitates shows them to be mainly Nb-rich MC type carbides containing Ti and C. As laves phase can be eliminated by high temperature homogenization and thermo–mechanical processing of wrought Inconel 718 and 718 Plus type of alloys, it is therefore not observed in the microstructure of the as received material. The delta phase, which is commonly observed in Inconel 718, is not observed in the as received microstructure of 718 plus alloy (Vishwakarma et al, 2007).

Heat treated at 950ºC for 1 hour microstructure of 718 plus alloy, grain size of 54 µm, which has normal B and P concentrations, has shown in Fig. 13. It can be seen that needle like δ phase is observed on the grain boundaries and occasionally intra-granularly on the twin boundaries. And also seen in the microstructure, round and blocky shaped MC type carbide particles are randomly distributed. Ti-rich carbo-nitride particles can be also observed (Vishwakarma & Chaturvedi, 2008). Intermetallic phases like FCC γ' and BCT γ'' are expected to form in 718 plus alloys but γ' is the main strengthening phase in these superalloys (Cao & Kennedy, 2004).

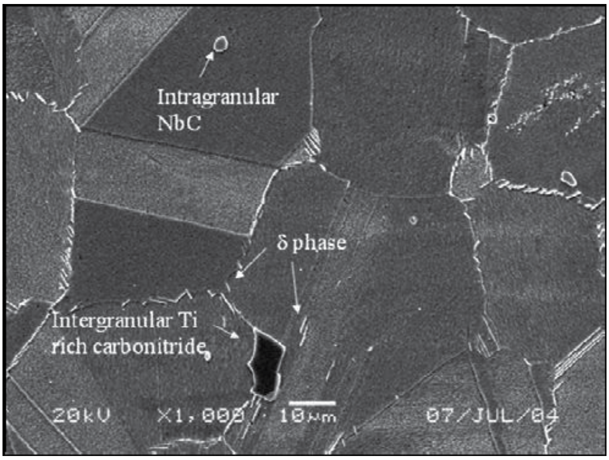

Intragranular NbC
δ phase
Intergranular Ti rich carbonitride
20kV X1,000 10µm 07/JUL/04

Fig. 13. Microstructure of 718 plus alloy heat treated at 950ºC for 1 h (Vishwakarma & Chaturvedi, 2008).

4.4 Mechanical properties

Alloy 718Plus™ has a significant strength advantage over alloy 718 at temperatures above 650°C and over the entire temperature range compared to Waspaloy and A286. Elongation for alloy 718Plus™ over the entire temperature range remained high at 18% minimum. These data are consistent with comparisons of alloy 718, Waspaloy, A286 and alloy 718Plus™ in other product forms, including billet, rolled rings, forgings and sheet (Bond & Kennedy, 2005). In Fig. 14 shows the effect of temperature on room temperature ultimate tensile strength for several alloys. The tensile strength of ALLVAC 718 plus and Waspaloy is shown in Fig. 15.

It is reported by ATI ALLVAC Cooperation that extensive studies demonstrated that this alloy has shown superior tensile and stress rupture properties to alloy 718 and comparable properties to Waspaloy at the temperature up to 704°C. However, relatively speaking, the data on fatigue crack propagation (FCP) resistance of this alloy are still insufficient. Alloys 718Plus, 718 and Waspaloy have similar fatigue crack growth rates under 3 seconds triangle loading at 650°C with 718Plus being slightly better. Waspalloy shows the best resistance to fatigue crack growth under hold time fatigue condition while the resistance of 718Plus is better than that of Alloy 718 (Liu et al, 2005).

Examination of the fatigue fracture surfaces by scanning electronic microscope (SEM) revealed transgranular crack propagation with striations for 718Plus at room temperature. The fracture mode at 650ºC is the mixture of intergranular and transgranular modes (Liu et al, 2004).

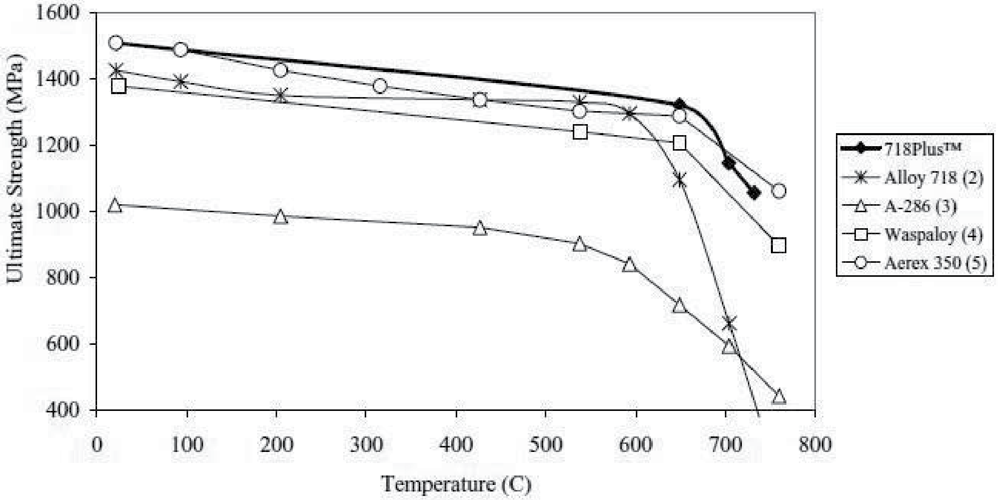


Fig. 14. Effect of test temperature on room temperature tensile ultimate tensile strength for several alloys (Bond & Kennedy, 2005).

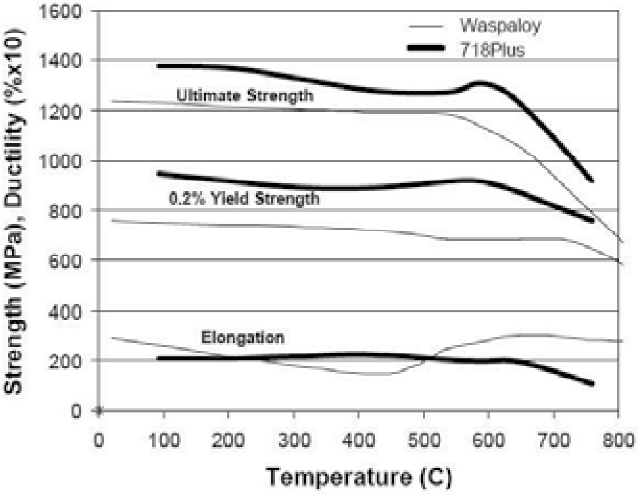


Fig. 15. The tensile strength of Alloy 718 plus and Waspaloy (Otti et al, 2005).

Direct aging can be effectively applied to alloy 718Plus to improve its mechanical properties, including strength and stress rupture life of alloy 718Plus. Considering the fine grain size and high strength resulting from direct aging, the low cycle fatigue resistance of this alloy should also be significantly improved although further experimental verification is necessary. DA processing of this alloy is also different from Waspaloy in that hot working at temperatures above the γ′ solvus can achieve a good, direct age response (Cao & Kennedy, 2005).

4.5 Weldability

There are numerous types of superalloys with a difference in weldability among the types. The solid solution alloys are the easiest to weld because they don't undergo drastic metallurgical changes when heated and cooled. Because of their limited strength, however, they are only used in certain areas of a gas turbine, such as the combustor.

The precipitation-strengthened alloys are more demanding during welding and post welding because of the precipitation of the hardening phase that usually contains aluminum, titanium, or niobium. These elements oxidize very easily and, therefore, alloys that contain them need better gas protection during welding. A third type of superalloy is the mechanically alloyed materials that cannot be welded without suffering a drastic drop in strength. These alloys are usually joined by mechanical means or diffusion bonding. In addition to those elements that enable a superalloy to undergo precipitation hardening, such as aluminum, titanium, and niobium, other elements are added to enhance mechanical properties or corrosion resistance. These include boron and zirconium, which are often intentionally added to some alloys to improve high temperature performance but at a cost to weldability. There are numerous other elements that are not intentionally added but can be present in very small quantities that are harmful, such as lead and zinc. These are practically insoluble in superalloys and can cause hot cracking during solidification of the welds. Small quantities of these elements on the surface of a metal can cause localized weld cracking. Sulfur is considered detrimental if present in too large a quantity, but can cause low weld penetration problems if present in very low amounts (Donald & Tillack, 2007).

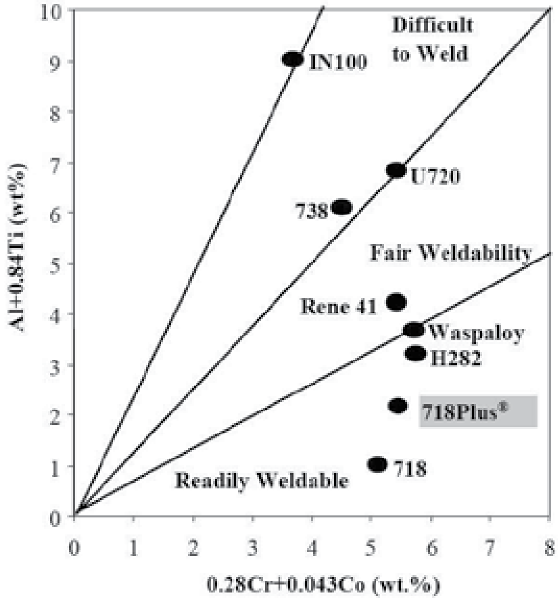


Fig. 16. Effect of chemistry on post-weld heat cracking (Jeniski & Kennedy, 2006).

It is reported that limited weldability testing has been conducted on 718Plus alloy but results have been encouraging. Weldability of alloy 718Plus is believed to be quite good, at least intermediate to alloys 718 and Waspaloy (Kennedy, 2005). Improved weldability over Waspaloy is one of the primary drivers for 718Plus alloy in engine applications. Figure 16 shows the weld cracking tendency for a number of well known commercial alloys and illustrates the good welding characteristics expected with 718Plus alloy based on its chemistry (Jeniski & Kennedy, 2006). Some micrographs of the Electron Beam Welded ALLVAC 718 plus, Inconel 718 and Waspaloy rings are shown in Fig. 17.

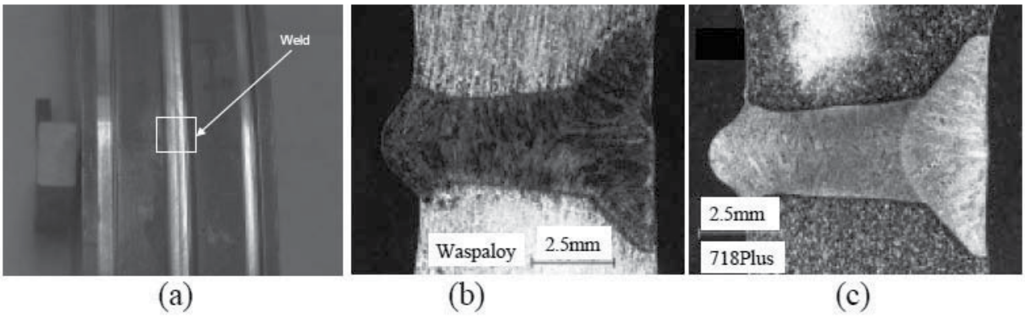


Fig. 17. EB welding of 890 mm diameter rolled rings, (a) typical weld location for 718, Waspaloy, and 718Plus welds and typical welds for (b) Waspaloy, and (c) 718Plus (Otti et al, 2005).

4.6 Cost and Applications

The first commitment to a production use of alloy 718Plus has been made for a high temperature tooling application, replacing Waspaloy as a hot shear knife. Other applications include aero and land-base turbine disks, forged compressor blades, fasteners, engine shafts and fabricated sheet/plate components. Product forms include rolled or flash butt welded rings, closed die forgings, bar, rod, wire, sheet, plate and castings.

The alloy also can be used for flash-butt welded ring applications. Sheet form of the alloy is being considered for fabricated engine parts such as turbine exhaust cases and engine seals. Fasteners remain another potential application for 718Plus alloy. The property advantages for 718Plus alloy have also led to its being considered for rotating parts. Cao and Kennedy have shown that 718Plus alloy is capable of direct aging (DA), low temperature working followed by aging with no prior solution heat treatment. DA processing resulted in the production of very fine grain material with yield strength improvement at 704°C of 70-100 MPa. The alloy is also being considered for blading applications in areas where alloy 718 is limited due to elevated operating temperatures.

The alloy has also other applications outside of the jet and power turbine engines. Any application that currently uses alloys 718, Waspaloy, René 41 or other nickel-based superalloys can consider 718Plus alloy as a substitute for reasons of cost savings or increased temperature capability. Other markets where 718Plus alloy has potential are automotive turbo-chargers or industrial markets like chemical process or oil and gas where alloy 718 is used (Jeniski & Kennedy, 2006).

The cost of finished components of alloy 718 Plus is expected to be intermediate to alloys 718 and Waspaloy (Kennedy, 2005).

5. Acknowledgment

The author would like to thank Eskisehir Osmangazi University for supporting the research study with Eskisehir Osmangazi University Research Funding Project, Project No: 2011/15020.

6. Conclusion

A superalloy is a metallic alloy which is developed to resist most of all high temperatures, usually in cases until 70 % of the absolute melting temperature. All of these alloys have an excellent creep, corrosion and oxidation resistance as well as a good surface stability and fatigue life.

The main alloying elements are nickel, cobalt or nickel – iron, which can be found in the VIII· group of the periodic system of the elements. Fields of application are found particularly in the aerospace industry and in the nuclear industries, e.g. for engines and turbines.

The development of these advanced alloys allows a better exploitation of engines, which work at high temperatures, because the Turbine Inlet Temperature (TIT) depends on the temperature capability of the material which forms the turbine blades. Nickel-based superalloys can be strengthened through solid-solution and precipitation hardening.

Nickel-based superalloys can be used for a higher fraction of melting temperature and are therefore more favourable than cobal-based and iron-nickel-based superalloys at operating temperatures close to the melting temperature of the materials.

The newly innovated nickel based ALLVAC 718 Plus superalloy which is the last version of Inconel 718 has been proceeding in the way to become a material that aerospace and defense industries never replace of any other material with combining its good mechanical properties, easy machinability and low cost.

7. References

Bergstrom D. S., and Bayhan. T. D., (2005) Properties and Microstructure Of ALLVAC 718 Plus Alloy Rolled Sheet, *Superalloys 718, 625, 706 and Derivatives Edited by E.A. Loria TMS (The Minerals, Metals & Materials Society).*

Bond. B.J. and Kennedy. R.L., (2005) Evaluation of ALLVAC 718 plus Alloy In the Cold Worked and Heat Treated Condition, *Superalloys 718, 625, 706 and Derivatives Edited by E.A. Loria TMS (The Minerals, Metals & Materials Society).*

Campbell F. C. (2008). *Elements Of Metallurgy And Engineering Alloys,* ASTM International., ISBN: 978-0-87170-867-0, USA.

Cao. W., (2005) Solidification and Solid State Phase Transformation of ALLVAC 718 Plus Alloy, *Superalloys 718, 625, 706 and Derivatives 2005, TMS.*

Cao. W. and Kennedy. R., (2004) Role of Chemistry in 718-Type Alloys- Alloy 718 plus Development, *Edited by K.A. Green, T.M. Pollock, H. Harada T.E. Howson, R.C. Reed, J.J. Schirra, and S, Walston, Superalloys, TMS (The Minerals, Metals & Materials Society)*

Cao. W., and Kennedy. R. L., (2005) Application Of Direct Aging To ALLVAC 718 Plus Alloy For Improved Performance, *Superalloys 718, 625, 706 and Derivatives Edited by E.A. Loria TMS (The Minerals, Metals & Materials Society).*

Carlos A. E. M., (2007) New Technology Used In Gas Turbine Blade Materials, *Scientia et Technica Ano XIII, No:36, ISSN: 0122-1701*

Donald. T. J., (2007) Welding Superalloys For Aerospace Applications, *Welding Journal.* pp 28-32 January.

Eliaz. N., Shemesh. G., Latanision. R.M., (2002) Hot Corrosion in Gas Turbine Components, *Engineering Failure Analysis* 9 31–43

Furrer. D., Fecht. H., (1999) Ni-Based Superalloys For Turbine Discs, *JOM*

Idowu, O.A. , Ojo, O.A., Chaturvedi, M.C. (2007) Effect of heat input on heat affected zone cracking in laser welded ATI Allvac 718Plus superalloy. *Materials Science and Engineering.* Vol. A No. 454–455 pp.389–397.

Jeniski. R. A., Jr. and Kennedy. R. L., (2006) Development of ATI Allvac 718Plus Alloy and Applications, *II. Symposium on Recent Advantages of Nb-Containing Materials in Europe*

Jonšta Z., Jonšta P., Vodárek V., Mazanec K. (2007) Physical-Metallurgical Characteristics Of Nickel Super-Alloys Of Inconel Type. *Acta Metallurgica Slovaca*, 13, 4 (546 - 553).

Jovanović T. M., Lukic. B., Miskovic. Z., Bobic. I., Ivana, Cvijovic. B. D., Processing And Some Applications Of Nickel, Cobalt And Titanium-Based Alloys, *Association of Metallurgical Engineers of Serbia Review paper,* MJoM *Metalurgija Journal Of Metallurgy*

Kennedy. R. L., (2005) Allvac® 718plus™ Superalloy For The Next Forty Years, *Superalloys 718, 625, 706 and Derivatives 2005, TMS.*

Liu. X., Xu. J., Deem. N., Chang. K., Barbero. E., Cao. W., Kennedy. R. L., Carneiro. T., (2005) Effect Of Thermal-Mechanical Treatment On The Fatigue Crack Propagatıon Behavior Of Newly Developed Allvac 718plus Alloy, *Superalloys 718, 625, 706 and Derivatives 2005, TMS.*

Liu. X., Rangararan. S., Barbero. E., Chang. K., Cao. W., Kennedy. R.L., and Carneiro. T., (2004) Fatigue Crack Propagatıon Behavıors Of New Developed Allvac 718plus Superalloy, *Superalloys 2004, TMS.*

Otti. E.A., Grohi. J. And Sizek. H., (2005) Metals Affordability Initiative: Application of Allvac Alloy 718Plus for Aircraft Engine Static Structural Components, *Superalloys 718, 625, 706 and Derivatives Edited by E.A. Loria TMS (The Minerals, Metals & Materials Society).*

Philip A. Schweitzer, P. E., (2003). *Metallic Materials: Physical, Mechanical and Corrosion Properties,* Marcel Dekker, Inc., ISBN: 0-8247-0878-4, USA.

Sajjadi, S.A., Zebarjad, S.M. (2006) Study of fracture mechanisms of a Ni-Base superalloy at different temperatures. *Journal of Achievements in Materials and Manufacturing Engineering*. Vol. 18 Issue 1-2.

Schreiber. K., Loehnert. K., Singer. R.F., (2006) Opportunities and Challenges for the New Nickel-Based Alloy 718 Plus, *II. Symposium on Recent Advantages of Nb-Containing Materials in Europe.*

Sims. C.T., Stoloff. N.S. and Hagel. W.C., (1987) Superalloys II- High Temperature Materials for Aerospace and Industrial Power, *John Wiley & Sons Inc.,* USA.

Smallman R. E. Ngan, A. W. H., (2007). *Physical Metallurgy and Advanced Materials,* Elsevier Ltd., ISBN: 978 0 7506 6906 1, UK.

Smallman R. E. Bishop, R. J., (1999). *Modern Physical Metallurgy and Materials Engineering,* Reed Educational and Professional Publishing Ltd., ISBN: 0 7506 4564 4, UK.

Thomas. A., El-Wahabi. M., Cabrera. J.M., Prado. J. M., (2006) High Temperature Deformation of Inconel 718, *Journal of Materials Processing Technology* 177 469–472.

Vishwakarma. K.R., Richards. N.L., Chaturvedi. M.C., (2007) Microstructural Analysis of Fusion and Heat Affected Zones in Electron Beam Welded ALLVAC® 718PLUSTM superalloy, *Materials Science and Engineering* A 480 517–528

Vishwakarma. K.R. and Chaturvedi. M.C., (2007) A Study Of Haz Mıcrofıssurıng In A Newly Developed Allvac® 718 Plus Tm Superalloy.

Yaman. Y.M., Kushan. M.C., (1998) Hot Cracking Susceptibilities In the Heat Affected Zone of Electron Beam Welded Inconel 718, *Journal of Materials Science Letters* 17, 1231-1234.

Xie. X., Wang. G., Dong. J., Xu. C., Cao. W., Kennedy. R. L., (2005) Structure Stability Study On A Newly Developed Nickel-Base Superalloy- ALLVAC 718 Plus, *Superalloys 718, 625, 706 and Derivatives Edited by E.A. Loria TMS (The Minerals, Metals & Materials Society).*

5

Potential of $MoSi_2$ and $MoSi_2$-Si_3N_4 Composites for Aircraft Gas Turbine Engines

Melih Cemal Kushan[1], Yagiz Uzunonat[2],
Sinem Cevik Uzgur[3] and Fehmi Diltemiz[4]
[1]Eskisehir Osmangazi University
[2] Anadolu University
[3]Ondokuz Mayis University
[4]1st Air Supply and Maintenance Base
Turkey

1. Introduction

It has been expected that gas turbine engines in high temperature environments where aggressive mechanical stresses may occur and a good surface stability is needed should operate more efficiently. So the investigations about the materials which will be able to carry the aviation technology to the next level are beginning to accelerate in this direction. And also it expected that those new materials using in gas turbine engines as a high temperature structural material will exceed the superalloys' mechanical and physical limits. The intended development can only be achieved by providing the improvement of the essential properties of the structural materials such as thermal fatigue, oxidation resistance, strength/weigth ratio and fracture toughness. There are two different type of materials which are candidate to resist the operating conditions about 1200°C; first one is structural ceramics such as SiC, Si_3N_4 and the second one is structural silicides such as $MoSi_2$.

After the propulsion systems with high strength/weight ratio, it observed that development of new materials with high strength and low density was necessary, thus the studies about the intermetallics began. The most important ones of these intermetallic compounds are silicides and aluminides. By the oxide layers in Al_2O_3, it can be used as a protective material in high temperature applications. Moreover, aluminides such as FeAl, TiAl, Ni_3Al, can be suitable for some special applications in low and medium temperatures. In spite of these advantages, they remain inadequate above the temperatures 1200°C for their melting points with 1400-1600°C. Their low strength and creep resistance is not suitable for the temperatures above 1000°C. For this reason, it seems that silicides and aluminides are the proper materials for high service applications (Vaseduvan & Petrovic, 1992).

2. Superalloys and their limitations at elevated temperatures

In aviation applications, advanced gas turbine elements are exposed to several mechanic, thermal and corrosive environments and intensive studies for the developing of these parts are still continuing. However, these alloys are needed to be cooled during the operation of the turbine engine and the practical temperature limits for metallic alloys remain below 1100ºC. But in this situation, the elevation of turbine inlet temperature will be quite difficult and expansive. Because of these given limitations, there is not any important improvements on nickel based superalloys since 1985 (Soetching, 1995).

The basic facts that can directly effect the performance of superalloys in high temperatures are oxidation, hot corrosion and thermal fatigue. These effects cause the superalloy elements' surfaces may react with hot gases easier, and then their surface stability decreases (Bradley, 1988). Furthermore, during operation and stand-by period of turbine, there occurs a oscillation motion in the hot section elements respectively. This causes thermal fatigues on the superalloy parts.

2.1 Oxidation

Oxidation is one of the most serious factors acting on the gas turbine's service life and can be determined as the reaction of materials with oxygen in 2-4 atm. partial pressure (Tein & Caulfield, 1989). Mostly the uniform oxidation is not accepted as a considerable problem in relatively low temperatures (870ºC and below). But in temperatures about 1100ºC, the aluminum content in the form of Al_2O_3 as a protective oxide can not provide the expected protection in long term periods. For this reason, it is necessary to use the silicide based structural composite materials or to make protective coatings with respect to the segment's location in gas turbine.

2.2 Hot corrosion

The process of hot corrosion contains a structural element and the reactions occurring in its surroundings. In operating conditions at high temperatures there is a possible accelerated oxidation for superalloys. Another name for this reaction is hot corrosion and it consists of two different mechanisms as low temperature (680-750ºC) and high temperature (900-1050ºC) hot corrosion (Akkuş, 1999).

The basic principle to avoid from the hot corrosion in superalloys is using of the high content of chrome (≥ %20) during the manufacturing of material. But only a few types of nickel based superalloys have this rate for their high proportion of γ' and γ'' structure.

2.3 Thermal fatigue

Heating with non-uniform distribution make interior stresses in the zones hotter than the average temperature of the turbine, and tension stresses in the colder zones. Superalloy turbine vanes are the good examples of elements exposed to thermal fatigue in aeroplane jet engines. During the acceleration, inlet and outlet edges of the turbine vane can heat and expand easier than the medium part under cooling. But in deceleration, inlet and outlet parts can quietly cool off than the medium parts. This case results as fatigue crack at the edges.

3. Physical and mechanical properties of $MoSi_2$

$MoSi_2$ is a potential material for high temperature structural applications primarily due to its high melting point (2020ºC), lower density (6.3 g/cm^3) compared with superalloys, excellent oxidation resistance, high thermal conductivity, and thermodynamic compatibility with many ceramic reinforcements. However, low fracture toughness at near-ambient temperatures, low strength at elevated temperatures in the monolithic form and tendency to pest degradation at ~500°C have seriously limited the development of $MoSi_2$-based structural materials. Several recent studies have attempted to address these issues and have shown promising results. For example, pest resistant $MoSi_2$-based materials have been developed using silicon nitride reinforcement or alloying with Al.

For general polycrystalline ductility five independent deformation modes are necessary. Changing the critical resolved shear stress of the slip systems through alloying may be a way to activate all three slip vectors, and obtain polycrystalline ductility. In fact, solid solution softening has been observed at room tempera ture in $MoSi_2$ alloyed with Al and transition metals such as Nb, V and Ta. The mechanism of softening is not clearly understood, although first principles calculations indicate that solutes such as Al, Mg, V and Nb may change the Peierls stress so as to enhance relative to cleavage. Clearly, more work is needed to understand how alloying may influence the mechanical behavior of $MoSi_2$.

With regard to elevated temperature strengthening of $MoSi_2$, both alloying with W to form C11b (Mo, W)Si_2 alloys and composites with ceramic re inforcements such as SiC have been tried. A (Mo, W)Si_2./20 vol.% SiC composite was shown to have significantly higher strength than Mar-M247 superalloy at temperatures above 1000°C. However, the strength of the (Mo, W)Si_2./20 vol.% SiC composite dropped by almost an order of mag -nitude from 1200 to 1500°C; the yield strength at 1500°C was only ~75 MPa. A simpler and more e€ective way of strengthening $MoSi_2$ at elevated temperatures is needed where the strength can be better retained with increasing temperature above 1200°C. Our preliminary studies using hot hardness experiments have shown that Re addition to $MoSi_2$ caused signifcant hardening up to 1300°C. Further, it has been reported that alloying with Re, perhaps in synergism with carbon, increased the pesting resistance in the temperature range of 500 ± 800°C. In another preliminary study, polycrystalline (Mo, Re)Si_2 alloys exhibited a minimum creep rate of ~5 x 10^{-6}/s at 100 MPa applied stress at 1400° C as compared with the ~1 x 10^{-4}/s creep rate exhibited by $MoSi_2$. No detailed mechanistic study has been performed to understand the effects of Re alloying on the elevated temperature mechanical behavior of $MoSi_2$. In the present investigation, we have evaluated the mechanical properties, in compression, of arc-melted polycrystalline $MoSi_2$ and (Mo, Re) Si_2 alloys. We find that significant strengthening is achieved up to 1600°C by only small additions of Re. The mechanisms of elevated temperature solid solution strengthening are elucidated by considering the generation of constitutional Si vacancies that may pair with Re substitutionals to form tetragonally distorted point defect complexes. Characteristics of $MoSi_2$ make it an interesting material as high temperature structural silicide. Not only it has a low density and a high melting point but also it can excellently resist the free oxygen of air in high temperature environments for a long time period. On the other hand, researchers noticed its potential as a structural material due to its electrical resistance increasing after every use and high modulus of elasticity at high temperatures. This makes $MoSi_2$ a candidate material for structural high temperature applications particularly in gas turbine engines. $MoSi_2$ and its composites offer a higher rate of resistance to oxidizing and aggressive environments during the combustion processes with their high melting points.

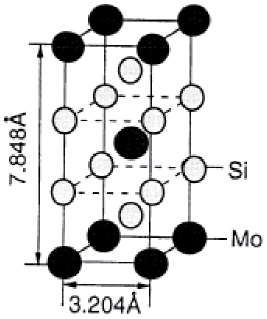


Fig. 1. Unit cell of the body-centered tetragonal C11b structure of $MoSi_2$. (Misra et al., 1999).

Fracture toughness of the material shows similarities with the other silicon based ceramics and yet it receives a brittle fracture resulted with low toughness. Table 1. shows the considerable characteristics of $MoSi_2$.

	Metric	English
Density	6.23 g/cm^3	0.225 lb/in^3
Molecular Weigth	152.11 g/mol	152.11 g/mol
Electrical Resistance (20°C)	3.5x10^{-7} ohm-cm	3.5x10^{-7} ohm-cm
Electrical Resistance(1700°C)	4.0x10^{-6} ohm-cm	4.0x10^{-6} ohm-cm
Thermal Capacity	0.437 J/g-°C	0.104 BTU/lb-°F
Thermal Conductivity	66.2 W/m-K	459 BTU-in/hr-ft^2-°F
Melting Point	2030°C	4046°F
Maximum Service Temp.	1600°C	2912°F
Crystal Structure	Tetragonal	Tetragonal

Table 1. Basic characteristics of $MoSi_2$.

The figure below shows the tetragonal lattice structure directions, red and blue points inidicate silicon and molybdenum atoms respectively.

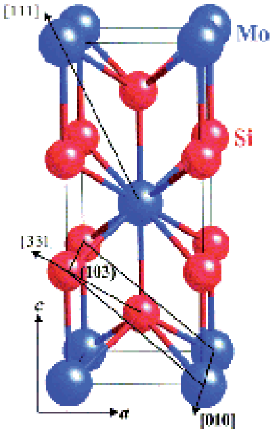


Fig. 2. Tetragonal $MoSi_2$ lattice structure.

One of the most considerable limitations in $MoSi_2$ applications is the structural disintegration during the low temperature oxidation which is known as pesting oxidation (Meschter, 1992). Previously, we noted that $MoSi_2$ has an excellent oxidation resistance above the 1000°C, but at the temperatures about 500°C as it is presented Figure 3., the oxidation mechanism accelerates because of the volume expansion, MoO_3 crystals, amorph-shaped SiO_2 bulks and $MoSi_2$ particles residual from the reaction. If the material is porous and the surface accuracy is low, this state can be observed along the cracks or grain boundaries, and granular oxide particles occur as a result.

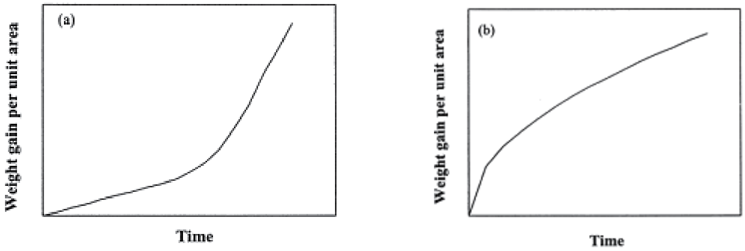


Fig. 3. Isothermal oxidation curves (a) at room temperature, (b) at the temperatures above 1000° C (Liu et al., 2001).

This fact was discovered in 1955 and predicted as the grain boundary fracture due to solution oxygen at the grain boundaries after the short-term cyclic diffusion, even though its complete nature is still a phenomena (Chou & Nieh, 1992, 1993). Methods for preventing from the pest effect are continuing. These methods are; making a protective SiO_2 coating on the material and increasing the relative density of $MoSi_2$ in the structure (Wang et al., 2003)

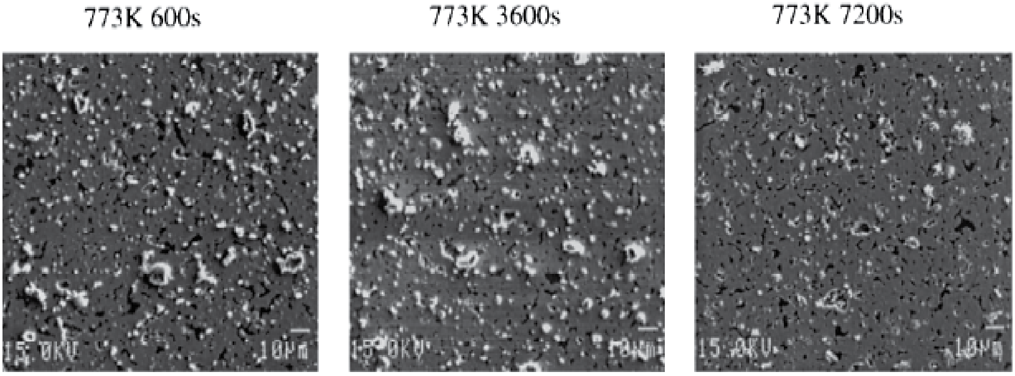


Fig. 4. Surfaces oxidized at 773K for 600-7200s in O_2 (Chen et al., 1999).

During the oxidation reactions above 600°C, no pesting effect can be observed. $MoSi_2$ based composites have considerably higher isothermal oxidation resistance than any other titanium, niobium or tantalum based composites, intermetallic compounds and nickel based superalloys, $MoSi_2$ perfectly keeps this condition to 1600°C (Vaseduvan & Petrovic, 1992).

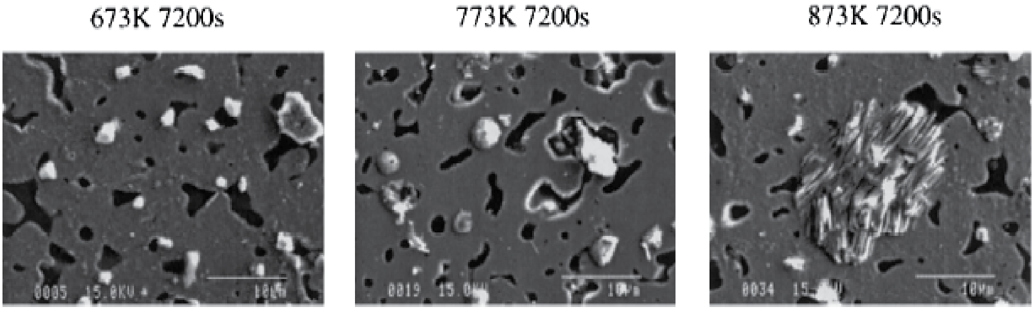


Fig. 5. Surfaces oxidized at 673-873*K* for 7200s in O_2 (Chen et al., 1999).

Despite excellent oxidation resistance, high melting point, and low density, the potentials of molybdenum disilicide as a high temperature structural material have not been utilized due to its brittleness at low temperatures and low strength at high temperatures . For example, below 900° C, the fracture toughness of $MoSi_2$ is in the range of 2–4 $MPam^{1/2}$, and the 0.2% offset yield strength of $MoSi_2$ at 1600° C is about 20 MPa. Alloying or reinforcing with a second phase may lower the brittle to ductile transition temperature (BDTT) of $MoSi_2$. However, ductile-phase toughening with metallic phases has limited applicability in $MoSi_2$ due to the chemical reaction with silicon to form silicides, and reinforcing with ceramic second phases such as SiC and ZrO_2 has only a modest effect on enhancing plastic flow and increasing toughness.

First principles calculations indicate that alloying of $MoSi_2$ while maintaining its body-centered tetragonal (C11b) structure may result in improved mechanical properties. For example, Al and Nb may enhance ductility and Re may increase strength. Improvements in both low and high temperature mechanical properties of $MoSi_2$ have been reported by alloying $MoSi_2$ with small amounts of Al, Nb, and Re (<2 at.%). During alloying, below the solubility limits of alloying elements in the C11b structure of $MoSi_2$, Al substitutes for Si, whereas Re and Nb substitute for Mo. The solubility limits of Re, Nb and Al in $MoSi_2$ have been reported as ~2.5, 1.3 and 2.7 at.%, respectively. Although improvements in the ambient temperature toughness have been reported by alloying of $MoSi_2$ beyond the solubility limits with Nb and Al, the rates of improvement per fraction of solute are not as considerable as those observed in single phase alloys. Furthermore, the presence of secondary phases with a lower high temperature strength than the matrix alloy would degrade the mechanical properties at high temperature (>1500°C) for which applications $MoSi_2$ is an excellent candidate. The aim of this investigation was to explore the possibility of obtaining concurrently enhanced room temperature ductility and high temperature strength in single-phase $MoSi_2$ by combining the high temperature hardening and the low temperature softening effects of Re, Al, and Nb. Hardness testing at room temperature and compression testing at 1600° C are conducted on unalloyed and alloyed $MoSi_2$ samples in order to study both low and high temperature effects of each alloying composition on the mechanical properties of $MoSi_{2.}$ (Sharif et al., 2001).

4. Effects of alloying

4.1 Hardness

Microhardness testing performed on stoichiometric samples obtained from melting (Mo, Re or Nb)$(Si, Al)_{2.01}$ samples indicated that unalloyed $MoSi_2$ had an average Vickers hardness value of 89968 Hv. Samples containing 2 at.% Al or 1 at.% Nb had average Vickers hardness values of 72928 Hv and 72950 Hv, respectively. 2.5 at.% Re containing samples had the highest hardness value of 103971 Hv. Samples containing 1 at.% Re+2 at.% Al had an average hardness value of 74230 Hv, slightly higher than Al containing samples but significantly lower than both $MoSi_2$ and (Mo, 1 at.% Re)Si_2 samples. Slip lines were observed around indentations in all samples except the unalloyed $MoSi_2$ and (Mo, 2.5 at.% Re)Si_2 samples. Samples containing 1 at.% Nb+2 at.% Al did not exhibit any improvements in the mechanical properties and were excluded from further considerations.

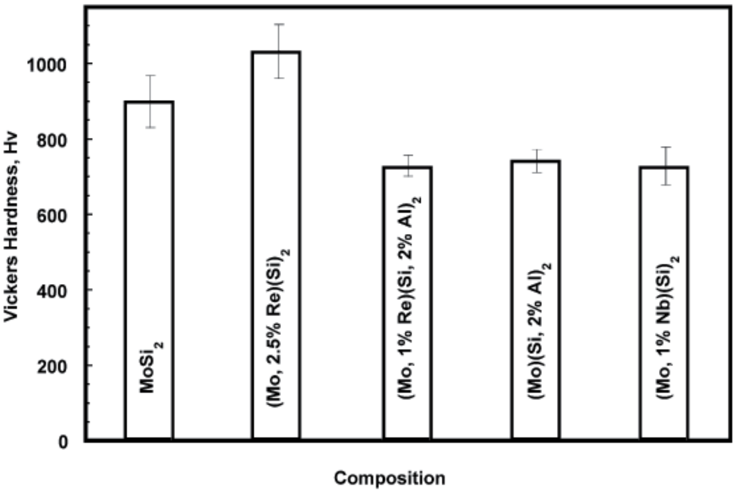


Fig. 6. Vickers hardness values for polycrystalline samples of all composition under investigation. Hardness of polycrystalline Al containing samples is also compared to the values obtained on monocrystalline Al containing samples on (100) and (001) surfaces of the crystal.

4.2 Yield strength

Compression testing at room temperature and at 1600°C was used to determine the 0.2% offset yield strength for all polycrystalline samples. Unalloyed $MoSi_2$ and 2.5 at.% Re containing samples could not be deformed plastically below 900 and 1200° C, respectively. Below these temperatures, the aforementioned samples would undergo brittle fracture during compression testing. The addition of 2.5 at.% Re increased the BDTT, in compression, of $MoSi_2$ by about 300° C while increasing its yield strength from 14 MPa to 170 at 1600° C. Among alloying elements investigated here, 2.5 at.% Re was most effective in increasing strength at 1600° C. The addition of 2 at.% Al was effective in both increasing the high temperature strength to 55 MPa and lowering the BDTT to 425° C. Mo(Si, 2 at.% Al)$_2$ samples exhibited the lowest room temperature yield strength of 415 MPa. Addition of 1 at.% Re+2 at.% Al combined the beneficial effects of both alloying elements and resulted in enhanced ambient temperature compressive plasticity and high temperature strength compared to the unalloyed samples.

However, the improvements in room temperature plasticity was less than that of samples alloyed with 2 at.% Al alone as evident from the value of the room temperature yield strength of (Mo, 1 at.% Re)(Si, 2 at.% Al)$_2$ alloy, 670 MPa. Similarly, the enhancement in high temperature strength was less than that of only Re containing samples but greater than that of only Al containing samples. The effects of 1 at.% Nb as an alloying element by itself in lowering BDTT and enhancing high temperature strength of MoSi$_2$ was more pronounced than the combined effects of 1 at.% Re and 2 at.% Al. The room temperature yield strength of (Mo, 1 at.% Nb)Si$_2$, 500 MPa, was higher than that of 2 at.% Al containing samples and lower than that of (Mo, 1 at.% Re)(Si, 2 at.% Al)$_2$ samples. The 0.2% offset strength of (Mo, 1 at.% Nb)Si$_2$ samples at 1600°C, 143 MPa, was an order of magnitude greater than that of unalloyed MoSi$_2$ (Sharif et al., 2001).

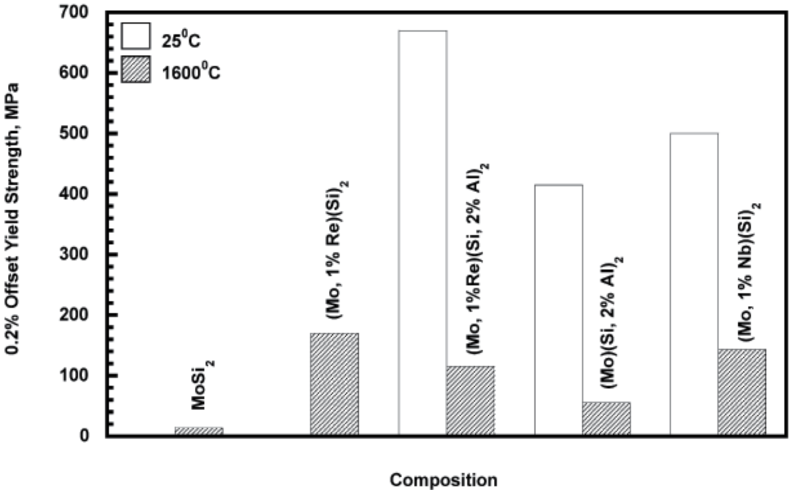


Fig. 7. Effects of alloying on the room temperature and high temperature (1600°C) strength of MoSi$_2$.

4.3 Ductility

Molybdenum disilicide crystallizes in an ordered body centered tetragonal structure with a=0.320 nm and c=0.785 nm, formed by alternate stacking of single Mo and double Si (001) layers. With its high temperature ductility and exceptional resistance to corrosion and fatigue crack growth, MoSi$_2$ combines the toughness of a metal with the strength of a ceramic and is a promising candidate to replace nickel alloys in the next generation of high-temperature gas turbines. Unfortunately, it undergoes a ductile–brittle transition (DBT) at 1200°C, with the fracture toughness dropping to 2–3 MPa m$^{1/2}$, well below the minimum of 20 MPa m$^{1/2}$ required for engine applications. This brittleness at low temperature means that MoSi$_2$ must be formed by costly electro-discharge machining and places a severe limitation on its potential technological utility. However, there is a reasonable chance that the DBT in MoSi$_2$ may be manipulated or even eliminated. Many of the slip systems in MoSi$_2$ are ductile and it is only for a stress axis near [001] that a DBT is observed.

It is desirable therefore, to alter the properties of MoSi$_2$ in very specific ways. This can be, and has in the past been, attempted by heuristically changing the composition or structure of the material and studying experimentally the effect of these changes. It will be argued in this paper that advances in the theory of bonding in solids, based on quantum mechanical density functional calculations, offer an alternative route which can be used as a cost-effective precursor to experiment. The need, in the case of $MoSi_2$, is for an element or elements which can be introduced at microalloy levels (less than 5%) and which will perturb the brittle–ductile behavior in favor of ductility without adversely affecting the advantageous physical properties. While the method of choice would normally be an atomistic calculation, bonding in $MoSi_2$ is known to have hybrid metallic and covalent character. Determination of the effects of alloying on such bonding requires accurate quantum mechanical treatment of the electrons, and generation of reliable interatomic potentials, which are an essential prerequisite to atomistic methods, is impractical1. Instead, use is made of recent advances in the theory of dislocation nucleation and mobility which provide approximate links between these properties and the generalized stacking fault energy surface, which can be calculated accurately using first principles quantum mechanical techniques. A similar approach has been used successfully by two of the present authors to investigate the DBT in silicon. Even with these gross approximations, the numerical work is intensive. The calculations are restricted to small supercells, with correspondingly large alloy content, and the effects of true microalloying must be estimated by interpolation. An overview of the experimental background will be presented in the next section.

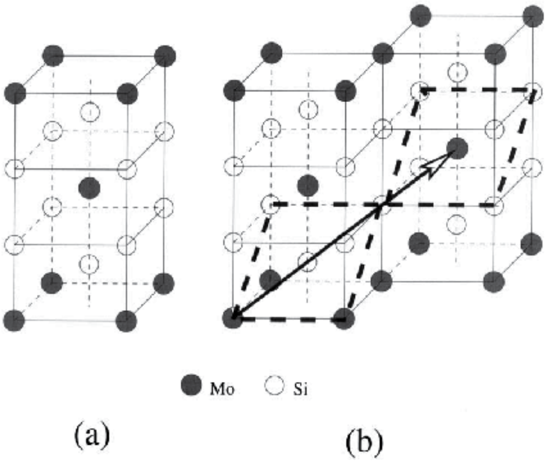


Fig. 8. Crystal structure of $MoSi_2$. (a) Unit cell for the body centered C11b structure; solid circles represent Mo atoms and open circles represent Si atoms. (b) (013) plane and the Burgers vector for {013}⟨331⟩ slip systems (Waghmare et al., 1999).

4.4 Creep

The creep behavior of $MoSi_2$-based materials has been extensively studied. It has been observed that the grain size has a large effect on creep resistance of monolithic $MoSi_2$. Reinforcing with SiC also refined grain size that enhanced creep rates overshadowing any beneficial effects of reinforcement Increased creep resistance has been noted only when volume fractions of SiC are above 20%. Another important factor strongly affecting the creep strain rate of $MoSi_2$ is the presence of silica particles (SiO_2). During high temperature deformation, the SiO_2 particles at the grain boundaries flow to form intergranular film, which slides or cracks. A high volume fraction of SiO_2 and reduction in grain size, both enhance creep rates; but it is of interest to examine how the two are interrelated. Alloying of polycrystalline $MoSi_2$ with Al and C converts SiO_2 to Al_2O_3 and SiC, respectively, which leads to the enhancement in creep resistance. When C is added, oxygen is got rid off in the form of CO or CO_2, which may leave behind fine pores, which are difficult to close. When Al is added, the reaction between Al and SiO_2 forms Al_2O_3 and Si. The Si may remain in elemental form or react with Mo_5Si_3 particles that are present in small volume fractions. These particles form due to partial oxidation of $MoSi_2$ during hot pressing, particularly when vacuum is low. The probability of the reaction between free Si and residual Mo_5Si_3 in the present study is high, as free Si has been observed in the microstructure only very rarely. The figure below presents the effect of alloying of single and polycrystalline $MoSi_2$ with Al, on the creep rates at 1300°C. The single crystals of $Mo(Si_{0.97}\ Al_{0.03})_2$, with hexagonal C40 or hP9 (Pearson's symbol) structure, have shown higher creep rates, compared to those of single crystals of $MoSi_2$ along the [0 15 1] orientation of stress axis. However, the trend reverses with change of stress-axis to [001] direction. On the other hand, polycrystalline $MoSi_2$–5.5Al alloy has shown improvement in creep resistance, compared to polycrystalline $MoSi_2$ at 1300°C. Unlike $Mo(Si_{0.97}\ Al_{0.03})_2$, the matrix phase of $MoSi_2$–5.5Al has tetragonal, C11b structure.

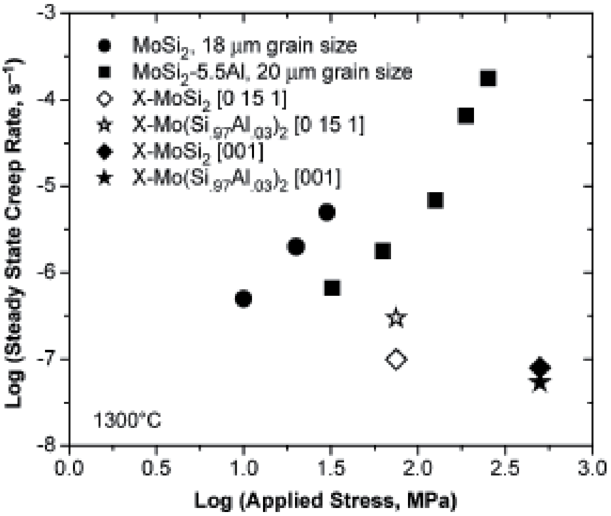


Fig. 9. Comparison of steady state creep rates, measured at 1300°C on $MoSi_2$ and $MoSi_2$–5.5Al alloy, as well as single crystals of $MoSi_2$ and $Mo(Si_{0.97}\ Al_{0.03})_2$ tested with [0 15 1] and [001] orientations. Single crystals are marked as X (Mitra et al., 2004).

As expected, the creep rates of polycrystalline $MoSi_2$ and $MoSi_2$–5.5Al alloy are higher compared to those of single crystals at 1300°C, because of the role of grain boundaries at 0.68 Tm: In the present investigation, samples of $MoSi_2$ with varying grain sizes and SiO_2 contents, as well as those of $MoSi_2$–20 vol% SiC composite and $MoSi_2$–Al alloys have been creep tested at 1200°C and their behaviors analyzed. The values of activation volume and threshold stress have been calculated. These provide an insight into the ratecontrolling and strengthening mechanisms. The creep behavior of the above materials has also been compared with deformation behavior under constant strain rate tests.

4.5 Plastic deformation

Monolithic $MoSi_2$ exhibits only a modest value of fracture toughness at low temperatures and inadequate strength at high temperatures. Thus, many of recent studies on the development of $MoSi_2$-based alloys have focused on improving these poor mechanical properties through forming composites with ceramics and with other silicides. These properties have recently been reported to be significantly improved in composites formed with Si_3N_4 and SiC. However, the volume fraction of Si_3N_4 and SiC ceramic reinforcements in these $MoSi_2$-composites generally exceeds 50%. Further improvements in mechanical properties of these composites will be achieved if those of the $MoSi_2$ matrix phase are improved. The present study was undertaken to achieve this by alloying additions to $MoSi_2$. Transition-metal atoms that form disilicides with tetragonal C11b, hexagonal C40 and orthorhombic C54 structures are considered as alloying elements to $MoSi_2$. These three structures commonly possess (pseudo-) hexagonally arranged $TMSi_2$ layers and differ from each other only in the stacking sequence of these $TMSi_2$ layers; the C11b, C40 and C54 structures are based on the AB, ABC and ADBC stacking of these layers, respectively. W and Re have been known to form a C11b disilicide with Si and they are believed to form a complete C11b solid-solution with $MoSi_2$, although recent studies have indicated that the disilicide formed with Re is an off-stoichiometric (defective) one for-mulated to be $ReSi_{1.75}$ having a monoclinic crystal structure. The details of our crystal structure assessment for $ReSi_{1.75}$ as well as phase equilibria in the $MoSi_2$±$ReSi_{1.75}$ pseudobinary system will be published elsewhere. Large amounts of alloying additions are possible for these alloying elements, and high temperature strength is expected to be improved through a solid solution hardening mechanism since the hardness of both WSi_2 and $ReSi_{1.75}$ is reported to be larger than that of $MoSi_2$. The yield strength of $MoSi_2$ powder compacts is greatly increased when WSi_2 is alloyed with $MoSi_2$ by more than 50 vol%. In addition, our previous study on single crystals of $MoSi_2$±WSi_2 solid solutions has indicated that the compression yield stres above 1200^o C greatly increases when the WSi_2 content in the solutions exceeds 50 vol.%. However, low temperature deformability may be declined upon alloying with these elements because of the increased strength. Indeed, the room temperature hardness of both $MoSi_2$± WSi_2 and $MoSi_2$± $ReSi_{1.75}$ solid solutions is reported to monotonically increase with the increase in either WSi_2 or $ReSi_{1.75}$ content.

V, Cr, Nb and Ta have been known to form a C40 disilicide with Si. Al is also known to transform $MoSi_2$ from the C11b to the C40 structures by substituting it for Si. Of the five slip systems identified to be operative in $MoSi_2$, slip on {110}<111> is operative from 500^oC. 1/2<111> dislocations of this slip system are reported to dissociate into two identical 1/4<111> partials separated by a stacking fault. The stacking across the fault is ABC and resembles the stacking of (0001) in the C40 structure. Hence, the addition of elements that form a C40 disilicide may cause the energy difference between C11b and C40 structures to decrease so that the energy of the stacking fault would also be decreased, although the solid solubility of these alloying elements in $MoSi_2$ has been reported to be rather limited to the level of a few atomic %. From this point of view, we may expect that the deformability of $MoSi_2$ at low temperatures increases upon alloying with elements that form a C40 disilicide. This is consistent that V and Nb may enhance the ductility of $MoSi_2$. Indeed, room temperature hardness of $MoSi_2$ polycrystals decreases upon alloying with Cr, Nb, Ta and Al and similar observations were made for Al-bearing $MoSi_2$ polycrystals. Compression deformation experiments made so far on ternary $MoSi_2$ single crystals containing these elements have focused attention to the high temperature deformation behavior. The yield stress of $MoSi_2$ increases upon allying with Cr above 1100ºC. A similar observation was made for Nb-bearing $MoSi_2$ at 1400ºC. However, since these compression experiments were made only at high temperatures above 1100ºC, almost nothing is know about the low temperature strength and deformability of these ternary $MoSi_2$ single crystals. V, Cr, Nb and Al that form a disilicide with the C40 structure and W and Re that form a disilicide with the C11b structure as alloying elements to $MoSi_2$, and investigated the deformation behavior of single crystals of $MoSi_2$ containing these elements in a wide temperature range from room temperature to 1500ºC. The crystal orientations investigated were the [0 15 1] orientation, in which slip on {110}<111> is operative, and the [001] orientation, in which the highest strength is obtained at high temperatures for binary $MoSi_2$.

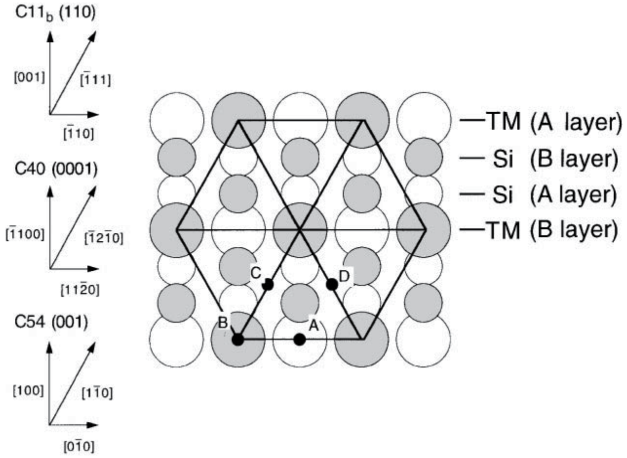


Fig. 10. Atomic arrangement on $TMSi_2$ layers corresponding to {110}, (0001) and (001) planes in the C11b, C40 and C54 structures, respectively. The stacking positions of A±D and crystallographic directions with respect to these three structures are indicated (Inui et al., 2000)

5. Development of $MoSi_2$ – Si_3N_4 composites

Interest in fiber reinforced ceramic matrix composites (FRCMCs) has increased steadily over the past 15 years, and several refined silicon-base composite systems are now being produced commercially. These composites offer very good structural stiffness, high specific strength to weight, and good high temperature environmental resistance. Industrial applications include, hot gas filters, shrouds, and combuster liners. In addition, silicon-base ceramic composites are being considered for gas turbine hot gas flow path components, e.g. combusters transition pieces, and nozzles. The manufacturers of liquid rocket engines are also looking to ceramic composites in hopes of obtaining better efficiency in the next generation of designs. Applications include inlet nozzles, fuel turbopump rotors, injectors, combustion chambers, nozzle throats, and nozzle extensions. In order to maximize properties, materials developers have now begun to pay more attention to engineered interfaces between the matrix and the fiber reinforcement. If the interfacial debonding energy and sliding resistance is low, the fibers can pull away or out of the matrix and form bridges behind the advancing crack front which renders these otherwise brittle materials acceptably compliant. Unfortunately, most of the incremental toughening and attended fiber pullout occurs at engineering strains that exceed the strain which occurs at the ultimate strength of the material, and the toughening benefit is, therefore, not useful for design purposes. Thus, the degree of incremental toughening that occurs during the inelastic portion of the stress-strain curve up to the ultimate strength will have to be improved before increased use of FRCMCs can be realized.

It has also been shown that $MoSi_2$ offers the potential of combining the effects of second phase reinforcements with metallurgical alloying to improve mechanical properties without degrading oxidation resistance. For example, high temperature (1200°C) creep was reduced by a factor of 10 by alloying with WSi_2 and by another factor of 10–15 with the addition of SiC whiskers. Additions of carbon and zirconia have also proven to be beneficial. Carbon reacts with the oxygen impurities in $MoSi_2$ to improve the toughness at high temperatures by removing SiO_2, and leaves behind a compatible SiC phase. Zirconia, which is thermochemically stable with $MoSi_2$, can also be used to increase fracture toughness. a 20 v/o loading of particulate ZrO_2 increased the low temperature fracture toughness of $MoSi_2$ by a factor of four. The toughening transformation occurs above the ductile–brittle transition and therefore enhances the low temperature properties. In an attempt to reinforce $MoSi_2$ with SCS-6 silicon fibers, it was discovered that the large thermal expansion difference between the matrix and fiber introduced matrix cracking upon cooling from the densification temperature. This problem was solved by adding Si_3N_4 to form a two phase composite matrix with a coefficient of expansion that more closely matches the fiber. There is no reaction between Si_3N_4 and $MoSi_2$, even at fabrication temperatures as high as 1750°C. No gross cracking occurs on cool down, although some microcracking has been observed and is a function of grain size. The critical particle size below which microcracking will not occur was calculated to be 3 mm. Coarse phase $MoSi_2$– Si_3N_4 composites also exhibit higher room temperature toughness than fine phase material, reaching values of 8 $MPam^{1/2}$. Fracture toughness also increases with temperature and the trend is quite significant above 800°C with toughness values exceeding 10 $MPam^{1/2}$.

However, the fine phase materials are stronger than the coarse phase materials with bend strengths reaching 1000 MPa. The $MoSi_2$– Si_3N_4 composites have also been shown to exhibit R-curve behavior, and crack deflection and particle pullout have been observed. Molybdenum disilicide does not have good creep resistance at high temperatures above its brittle-to-ductile transition. When high volume fractions of Si_3N_4 are added, creep is improved significantly and the activation energy is comparable to monolithic silicon nitride.

Additions of carbon can also improve creep resistance as well as toughness. In situ processing with carbon additions have produced material with creep resistance comparable to Ni-base superalloys. The silicon-base composite systems of current interest typically utilize carbon or silicon carbide fibers and silicon nitride or silicon carbide matrices. A popular designation is to display the fiber first followed by the matrix phase, e.g. C/SIC, SiC/SiC, and SiC/Si_3N_4. Mixtures composed of $MoSi_2$ and Si_3N_4 form two phase composites that are also candidates as matrices in C or SiC fiber reinforced composite systems. The combination of a fiber reinforced composite with a composite matrix becomes a little confusing, but can be represented by SiC/ $MoSi_2$– Si_3N_4.

The matrix properties for several silicon-base composite systems and their fiber properties are presented below. The high coefficient of thermal expansion of SCS-6, i.e. 4.8 x 1^{-6} °C^{-1}, indicates that this fiber will have a larger expansion coefficient than the matrix phase for SiC/SiC and SiC/ Si_3N_4 fiber reinforced composite systems. It is generally more difficult to weave and fabricate structural components from large diameter fibers. The size can also influence properties. For example, toughness scales directly with fiber radius while the matrix cracking strength is inversely proportional to the radius. The properties for $MoSi_2$–Si_3N_4 compare quite favorably with both SiC and Si_3N_4. The fracture toughness is slightly above the middle range for SiC and comparable to Si_3N_4. The highest matrix toughness values are on the order of 10 $MPam^{1/2}$ for in situ toughened silicon nitride. The toughness for all the candidate matrix materials will depend upon processing conditions and microstructure. It is anticipated that in situ toughening of the silicon nitride phase in the two phase $MoSi_2$–Si_3N_4 composites should yield further improvements for this matrix candidate.

The onset of nonlinear behavior s often found in tension tests marked by a distinct load drop, indicating the initiation of matrix cracking, whereas in flexure, this important feature may go undetected. It is the region between matrix cracking and fiber bundle failure at maximum load where matrix enhancement can make the greatest contribution. Many FRCMC composites exhibit much of their toughness beyond this point, because as the fibers pull away from the matrix, they bridge cracks and impose traction forces that retard crack growth. Pullout toughening extends life after failure, and adds a margin of safety from the catastrophic nature of failure often found in brittle materials, but this phenomenon is not useful as a design property. Three of the composites exhibit matrix cracking stresses in the range of 150–175 MPa. Of the four systems, the SiC/Si_3N_4 has the highest matrix cracking stress, which is near 350 MPa. This composite also has the largest coefficient of thermal expansion mismatch, CTE, with the fiber having the larger value.

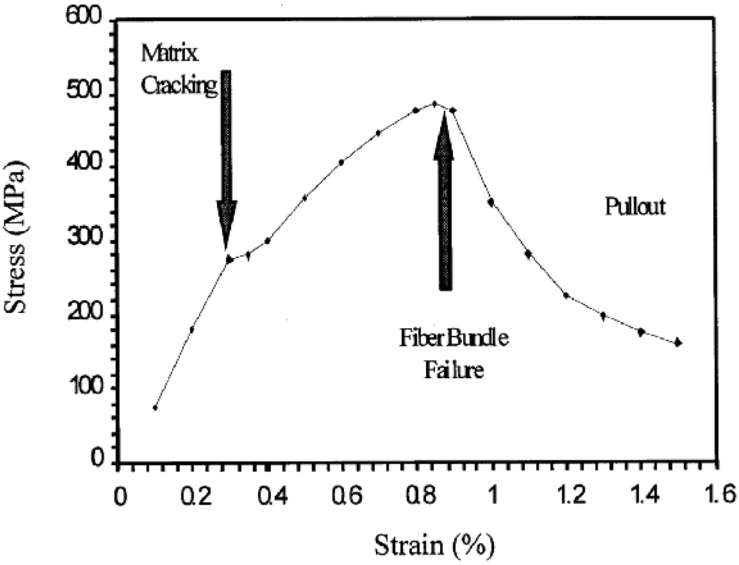


Fig. 11. Stress–strain behaviour for ceramic fiber reinforced: ceramic matrix composites.

Upon cooling from the consolidation temperature, the fiber can theoretically contract and debond from the matrix unless there is enough surface roughness for asperity contact. Upon reloading, the matrix will not efficiently transfer stress to the fiber unless it is in intimate contact. However, the high elastic modulus, as indicated by the stress– strain curve, suggests that good load transfer is occurring in this system. Inelastic behavior starts at about 350 MPa, but the ultimate is reached at about 450 MPa which is well below the expected fiber bundle failure. The ultimate strength for the Silcomp matrix composite is on the order of 650 MPa and is in reasonable agreement with bundle fiber failure, as is the $MoSi_2$–$0.5Si_3N_4$ matrix composite.

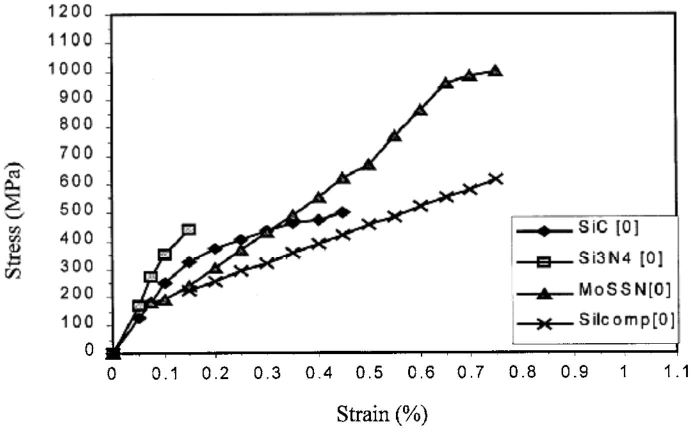


Fig. 12. Tensile stress–strain curves for uniaxial SCS-6 reinforced silicon-base composite systems (Courtright, 1999).

A number of composite approaches have been developed to toughen brittle high temperature structural ceramic materials. Many of these approaches have also been applied to high temperature structural silicides.

The $MoSi_2$–Si_3N_4 composite system is an interesting and important one. Si_3N_4 is considered to be the most important structural ceramic, due to its high strength, good thermal shock resistance, and relatively high (for a structural ceramic) room temperature fracture toughness. Si_3N_4 and $MoSi_2$ are thermodynamically stable species at elevated temperatures.

Property	**$MoSi_2$**	**Si_3N_4**
Density(g/cm^3)	6.2	3.2
Thermal expansion coefficent (10^{-6}/^{o}C)	7.2	3.8
Thermal conductivity (W/mK)	65	37
Melting point(^{o}C)	2030	2100
Creep resistance (^{0}C)	1200	1400
Toughness	Low	Low
Oxidation resistance	Good	Excellent
Structural stability	Good	Good
Intricate machinability	Good	Difficult
Cost	Low	High

Table 2. Some physical and thermal properties of $MoSi_2$ and Si_3N_4 (Nathesan & Devi, 2000).

When composites were synthesized with elongated Si_3N_4 grains toughness can reach to 15 MPa $m^{1/2}$ (Nathal & Hebsur, 1997).

Designation	Microstructure
MS-60	Fully dense β-Si_3 N_4, with long whisker-type morphology
MS-70	Fully dense β- Si_3 N_4, with long whisker-type morphology
MS-80	Not fully dense β- Si_3 N_4, with blocky morphology
MS-50	Fully dense α- Si_3 N_4, with blocky morphology
MS-40	Not fully dense α- Si_3 N_4, fine grained $MoSi_2$ and blocky α- Si_3 N_4

Fig. 3. Microstructures of different $MoSi_2$– Si_3N_4 composites.

However, a drawback of transformation toughening is that toughness decreases with increasing temperature, due to the thermodynamics of the phase transformation. Discontinuously reinforced ceramic composites have typically employed ceramic whiskers or particles as the reinforcing phases. An example is SiC whisker reinforced Si_3N_4. Toughening mechanisms here are crack deflection and crack bridging. Discontinuous ceramic composites can reach toughness levels of 10 $MPam^{1/2}$. One important variant of this approach is the in-situ toughening of Si_3N_4 due to the presence of elongated Si_3N_4 grains. By way of comparison to structural ceramics, the room temperature fracture toughness of polycrystalline $MoSi_2$ is approximately 3 $MPam^{1/2}$, while the room temperature fracture toughness of equiaxed polycrystalline Si_3N_4 which is densified without densification aids is also 3 $MPam^{1/2}$ (Petrovic, 2000). For comparison, two monolithic ceramics SiC and Si_3N_4 are also included in the figure. Further improvement in room temperature fracture can be achieved by microalloying $MoSi_2$ with elements like Nb, Al an Mg or by randomly oriented long whisker type β-Si_3N_4 grains (Hebsur, 1999).

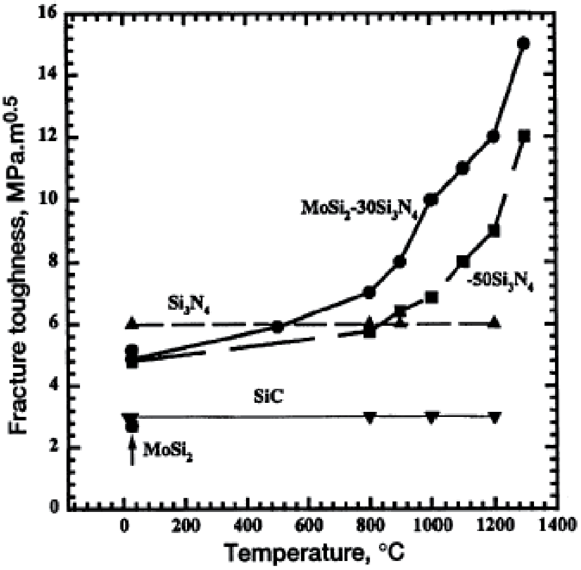


Fig. 13. Temperature dependence of fracture toughness of $MoSi_2$-based materials compared with ceramic matrices (Hebsur, 1999).

$MoSi_2$–Si_3N_4–SiC hybrid discontinuous particle-continuous fiber composites have been developed with excellent room temperature fracture toughness, thermal shock resistance, and thermo-mechanical impact behavior. These hybrid composites consist of $MoSi_2$–Si_3N_4 particulate composites which form the matrix for SiC continuous fibers. The $MoSi_2$–Si_3N_4 portion of the hybrid composites has two functions. First, additions of 30-50.% Si_3N_4 to the $MoSi_2$ completely eliminates the oxidation pest behavior at the intermediate 500°C temperature. Second, the Si_3N_4 addition aids to match the thermal expansion coefficient of the matrix to that of the SiC fibers. This prevents thermal expansion coefficient mismatch cracking in the hybrid composite matrix.

Figure 14.(a) shows a SEM back scattered image of a fully dense $MoSi_2$-βSi_3N_4 composite (MS-70). During processing, the original α-Si_3N_4 powder particles are transformed into randomly oriented whiskers of β-Si_3N_4. These long whiskers are well dispersed throughout the material and appear to be quite stable, with very little or no reaction with the $MoSi_2$, even at 1900°C. In some isolated areas, the Mo_5Si_3 phase is visible. Figure 14.(b) shows a back scattered image of $MoSi_2$-βSi_3N_4 (MS-80) with the β-Si_3N_4 exhibiting a blocky aggregate-type morphology.

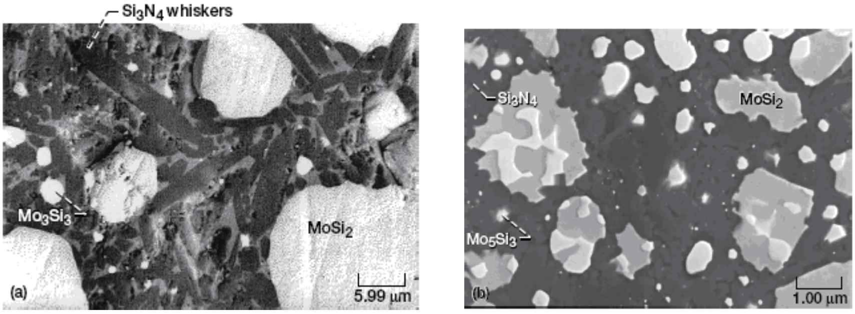


Fig. 14.(a) randomly oriented in-situ grown long whiskers of β-Si_3N_4 and large $MoSi_2$ particle size, (b) Si_3N_4 has a blocky particulate structure (Hebsur et al., 2001).

Density of (MS-70) is 4.57±0.01g/cm^3 and Vickers microhardness is 10.7±0.6GPa. Figure 15. shows the coefficient of thermal expansion as a function of temperature for (MS-70). From this data the average coefficient for expansion of this composite material is about 4.0ppm/°C.

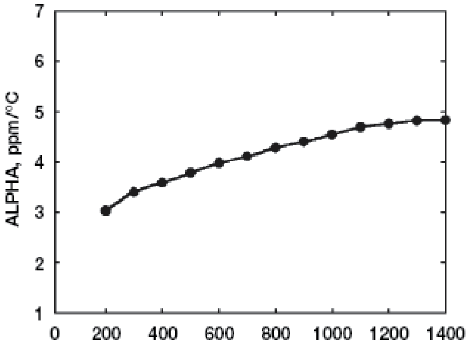


Fig. 15. The coefficient of linear expansion for $MoSi_2$-βSi_3N_4 (Hebsur et al., 2001).

The oxidation behaviour of a MoSiB alloy is also included for comparison. 500°C is the temperature for maximum accelerated oxidation and pest for $MoSi_2$-base alloys. There is interest in this alloy, over $MoSi_2$, for structural aerospace applications due to its attractive high temperature oxidation resistance (Bose, 1992; Berczik, 1997).

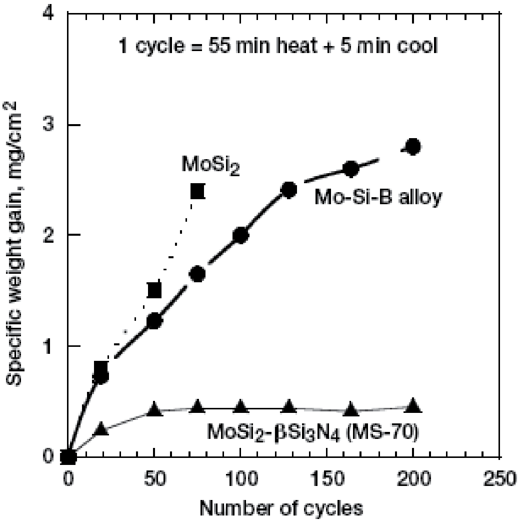


Fig. 16. Specific weight gain versus number of cycles of (MS-70) at 500°C (Hebsur, 2001).

However, the (MS-70) shows very little weight gain compared to binary $MoSi_2$ and the MoSiB alloy, indicating the absence of accelerated oxidation. In contrast the binary $MoSi_2$ and MoSiB alloys exhibits accelerated oxidation followed by pesting.

6. Conclusion

Based on the cyclic oxidation properties at 900°C, the family of $MoSi_2$-Si_3N_4 composites show promise for aircraft applications. The composites do not exhibit the phenomena of pesting, and the weight gain after 500h is negligible and superior to base line hybrid composites.

A wide spectrum of mechanical and environmental properties have been measured in order to establish feasibility of an $MoSi_2$ composite with Si_3N_4 particulate. The high impact resistance of the composite is of particular note, as it was a key property of interest for engine applications. Processing issues have also been addressed in order to lower cost and improve shape making capability. These results indicate that this composite system remains competitive with other ceramics as potential replacement for superalloys.

7. References

Vaseduvan, A.K. & Petrovic, J.J. (1992). A Comparative Overview of Molybdenum Disilicide Composites. *Materials Science and Engineering*, A155, pp. 1-17

Tein, J. K. & Caulfield, T. (1989). *Superalloys, Supercomposites, and Superceramics*, Boston Academic Press, New York

Soetching, F.O. (1995). A Design Perspective on Thermal Barrier Coatings. *Proceedings of a Conference at NASA Lewis Research Center*, September 1994

Bradley, E.F. (1988). *Superalloys a Technical Guide*, Metals Park, Ohio

Misra, A.; Sharif, A.A. & Petrovic, J.J. (2000). Rapid Solution Hardening at Elevated Temperatures by Substitional Re Alloying in $MoSi_2$, *Acta Mater*, Vol.48, pp. 925-932

Akkus, I. (1999). The Aluminide Coating of Superalloys with Pack Cementation Method. *Journal of Institution of Science Osmangazi University*, Vol. 18, pp. 27-28

Meschter, P.J. (1992). Low Temperature Oxidation of Molybdenum Disilicide, *Metallurg. Trans. A*, Vol. 23A, pp. 1763-1772

Liu, Q.; Shao G. & Tsakiropoulos, P. (2001). On the Oxidation Behaviour of $MoSi_2$. *Intermetallics*, Vol. 8, pp. 1147-1158

Chou, T.C. & Nieh, T.G. (1992). New Observation of $MoSi_2$ Pest at 500°C. *Script. Metallurg. Mater.*, Vol. 26, pp. 1637-1642

Chou, T.C. & Nieh, T.G. (1993). Pesting of the High Temperature Intermetallic $MoSi_2$. *Journal of Materials*, Vol. 30, pp. 15-22

Wang, G.; Jiang, W. & Bai, G. (2003). Effect of Addition of Oxides on Low-Temperature Oxidation of Molybdenum Disilicide. *Journal of American Ceramic Society*, Vol. 86, pp. 731-734

Chen, J.; Li, C.; Fu, Z.; Tu, X.; Sundberg, M. & Pompe, R. (1999). Low Temperature Oxidation Behaviour of $MoSi_2$Bbased Material. *Materials Science and Engineering*, Vol. A26, pp. 239-244

Sharif, A.A.; Misra, A. & Petrovic, J.J. (2001). Rapid Solution Hardening at Elevated Temperatures by Substitional Re Alloying in $MoSi_2$, *Acta Mater*, Vol. 48, pp. 925-932

Waghmare, U.V.; Bulatov, V. & Kasiras, E. (1999). Microalloying for Ductility in Molybdenum Disilicide, *Materials Science and Engineering*, Vol. A261, pp. 147-157

Mitra, R.; Sadananda, K. & Feng, C.R. (2004). Effect of microstructural parameters and Al Alloying on Creep Behavior, Threshold Stress and Activation Volumes of Molybdenum Disilicides. *Intermetallics*, Vol. 12, pp. 827-836

Inui, H.; Ishikawa, K. & Yamaguchi, M. (2000). Effects of Alloying Elements on Plastic Deformation of Single Crystals of $MoSi_2$. *Intermetallics*, Vol. 8, pp. 1131-1145

Courtright, E.R. (1999). A Comparison of $MoSi_2$ Matrix Composites with Other Silicon-Base Composite Systems. *Materials Science and Engineering*, Vol. A261, pp. 53-63

Natesan, K. & Deevi, S.C. (2000). Oxidation Behaviour of Molybdenum Disilicides and Their Composites. *Intermetallics*, Vol. 8, pp. 1147-1158

Nathal, M.V. & Hebsur, M.G. (1997). Strong, Tough, and Pest-Resistant $MoSi_2$-Base Hybrid Composite for Structural Applications. *Structural Intermetallics*, Warrendale (USA): TMS, pp. 949-953

Petrovic, J.J. (2000). Toughening Strategies for $MoSi_2$-Based High Temperature Structural Silicides, *Intermetallics*, Vol. 8, pp. 1175-1182

Hebsur, M.G. (1999). Development and Characterization of $SiC_{(f)}/MoSi_2$-$Si_3N_{4(p)}$ Hybrid Composites. *Material Science and Engineering*, Vol. A261, pp. 24-37

Hebsur, M.G.; Choi, S.R.; Whittenberger, J.D.; Salem, J.A. & Noebe, R.D. (2001). Development of Tough, Strong, and Pest-Resistant $MoSi_2$-βSi_3N_4 Composites for

High Temperature Structural Applications, *International Symposium on Structural Intermetallics*, NASA, 2001

Bose, S. (1992). *High Temperature Silicides*, North-Holland, NY

Berczik, D.M. (1997). Oxidation Resistant Molybdenum Alloy, U.S. Patent, No. 5, 696, 150

Part 2

Aircraft Control Systems

6

An Algorithm for Parameters Identification of an Aircraft's Dynamics*

I. A. Boguslavsky
State Institute of Aviation Systems, Moskow Physical Technical Institute
Russia

1. Introduction

Development of efficient parameter identification methods for the model of a dynamic system based on real-time measurements of some components of its state vector should be taken as one of the most important problems of applied statistics and computational mathematics. Calculating the motion of the system given the initial conditions and its mathematical model is conventionally called the direct problem of dynamics. Then, the inverse problem of dynamics would be the problem of identifying the system model parameters based on measurements of certain components of the state vector provided that the general structural scheme of the model is known from physical considerations. Such an inverse problem corresponds to identification problem for the dynamic system representing an aircraft. In this case, the general structural scheme of the model (motion equations) follows from the fundamental laws of aerodynamics.

In many cases, modern computational methods and wind tunnel experiments can provide sufficient data on nominal parameters of the mathematical model - nominal aerodynamic characteristics of the aircraft. Nevertheless, there exist problems [1] that require correcting nominal parameters based on measurements taken in real flights. These imply

(1) verifying and interpreting theoretical predictions and results of wind tunnel experiments (flight data can also be used to improve ground prediction methods),

(2) obtaining more exact and complete mathematical models of the aircraft dynamics to be applied in designing stability enhancement methods and flight control systems,

(3) designing flight simulators that require more accurate dynamic aircraft profile in all flight modes (many motions of aircrafts and flight conditions can be neither reconstructed in the wind tunnel nor calculated analytically up to sufficient accuracy or efficiency),

(4) extending the range of flight modes for new aircrafts, which can include quantitative determination of stability and impact of control when the configuration is changed or when special flight conditions are realized,

(5) testing whether the aircraft specification is compliant.

*This work has been supported by the Russian Foundation for Basic Research.

Furthermore, dimensionless numbers at the nodes of one-or two-dimensional tables found in wind tunnel experiments serve as nominal values in the aerodynamic parameter identification problem of the aircraft. This causes the vector that corrects these parameters determined by the algorithm processing digital data flows received from the aircraft sensors to have a significant dimension of the order about several tens or hundreds.

It is worth noting that the USA (NASA) is doing extensive work on theoretical and practical aircraft identification by test flights. In 2006 alone, in addition to many journal publications, American Institute of Aeronatics and Astronatics (AIAA) published three fundamental monographs [1-3] on the subject. An implementation of multiple NASA recommended algorithms for identification problems, SIDPAS (Systems Identification Programs for Aircraft) software package written in MATLAB M-files language is available on the Internet as an appendix to [1]. Various existing identification methods published in monographs on statistics and computational mathematics are widely reviewed in [1].

For the most general identification method, one should take the known nonlinear least squares method [4] that forms the sum of errors squared - differences between the real measurements and their calculated analogues obtained by numerical integration of motion equations of the system for some realization of the vector of unknown parameters.

Successful identification yields the vector of parameters that delivers the global minimum to the above mentioned sum of errors squared. Still, this criterion is statistically valid only for linear identification problems, in which measurements are linear with respect to the unknown vector of parameters.

Implementing the nonlinear least squares method to correct nominal parameters of the aircraft based on its test flight data involves computational challenges. These arise when the dimension of the correction vector is big and the sum of errors squared as the function of the correction vector has multiple relative minimums or when variations of the Newton's method are applied, with the sequence of local linearizations performed to find stationary points of this function. In [1], the regression method supported by *lesq.m, smoo.m, derive.m, and xstep.m* files in SIDPAS is recommended for practical applications.

Suppose the motion equations of the system and the sequence of measurements have the form

$$dx/dt = f(x, \vartheta + \eta, u), ...(0.1)$$

$$y_k = H_k(x(t_k)) + \xi_k, ...(0.2)$$

where $x(t_k)$ is the $n \times 1$-dimensional vector of the system states at the current instant t and at the given instants $t_k, k = 1, ..., N, \vartheta$ is the $r \times 1$-vector of nominal (known) parameters of the system, η is the vector of unknown parameters that serves as the correction vector for the nominal vector ϑ after the results of measurements are stochastically processed, u is the control vector of the system, $f(...)$ is the given vector-function, y_k is the sequence of vectors-results of measurements, $H_k(...)$ is the given vector function, and $\xi_k, k = 1, ..., N$ is the sequence of random vectors-errors of measurements with the given random generator for the mathematical simulation.

We can state the identification problem for the vector η as follows. Find the estimate as the function of the vector Y_N formed of the results of all measurements $y_1, ...yN$.

The regression method given in [1] solves this problem under the following limitations

(1) all components of the state vector can be measured : $y_k = x(t_k) + \xi_k$,

(2) at the measurement instants t_k, the algorithm constructs the estimate of the vector of derivatives dx/dt,

(3) the vector function $f(x, \vartheta + \eta, u)$ linearly depends on the vector η.

These fundamental limitations of the regression method duplicate features of the identification algorithm from [5]. The substantial drawback of the algorithm [5] and the algorithm of the regression method is that they do not allow using the mathematical model to analyze theoretically (without applying the Monte-Carlo method) observability conditions of components of the vector of parameters to be identified for the preliminary given control law for the test flight of the aircraft and information on random errors of its sensors. Note that this is the drawback of all known numerical methods that solve nonlinear identification problems.

Relations (0.1) and (0.2) show that when conditions (1)-(3) are met and N is sufficiently big, the estimation vector satisfies the overdetermind system of linear algebraic equations, with methods to solve it being well known. The given conditions seem to be rather rigid and may be hard-to-implement. For instance, it is arguable whether one can construct the vector of derivatives dx/dt sufficiently accurately given the real turbulent atmosphere conditions, which imply that the outputs of the angle of attack and sideslip sensors inevitably include random and unpredictable frequency components.

All this justifies the development of new identification algorithms that can be applied to dynamic systems of a rather general class and do not possess drawbacks of NASA algorithms. The proposed multipolynomial approximation algorithm (MPA algorithm) serves as such a new identification algorithm.

2. Statement of the problem and basic scheme of the proposed identification algorithm

The general scheme for identifying aerodynamic characteristics of the aircraft by the test flight data is as follows [1]. Motion equations of the aircraft (0.1) and system (0.2) of measurements of motion characteristics of the aircraft are given. The vector ϑ is the vector of nominal aerodynamic parameters determined in the wind tunnel experiment. Calculated by the results of real (test) flight, the vector η is used to correct the vector ϑ.

When the aircraft flies, its computer fixes the digital array of initial conditions and time functions, viz. current control surface angles and measurements of some motion parameters of the aircraft (some components of the vector x(t) of the state of the aircraft) received from its sensors. Note that selecting the criterion for optimal or, at least, rational mode to control the test flight is a separate problem and lies beyond our further consideration. The current motion characteristics measured as the time function such as angles of attack and sideslip and components of the vector of angular velocity and g-load obtained by the inertial system of the aircraft are registered for real (not known for sure) aerodynamic parameters of the aircraft (parameters $\vartheta + \eta$) and can be called measured characteristics of the perturbed motion.

Once the flight under the mentioned (given) initial conditions and time functions (control surface angles) is completed, nominal motion equations (equations of form (1) for $\eta = 0$) are integrated numerically for the nominal aerodynamic parameters of the aircraft. For the calculated characteristics of the nominal motion of the aircraft one should take the obtained data - components of the state vector of the aircraft as the function of discrete time. Differences between measurable characteristics of the perturbed motion and calculated characteristics of the nominal motion serve as carriers of data on the unknown vector η that shows the difference between real and nominal aerodynamic parameters.

The input of the MPA identification algorithm receives the vector of initial conditions and control surface angles as functions of time and arrays of characteristics of nominal and perturbed motions.

The output of the algorithm is $\hat{\eta}(Y_N)$, which is the correction vector for nominal aerodynamic parameters.

The identification algorithm is efficient if the motion equations integrated numerically with the corrected aerodynamic parameters yield such motion characteristics $\vartheta + \hat{\eta}(Y_N)$ (*corrected* characteristics, in what follows) that are close to real (measurable) characteristics.

In this work, we consider the technology of applying the Bayes MPA algorithm [6, 7] to solve identification problems on the example of the aircraft, for which nominal aerodynamic parameters of the pitching motion are the nominal parameters of one of an "pseudo" F-16 aircraft.

We replace real flights by mathematical simulation, with characteristics of the perturbed motion obtained by integrating the motion equations of the aircraft numerically. In these equations, nominal aerodynamic parameters at the nodes of the corresponding tables are changed to random values that do not exceed in modulus the given $25 \div 50$ percents of nominal values at these nodes.

Fundamentally, the MPA algorithm assumes that the vector of unknown parameters η is random on the set of possible flights. We assume that the a priori statistical-generator for computer generated random vectors η and ξ_k is given. This generator makes the algorithm estimating components of the vector η (the identification algorithm) Bayesian. Further, for particular calculations, we assume that random components of the mentioned vectors are distributed uniformly and can be called by the standard Random program in Turbo Pascal.

The MPA algorithm provides the approximation method we implement with the multidimensional power series of the vector $E(\eta|Y_N)$ of the conditional mathematical expectation of the vector η if the vector of measurements Y_N is fixed and a priori statistical data on random vectors η and ξ_k are given.

The vector $E(\eta|Y_N)$ is known to be optimal, in root-mean-square sense, estimate of the random vector η.

We describe the steps of operation of the MPA algorithm when it identifies the vector η[6, 7].

Step 1. Suppose d is a given positive integer number and the set of integer numbers $a_1, ..., a_N$ consists of all nonnegative solutions of the integer inequality $a_1 + ... + a_N \leq d$, the number of which we denote by $m(d, N)$. The value $m(d, N)$ is given by the recurrent formula proved by induction.

$$m(d, N) = m(d-1, N) + (N + d - 1) \cdots N/d!, m(1, N) = N.$$

We obtain the vector $W_N(d)$ of dimension $m(d, N) \times 1$, the components $w_1, ..., w_m(d, N)$ of which are all possible values $y_1^{a_1} ... y_N^{a_n}$ of the form that represent the powers of measurable values.

Then, we construct the base vector $V(d, N)$ of dimension $(r + m(d, N)) \times 1, V(d, N) = \|\eta W_N(d)\|$.

Step 2. We use a known statistical generator of random vectors η and ξ_k to solve repeatedly the Cauchy problem for $Eq.(1)$ for given initial conditions $x(0)$, a control law $u(t)$ and various realizations of random vectors η and xi_k.

We apply the Monte-Carlo method to find the prior first and second statistical moments of the vector $V(d, N)$, i.e., the mathematical expectation $\bar{V}(d, N)$, and the covariance matrix $C_V(d, N) = E((V(d, N) - \bar{V}(d, N))(V(d, N) - \bar{V}(d, N))^T)$.

Implementation of step 2 is a learning process for the algorithm, adjusting it to solve the particular problem described by Eqs. (1) and (2).

Step 3. For given d and N and a fixed vector Y_N, we assign the vector $\widehat{\eta}(W_N(d))$ to be the solution to the estimation problem. This vector gives an approximate estimate of the vector $E(\eta|Y_N)$ that is optimal in the root-mean-square sense on the set of vector linear combinations of components of the vector $W_{N_1}(d)$

$$\widehat{\eta}(W_N(d)) = \sum_{a_1+...+a_N \leq d} \lambda(a_1, ..., a_N) y_1^{a_1} \cdots y_N^{a_N}. \quad (1.1)$$

The vector $\bar{V}(d, N)$ and the matrix $C_V(d, N)$ are the initial conditions for the process of recurrent calculations that realizes the principle of observation decomposition [6] and consists of $m(d, N)$ steps. Once the final step is performed, we obtain vector coefficients $\lambda(a_1, ..., a_N)$ for (1.1). Moreover, we determine the matrix $C(d, N)$, which is the covariance matrix of the estimation errors for the vector $E(\eta_N|Y_N)$ of conditional mathematical expectation estimated by the vector $\widehat{\eta}(W_N(d))$.

Calculating the elements of the matrix $C(d, N)$, we have the method of preliminary (prior to the actual flight) analysis of observability of identified parameters for the given control law, structure of measurements and their expected random errors. Recurrent calculations do not require matrix inversion and indicate the situations when the next component of the vector $W_N(d)$ is close to linear combination of its previous components. To implement the recursion, we process the components of the vector $W_N(d)$ one after another. However, the adjustment of the algorithm performed by applying the Monte-Carlo method to find the vector $\bar{V}(d, N)$ and the matrix $C_V(d, N)$ takes into account a priori ideas on stochastic structure of components of the whole set of possible vectors $W_N(d)$ that can appear in any realizations of the random vectors η and ξ_k allowed by the a priori conditions.

This adjustment is the price we have to pay if we want the MPA algorithm to solve nonlinear identification problems efficiently. This is what makes the MPA algorithm differ fundamentally from, for instance, the standard Kalman filter designed to solve linear identification problems only or from multiple variations of algorithms resulted from attempts to extend the Kalman filter to nonlinear filtration problems.

In [6], a multidimensional analogue of the K. Weierstrass theorem (the corollary of the M. Stone theorem [9]) is used to prove that when the integer d is increases then the error estimates of the vector $E(\eta|Y_N)$ the vector $|\widehat{\eta}(W_N(d)) - E(\eta|Y_N)|$ tend to zero uniformly on some region. Formulas of the recurrent algorithm are given and justified in [6, 7] and in the Appendix.

This scheme for the MPA algorithm operation shows that it can be applied to identify parameters of almost any dynamic system provided that the structures of the motion equations and measurements of form (0.1) and (0.2) and prior statistical generators of random unknown parameters and errors of measurements are given. The MPA algorithm is devoid of the above listed limitations and drawbacks, which gives it substantial advantages over NASA identification algorithms. Apart from errors of computations, the algorithm does not add any other errors (such as errors due to linearization of nonlinear functions) into the identified parameters. Therefore, one should expect that the priori spread of identifiable parameters to be always greater than the posterior spread. This is why we can use iterations.

Let us compare the sequential steps of the standard discrete Kalman filter and the MPA algorithm.

(1) The Kalman filter identifies the vector η, which can be represented by part of components of the state vector of the linear dynamic system for the observations that linearly depend on state vectors. The a priori data are the first and second moments of components of random initial state vectors, uncorrelated random vectors of perturbations and observation errors. We need these data for sequential (recurrent) construction of the estimation vector that is root-mean-square optimal. Usually assigned, a priori data can be also determined by the Monte-Carlo method if the complex mechanism of their appearance is given.

(2) To find an asymptotic solution to the nonlinear identification problem, the MPA algorithm, unlike the Kalman filter, requires a priori statistical data on both the initial and all hypothesized future state vectors of the dynamic system and observations. These a priori data are represented by the first and second statistical moments for the random vector V(d, N): the vector $\overline{V}(d, N))$ and the matrix $C_V(d, N)$. These moments are calculated using the Monte-Carlo method. However, there are cases when they can be obtained by numerical multidimensional region integration.

(1.1) Once conditions from (1) are met, the Kalman filter constructs the recurrent process, at every step of which the current estimation vector optimal in the root-mean-square sense and the covariance matrix of errors of the estimate are calculated.

(2.1) Based on (2), the MPA algorithm implements the recurrent computational process that do not require matrix inversion. At each step of the process, we construct

i. the current estimation vector $\hat{\eta}(W_N(d))$ linear with respect to components of the vector $W_N(d)$ and optimal in the root-mean-square sense on the set of linear combinations of components of this vector; moreover, the uniform convergence $\hat{\eta}(W_N(d)) \to E(\eta|Y_N), d \to \infty$. is attained on some region,

ii. the current covariance matrix of estimation errors (we emphasize that known numerical methods of constructing approximations of the vector of nonlinear estimates cannot calculate current covariance matrices of estimation errors).

Implementation of items 2 and 2.1 makes the MPA algorithm more efficient than any known linear identification algorithm since it

i. does not involve linearization,

ii. does not apply variants of the Newton method to solve systems of nonlinear algebraic equations,

iii. forms the estimation vector that tends uniformly to the vector of conditional mathematical expectation for the growing integer d,

iiii. obtains the covariance matrix of estimation errors.

It is worth emphasizing that in this work we just develop the fundamental ground of computational technique for solving the complex problem of aircraft parameter identification.

3. Testing the MPA algorithm: Problem reconstruction (identification of parameters) for the attractor from units of an electrical chain

We consider the boundary inverse problem for the attractor whose equations are presented in [8]. The three parameters are the initial conditions: $X_1[0] = \eta_1, X_2[0] = \eta_2, X_3[0] = \eta_3$. The six parameters $\eta_{3+i}, i = 1, ..., 6$ correspond to combinations of the inductance, the resistances and the two capacitances of a circuit.

The equations of the mathematical model of the circuit take the following form [8]:

$$X_1[k-1] < -\eta_{3+6} : f = \eta_{3+5};$$

$$-\eta_{3+6} < X_1[k-1] < \eta_{3+6} : f = X_1[k-1](1 - X_1[k-1]^2);$$

$$X_1[k-1] > \eta_{3+6} : f = -\eta_{3+5};$$

$$X_1[k] = X_1[k-1] + \delta X_2[k-1];$$

$$X_2[k] = X_2[k-1] + \delta(-X_1[k-1] - \eta_{3+1}X_2[k-1] + X_3[k-1]);$$

$$X_3[k] = X_3[k-1] + \delta(\theta_{3+2}(\eta_{3+3}f - X_3[k-1]) - \eta_{3+4}X_2[k-1]);$$

where $X_1[k]$ corresponds to a voltage, $X_2[k]$ to a current and $X_3[k]$ to another voltage.

We suppose that by $i = 1, 2, 3$

$$\eta_i \in 1 + (\varepsilon_i - 0.5).$$

We also suppose [8] that

$$\eta_{3+1} \in 0.5(1 + (\varepsilon_1 - 0.5)); \eta_{3+2} = 0.3(1 + (\varepsilon_2 - 0.5)); \eta_{3+3} = 15(1 + (\varepsilon_3 - 0.5));$$

$$\eta_{3+4} \in 1.5(1 + (\varepsilon_4 - 0.5)); \eta_{3+5} = 0.5(1 + (\varepsilon_5 - 0.5)); \eta_{3+6} = 1.2(1 + (\varepsilon_6 - 0.5));$$

$$y_k = X_1(t_k)) + \zeta_k$$

$\delta = 0.01, k = 1, ..., N = 1200, z_1 = \sum_{k=1}^{k=N/T} y_k, z_2 = \sum_{k=1+N/T}^{k=2\times N/T} y_k, ...$

The algorithm uses approximations of parameters by means of linear combinations of the constructed values $z_i(d = 1)$. Values z_1, z_2 - are the sums of values of flowing observations - serve as inputs of MPA algorithm

The relative errors of the boundary problem are

i	1	2	3
$T = 24$			
Δ_i	0.025	0.264	0.272
$T = 48$			
Δ_i	0.0007	-0.003	0.046 .
$T = 120$			
Δ_i	0.00005	-0.00264	0.01687
$T = 240$			
Δ_i	0.00001	-0.00049	0.02686

The relative errors of the inverse problem are

i	1	2	3	4	5	6
$T = 24$						
Δ_{3+i}	-0.347	0.198	-0.250	0.097	0.095	0.136
$T = 48$						
Δ_{3+i}	-0.140	0.234	-0.222	0.104	0.143	0.133 .
$T = 120$						
Δ_{3+i}	-0.169	0.205	-0.167	0.094	0.179	0.097
$T = 240$						
Δ_{3+i}	-0.042	0.129	-0.031	-0.0001	0.151	0.146

The resulted tables show, that corresponding adjustment the MPA algorithm - a corresponding selection of value T allows to make small relative errors of an estimation of parameters of the non-linear dynamic system.

4. Identification of aerodynamic coefficients of the pitching motion for an pseudo f-16 aircraft

We illustrate efficiency of offered MPA algorithm on an example of identification of 48 dimensionless aerodynamic coefficients for the aircraft of near F-16. The aircraft we shall conditionally name " pseudo F-16 ". The term "near" is justified by that, what is the coefficients are taken from SIDPAS [1], but are perturbed by addition of some random numbers.

The tables resulted below, show, that errors of identification are small also modules of their relative values do not surpass several hundredth. The considered problem corresponds to minimization of object function of 48 variables, which it is made of the sum of squares of differences of actual and computational angles of attack, g-load, pitch angles, observable with frequency 10 hertz during 25 sec. flight of the aircraft maneuvering in a vertical plane.

number α_i	$C_{Z_0}(\alpha_i)$	$C_{m_0}(\alpha_i)$	$C_{Z_q}(\alpha_i)$	$C_{m_q}(\alpha_i)$
1	0.7700	−0.1740	−8.8000	−7.2100
2	0.2410	−0.1450	−25.8000	−5.4000
3	−0.1000	−0.1210	−28.9000	−5.2300
4	−0.4160	−0.1270	−31.4000	−5.2600
5	−0.7310	−0.1290	−31.2000	−6.1100
6	−1.0530	−0.1020	−30.7000	−6.6400
7	−1.3660	−0.0970	−27.7000	−5.6900
8	−1.6460	−0.1130	−28.2000	−6.0000
9	−1.9170	−0.0870	−29.0000	−6.2000
10	−2.1200	−0.0840	−29.8000	−6.4000
11	−2.2480	−0.0690	−38.3000	−6.6000
12	−2.2290	−0.0060	−35.3000	−6.0000

Table 1. Nominal values of the functions $C_{Z_0}(\alpha), C_{m_0}(\alpha), C_{Z_q}(\alpha), C_{m_q}(\alpha)$

4.1 Pitching motion equations

We use the rectangular coordinate system XYZ adopted in NASA. Then for the unperturbed atmosphere and conditions $V = const$, pitching motion equations have the form [1]:

$$\begin{aligned}
d\alpha/dt &= \omega_Y + (g/V)(N_Z + cos(\theta - \alpha)),\\
d\omega_Y/dt &= M_Y/J_Y,\\
d\theta/dt &= \omega_Y,\\
N_Z &= C_Z(\alpha, \delta_s)qS/G,\\
M_Y &= C_m(\alpha, \delta_s)qSb,
\end{aligned}$$

where V=300 ft/sec,H=20000 ft, α is the angle of attack, N_Z is the g-load, which is the vector of aerodynamic forces projected onto the axis Z and divided by the weight of the aircraft, M_Y is the vector of the moment of aerodynamic forces projected onto the axis Y, ω is the vector of the angular velocity of the aircraft projected onto the axis Y,θ is the angle between the the axis X and the horizontal plane, q is the value of the dynamic pressure, G is the weight, J_Y is the moment of inertia with respect to the axis Y, S is the area of the surface generating aerodynamic forces, b is the mean aerodynamic of the wing, $C_Z(\alpha, \delta)$ and $C_m(\alpha, \delta)$ are dimensionless coefficients of the aerodynamic force and moment,δ_s is the angle of the stabilator devlection measured in degrees.

Functions $C_Z(\alpha, \delta_s)$ and $C_m(\alpha, \delta_s)$ are given by the relations [1],:

$$\begin{aligned}
C_Z(\alpha, \delta_s) &= C_{Z_0}(\alpha) - 0.19(\delta_s/25) + C_{Z_q}(\alpha)(b/(2V))\omega,\\
C_m(\alpha, \delta_s) &= C_{m_0}(\alpha)\delta_s + C_{m_q}(\alpha)(b/(2V))\omega + 0.1C_Z.
\end{aligned}$$

4.2 Parametric model aerodynamic forces and moments

Nominal values of 4 functions of the angle of attack $C_{Z_0}(\alpha), C_{m_0}(\alpha), C_{Z_q}(\alpha), C_{m_q}(\alpha)$ are given with the argument step $(55 - 1)/12$ degree at 12 nodes (Table 1) in range $-10^\circ \leq \alpha \leq 45^\circ$.

To determine values of functions between the nodes, we use linear interpolation. Having analyzed Table 1, we can see that functions $C_{Z_0}(\alpha_i), C_{m_0}(\alpha_i), C_{Z_q}(\alpha_i), C_{m_q}(\alpha_i)$ are essentially

number α_i	$\Delta(C_{Z_0}(\alpha_i))$	$\Delta(C_{m_0}(\alpha_i))$	$\Delta(C_{Z_q}(\alpha_i))$	$\Delta(C_{m_q}(\alpha_i))$
1	0.7700	−0.1740	−8.8000	−7.2100
2	−0.5290	0.0290	−17.0000	1.8100
3	−0.3410	0.0240	−3.1000	0.1700
4	−0.3160	−0.0060	−2.5000	−0.0300
5	−0.3150	−0.0020	0.2000	−0.8500
6	−0.3220	0.0270	0.5000	−0.5300
7	−0.3130	0.0050	3.0000	0.9500
8	−0.2800	−0.0160	−0.5000	−0.3100
9	−0.2710	0.0260	−0.8000	−0.2000
10	−0.2030	0.0030	−0.8000	−0.2000
11	−0.1280	0.0150	−8.5000	−0.2000
12	0.0190	0.0630	3.0000	0.6000

Table 2. Nominal values of increment $\Delta(C_{Z_0}(\alpha_i)), \Delta(C_{m_0}(\alpha_i)), \Delta(C_{Z_q}(\alpha_i)), \Delta(C_{m_q}(\alpha_i))$

nonlinear. Table 2 confirms this visual impression. In it increments are presented 4 functions on each step of Table 1. Apparently, increments noticeably vary.

We study the identification problem for the perturbed analogues of the functions $C_{Z_0}(\alpha), C_{m_0}(\alpha), C_{Z_q}(\alpha), C_{m_q}(\alpha)$. The number of nominal coefficients that determine these functions is 12+12+12+12 = 48. Let us single out the problem which is the most complex for the MPA algorithm, when the actual coefficients differs from the nominal coefficients by the unknown bounded by the prior limits value η_i at each point of the table. Then, for accumulated results of measurements of parameters of the perturbed motion, the MPA algorithm is to estimate 48 components of the vector of random estimates, - the vector of differences between actual and nominal coefficients.

Suppose ϑ_i and B_i are the i-th components of the nominal and actual (perturbed) vectors of aerodynamic coefficients , $i = 1, ..., 48$, i.e. the number of actual coefficients to be identified is 48 in this case. We assume that the parametric model

$$B_i = \vartheta_i + \eta_i.$$

holds. The vector η serves as the vector of perturbations of nominal data errors of aerodynamic parameters, and identification yields the estimates of its components. We give the structure of these components by the formula $\eta_i = \vartheta_i \rho_i \varepsilon_i, 0 < \rho_i < 1, -1 < \varepsilon_i < 1$. The positive number ρ_i gives the maximum value that, by identification conditions, can be attained by the ratio of the absolute values of the random value of perturbations η_i and nominal coefficients ϑ_i .

4.3 Transient processes of characteristics of nominal motions

We wish to identify-estimate - during one test flight the 48 unknown aerodynamic coefficients for 12 nodes-12 the set angles of attack $\alpha_i, i = 1..., 12$. For a testing maneuver the characteristics $\alpha(t), N_Z(t), \theta(t)$ of Transient Processes are carrier of information of the the identified coefficients. Therefore during flight the aircraft should "visit" vicinities of angles of attack $-10^\circ \leq \alpha \leq 45^\circ$

number. obs. k	$\delta_s(k)$	$\alpha(k)$	$N_Z(k)$	$\theta(k)$
1	−0.0200	3.6820	0.1021	0.0132
3	−0.0600	5.1462	−0.3525	0.1388
5	−0.1000	6.0119	−0.2956	0.1689
7	−0.1400	6.7707	−0.2493	0.0300
9	−0.1800	7.5061	−0.2085	−0.1851
11	−0.2200	8.2964	−0.1685	−0.3945
13	−0.2600	9.2186	−0.1253	−0.5227
15	−0.3000	10.2083	−0.6016	−0.5119
17	−0.3400	10.6145	−0.5691	−0.6187
19	−0.3800	10.8889	−0.5477	−0.8891
21	−0.4200	11.0977	−0.5334	−1.2461
23	−0.4600	11.2993	−0.5223	−1.6252
25	−0.5000	11.5494	−0.5114	−1.9688
27	−0.5400	11.9047	−0.4974	−2.2222
29	−0.5800	12.4277	−0.4774	−2.3286
31	−0.6200	13.1919	−0.4477	−2.2247
33	−0.6600	14.2870	−0.4043	−1.8352
35	−0.7000	15.4810	−0.8822	−1.1481
37	−0.7400	16.1493	−0.8343	−0.6582
39	−0.7800	16.6530	−0.7993	−0.3828
41	−0.8200	17.0629	−0.7728	−0.2382
43	−0.8600	17.4401	−0.7511	−0.1552
45	−0.9000	17.8400	−0.7310	−0.0747
47	−0.9400	18.3168	−0.7095	0.0571
49	−0.9800	18.9260	−0.6838	0.2926
51	−1.0200	19.7287	−0.6512	0.6855
53	−1.0600	20.1833	−1.1389	1.1343
55	−1.1000	20.1954	−1.1266	1.3075
57	−1.1400	19.9812	−1.1273	1.2486
59	−1.1800	19.9888	−1.1293	1.1572
61	−1.2200	19.9789	−1.1308	1.1146
63	−1.2600	20.0083	−1.1316	1.1129
65	−1.3000	20.0049	−1.1325	1.1437
67	−1.3400	20.0371	−1.1330	1.2113
69	−1.3800	20.0328	−1.1338	1.3073
71	−1.4200	20.0598	−1.1344	1.4359
73	−1.4600	20.0636	−1.1346	1.6053
75	−1.5000	20.0993	−1.1344	1.8069
77	−1.5400	20.1945	−1.1326	2.0671
79	−1.5800	20.3760	−1.1278	2.4096
81	−1.6200	20.6696	−1.1186	2.8558
83	−1.6600	21.1005	−1.1037	3.4247
85	−1.7000	21.6926	−1.0820	4.1331
87	−1.7400	22.4690	−1.0523	4.9948
89	−1.7800	23.4509	−1.0133	6.0212

91	−1.8200	24.6576	−0.9641	7.2201
93	−1.8600	25.7086	−1.3912	8.6103
95	−1.9000	26.9187	−1.3506	10.2913
97	−1.9400	28.6178	−1.2942	12.4085
99	−1.9800	30.7696	−1.6768	15.0754
101	−2.0200	32.4354	−1.6031	17.5706
103	−2.0600	33.8743	−1.5431	19.7731
105	−2.1000	35.1432	−1.8357	21.7958
107	−2.1400	35.7317	−1.8014	23.4808
109	−2.1800	35.9655	−1.7815	24.7918
111	−2.1800	35.9159	−1.7715	25.8128
113	−2.1400	35.6010	−1.7680	26.5691
115	−2.1000	34.9990	−1.7695	27.0436
117	−2.0600	34.7166	−1.4423	27.4338
119	−2.0200	34.4743	−1.4525	27.8824
121	−1.9800	34.2309	−1.4610	28.3457
123	−1.9400	33.9488	−1.4692	28.7851
125	−1.9000	33.5921	−1.4785	29.1649
127	−1.8600	33.1250	−1.4900	29.4506
129	−1.8200	32.5103	−1.5050	29.6075
131	−1.7800	31.7080	−1.5248	29.5992
133	−1.7400	30.6740	−1.5505	29.3864
135	−1.7000	29.6897	−1.1403	29.0369
137	−1.6600	29.5906	−1.1751	29.2924
139	−1.6200	30.0407	−1.6327	30.1792
141	−1.5800	30.0324	−1.1643	30.9954
143	−1.5400	30.0007	−1.1623	31.6784
145	−1.5000	29.9971	−1.1623	32.3405
147	−1.4600	29.9834	−1.6165	32.9999
149	−1.4200	29.9916	−1.1620	33.6324
151	−1.3800	29.9805	−1.1621	34.2532
153	−1.3400	30.0190	−1.1626	34.8756
155	−1.3000	29.9687	−1.6164	35.4719
157	−1.2600	30.0181	−1.1623	36.0635
159	−1.2200	29.9808	−1.1614	36.6311
161	−1.1800	29.9772	−1.1614	37.1835
163	−1.1400	29.9490	−1.6153	37.7184
165	−1.1000	29.9574	−1.1606	38.2407
167	−1.0600	29.9417	−1.6141	38.7453
169	−1.0200	29.9178	−1.1614	39.1922
171	−0.9800	29.8635	−1.1625	39.6161
173	−0.9400	29.7346	−1.1650	39.9729
175	−0.9000	29.4842	−1.1711	40.2178
177	−0.8600	29.0596	−1.1829	40.3020
179	−0.8200	28.3990	−1.2028	40.1699

181	−0.7800	27.4286	−1.2336	39.7559
183	−0.7400	26.0585	−1.2787	38.9816
185	−0.7000	24.4358	−0.8717	37.7788
187	−0.6600	22.7903	−0.9362	36.2669
189	−0.6200	20.7622	−1.0163	34.4409
191	−0.5800	18.7935	−0.5743	32.3636
193	−0.5400	17.0475	−0.6702	30.3402
195	−0.5000	15.1676	−0.7674	28.2787
197	−0.4600	13.9500	−0.3141	26.3775
199	−0.4200	13.1154	−0.3750	24.8646
201	−0.3800	12.4703	−0.4193	23.5913
203	−0.3400	11.9129	−0.4537	22.4424
205	−0.3000	11.3566	−0.4837	21.3242
207	−0.2600	10.7223	−0.5143	20.1561
209	−0.2200	9.9324	−0.5497	18.8636
211	−0.1800	9.8365	−0.0427	17.6449
213	−0.1400	9.9588	−0.0463	16.6494
215	−0.1000	9.9853	−0.0452	15.7456
217	−0.0600	9.9853	−0.0437	14.8087
219	−0.0200	10.0132	−0.0415	13.8278
221	0.0200	9.9999	−0.0398	12.7981
223	0.0600	9.9409	−0.5639	11.7156
225	0.1000	9.9557	−0.0371	10.5715
227	0.1400	9.9311	−0.0359	9.3815
229	0.1800	9.8312	−0.0369	8.1116
231	0.2200	9.6174	−0.0423	6.7279
233	0.2600	9.2466	−0.0540	5.1941
235	0.3000	8.6685	−0.0746	3.4700
237	0.3400	7.8225	−0.1068	1.5084
239	0.3800	6.6340	−0.1539	−0.7475
241	0.4200	5.0095	−0.2203	−3.3684
243	0.4600	3.6383	0.2178	−6.2070
245	0.5000	2.2053	0.1484	−9.0796
247	0.5400	0.5330	0.0717	−12.0920
249	0.5800	−0.9453	0.5373	−15.2658

Table 3. The characteristics $\alpha(t), N_Z(t), \theta(t)$ of the nominal motions for the chosen control law $\delta_s(t)$.

4.4 Estimating identification accuracy of 48 errors of aerodynamic parameters of the aircraft

Primary task of MPA algorithm consists in identification - estimation-48 increments of 4 functions. If entry conditions and increments are determined, values of the unknown coefficients follow from obvious recurrent formulas.

To estimate the accuracy, we assume that the current values of α, N_Y, θ are measured every 0.1 sec. during 25 seconds .We assume that random errors of measurement represent the discrete white noise bounded by the true measurable value multiplied by the given value ϵ. An amount of the primary observations equal 3*250=750.

number α_i	nom.koef. $C_{Z_0}(\alpha_i)$	perturb.koef. $C_{Z_0}(\alpha_i)$	$\delta(C_{Z_0}(\alpha_i))$
1	0.6512	0.6326	0.02854
2	0.0205	0.0260	−0.26410
3	−0.3778	−0.3646	0.03491
4	−0.7395	−0.7213	0.02456
5	−1.0610	−1.0657	−0.00443
6	−1.4038	−1.4016	0.00159
7	−1.7679	−1.7424	0.01444
8	−2.0582	−2.0453	0.00627
9	−2.2774	−2.3388	−0.02693
10	−2.4568	−2.5459	−0.03625
11	−2.5639	−2.6698	−0.04130
12	−2.5404	−2.6505	−0.04334

Table 4. The Relative errors of the identifications of $C_{Z_0}(\alpha_i)$ by $\rho = 0.25$

number α_i	nom.koef. $C_{m_0}(\alpha_i)$	perturb.koef. $C_{m_0}(\alpha_i)$	$\delta(C_{m_0}(\alpha_i))$
1	−0.2130	−0.2054	0.03582
2	−0.1816	−0.1783	0.01851
3	−0.1567	−0.1550	0.01061
4	−0.1618	−0.1611	0.00439
5	−0.1634	−0.1631	0.00209
6	−0.1427	−0.1388	0.02754
7	−0.1372	−0.1338	0.02439
8	−0.1495	−0.1502	−0.00467
9	−0.1175	−0.1220	−0.03771
10	−0.1139	−0.1190	−0.04484
11	−0.0957	−0.1043	−0.08937
12	−0.0399	−0.0394	0.01236

Table 5. The Relative errors of the identifications of $C_{m_0}(\alpha_i)$ by $\rho = 0.25$

We compress primary observations for a smoothing the high-frequency errors and reduction of a dimension of matrixes covariance . The file of the primary observations is divided into 12 groups and as an input of the algorithm of the identification the vector of the dimension 12 × 1 serves. Components of this vector are the sums of elements of each of 12 groups.

To characterize the accuracy of identification of the random parameter η_i the degree of perturbation of the aerodynamic coefficients ϑ , we determine the relative errors of estimation $(\eta_i - \hat{\eta}_i)/\eta_i$ for every component the identifiable functions . The relative errors designate $\delta(C_{Z_0}(\alpha_i)), \delta(C_{m_0}(\alpha_i)), \delta(C_{Z_q}(\alpha_i)), \delta(C_{m_q}(\alpha_i)), i = 1, ..., 12.$

Apparently, relative errors of identification are small and do not surpass several hundredth at $\rho = 0.25$

number α_i	nom.koef. $C_{Z_q}(\alpha_i)$	perturb.koef. $C_{Z_q}(\alpha_i)$	$\delta(C_{Z_q}(\alpha_i))$
1	-9.9636	-8.8984	0.10691
2	-25.2235	-26.1655	-0.03735
3	-28.4644	-29.2857	-0.02885
4	-31.4821	-31.8270	-0.01096
5	-31.3125	-31.6274	-0.01006
6	-30.8417	-31.1249	-0.00918
7	-27.5461	-28.0921	-0.01982
8	-28.1388	-28.6036	-0.01652
9	-28.9682	-29.4069	-0.01515
10	-29.7908	-30.2114	-0.01412
11	-38.6789	-38.7933	-0.00296
12	-35.7355	-35.8053	-0.00195

Table 6. The Relative errors of the identifications of $C_{Z_q}(\alpha_i)$ by $\rho = 0.25$

number α_i	nom.koef. $C_{m_q}(\alpha_i)$	perturb.koef. $C_{m_q}(\alpha_i)$	$\delta(C_{m_q}(\alpha_i))$
1	-5.5807	-6.1771	-0.10686
2	-4.1294	-4.3066	-0.04291
3	-3.9913	-4.1368	-0.03645
4	-4.0250	-4.1662	-0.03510
5	-4.9363	-5.0012	-0.01315
6	-5.5024	-5.5314	-0.00527
7	-4.5272	-4.5870	-0.01320
8	-4.8711	-4.8936	-0.00462
9	-5.0970	-5.0915	0.00108
10	-5.3245	-5.2912	0.00626
11	-5.5637	-5.4908	0.01310
12	-4.8726	-4.8937	-0.00434

Table 7. The Relative errors of the identifications of $C_{m_q}(\alpha_i)$ by $\rho = 0.25$

number α_i	nom.koef. $C_{Z_0}(\alpha_i)$	perturb.koef. $C_{Z_0}(\alpha_i)$	$\delta(C_{Z_0}(\alpha_i))$
1	0.5324	0.4255	0.20083
2	-0.1999	-0.2092	-0.04637
3	-0.6556	-0.5969	0.08959
4	-1.0629	-0.9303	0.12481
5	-1.3911	-1.2772	0.08188
6	-1.7546	-1.6697	0.04839
7	-2.1699	-2.0331	0.06304
8	-2.4704	-2.3417	0.05209
9	-2.6379	-2.6342	0.00138
10	-2.7936	-2.8450	-0.01839
11	-2.8799	-2.9723	-0.03208
12	-2.8518	-2.9530	-0.03548

Table 8. The Relative errors of the identifications of $C_{Z_0}(\alpha_i)$ by $\rho = 0.50$

number α_i	nom.koef. $C_{m_0}(\alpha_i)$	perturb.koef. $C_{m_0}(\alpha_i)$	$\delta(C_{m_0}(\alpha_i))$
1	−0.2520	−0.2441	0.03123
2	−0.2183	−0.2166	0.00781
3	−0.1924	−0.1934	−0.00523
4	−0.1966	−0.1994	−0.01457
5	−0.1979	−0.2014	−0.01792
6	−0.1834	−0.1747	0.04747
7	−0.1773	−0.1697	0.04301
8	−0.1860	−0.1858	0.00145
9	−0.1481	−0.1599	−0.08004
10	−0.1438	−0.1569	−0.09149
11	−0.1225	−0.1420	−0.15942
12	−0.0738	−0.0811	−0.09934

Table 9. The Relative errors of the identifications of $C_{m_0}(\alpha_i)$ by $\rho = 0.50$

number α_i	nom.koef. $C_{Z_q}(\alpha_i)$	perturb.koef. $C_{Z_q}(\alpha_i)$	$\delta(C_{Z_q}(\alpha_i))$
1	−11.1272	−8.6840	0.21957
2	−24.6470	−25.6672	−0.04139
3	−28.0288	−28.8049	−0.02769
4	−31.5642	−31.3356	0.00724
5	−31.4249	−31.1306	0.00937
6	−30.9833	−30.6296	0.01142
7	−27.3921	−27.6113	−0.00800
8	−28.0776	−28.1104	−0.00117
9	−28.9364	−28.9144	0.00076
10	−29.7817	−29.7303	0.00172
11	−39.0577	−38.2130	0.02163
12	−36.1709	−35.1346	0.02865

Table 10. The Relative errors of the identifications of $C_{Z_q}(\alpha_i)$ by $\rho = 0.25$

number α_i	nom.koef. $C_{m_q}(\alpha_i)$	perturb.koef. $C_{m_q}(\alpha_i)$	$\delta(C_{m_q}(\alpha_i))$
1	−3.9514	−6.9359	−0.75528
2	−2.8588	−5.1596	−0.80480
3	−2.7526	−4.9893	−0.81258
4	−2.7899	−5.0189	−0.79894
5	−3.7625	−5.8672	−0.55939
6	−4.3649	−6.3936	−0.46477
7	−3.3644	−5.4530	−0.62079
8	−3.7422	−5.7627	−0.53993
9	−3.9940	−5.9631	−0.49299
10	−4.2490	−6.1601	−0.44976
11	−4.5274	−6.3594	−0.40464
12	−3.7451	−5.7631	−0.53882

Table 11. The Relative errors of the identifications of $C_{m_q}(\alpha_i)$ by $\rho = 0.50$

5. Conclusions

The presented data show that the multipolynomial approximation algorithm can provide a computational ground for developing an efficient parameter identification technique for the nonlinear dynamic system, including identification of aerodynamic parameters of an aircraft. We emphasize that tables characterizing a sufficiently high accuracy of aerodynamic parameter identification are obtained when there are no iterations and d = 1, which corresponds to the case when the estimation vector $(\vartheta \hat{+} \eta)(W_N(d))$ is represented by the vector linear combination of measured data that is optimal on the family of linear operators over the vector of measurements. This is due to good (in terms of the identification problem) properties of the parametric system of equations of the pitching motion of the "pseudo F-16 " aircraft. It can become much more complicated when it comes to the identification problem of the parametric system of equations of complete (spatial) motion of the aircraft. In such case, we may need to use polynomials of the power d > 1 and increase requirements on the computer performance and RAM. This was the case for identification attempts made for some parameters of F-16 complete motion equations. We emphasize that the inputs of the MPA algorithm we considered were not real (were not the results of operation of real sensors of the aircraft during its test flight); they were determined by mathematical simulation - by means the numerically integrations motion equations for perturbed parameters of aerodynamic forces and moments.

6. Appendix A: An estimate of the vector of the conditional mathematical expectation that is optimal in the root-mean-square sense

A.1. An algorithm fundamental (AF)

We consider the algorithm fundamental (AF) for solving the problem of finding the estimate of the vector $E(\eta|Y_N)$ that is optimal in the root-mean-square sense. This vector is known to be the estimate optimal in the root-mean-square sense of the vector η once the vector Y_N is fixed. Therefore, it is justified that it is the vector of conditional expectation that AF tends to estimate.

We construct AF that ensures polynomial approximation of the vector $E(\eta|Y_N)$. To do this, we find the approximate estimate of the vector $E(\eta|Y_N)$, which is linear with respect to components of the vector $W_N(d)$ and optimal in the root-mean-square sense. We denote the vector of this estimate by $\widehat{\eta}(W_N(d))$. To obtain the explicit expression for the estimation vector, we calculate elements of the vector $\overline{V}(d,N)$ and the covariance matrix $C_V(d,N)$ that are the first and second (centered) statistical moments for the vector $V(d,N)$. These vector and matrix can be divided into blocks of the following structure

$$E(E(\eta|Y_N)) = E(\eta);$$

$$E(E(W(d,N)|Y_N)) = E(W(d,N));$$

$$= E((E(\eta|Y_N) - E(\eta))(E(\eta|Y_N) - E(\eta))^T) =$$
$$E((\eta - E(\eta))(\eta - E(\eta))^T).$$

$$L_N(d) = E((E(\eta|Y_N) - E(E(\eta|Y_N(d)))(W_N(d) - E(W_N(d)))^T) =$$
$$E(\eta)W_N(d)^T) - E(\eta)E(W_N(d))^T,$$

$$Q_N(d) = E((W_N(d) - E(W_N(d)))(W_N(d) - E(W_N(d)))^T);$$

The right-hand sides of these blocks are the first and second (centered) statistical moments calculated by the Monte-Carlo method. However, their left-hand sides also serve as the first and second (centered) statistical moments of components of the vector of conditional mathematical expectations. Hence, we can use mathematical models of form (0.1) and (0.2) to find these statistical moments experimentally for vectors of conditional expectations as well. This obvious proposition gives us the basis for practical implementation of the computational procedure of estimating the vector of the conditional expectation.

We introduce

$$\widehat{\eta}(W_N(d)) = E(\eta) + \Lambda_N(d)(W_N(d) - E(W_N(d))), \quad (A.1)$$

where the matrix $\Lambda_N(d), r \times m(d,N)$ satisfies the equation

$$\Lambda_N(d)Q_N(d) = L_N(d).$$

We also introduce

$$\widetilde{\eta}(W_N(d)) = z + \widetilde{\Lambda}_N(d)(W_N(d) - E(W_N(d))), \quad (A.2)$$

where z and $\widetilde{\Lambda}_N(d)$ are the arbitrary vector and matrix of dimensions $r \times 1$ and $r \times m(d,N)$. Suppose $C(d,N)$ and $\widetilde{C}(d,N)$ are the covariance matrices of estimation errors for the vector $E(\eta|Y_N)$ generated by the estimates $\widehat{\eta}(W_N(d))$ and $\widetilde{\eta}(W_N(d))$.

Lemma. The matrix $\widetilde{C}(d,N) - C(d,N)$ is a nonnegative definite matrix : $C(d,N) \leq \widetilde{C}(d,N)$.

The lemma follows from the identity

$$\widetilde{C}(d,N) = C(d,N) + (\Lambda_N(d) - \widetilde{\Lambda}_N(d))(\Lambda_N(d) - \widetilde{\Lambda}_N(d))^T +$$
$$(\Lambda_N(d)Q_N(d) - L_N(d))(\widetilde{\Lambda}_N(d) - \Lambda_N(d))^T +$$
$$((\Lambda_N(d)Q_N(d) - L_N(d))(\widetilde{\Lambda}_N(d) - \Lambda_N(d))^T)^T + (z - E(\eta)(z - E\eta)^T. \quad (A.3)$$

Corollary of the lemma. For the vector $E(\eta|Y_N$, the vector $\widehat{\eta}(W_N(d))$ is the estimate optimal in the root-mean-square sense among the set of estimates linear with respect to components of the vector $W_N(d)$. If $Q_N(d) > 0$, the estimation vector is unique and

$$\widehat{\eta}(W_N(d)) = E(\eta) + L_N(d)Q_N(d)^{-1}(W_N(d) - E(W_N(d))). \quad (A.4)$$

The covariance matrix $C(d,N)$ of estimation errors of the vector $E(\eta|Y_N)$ is given by the formula

$$C(d,N) = C_\eta - \Lambda_N(d)L_N(d).(A.5)$$

If $Q_N(d) \geq 0$, the vectors that provide linear and optimal in the root-mean-square sense estimate are not unique; however, the variances of components of the difference between these vectors are zeros.

Formula (A.1) gives explicit expressions for the vector coefficients of the form $\lambda(a_1, ..., a_N)$ in (1.1). To find these relations, we open the explicit expressions for components of the vector

$W_N(d)$ and the right-hand side of (A.1) and equate them to the right-hand side of formula (1.1).

We consider asymptotic estimation errors when we use (A.1). Suppose the vector Y_N is fixed. We assume that the vector $E(\eta|Y_N$ is given by the function of Y_N on some a priori region that is a compact; the function is continuous on this region. Then, the following theorem holds. Theorem.

$$Sup \quad _{Y_N \in \Omega_{Y_N}} |\widehat{\eta}(W_N(d)) - E(\eta|Y_N)| \Rightarrow 0, d \Rightarrow \infty. \quad (A.6)$$

Proof. The multidimensional analogue of the K. Weierstrass theorem, which is the corollary of the M. Stone theorem [9], states that for any number $\varepsilon > 0$ there exists a multidimensional polynomial $P(W_N(d_\varepsilon))$ such that

$$Sup \quad _{Y_N \in \Omega_{Y_N}} |P(W_N(d_\varepsilon)) - E(\eta|Y_N)| < \varepsilon.$$

We can rewrite this relation as

$$Sup \quad _{Y_N \in \Omega_{Y_N}} |P(W_N(d)) - E(\eta|Y_N)| \Rightarrow 0, \quad d \Rightarrow \infty. \quad (A.7)$$

We assume that C is the covariance matrix of the random vector $P(W_N(d) - E(\theta|Y_N):$

$$C = E((P(W_N(d)) - E(\eta|Y_N))(P(W_N(d)) - E(\eta|Y_N))^T.$$

It follows from (A.7) that

$$C \Rightarrow 0_n, \quad d \Rightarrow \infty \quad (A.8).$$

By construction, the vector $\widehat{\eta}(W_N(d))$ provides the estimate of the vector η that is linear with respect to components $W_N(d)$ and optimal in the root-mean-square sense. However, it follows from the lemma that for any other non-optimal linear estimate, including estimates of the form $P(W_N(d)$, the relation $C \geq C(d, N)$ holds. Hence, taking into account (A.8), we obtain

$$C(d, N) \Rightarrow 0_n, d \Rightarrow \infty. \quad (A.9)$$

Proposition (A.9) is equivalent to (A.6) if we recall that

$$C(d, N) = \int (Z_{E(\eta|Y_N)}(W_N(d)) - E(\eta|Y_N))(Z_{E(\eta|Y_N)}(W_N(d)) - E(\eta|Y_N))^T$$

$$p(\eta, Y_N) d\eta dY_N,$$

where $p(\eta, Y_N)$ is the joint probability density of the random vectors η and Y_N. The theorem is proved.

Thus, by (A.1), AF determines the vector series that, with the increasing number $m(d, N)$ of its terms, approximates the vector of conditional mathematical expectation of the vector θ of the estimated parameters with an arbitrary uniformly small root-mean-square error.

A.2. Recurrent (Realizable) MPA algorithm

To use formula (A.1), we need to find the matrix inverse to the matrix $Q_N(d)$. When the dimension $m(d, N) \times m(d, N)$ of the matrix $Q_N(d)$ is high and $Q_N(d)$ is close to the singular matrix, it is difficult to calculate elements of the inverse matrix. Below, we give the recurrent computational process based on the principle of decomposing observations, described in [6, 7]. Above, we specified the vector $W_N(d)$ of dimension $m(d,N) \times 1$ with the components $w_1, ..., w_{m(d,N)}$. The computational process consists of $m(d,N)$ successive steps. At each step, we use new updated prior data to find the new estimate of the vector θ and perform the prediction, which provides estimates for the rest part of the observation vector. At the same time, we determine the covariance matrix of the estimation errors attained at this step. There is no prediction at the last $m(d,N)-$ th step, and the vector θ is refined for the last time.

Let us construct the recurrent algorithm (the MPA algorithm) that does not calculate inverse matrices and consists of $m(d,N)$ steps of calculating the first and second statistical moments for the sequence of special vectors $V_1, ..., V_i, ..., V_{m(d,N)}$ performed after prior moments $\bar{V}(d,N)$ and $C_V(d,N)$ are found for the basic vector $V(d,N)$. We assume that V_1 is composed of $r + m(d,N) - 1$ components of the basic vector $V(d,N)$ left after the component w_1 was excluded, $w_1, ...$; V_i composed of components of the vector V_{i-1} left after the component w_i was excluded, etc. The component $w_{m(d,N)}$ is the last component of the vector $W_N(d)$, and once we exclude it, the resulting vector $V_{m(d,N)}$ turns out to equal the estimation vector $\hat{\eta}(W_N(d))$.

At step 1, we use the particular case of formulas of form (A.1) and (A.5) to calculate the vector $\bar{V}_1$ that estimates the vector V_1 and is optimal in the root-mean-square sense and linear with respect to w_1, and the covariance matrix of the estimation errors $C(V_1)$.

The estimation vector is formed of the estimate of the vector of conditional mathematical expectation $E(\eta|Y_N)$ and the vector of dimension $m(d,N) - 1) \times 1$. Once we fix the value w_1, the latter becomes the vector of statistical prediction of the mean values of "future" values $w_2, ..., w_{m(d,N)}$. We emphasize that calculations performed at step 1 are based on the preliminary found prior , $\bar{V}(d,N), C_V(d,N)$.

Suppose steps $1, ..., i$ of the computational process yielded the vector $\bar{V}(d,N)$ and the matrix $C(V_i)$ after the values $w_1, ...w_i$, wi were fixed. At step $i+1$, we have from the particular case of formulas (A.1) and (A.5) the vector $\bar{V}_{i+1}$ that estimates the vector V_{i+1} and is optimal in the root-mean-square sense and linear with respect to $w_1, ..., w_{i+1}$, and the covariance matrix $C(V_{i+1})$ of estimation errors. The vector $\bar{V}_{i+1}$ is still formed of the estimate of the vector of conditional mathematical expectation $E(\eta|w_1, ..., w_{i+1}$ (first r components of the vector $\bar{V}_{i+1}$) and the vector of statistical prediction of mean values of "future" - h values $w_{i+2}, ..., w_{m(d,N)}$ after $w_1, ..., w_{i+1}$ (the rest $m(d,N) - (i+1)$ components of the vector $\bar{V}_{i+1}$) are fixed. We emphasize that calculations at step $i+1$ are based on the preliminary found $\bar{V}_i$and $C(V_i)$, which can be naturally called the first and second statistical moments for "future" random values $w_{i+1}, ..., w_{m(d,N)}$. These vectors and matrices represent a priori data on statistical moments of components of the vector V_{i+1} before the algorithm receives the value w_{i+1} at its input.

Recurrent formulas that corresponds exactly to the above given qualitative description of the computational process have the form

$$\overline{V}_{i+1} = \overline{V}_i^1 + q_i^{-1} b_i (w_{i+1} - z_{w_{i+1}}), \quad (A.10)$$

$$C(V_{i+1}) = C(V_i)^1 - q_i^{-1} b_i b_i^T, \quad (A.11)$$

where the scalar $z_{w_{i+1}}$ is the $(r+1)$-th component of the vector $\bar{V}_i$, the scalar is the linear and optimal in the root-mean-square sense estimate of the component after the algorithm has processed the components $w_1, ..., w_i, \bar{V}_i^{-1}$ is the vector obtained from the vector $\bar{V}_i$ by eliminating its component $z_{w_{i+1}}$, the scalar q_i is the $(r+1)$-th diagonal element of the matrix $C(V_i)$, which is the variance of the estimation error of the component w_{i+1} after components $w_1, ..., w_i$ were processed, $C(V_i)^1$ is the matrix formed of $C(V_i)$ after the $(r+1)$-th row vector and $(r+1)$-th column vector were excluded, and b_i is the $(r+1)$-th column vector of the matrix $C(V_i)$ with its $(r+1)$-th component deleted.

If the scalar q_i turned out to be close to zero, the component w_{i+1} corresponds to a linear combination of components $w_1, ..., w_i$. Then, w_{i+1} do not give any new information on θ and should be excluded from the computational process. Note that the sequence of random variables like $(w_{i+1} - z_{w_{i+1}})$ forms an updating sequence. The upper left block of the $(r \times r)$-matrix $C(V_i)$ includes the covariance matrix $C(d,i)$ of estimation errors of the vector $E(\eta | w_i, ..., w_i)$ after the algorithm processed the vector $W_i(d)$.

We assume that $l(i)$ is the vector composed of r first components of the vector b_i. The formula representing the evolution of the covariance matrix $C(d,i)$ in the function of the number i of observable components of the vector $W_N(d)$ has the form

$$C(d,i) = C_\eta(0) - q_1^{-1} l(1) l(1)^T - ... - q_i^{-1} l(i) l(i)^T. \quad (A.12)$$

To test this MPA algorithm, we solved numerically several problems of estimating the components of the state vector for essentially nonlinear dynamic systems. The estimated components are unknown random constant parameters $\eta_1, ..., \eta_r$ of the dynamic system.

As for particular applied problems, we considered smoothing problems and the filtration problem.

In the above examples, we applied the Monte-Carlo method for the number of random realizations lying within 5000 - 10000. This number does not affect the estimation errors provided by the MPA algorithm significantly. The estimated random parameters are assumed to be statistically independent and are a priori uniformly distributed. The value of the root-mean- square deviation $\sigma(i, theo)$ is determined theoretically by calculating variances :the diagonal elements of the covariance matrix $C(d,N)$. The value of the root-mean-square deviation $\sigma(i, exp)$ is obtained experimentally by applying the Monte-Carlo method for 5000 realizations. Experimental and theoretical root-mean-square deviations almost coincide, which proves that the above given formulas of the MPA algorithm are correct.

7. Acknowledgments

This study was financially supported by the Russian Foundation for Basic Research (grant no. 10-08-00415a).

8. References

[1] V. Klein and A. G. Morelli, Aircraft System Identification: Theory and Methods (American Institute of Aeronautics and Astronautics, Reston 2006).

[2] M. B. Tischle and R. K. Remple, Aircraft and Rotorcraft System Identification: Engineering Methods with Flight Test Examples (American Institute of Aeronautics and Astronautics, Reston, 2006).

[3] R. Jategaonkar, Flying Vehicle System Identification: A Time Domain Methodology (American Institute of Aeronautics and Astronautics, Reston, 2006).

[4] L. Ljung, System Identification: Theory for the USER (Prentice Hall, 1987).

[5] B. K. Poplavskii and G. N. Sirotkin, Integrated Approach to Analysis of Processes of Identification of Model Parameters of a Spacecraft, in Transactions No. 429 of p/ya V-8759 [in Russian].

[6] J. A. Boguslavskiy, Bayes Estimators of Nonlinear Regression and Allied Issues, Izv. Ross. Akad. Nauk, Teor. Sist. Upr., No. 4, 14.24 (1996) [Comp. Syst. Sci. 35 (4), 511.521 (1996)].

[7] J. A. Boguslavskiy, A Polynomial Approximation for Nonlinear Problems of Estimation and Control (Fizmatlit,Moscow, 2006) [in Russian].

[8] J. Timmer, H.Rust, W.Horbelt, H.U. Voss, Parametric, nonparametric and parametric modelling of a chaotic circuit time series, Physics Letters A 274 (2000) 123 - 134

[9] A. F. Timan, The Theory of Approximation of Functions of a Real Variable (Nauka, Moscow, 1960) [in Russian].

Influence of Forward and Descent Flight on Quadrotor Dynamics

Matko Orsag and Stjepan Bogdan
LARICS-Laboratory for Robotics and Intelligent Control Systems
Department of Control and Computer Engineering,
Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb
Croatia

1. Introduction

The focus of this chapter is an aircraft propelled with four rotors, called the quadrotor. Quadrotor was among the first rotorcrafts ever built. The first successful quadrotor flight was recorded in 1921, when De Bothezat Quadrotor remained airborne for two minutes and 45 seconds. Later he perfected his design, which was then powered by 180-horse power engine and was capable of carrying 3 passengers on limited altitudes. Quadrotor rotorcrafts actually preceded the more common helicopters, but were later replaced by them because of very sophisticated control requirements Gessow & Myers (1952). At the moment, quadrotors are mostly designed as small or micro aircrafts capable of carrying only surveillance equipment. In the future, however, some designs, like Bell Boeing Quad TiltRotor, are being planned for heavy lift operations Anderson (1981); Warwick (2007).

In the last couple of years, quadrotor aircrafts have been a subject of extensive research in the field of autonomous control systems. This is mostly because of their small size, which prevents them to carry any passengers. Various control algorithms, both for stabilization and control, have been proposed. The authors in Bouabdallah et al. (2004) synthesized and compared PID and LQ controllers used for stabilization of a similar aircraft. They have concluded that classical PID controllers achieve more robust results. In Adigbli et al. (2007); Bouabdallah & Siegwart (2005) "Backstepping" and "Sliding-mode" control techniques are compared. The research presented in Adigbli et al. (2007) shows how PID controllers cannot be used as effective set point tracking controller. Fuzzy based controller is presented in Varga & Bogdan (2009). This controller exhibits good tracking results for simple, predefined trajectories. Each of these control algorithms proved to be successful and energy efficient for a single flying manoeuvre (hovering, liftoff, horizontal flight, etc.).

This chapter examines the behaviour of a quadrotor propulsion system focusing on its limitations (i.e. saturation and dynamic capabilities) and influence that the forward and descent flights have on this propulsion system. A lot of previous research failed to address this practical problem. However, in case of demanding flight trajectories, such as fast forward and descent flight manoeuvres, as well as in the presence of the In Ground Effect, these aerodynamic phenomena could significantly influence quadrotor's dynamics. Authors in Hoffmann et al. (2007) show how control performance can be diminished if aerodynamic effects are not considered. In these situations control signals could drive the propulsion system well within the region of saturation, thus causing undesired or unstable quadrotor behaviour. This effect is especially important in situations where the aircraft is operating at its limits (i.e. carrying heavy load, single engine breakdown, etc.).

The proposed analysis of propulsion system is based on the thin airfoil (blade element) theory combined with the momentum theory Bramwell et al. (2001). The analysis takes into account the important aerodynamic effects, specific to quadrotor construction. As a result, the chapter presents analytical expressions showing how thrust, produced by a small propeller used in quadrotor propulsion system, can be significantly influenced by airflow induced from certain manoeuvres.

2. Basic dynamic model

This section introduces the basic quadrotor dynamic modeling, which includes rigid body dynamics (i.e. Euler equations), kinematics and static nonlinear rotor thrust equation. This model, based on the first order approximation, has been successfully utilized in various quadrotor control designs so far. Nevertheless, recent shift in Unmanned Aerial Vehicle research community towards more payload oriented missions (i.e. pick and place or mobile manipulation missions) emphasized the need for a more complete dynamic model.

2.1 Kinematics

Quadrotor kinematics problem is, actually, a rigid-body attitude representation problem. Rigid-body attitude can be accurately described with a set of 3-by-3 orthogonal matrices. Additionally, the determinant of these matrices has to be one Chaturvedi et al. (2011). Since matrix representation cannot give a clear insight into the exact rigid body pose, attitude is often studied using parameterizations Shuster (1993). Regardless of the choice, every parameterization at some point fails to fully represent rigid body pose. Due to the gimbal lock, Euler angles cannot globally represent rigid body pose, whereas quaternions cannot define it uniquely.Chaturvedi et al. (2011)

Although researchers proved the effectiveness of using quaternions in quadrotor control Stingu & Lewis (2009), Euler angles are still the most common way of representing rigid body pose. To uniquely describe quadrotor pose using Euler angles, a composition of 3 elemental rotations is chosen. Following $X-Y-Z$ convention, a world reference coordinate system is first rotated Ψ degrees around X axis. After this, a Θ degree rotation around an intermediate Y axis is applied. Finally, a Φ degree rotation around a newly formed Z axis is applied to yield a transformation matrix from the world coordinate system $\mathfrak{W}$ to the body frame $\mathfrak{B}$, as shown in figure 1. Equations 1 and 2 formalize this procedure:

$$Rot\left(\Phi,\Theta,\Psi\right) = Rot\left(z^{w},\Phi\right)Rot\left(y^{w},\Theta\right)Rot\left(x^{w},\Psi\right) \tag{1}$$

$$Rot\left(\Phi,\Theta,\Psi\right) = \begin{bmatrix} c\phi c\theta & c\phi s\theta s\psi - s\phi c\psi & c\phi s\theta c\psi + s\phi s\psi \\ s\phi c\theta & s\phi s\theta s\psi + c\phi c\psi & s\phi s\theta c\psi - c\phi s\psi \\ -s\theta & c\theta s\psi & c\theta c\psi \end{bmatrix} \tag{2}$$

where $c\phi$ and $s\phi$ stand for $cos(\phi)$ and $sin(\phi)$, respectively. The same abbreviations are applied to other angles as well.

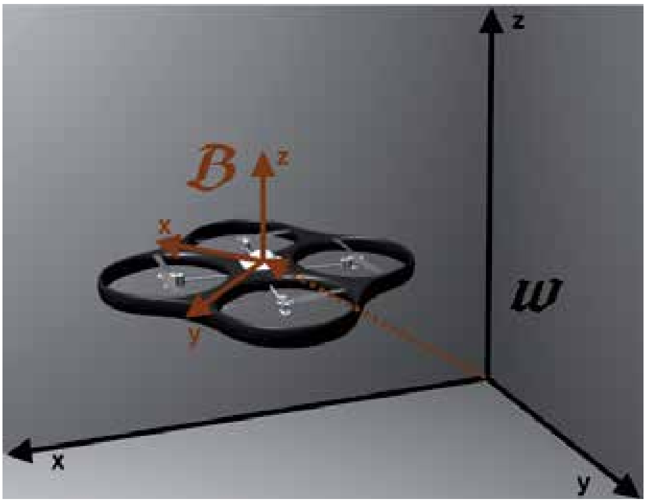


Fig. 1. Transformation from the body frame to the world frame coordinate system

2.2 Dynamic motion equations

Forces and torques, produced from the propulsion system and the surroundings, move and turn the quadrotor. In this paragraph, the quadrotor is viewed as a rigid body with linear and circular momentum, $\overrightarrow{L}$ and $\overrightarrow{M}$ respectively. According to the 2nd Newtons law, the force applied to the body equals the change of linear momentum. Using the principal of the change of momentum used in Jazar (2010), the following equation maps the change of quadrotor's position with respect to the applied force:

$$\begin{aligned} \overrightarrow{\mathbf{F}} &= \frac{\partial \overrightarrow{\mathbf{L}}}{\partial t} = \frac{\partial \overrightarrow{\mathbf{v}}}{\partial t} m_q + \frac{\partial m_q}{\partial t} \overrightarrow{\mathbf{v}} \\ \overrightarrow{\mathbf{F}} &= \frac{\partial \overrightarrow{\mathbf{v}}}{\partial t} m_q = m_q \begin{bmatrix} \frac{\partial^2 x}{\partial t^2} & \frac{\partial^2 y}{\partial t^2} & \frac{\partial^2 z}{\partial t^2} \end{bmatrix}^T \end{aligned} \tag{3}$$

where m_q represents quadrotor mass and $\overrightarrow{\mathbf{v}}$ its velocity vector. Due to the fact that most unmanned quadrotors are electrically driven, it is safe to assume that quadrotor mass does not change over time, resulting in a simple equation 3.

Same analysis can be applied to angular momentum, having in mind, the angular momentum is produced from the quadrotor motion as well as from the rotors spinning to produce the desired thrust. There are four important variables concerning angular momentum: quadrotor angular speed vector - $\overrightarrow{\omega}$, rotor angular speed vector - $\overrightarrow{\Omega}$, quadrotor inertia tensor - $\mathbf{I_q}$ and rotor inertia tensor - $\mathbf{I_r}$. Angular motion equations can be derived as follows:

$$\begin{aligned} \overrightarrow{\mathbf{M}} &= \overrightarrow{\omega} + \mathbf{I_r}\overrightarrow{\Omega} \\ \overrightarrow{\mathbf{T}} = \frac{\partial \overrightarrow{\mathbf{M}}}{\partial t} + \overrightarrow{\omega} \times \overrightarrow{\mathbf{M}} &= \frac{\partial \overrightarrow{\omega}}{\partial t}\mathbf{I_q} + \overrightarrow{\omega} \times \mathbf{I_q}\overrightarrow{\omega} + \overrightarrow{\omega} \times \mathbf{I_r}\overrightarrow{\Omega} \end{aligned} \tag{4}$$

Quadrotors are normally constructed to be completely symmetric. Therefore, their tensor of inertia is a diagonal matrix 5. The same rule applies for rotors as well(otherwise they would be misbalanced and completely useless). Furthermore, rotors spin in one direction only, so the rotor angular speed vector $\overrightarrow{\Omega}$ has only one component Ω_z. Evaluating 3 yields a circular motion equation 6.

$$\mathbf{I_q} = \begin{bmatrix} I_{xx} & 0 & 0 \\ 0 & I_{yy} & 0 \\ 0 & 0 & I_{zz} \end{bmatrix} \tag{5}$$

$$\begin{aligned} M_x &= I_{xx}\frac{\mathrm{d}\omega_x}{\mathrm{d}t} - \left(I_{yy} - I_{zz}\right)\omega_y\omega_z + I_r\omega_y\Omega_z \\ M_y &= I_{yy}\frac{\mathrm{d}\omega_y}{\mathrm{d}t} - \left(I_{zz} - I_{xx}\right)\omega_x\omega_z - I_r\omega_x\Omega_z \\ M_z &= I_{zz}\frac{\mathrm{d}\omega_z}{\mathrm{d}t} - \left(I_{xx} - I_{yy}\right)\omega_x\omega_y \end{aligned} \tag{6}$$

Equation 6 calculates rotation speeds in the body frame coordinate system. To transform these body frame angular velocities into world frame rotations, one needs a transformation matrix 8. This matrix is derived from successive elemental transformations 7 similarly as kinematics equation 2. Infinitesimal changes in Euler angles, affect the rotation vector in a way that the first Euler angle Ψ undergoes two additional rotations, the second angle Θ only one additional rotation, and the final Euler angle Φ no additional rotations Jazar (2010):

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}^{\mathfrak{B}} = \begin{bmatrix} 0 \\ 0 \\ \dot{\Phi} \end{bmatrix}^{\mathfrak{W}} + Rot\left(\Phi, z^{\mathfrak{W}}\right)^T \begin{bmatrix} 0 \\ \dot{\Theta} \\ 0 \end{bmatrix}^{\mathfrak{W}} + Rot\left(\Phi, z^{\mathfrak{W}}\right)^T Rot\left(\Theta, y^{\mathfrak{W}}\right)^T \begin{bmatrix} \dot{\Psi} \\ 0 \\ 0 \end{bmatrix}^{\mathfrak{W}} \tag{7}$$

$$\mathbf{J} = \begin{bmatrix} cos(\Psi)/cos(\Theta) & sin(\Psi)/cos(\Theta) & 0 \\ -sin(\Psi) & cos(\Psi) & 0 \\ cos(\Psi)tan(\Theta) & sin(\Psi)tan(\Theta) & 1 \end{bmatrix} \tag{8}$$

2.3 Rotor forces and torques

Four quadrotor blades are placed in a square shaped form. Blades that are next to each other spin in opposite directions, thus maintaining inherent stability of the aircraft. The same four blades that make the quadrotor hover enable it to move in the desired direction. Therefore, in order for quadrotor to move, it has to be pitched and rolled in the desired direction. To pitch and roll the quadrotor, some blades need to spin faster, while others spin slower. This produces the desired torques, which in term affect aircraft attitude and position Orsag et al. (2010).

Depending on the orientation of the blades, relative to the body coordinate system, there are two basic types of quadrotor configurations: cross and plus configuration shown in figure 2. In the plus configuration, a pair of blades spinning in the same direction, are placed on x and y coordinates of the body frame coordinate system. With this configuration it is easier to control the aircraft, because each move (i.e. x or y direction) requires a controller to disbalance only the speeds of two blades placed on the desired direction.

The cross configuration, on the other hand, requires that the blades are placed in each quadrant of the body frame coordinate system. In such a configuration each move requires all four blades to vary their rotation speed. Although the control system seems to be more complex, there is one big advantage to the cross construction. Keeping in mind that the amount of torque needed to rotate the aircraft is very similar for both configurations, it takes less change per blade if all four blades change their speeds. Therefore, when the aircraft carries

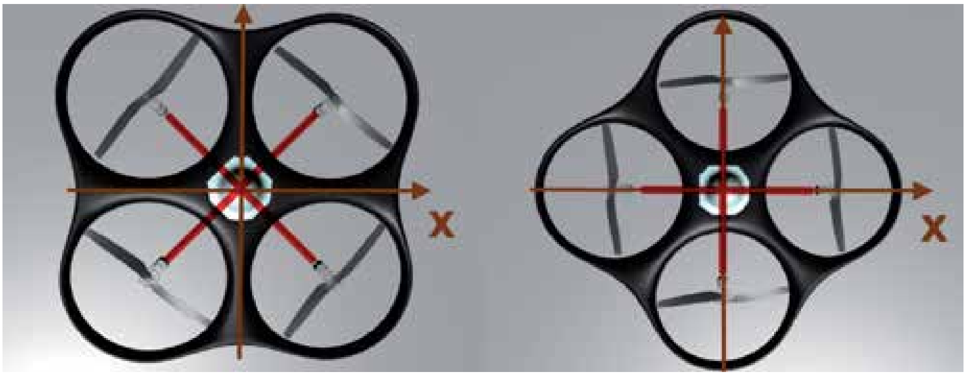


Fig. 2. A side by side image of X and Plus quadrotor configurations

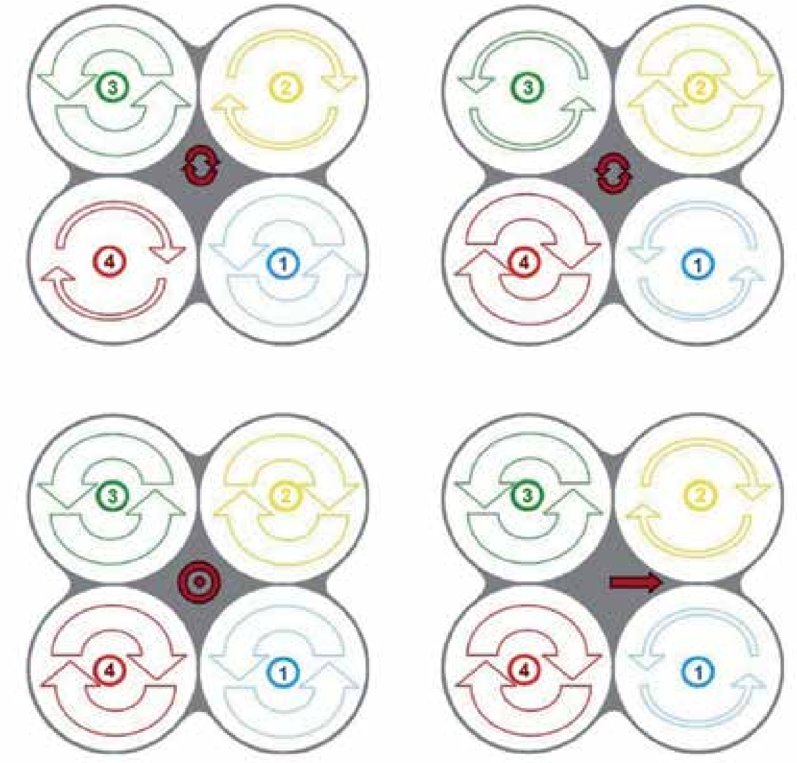


Fig. 3. Plus configuration control inputs for rotation, lift and forward motion. Arrow thickness stands for higher speed.

payload and operates near the point of saturation, it is wiser to use the cross configuration. Changing the speed of each blade for a small amount, as opposed to changing only two blades but doubling the amount of speed change, will keep the engines safe from saturation point. Basic control sequences of cross configuration are shown in figure 3. First approximation of rotor dynamics implies that rotors produce only the vertical thrust force. As the rotors are displaced from the axis of rotation (i.e. x and y axis) they produce corresponding torques,

$\overrightarrow{\mathbf{M_x}} = \overrightarrow{\mathbf{F}} \times \overrightarrow{\mathbf{r_x}}$ and $\overrightarrow{\mathbf{M_y}} = \overrightarrow{\mathbf{F}} \times \overrightarrow{\mathbf{r_y}}$ respectively. Torque $\overrightarrow{\mathbf{M_z}}$ comes from the spinning of each rotor blade $\mathbf{I_r}\overrightarrow{\boldsymbol{\Omega}}$. Adding the corresponding thrust forces and torques yields the following equation:

$$\begin{aligned}\overrightarrow{\mathbf{F_{tot}}} &= \overrightarrow{\mathbf{T_1}} + \overrightarrow{\mathbf{T_2}} + \overrightarrow{\mathbf{T_3}} + \overrightarrow{\mathbf{T_4}} \\ M_x^{tot} &= M_x^2 + M_x^3 - M_x^1 - M_x^4 \\ M_y^{tot} &= M_y^3 + M_y^4 - M_y^1 - M_y^2 \\ M_z^{tot} &= M_z^2 + M_z^4 - M_z^1 - M_z^3\end{aligned} \tag{9}$$

3. Aerodynamics

As the quadrotor research shifts to new research areas (i.e. Mobile manipulation, Aerobatic moves, etc.) Korpela et al. (2011); Mellinger et al. (2010), the need for an elaborate mathematical model arises. The model needs to incorporate a full spectrum of aerodynamic effects that act on the quadrotor during climb, descent and forward flight. To derive a more complete mathematical model of a quadrotor, one needs to start with basic concepts of momentum theory and blade elemental theory.

3.1 Combining momentum and blade elemental theory

The momentum theory of a rotor, also known as classical actuator disk theory, combines rotor thrust, induced velocity (i.e. airspeed produced in rotor) and aircraft speed into a single equation. On the other hand, blade elemental theory is used to calculate forces and torques acting on the rotor by studying a small rotor blade element modeled as an airplane wing so that the airfoil theory can be applied.Bramwell et al. (2001) A combination of these two views, macroscopic and microscopic, yields a base ground for a good approximative mathematical model.

3.1.1 Momentum theory

Basic momentum theory offers two solutions, one for each of the two operational states in which the defined rotor slipstream exists. The solutions refer to rotorcraft climb and descent, the so called helicopter and the windmill states. Quadrotor in a combined lateral and vertical move is shown in figure 4. The figure shows the most important airflows viewed in Momentum theory: V_z and V_{xy} that are induced by quadrotor's movement, together with the induced speed v_i that is produced by the rotors.

Unfortunately, classic momentum theory implies no steady state transition between the helicopter and the windmill states. Experimental results, however, show that this transition exists. In order for momentum theory to comply with experimental results, the augmented momentum theory equation 10 is proposed Gessow & Myers (1952),

$$T = 2\rho R^2 \pi v_i \sqrt{(v_i + V_z)^2 + {V_{xy}}^2 + \frac{{V_z}^2}{7.67}} \tag{10}$$

where $\frac{V_z^2}{7.67}$ term is introduced to assure that the augmented momentum theory equation complies with experimental results, R stands for rotor radius and ρ is the air density. It is easy to show that in case of autorotation with no forward speed, thrust in equation 10 becomes

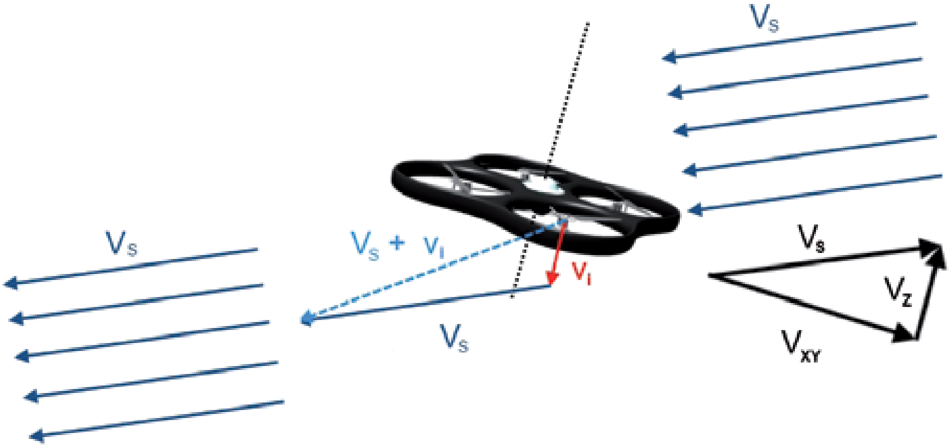


Fig. 4. Momentum theory - horizontal motion, vertical motion and induced speed total airflow vector sum

equal to the drag equation $D = \frac{1}{2}C_D\rho R^2\pi V_z^2$ of a free-falling plate with a drag coefficient $C_D = 1$.

3.1.2 Blade element theory

Blade element theory observes a small rotor blade element Δr 5. Figure 5 shows this infinitesimal part of quadrotor's blade together with elemental lift and drag forces it produces Bramwell et al. (2001). For better clarity angles are drawn larger than they actually are:

$$\begin{aligned}\frac{\Delta L}{\Delta R} &= \frac{1}{2}\rho V_{str} C_L S \\ \frac{\Delta D}{\Delta R} &= \frac{1}{2}\rho V_{str} C_D S\end{aligned} \tag{11}$$

where C_L and C_D are lift and drag coefficients, S is the surface of the element and V_{str} the airflow around the blade element. The airflow is mostly produced from the rotor spin ΩR and therefore depends on the distance of each blade element to the center of blade rotation. Adding to this airflow is the total air stream coming from quadrotor's vertical and horizontal movement, $V_S = V_{xy} + V_z$. Finally, blade rotation produces additional induced speed v_i. The ideal airfoil lift coefficient C_L can be calculated using equation 12 Gessow & Myers (1952).

$$C_L = a\alpha_{ef} = 2\pi\alpha_{ef} \tag{12}$$

where a is an aerodynamic coefficient, ideally equal to 2π. The effective angle of attack α_{ef}, is the angle between the airflow and the blade. Its value changes with the change of airflow direction and due to the blade twist.

Standard rotor blades are twisted because the dominant airflow coming from blade rotation increases linearly towards the end of the blade. According to equation 11 this causes the increase of lift and drag forces. The difference in forces produced near and far from the center of rotation would cause the blade to twist, and ultimately brake. To avoid that, a linear twist,

$$\alpha_m(r) = \Theta_0 - \frac{r}{R}Q_{tw} \tag{13}$$

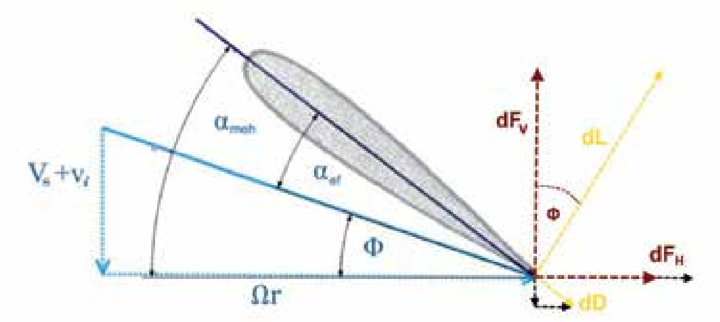


Fig. 5. Infitesimal rotor blade element Δr in surrounding airflow Orsag & Bogdan (2009)

is introduced to the blade design.

The effect of varying airflow can be calculated separating the vertical components $V_z + v_i$ and horizontal ones $V_{xy} + \Omega r$. The airflow direction angle Φ can be easily calculated from the equation

$$\Phi = \arctan\left(\frac{V_z + v_i}{V_x y + \Omega r}\right) \approx \arctan\left(\frac{V_z + v_i}{\Omega r}\right) \tag{14}$$

As lift and drag forces are not aligned with body frame of reference, horizontal and vertical projection forces need to be derived. Keeping in mind that $\Omega r \gg \{V_z, v_i, V_x y\}$ small angle approximations $\cos(\Phi) \approx 1$ and $\sin(\Phi) \approx \Phi$ can be used. Moreover, in a well balanced rotor blade, drag force should be negligible compared to the lift Gessow & Myers (1952). Applying this considerations to 11 and keeping in mind the relations from figure 5 enables the derivation of horizontal and vertical force equations 15.

$$\begin{aligned} \frac{\mathrm{d}F_V}{\mathrm{d}r} &= \frac{\mathrm{d}L}{\mathrm{d}r}\cos(\Phi) + \frac{\mathrm{d}D}{\mathrm{d}r}\sin(\Phi) \approx \frac{\mathrm{d}L}{\mathrm{d}r} = \rho V_{tot}^2 c\pi\alpha_{ef} \\ \frac{\mathrm{d}F_H}{\mathrm{d}r} &= \frac{\mathrm{d}L}{\mathrm{d}r}\sin(\Phi) + \frac{\mathrm{d}D}{\mathrm{d}r}\cos(\Phi) \approx \frac{1}{2}\rho V_{tot}^2 C_D S + \frac{1}{2}\rho V_{tot}^2 C_L S\Phi \end{aligned} \tag{15}$$

3.1.3 Applying blade element theory to quadrotor construction

This section continues with the observation of a small rotor blade element Δr from the previous section, placing it in real surroundings shown in figure 6. Since the blades rotate, the forces produced by blade elements tend to change both in size and direction. This is the reason why an average elemental thrust of all blade elements should be calculated.

Figure 6 shows the relative position of one rotor as it is seen from quadrotor's body frame. This rotor is displaced from the body frame origin and forms an angle of 45° with quadrotor's body frame x axis. Similar relations can be shown for other rotors. Accounting for the number of rotor blades N, the following equation for rotor vertical thrust force calculation is proposed Orsag & Bogdan (2009):

$$T = F_V = \frac{1}{2\pi}\int_0^{2\pi}\int_0^R N\frac{\Delta F_V}{\Delta R} dr d\psi \tag{16}$$

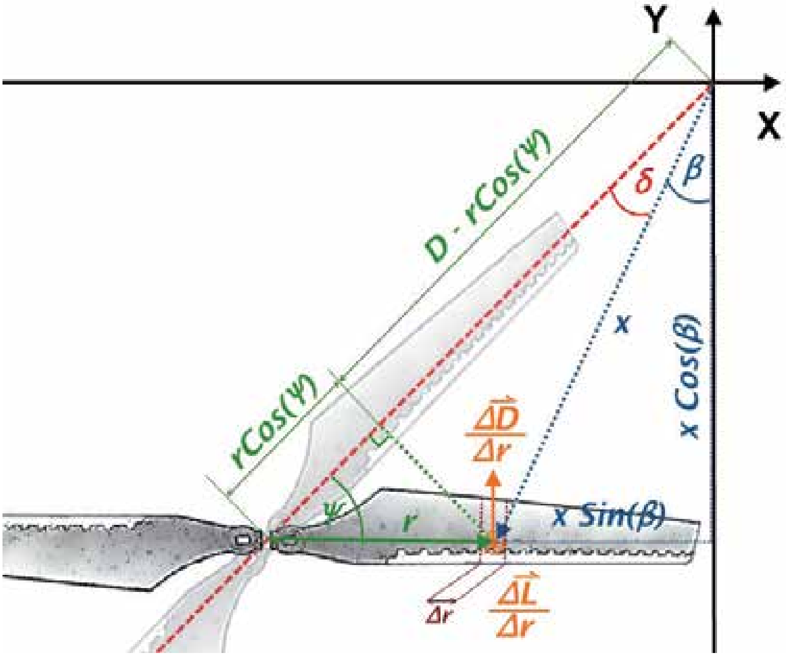


Fig. 6. Blade element in quadrotor coordinate system

where ψ is the blade angle due to rotation, taken at a certain sample time. Solving integral equation 16 yields the expression for rotor thrust (i.e. vertical force) Orsag & Bogdan (2009):

$$F_V = \frac{N\rho a\bar{c}R^3\Omega^2}{4}\left[\left(\frac{2}{3}+\mu^2\right)\Theta_0-\left(1+\mu^2\right)\frac{\Theta_{tw}}{2}-\lambda_i-\lambda_c\right] \tag{17}$$

The term inside the brackets of equation 17 is known as a thrust coefficient, and is given separately in 18.

$$C_T = \left(\frac{2}{3}+\mu^2\right)\Theta_0-\left(1+\mu^2\right)\frac{\Theta_{tw}}{2}-\lambda_i-\lambda_c \tag{18}$$

Variables μ,λ_i and λ_c are speed coefficients $\frac{V_{xy}}{R\Omega}$,$\frac{V_z}{R\Omega}$ and $\frac{v_i}{R\Omega}$ respectively. New constant $\bar{c}$ is the average cord length of the blade element shown in figure 5.

The same approach can be applied for the calculation of horizontal forces and torques produced within the quadrotor Orsag & Bogdan (2009). Calculated lateral force has x and y components, coming both from the drag and lift of the rotor, given in 19.

$$\begin{aligned} C_{Hx} &= \cos(\alpha)\,\mu\left[\frac{C_D}{a}+(\lambda_i+\lambda_c)\left(\Theta_0-\frac{\Theta_{tw}}{2}\right)\right] \\ C_{Hy} &= \sin(\alpha)\,\mu\left[\frac{C_D}{a}+(\lambda_i+\lambda_c)\left(\Theta_0-\frac{\Theta_{tw}}{2}\right)\right] \end{aligned} \tag{19}$$

In case of torque equations the angles between the forces and directions are easily derived from basic geometric relations shown in figure 6, resulting in the elemental torque equations

Orsag & Bogdan (2009):

$$
\begin{aligned}
\frac{\Delta M_z}{\Delta r} &= -\frac{\Delta F_H}{\Delta r}\left(D\cos\left(\Psi - \frac{pi}{4}\right) - r\right) \\
\frac{\Delta M_{xy}}{\Delta r} &= -\frac{\sqrt{2}\Delta F_V}{2\Delta r}\left(D - r\cos\left(\Psi - \frac{\pi}{4}\right) \pm r\sin\left(\Psi - \frac{pi}{4}\right)\right)
\end{aligned}
\tag{20}
$$

Using the same methods which were used for force calculation, the following momentum coefficients were calculated:

$$
\begin{aligned}
C_{Mz} &= R\left[\frac{1+\mu^2}{2a}C_D - C_T\left(\mu,\lambda,\lambda_i\right)|_{\mu=0}\right] \pm D\mu\cos\left(\frac{\pi}{4} + \phi\frac{C_{Hx}}{\cos(\phi)\mu}\right) \\
C_{Mx} &= D\frac{\sqrt{2}}{2}C_T \pm R\mu\sin(\phi)\left[\frac{2}{3}\Theta_0 - \frac{1}{2}\left(\Theta_{tw} + \lambda\right)\right] \\
C_{My} &= D\frac{\sqrt{2}}{2}C_T \pm R\mu\cos(\phi)\left[\frac{2}{3}\Theta_0 - \frac{1}{2}\left(\Theta_{tw} + \lambda\right)\right]
\end{aligned}
\tag{21}
$$

It is important to notice that equations 20 have two solutions, since the rotors spin in different directions, as seen in figure 3. Different rotational directions have the opposite effect on torques. This is why the ± sign is used in torque equations. These differences, induced from the specific quadrotor construction, along with the augmented momentum equation provide an improved insight to quadrotor aerodynamics. Regardless of the flying state of the quadrotor, by using these equations one can effectively model its behavior.

3.2 Building a more realistic rotor model

Building a more realistic rotor model begins with redefining its widely accepted static thrust equation 22 with real experimental results. No matter how precise, static equation is valid only when quadrotor remains stationary (i.e. hover mode). In order for the equation to be valid during quadrotor maneuvers, aerodynamic effects from 3.1 need to be incorporated into the equation.

$$
T \sim k_T\Omega^2 \tag{22}
$$

3.2.1 Experimental results

This section presents the experimental results of a static thrust equation for an example quadrotor. Most of researched quadrotors use DC motors to drive the rotors. Although new designs use brushless DC motors (BLDC), brushed motors are still used due to their lower cost. Some advantages of brushless over brushed DC motors include more torque per weight, more torque per watt (increased efficiency) and increased reliability Sanchez et al. (2011); Solomon & Famouri (2006); Y. (2003).

Quadrotor used in described experiments is equipped with a standard brushed DC motor. Experimental results show that quadratic relationship between rotor speed (applied voltage) and resulting thrust is valid for certain range of voltages. Moving close to saturation point (i.e. 11V-12V), the quadratic relation of thrust and rotor speed deteriorates. Experimental results are shown in figure 7 and in the table 1.

Voltage [V]	Rotation speedΩ [rpm]	Induced speed v_i [m/s]	Thrust [N]
4.04	194.465	1.5	0.16
5.01	241.17	2	0.29
5.99	284.105	2.45	0.44
6.99	328.82	2.7	0.58
8.00	367.357	3.2	0.72
8.98	403.171	3.5	0.94
10.02	433.540	3.8	1.16
10.99	464.223	4.05	1.34
12.05	490.088	4.3	1.42

Table 1. Data collected from the experiments

In order to use thrust equation 18, certain coefficients need to be known. Some of them like rotor radius R and cord length c can be measured. Others, like the mechanical angle Θ_0 have to be calculated. Solving thrust equation 18 for $\mu = 0$ and $\lambda_c = 0$ (i.e. static conditions) yields:

$$F_V = \frac{1}{2}\rho a c \omega^2 R^3 \left(\Theta_{\frac{3}{4}} - \lambda_i\right) \tag{23}$$

where $\Theta_{\frac{3}{4}}$ is a mechanical angle at the $\frac{3}{4}$ of the blade length R 13. Θ_{tw} can later be assessed from the blade construction. Rearranging equation 23 yields an equation for solving the mechanical angle problem 24.

$$\Theta_{\frac{3}{4}} = \frac{3}{2}\left(\frac{2F_V}{\rho a c \Omega^2 R^3} + \lambda_i\right) \tag{24}$$

Using experimental data from table 1 it is easy to calculate rotor angle $\Theta_{\frac{3}{4}}$. For given set of data the average $\lambda_i = \frac{v_i}{\Omega R} = 0.0766$. Therefore the mechanical angle $\Theta_{\frac{3}{4}} = 11.6291^o$, which is well between the expected boundaries.

Obtained data is piecewise linearized, in order to clearly demonstrate the differences between various voltage ranges. From Fig. 7 it can be seen how thrust declines near the point of saturation. This is important to notice, when deriving valid algorithms for quadrotor stabilization and control. Linearizaton coefficients are given in table 2.

Voltage [V]	Linear gain [N/V]
$[0-3]$	0
$[3-8]$	0.1433
$[8-11]$	0.2070
$[11-12]$	0.08

Table 2. Piecewise linearization coefficients

3.2.2 Applying aerodynamics to rotor dynamic model

To apply aerodynamic coefficient 18 to the static thrust experimental results, one needs to multiply experimental results with dynamic-to-static aerodynamic coefficient ratio 25.

$$T(\mu, \lambda_c, \lambda_i) = \frac{C_T(\mu, \lambda_c, \lambda_i)}{C_T(0,0,0)} T(0,0,0) \tag{25}$$

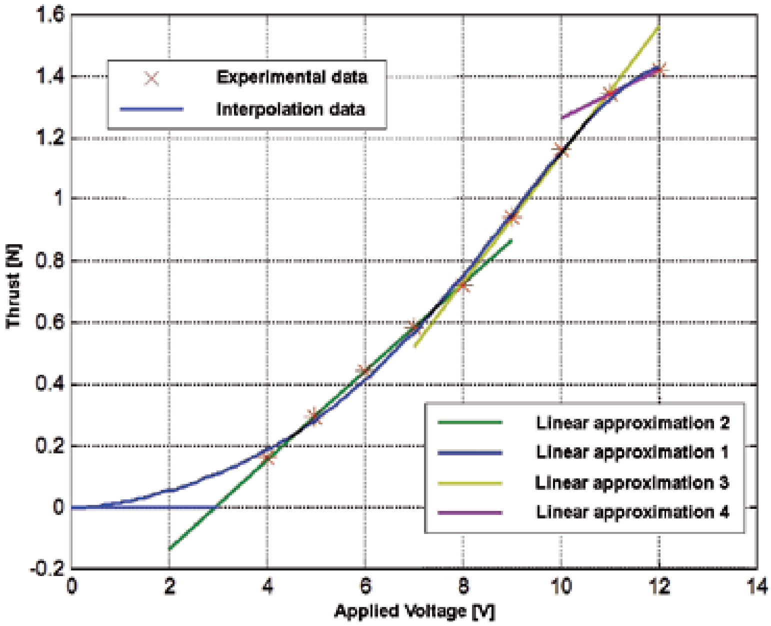


Fig. 7. Static rotor thrust experimental results with interpolation function and piecewise linear approximations

For the calculation of the aerodynamic coefficient C_T it is crucial to know three airspeed coefficients μ, λ_c and λ_i. Two of them, μ, λ_c, can easily be obtained from the available motion data V_{xy}, V_z and ΩR. λ_i however, is very hard to know, because it is impossible to measure the induced velocity v_i.

One way to solve this problem is to calculate the induced velocity coefficient λ_i from the two aerodynamic principals, momentum and blade element theories. The macroscopic momentum equation 10 and the microscopic blade element equation 17 provide the same rotor thrust using different physical approach:

$$T = \frac{1}{4}\rho a R^3 \Omega^2 c \left[\frac{2}{3}\alpha_{meh}\left(1+\frac{3}{2}\mu^2\right) - \lambda_i - \lambda_c\right] = 2\rho R^2 \pi \lambda_i \sqrt{(\lambda_c+\lambda_i)^2 + \mu^2 + \frac{\lambda_c^2}{7,67}} \quad (26)$$

When squared, equation 26 can be easily solved as a quadrinome:

$$\begin{aligned} \lambda_i^4 + p_3\lambda_i^3 + p_2\lambda_i^2 + p_1\lambda_i + p_0 &= 0 \\ p_0 &= -c_1 c_2^2 \\ p_1 &= 2c_1 c_2 \\ p_2 = \left(1+\frac{1}{7.67}\right)\lambda_c^2 + \mu^2 - c_1 \\ p_3 &= 2\lambda_c \\ c_1 &= \frac{a^2 s^2}{64} \\ c_2 = 2\Theta_0\left(1+1.5\mu^2\right)/3 - \lambda_c \end{aligned} \quad (27)$$

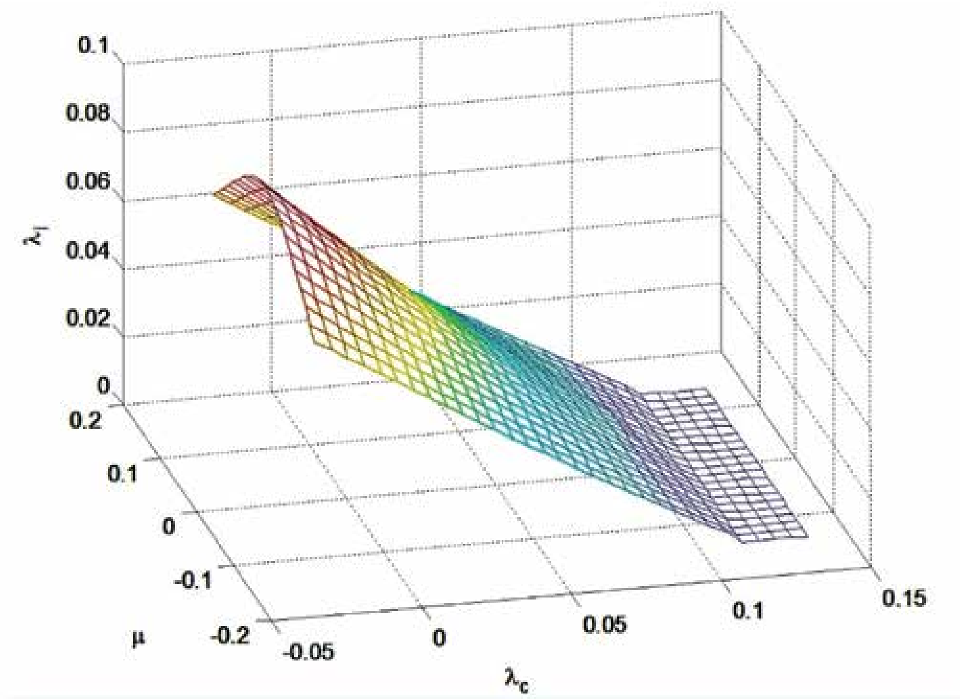


Fig. 8. 3D representation of λ_i change during horizontal and vertical movement

The results of solving this quadrinome can be shown in a 3D graph 8. Although equations 27 look straightforward to solve, it still requires a substantial amount of processor capacity. This is why an offline calculation is proposed. This way, the calculated data can be used during simulation without the need for online computation. By using calculated values of the induced velocity, it is easy to calculate the dynamic thrust coefficient from equation 18. The 3D representation of final results is shown in figure 9.

Due to an increase of airflow produced by quadrotor movement, the induced velocity decreases. This can be seen in figure 8. Although both movements tend to increase induced velocity, only the vertical movement decreases the thrust coefficient. As a result, during takeoff the quadrotor looses rotor thrust, but during horizontal movement that same thrust is increased and enables more aggressive maneuvers.

3.2.3 Quadrotor model

A complete quadrotor model, incorporating previously mentioned effects is shown in figure 10. A control input block feeds the voltage signals to calculate statics thrust, which is easily interpolated from the available experimental data, using an interpolation function as shown in figure 7.

Static rotor thrust is applied to equation 25 along with aerodynamic coefficient $C_T(\mu, \lambda_c, \lambda_i)$. Induced velocity and aerodynamic coefficient are calculated using inputs from the current flight data (i.e. λ_c, μ). This data is supplied from the Quadrotor Dynamics block. The calculation can be done offline, so that a set of data points from figure 9 can be used to

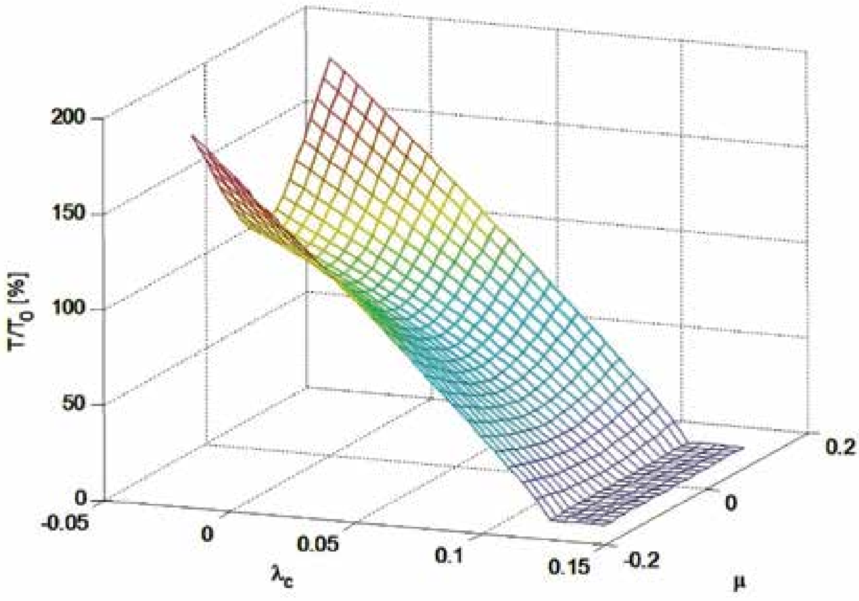


Fig. 9. 3D representation of $\frac{T(\lambda_i,\lambda_c,\mu)}{T(0,0,0)}$ ratio during horizontal and vertical movement

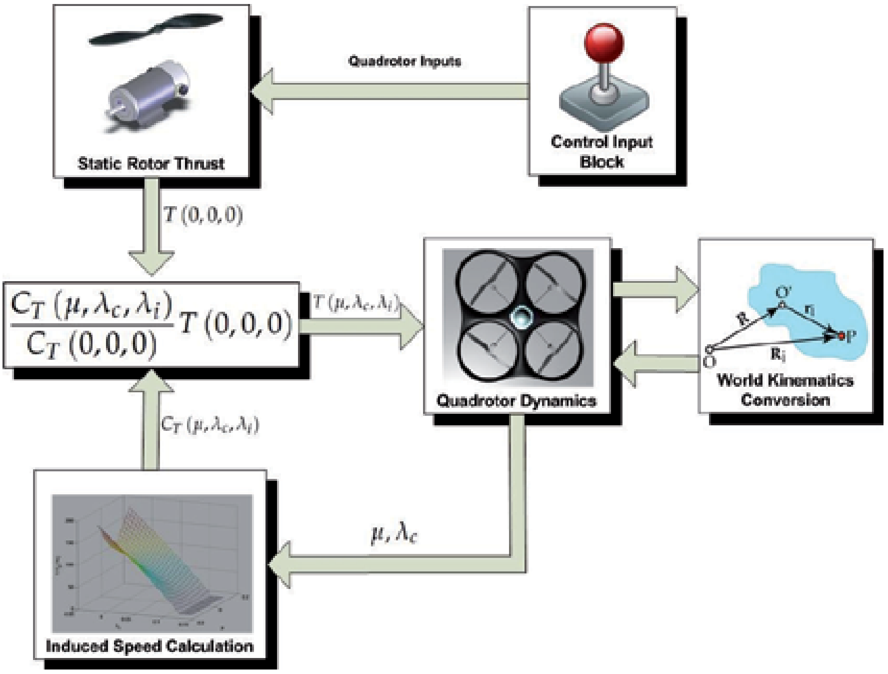

Quadrotor Inputs
Control Input Block
Static Rotor Thrust
$T\,(0,0,0)$
$\frac{C_T\,(\mu,\lambda_c,\lambda_i)}{C_T\,(0,0,0)}T\,(0,0,0)$
$T\,(\mu,\lambda_c,\lambda_i)$
Quadrotor Dynamics
O'
r_i
R
P
O
R_i
World Kinematics Conversion
$C_T\,(\mu,\lambda_c,\lambda_i)$
μ, λ_c
Induced Speed Calculation

Fig. 10. Quadrotor model

interpolate true aerodynamic coefficient. This speeds up the simulation, as opposed to solving the quadrinome problem online.

A combination of the results provided from these two blocks using equation 25 gives the true aerodynamic rotor thrust. The same procedure is used to calculate the induced speed from the data shown in figure 8. Once the exact induced speed is known it can be applied to horizontal coefficients 19 and torque coefficients 21. In this way, quadrotor dynamics block can calculate quadrotors angular and linear dynamics using equations 6 and 3.

Dynamics data is finally fed into the kinematics block, that calculates quadrotor motion in world coordinate system using transformation matrices 2 and 7.

4. Conclusion

As the unmanned aerial research community shifts its efforts towards more and more aggressive flying maneuvers as well as mobile manipulation, the need for a more complete aerodynamic quadrotor model, such as the one presented in this chapter arises.

The chapter introduces a nonlinear mathematical model that incorporates aerodynamic effects of forward and vertical flights. A clear insight on how to incorporate these effects to a basic quadrotor model is given. Experimental results of widely used brushed DC motors are presented. The results show negative saturation effects observed when using this type of DC motors, as well as the phenomenon of thrust variations during quadrotor's flight.

The proposed model incorporates aerodynamic effects using offline precalculated data, that can easily be added to existing basic quadrotor model. Furthermore, the model described in the paper can incorporate additional aerodynamic effects like the In Ground Effect.

5. References

Adigbli, P., Grand, C., Mouret, J.-B. & Doncieux, S. (2007). Nonlinear attitude and position control of a micro quadrotor using sliding mode and backstepping techniques, *7th European Micro Air Vehicle Conference (MAV07)*, Toulouse.

Anderson, S. B. (1981). Historical overview of v/stol aircraft technology, *NASA Technical Memorandum* .

Bouabdallah, S., Noth, A. & Siegwart, R. (2004). Pid vs lq control techniques applied to an indoor micro quadrotor, *Proc. of The IEEE International Conference on Intelligent Robots and Systems (IROS)*.

Bouabdallah, S. & Siegwart, R. (2005). Backstepping and sliding-mode techniques applied to an indoor micro quadrotor, *Proc. of The IEEE International Conference on Robotics and Automation (ICRA)*.

Bramwell, A., Done, G. & Balmford, D. (2001). *Bramwell's helicopter dynamics*, American Institute of Aeronautics and Astronautics.

Chaturvedi, N., Sanyal, A. & McClamroch, N. (2011). Rigid-body attitude control, *Control systems magazine* .

Gessow, A. & Myers, G. (1952). *Aerodynamics of the helicopter*, F. Ungar Pub. Co.

Hoffmann, G. M., Huang, H., Wasl, S. L. & Tomlin, E. C. J. (2007). Quadrotor helicopter flight dynamics and control: Theory and experiment, *In Proc. of the AIAA Guidance, Navigation, and Control Conference*.

Jazar, R. (2010). *Theory of Applied Robotics: Kinematics, Dynamics, and Control (2nd Edition)*, Springer.

Korpela, C. M., Danko, T. W. & Oh, P. Y. (2011). Mm-uav: Mobile manipulating unmanned aerial vehicle, *In Proc. of the International Conference on Unmanned Aerial Systems (ICUAS)*.

Mellinger, D., Michael, N. & Kumar, V. (2010). Trajectory generation and control for precise aggressive maneuvers with quadrotors, *Proceedings of the International Symposium on Experimental Robotics*.

Orsag, M. & Bogdan, S. (2009). Hybrid control of quadrotor, *Proc.of the 17. Mediterranean Conference on Control and Automation, (MED)*.

Orsag, M., Poropat, M. & S., B. (2010). Hybrid fly-by-wire quadrotor controller, *AUTOMATIKA: Journal for Control, Measurement, Electronics, Computing and Communications* .

Sanchez, A., García Carrillo, L. R., Rondon, E., Lozano, R. & Garcia, O. (2011). Hovering flight improvement of a quad-rotor mini uav using brushless dc motors, *J. Intell. Robotics Syst.* 61: 85–101.
URL: *http://dx.doi.org/10.1007/s10846-010-9470-3*

Shuster, M. D. (1993). Survey of attitude representations, *Journal of the Astronautical Sciences* 41: 439–517.

Solomon, O. & Famouri, P. (2006). Dynamic performance of a permanent magnet brushless dc motor for uav electric propulsion system - part i, *Proc.of IEEE Industrial Electronics, (IECON)*.

Stingu, E. & Lewis, F. (2009). Design and implementation of a structured flight controller for a 6dof quadrotor using quaternions, *Proc.of the 17. Mediterranean Conference on Control and Automation, (MED)*.

Varga, M. & Bogdan, S. (2009). FuzzyâĂŞlyapunov based quadrotor controller design, *Proc. of the European Control Conference,(ECC)*.

Warwick, G. (2007). Army looking at three configuration concepts for large cargo rotorcraft, *Flight International via www.flightglobal.com* .

Y., P. (2003). Brushless dc (bldc) motor fundamentals, *Microchip Application Notes, AN885* .

8

Advanced Graph Search Algorithms for Path Planning of Flight Vehicles

Luca De Filippis and Giorgio Guglieri
Politecnico di Torino
Italy

1. Introduction

Path planning is one of the most important tasks for mission definition and management of manned flight vehicles and it is crucial for Unmanned Aerial Vehicles (UAVs) that have autonomous flight capabilities. This task involves mission constraints, vehicle's characteristics and mission environment that must be combined in order to comply with the mission requirements. Nevertheless, to implement an effective path planning strategy, a deep analysis of various contributing elements is needed. Mission tasks, required payload and surveillance systems drive the aircraft selection, but its characteristics strongly influence the path. As an example, quad-rotors have hovering capabilities. This feature permits to relax turning constraints on the path (which represents a crucial problem for fixed-wing vehicles). The type of mission defines the environment for planning actions, the path constraints (mountains, hills, valleys, …) and the required optimization process. The need for off-line or real-time re-planning may also substantially revise the path planning strategy for the selected type of missions. Finally, the computational performances of the Remote Control Station (RCS), where the mission management system is generally running, can influence the algorithm selection and design, as time constraints can be a serious operational issue.

This chapter aims to cover three main topics:

- Describe the most important algorithms developed for path planning of flying vehicles in order to compare them and depict their merits and drawbacks.
- Focus on graph search algorithms in order to define their main characteristics and provide a complete overview of the most important methods developed.
- Present a new graph search algorithm (called Kinematic A*) that has been developed on the base of the well-known A* algorithm and aims to fill the relation gap between the path planned with classical graph search solutions and the aircraft kinematic constraints.

The chapter is structured as follow:

- General description of the most important path planning algorithms:
    - Introduction to first approaches to path planning: manual path planning and Dubins curves. Also some simple applications developed by this research group are presented.

  - General description of probabilistic and graph search algorithms.
  - General description of potential field and model predictive algorithms.
  - Introduction to some generic optimization algorithms.
- Study on graph search algorithms:
  - General description of commonality and differences between methods composing this family. Basic algorithm structure identification and introduction to the general features of these methods.
  - First graph search solutions focusing on the A* algorithm.
  - Introduction to dynamic-graph search and to the principal developed methods.
  - "Any heading" algorithms description focusing on Theta*.
  - Brief comparison between Theta* and A* on paths planned with the tools developed by this research group, focusing on the main improvements introduced with Theta*.
- Kinematic A*:
  - State space definition: in order to implement Kinematic A*, redefinition of the state space is needed.
  - Kinematic model description: the system of differential equation modelling the aircraft kinematic behaviour.
  - Introduction of wind in the kinematic model in order to take into account this disturbance on the path.
  - Formulation of the optimization problem solved with the graph search approach.
  - Constraints definition identifying the set of states evaluated to find the optimal path.
  - Algorithm description.
- Results presentation in order to identify new algorithm merits and drawbacks:
  - Algorithm test on a square map collecting four obstacles placed close to the four corners. A* path comparison with the Kinematic A* one planned with and without wind.
  - Algorithm test on a square map with one obstacle obstructing the path. This test is made to verify the algorithm search performances.
- Conclusion and future work description.

2. The path planning task

Generally, path planning aims to generate a real-time trajectory to a target, avoiding obstacles or collisions (assuming reference flight-conditions and providing maps of the environment), but also optimizing a given functional under kinematic and/or dynamic constraints. Several solutions were developed matching different planning requirements: performances optimization, collision avoidance, real-time planning or risk minimization, etc. Several algorithms were designed for robotic systems and ground vehicles. They took hints from research fields like physics for potential field algorithms, mathematics for probabilistic approaches, or computer science for graph search algorithms. Each family of algorithms has been tailored for path planning of UAVs, and future work will enforce the development of new strategies.

2.1 Manual path planning and Dubins curves

First studies on path planning of unmanned aircrafts evidenced task complexities, strict safety requirements and reduced technological capabilities that imposed as unique solution manual approaches for path planning of UASs. The waypoint sequences where based on the environment map and on the mission tasks, taking into account some basic kinematic constraints. The flight programs were then loaded on the aircraft flight control system (FCS) and the path tracking were monitored in real time. These approaches were overtaken researching on this problem, but some of them are still used for industrial applications where the plan complexity requires a human supervision at all stages. In these cases computer tools driving the waypoints allocation, the path feasibility verification and the waypoints-sequence conversion in formats compatible with the aircraft FCS assist the human agent.

The above-mentioned path-planning procedures were investigated (De Filippis et al., 2009) and some simple tools were developed in Matlab/Simulink and integrated into a single software package. This software (named PCube) handles geotiff and Digital Elevation Models to generate waypoint sequences compatible with the programming scripts of Micropilot commercial autopilots. The tool has a basic Graphical User Interface (GUI) used to manage the map and the path planning sequence. This tool can be used to:

- generate point and click waypoint sequences,
- choose predefined path shapes (square, rectangular and butterfly shapes),
- generate automatic grid type waypoint sequences (grid patterns for photogrammetric use).

If manual planning is the first and basic approach to path planning, it motivated research of more accurate solutions. In this direction optimization of the path with respect to some performance parameters was the challenge. Many approaches from optimal theory were studied and adapted to path planning and the Dubins curves are one of the most used and attractive solutions for their conceptual and implementation simplicity.

In a bi-dimensional space a couple of points each one associated to a unitary vector is given such that a vehicle is supposed to pass from these points with its trajectory tangent to the vector in that point. Dubins considered a non-holonomic vehicle moving at constant speed with limited turning capabilities and tried to find the shortest path between the two points under such constraints. He demonstrated that assuming constant turning radiuses this path exists and analysing each possible case a set of geodesic curves can be defined (Dubins, 1957). The same work was moved forward through successive studies on holonomic vehicles (Reeds & Shepp, 1990).

Dubins curves are used in PCube to take into account the UAVs turn performances and average flight speed, reallocating waypoints violating the constraints. For grid type patterns, the path generation is optimized for optical type payloads, specifying image overlaps and focal length. The package also allows the manipulation of maps and flight paths (i.e. sizing, scaling and rotation of mapped patterns). 3D surface and contour level plots are available for enhancing the visualization of the flight path. Coordinates and map formats can also be converted in different standards according to user specifications.

An example of manual planning using the point and click technique on a highland area is shown in **Figure 1**. Where the waypoint sequence defined by the user has been modified exploiting the Dubins curves. Manual path planning can generate paths with very simple logics when the optimization constraints do not affect the task and more complex solutions were developed and implemented.

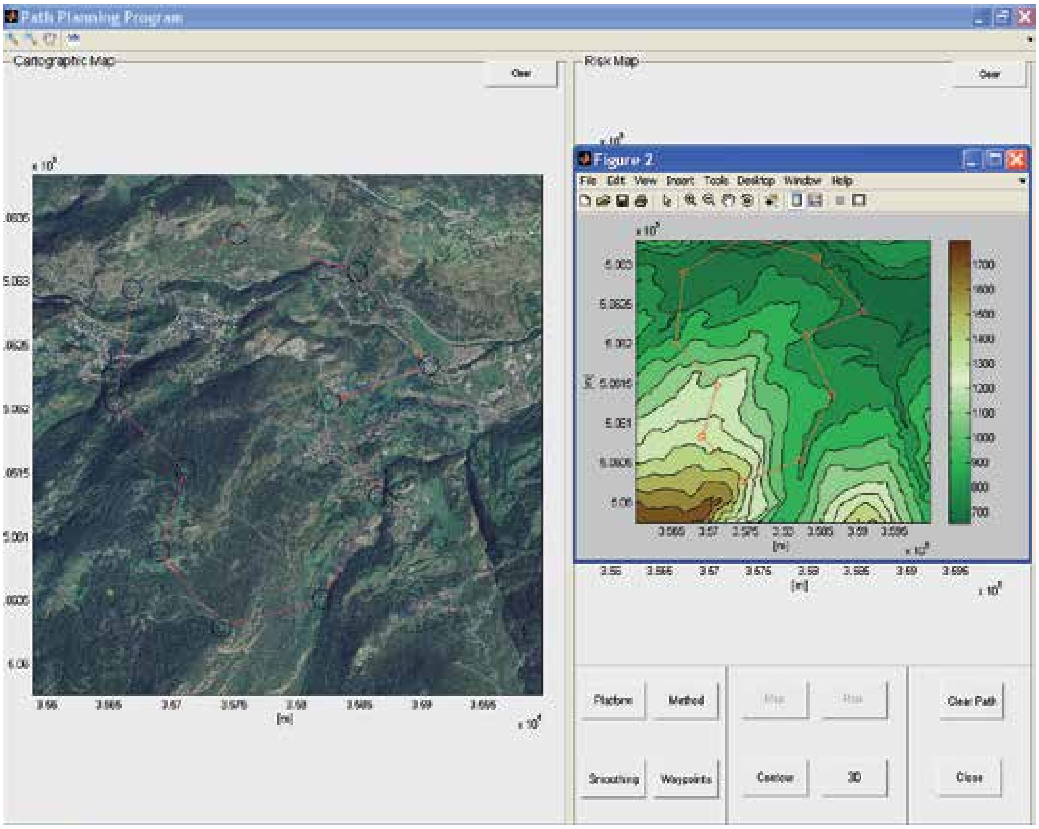


Fig. 1. Manula path planning with PCube Graphical User Interface (GUI).

2.2 Probabilistic and graph search algorithms

The problem of path planning is just an optimization problem made complex by the concurring parameters to be optimized on the same path. These parameters sometimes jar each other and they have to be balanced with respect to the mission tasks. All the more advanced algorithms developed for path planning try to identify the object of the optimization and reformulate the problem to cope with the prominent task, finding different approaches to optimize the parameter connected with this task. Many of them were developed for other applications and were modified to match with the problem of path planning. It's the case of the probability algorithms.

These algorithms generate a probability distribution connected with the parameter to be optimized and they implement statistic techniques to find the most probable path that optimizes this parameter (Jun & D'Andrea, 2002). Many implementations are related with the risk distribution of some kind of threat on a map (obstacles, forbidden areas, wind, etc.) and the algorithm creates probabilistic maps (Bertuccelli & How 2005) or look-up tables (that can be updated in real-time) modelling this distribution with various theories and logics (Pfeiffer, B et al., 2008). Markov processes are commonly used to introduce probabilistic uncertainties on the problem of path planning and Markov decision processes (MDPs) are defined all the approaches connecting this uncertainty with the taken action. These techniques are useful for all the cases where the optimization parameters are uncertain and can change in time and space, like conditions of flight, environment, and mission tasks.

Graph search algorithms are then interesting techniques coming from computer science. They were developed to find optimal plans to drive data exchanges on computer networks. These algorithms are commonly defined "greedy" as they generate a local optimal solution that can be quite far from the global optimal one. These algorithms are widely used in different fields thanks to their simplicity and small computational load and in the last five decades they evolved from basic approaches as Djikstra and Bellman-Ford algorithms to more complex solutions as D* Lite and Theta*. All of them differ in some aspects related to arc-weights definition and cost-function, but they are very similar in the implementation philosophy.

The main drawback of probabilistic and graph search algorithms resides in the lack of correlation between the aircraft kinematics and the planned path. Commonly, after the path between nodes of the graph has been generated with the minimum path algorithms, it has to be smoothed in order to be adapted to the vehicle flight performances. Indeed greedy algorithms provide a path constituted by line segments connected with edges that can't be followed by any type of flight vehicle. In order to obtain a more feasible and realistic path, refinement algorithms have to be used. This kind of algorithms can be very different in nature, starting from geometric curve definition algorithms also line flow smoothing logic can be used, but in any case at the end of this process a more realistic path is obtained, which better matches with autopilot control characteristics and flight performances.

Successive research on path planning algorithms brought to development of potential field based solutions. First potential field implementations came out to solve obstacles avoidance and formation flight problems, but in the last few years trajectory optimization under some performance constrains has been investigated.

2.3 Potential field and model predictive algorithms

Potential field algorithms come from robotic science and have been adapted to UASs simply modifying the kinematic models and the obstacles models. The environment is modelled to generate attractive forces toward the goal and repulsive ones around the obstacles (Dogan, 2003). The potential field model can be magnetic or electric (Horner & Healey, 2004), but the methods derived from aerodynamics provide the best choice in generation of trajectories for flight (Waydo & Murray, 2003). The vehicle motion is forced to follow the energy minimum respecting some dynamic constraints connected with its characteristics (Ford & Fulkerson, 1962). Two important aerodynamic field methods can be mentioned here: one obtained modelling path through propagation of pressure waves and another based on streamline modelling the motion field. The first method has been implemented supposing the fluid expanding from the target position through the starting one and modelling objects in the environment as obstacles. The second method instead models the environment like an aerodynamic field where obstacles are represented with singularities characterized by outgoing flow direction and target position like attracting singularities. The trajectory is chosen between all the streamlines defined in the field, to minimize the potential field gradient.

As a matter of fact these algorithms give smoothed and flyable paths, avoiding static and dynamic obstacles according with the field complexity. In the last years they have been widely investigated and interesting applications have been published. Even tough they are a promising solution for path planning and collision avoidance their application to some problems seemed hard due to their tendency to local minima on complex potential models.

The last and more advanced family of methods presented here, is based on technique coming from control science and applied to path planning and collision avoidance in the last decades. These algorithms apply model predictive control techniques to path planning problems linking a simplified model of the vehicle to some optimization parameters.

These algorithms solve in open loop an optimization problem constrained with a set of differential equations over a finite time horizon. The fundamental idea is to generate a control input that respects vehicle dynamics, environment characteristics and optimization constrains inside the defined time step and to repeat this process each step up to reach the goal. Sensors data can be integrated to update the model states so that these algorithms are used for collision avoidance in presence of active obstacles and particular harsh environments.

The big merit of model predictive solutions is the inclusion inside the optimization problem of the vehicle kinematics and dynamics in order to generate flyable trajectories. Model Predictive Control (MPC) or Receding Horizon Control (RHC) are the first techniques developed for industrial processes control that have been adapted to path planning (Ma & Castanon, 2006). Relation between control theory and path planning underlines another important characteristic of these methods. Indeed using the same logic for control and path planning opens the possibility to generate an integrated system that provides trajectories and control signals. On the other hand because complex sets of differential equations solved iteratively to generate the path are used in these methods, computation speed has been a real issue for these algorithms to spread. Also, as more as the problem complexity increases, as more the optimization space becomes complex and convergence to the optimal solution becomes an issue. Though, successive evolution of the model predictive technique is the Mixed-Integer Linear Programming (MILP). This algorithm applies the same logics of the model predictive one but allows inclusion of integer variables and discrete logics in a continuous linear optimization problem. Variables are used to model obstacles and to generate collision avoidance rules, while dynamics can be modelled with continuous constrains.

As it was stated previously, path planning is an optimization process then classical optimization techniques must be described to give a complete overview of the main tools developed to cope with this problem.

2.4 Generic optimization algorithms

Mathematical methods to solve optimization problems, known as indirect methods, are the most important and referenced techniques in this field. Algorithms based on Pontryagin minimum principle and Lagrange multipliers have been widely used to reduce optimization problems to a boundary condition one (Chitsaz & LaValle, 2007). Sequential Gradient Restoration Algorithm (SGRA) represents an indirect method used for several problems like space trajectories optimization (Miele & Pritchard, 1969, Miele, 1970). These techniques are elegant and reliable thanks to decades of research and application to thousands of different problems. They require a complex problem formulation and simplification in order to reach the required mathematical structure that ensures convergence. In some cases where complex and non-linear problems need to be solved these methods can result impracticable and other optimization techniques are needed (Sussmann & Tang, 1991).

Genetic algorithms are nowadays the most attractive solution in problems where constraints and optimization variables are the issue (Carroll,1996). They are based on the concept of natural selection, modelling the solutions like a population of individuals and evaluating evolution of this population over an environment represented by the problem itself. Using Splines or random threes to model the trajectory, these algorithms can reallocate the waypoint sequence to generate optimum solutions under constraints on complex environments (Nikolos et al., 2003). Being interesting and flexible, the evolutionary algorithms are spreading on different planning problems, but their solving complexity is paid with a heavy computational effort.

Finally, more advanced optimization techniques inspired to biological behaviours must be mentioned. These techniques recall biological behaviours to find the optimal solution to the problem. The key aspect of these solutions is the observation of biological phenomena and the adaptation to path planning problems. These algorithms permit to improve the system flexibility to changes in mission constraints and environmental conditions and with respect to genetic approaches these algorithms optimize the solution through a cooperative search.

3. The graph search algorithms for path planning

Graph search algorithms were developed for computer science to find the shortest path between two nodes of connected graphs. They were designed for computer networks to develop routing protocols and were applied to path planning through decomposition of the path in waypoint sequences. The optimization logics behind these algorithms attain the minimization of the distance covered by the vehicle, but none of its performances or kinematic characteristics is involved in the path search.

3.1 General overview

Basic elements common to each graph search method are (LaValle, 2006):

- a finite or countably infinite state space that collects all the possible states or nodes of the graph (X),
- an actions space that collects for each state the set of action that can be taken to move from a state to the next (U),
- a state transition function:

$$f: \ \forall \quad x \in X \quad and \quad u \in U \quad f(x,u) = x' \quad x' \in X \tag{1}$$

- an initial state $x_I \in X$,
- a goal state $x_G \in X$,

Classical graph search algorithms applied to path planning tasks then have other common elements:

- the state space is the set of cells obtained meshing the environment in discrete fractions,
- the action space is the set of cells reachable from a given cell,
- the transition function checks the neighbours of a given cell to determine whether motion is possible (i.e. for an eight connected mesh the transition function checks the eight neighbours of a given cell),
- the cost function evaluates the cost to move from a given cell to one of its neighbours.
- the initial state is the starting cell where the aircraft is supposed to be,
- the goal state is the goal cell where the aircraft is supposed to arrive.

Classical graph search algorithms treat each cell as a graph node and they search the shortest path with "greedy" logics. The algorithm applies the transition function to the current cell to move to the next one and it analyses systematically the state space from the starting cell trying to reach the goal one. Each analysed cell can be:

- Unexpanded: a cell that the algorithm has not been reached yet. When the algorithm reaches an unexpanded cell the cost to come to that cell is computed and the cell is stored in a list called *open list.*
- Expanded (a cell already reached):
    - Alive: a cell that the algorithm could reach from another neighbouring cell. A cell alive is yet in the open list. The algorithm computes the new cost to come and substitutes the new cost associated to the cell whether it is lower then the previous one.
    - Dead: a cell that the algorithm already reached and its cost to come cannot be reduced further. These cells are stored in a list called *closed list.*

For each cell together with its coordinates and the cost to come, the algorithm stores in the lists also the parent coordinates. The parent is the cell left to reach a current one (i.e. x_0 used in $f(x_0,u_0)$ is the x parent assuming that x is the current cell and $x' = f(x,u)$ is the x neighbour).

Main structure of any classical graph search algorithms is:

- Insert the starting cell in open list
- Searching cycle (this cycle breaks when the goal cell is reached or the open list is empty):
    - Check that the open list is not empty
        - True: go on
        - False: cycle break
    - Sort the open list with respect to the cost to come
    - Take the cell with the lower cost
    - Check that this cell is not the target one

        - True: go on
        - False: cycle break
    - Add this cell to the closed list
    - Cancel this cell from the open list
    - Cell expansion cycle (this cycle breaks when each new cell has been evaluated)
        - Use the transition function to find a new cell
        - Check inclusion of the new cell in the closed list
            - True: jump the state
            - False: go on
        - Check inclusion of the new cell in the open list
            - True:
                - Evaluate the cost to come
                - Check if the new cost is lower then the previous one:
                    - True: substitute the new cost and the cell parent
                    - False: jump the state
            - False:
                - Evaluate the cost to come
                - Add the cell to the open list
    - End of the new state evaluation cycle
- End of the searching cycle.

The algorithm expands systematically the cells up to reach the goal and the different solutions composing this family of algorithms differ each other because of the logics driving the expansion. However the algorithm breaks when the goal cell is reached without providing any guaranty of global optimality on the solution. More advanced algorithms include more complex cost functions driving the expansion in such a way to provide some guarantees of local optimality of the solution, but the “greedy” optimization logics characterizing these path planning techniques has in this one of its drawbacks.

From late 50s wide research activity was performed on graph-search algorithms within computer science, trying to support the design of computer networks. Soon after, the possibility of their application in robotics resulted evident and new solutions were developed to implement algorithms tailored for autonomous agents. As a consequence, research on graph-search methods brought new solutions and still continues nowadays. Therefore, an accurate analysis is required to understand advantages and drawbacks of each proposed approach, in order to find possible improvements.

3.2 From Dijkstra to A*

The Dijkstra algorithm (Dijkstra, 1959) is one of the first and most important algorithms for graph search and permits to find the minimum path between two nodes of a graph with positive arc costs (Chandler et al, 2000). The structure of this algorithm is the one reported in the previous section and it represents the basic code for all the successive developments. An evolution of the Dijkstra algorithm is the Bellman-Ford (Bellman, 1958) algorithm; this method finds the minimum path on oriented graphs with positive, but also negative costs (Papaefthymiou & Rodriguez, 1991). Another method arose by the previous two is the Floyd-Warshall algorithm (Floyd, 1962, Warshall, 1962), that finds the shortest path on a weighted graph with positive and negative weights, but it reduces the number of evaluated nodes compared with Dijkstra.

The A* algorithm is one of the most important solvers developed between 50s and 70s, explicitly oriented to motion-robotics (Hart et al., 1968). A* improved the logic of graph search adding a heuristic component to the cost function. Together with the evaluation of the cost to come (i.e. the distance between the current node and a neighbour), it also considers the cost to go (i.e. an heuristic evaluation of the distance between a neighbour and the goal cell). Indeed the cost function (F) exploited by the A* algorithm is obtained summing up two terms:

- The cost to go H: a heuristic estimation of the distance from the neighbouring cell x' to the goal x_G.
- The cost to come G: the distance between the expanded cell x and the neighbouring one x'.

The G-value is 0 for the starting cell and it increases while the algorithm expands successive cells. The H-value is used to drive the cells expansion toward the goal, reducing this way the amount of expanded cells and improving the convergence. Because in many cases is hard to determine the exact cost to go for a given cell, the H-function is an heuristic evaluation of this cost that has to be monotone or consistent. In other words, at each step the H-value of a cell has not to overestimate the cost to go and H has to vary along the path in such a way that:

$$H(x', x_G) \leq H(x, x_G) + G(x, x') \tag{2}$$

3.3 Dynamic graph search

The graph-search algorithms developed between 60s and 80s were widely used in many fields, from robotics to video games, assuming fixed and known positions of the obstacles on the map. This is a logic assumption for many planning problems, but represents a limit when robots move in unknown environments. This problem excited research on algorithms able to face with map modifications during the path execution. Particularly, results on sensing robots, able to detect obstacles along the path, induced research on algorithms used to re-plan the trajectory with a more effective strategy than static solvers were able to implement.

Dynamic re-planning with graph search algorithms was introduced. D* (Dynamic A*) was published in 1993 (Stentz, 1993) and it represents the evolution of A* for re-planning . When changes occur on the obstacle distribution some of the cell costs to come changes. Dynamic algorithms update the cost for these cells and replan only the portion of path around them keeping the remaining path unchanged. This way D* expands less cells than A* because it has not to re-plan the whole path through the end. D* focused was the evolution of D*, published by the same authors and developed to improve its characteristics (Stentz, 1995). This algorithm improved the expansion, reducing the amount of analysed nodes and the computational time.

Then, research on dynamic re-planning brought to the development of Lifelong Planning A* (LPA*) and D* Lite (Koenig & Likhachev, 2001, 2002). They are based on the same principles of D* and D* focused, but they recall the heuristic cost component of A* to drive the cell expansion process. They are very similar and can be described together. LPA* and D* Lite exploit an incremental search method to update modified nodes, recalculating only the start distances (i.e. distance from the start cell) that have changed or have not been calculated before. These algorithms exploit the change of *consistency* of the path to replan. When obstacles move, graph cells are updated and their cost to come changes. The algorithm records the cell cost to come before modifications and compares the new cost with the old one to verify consistency. The change in consistency of the path drives the algorithm search.

3.4 Any heading graph search

Dynamic algorithms allowed new applications of graph search methods to path planning of robotic systems. More recently, other drawbacks and possible improvements were discovered. Particularly, one of the most important drawbacks of A* and the entire dynamic algorithms resides on the heading constraints connected with the graph structure. The graph obtained from a surface map is a mesh of eight-connected cells with undirected edges. Moving from a given cell to the next means to move along the graph edge. The edges of these graphs are the straight lines connecting the centre of the current cell with the one of the neighbour. As a matter of fact the edges between cells of an eight connected graph can have slope *a* such that:

$$a = n \cdot \frac{\pi}{4} \quad 0 \leq n \leq 8 \quad n \in N \tag{3}$$

Then the paths obtained with A* and its successors is made of steps with heading defined in equation [3]. This limit is demonstrated prevents these algorithms to find the real shortest path between goal and start cells in many cases (it is easy to imagine a straight line connecting the start with goal cell having heading different from the ones of equation [3]). A* and dynamic algorithms generate strongly suboptimal solutions because of this limit, that comes out in any application to path planning. Suboptimal solutions are paths with continuous heading changes and useless vehicle steering (increasing control losses) that require some kind of post processing to become feasible. Different approaches were developed to cope with this problem, based on post-processing algorithms or on improvements of the graph-search algorithm itself. Very important examples are Field D* and Theta*. These algorithms refined the graph search obtaining generalized paths with almost “any” heading.

To exploit Field D*, the map must be meshed with cells of given geometry and the algorithm propagates information along the edges of the cells (Ferguson & Stentz, 2006). Field D* evaluates neighbours of the current cell like D*, but it also considers any path from the cell to any point along the perimeter of the neighbouring cell. A functional defines the point on the perimeter characterising the shortest path. With this method a wider range of headings can be achieved and shorter paths are obtained.

Theta* represents the cutting edge algorithm on graph search, solving with a simple and effective method the heading constraint issue (Nash et al., 2007). It evaluates the distance from the parent to one of the neighbours for the current cell so that the shortest path is obtained. When the algorithm expands a cell, it evaluates two types of paths: from the current cell to the neighbour (like in A*) and from the current-cell parent to the neighbour. As a conclusion, paths obtained by the Theta* solver are smoother and shorter than those generated by A*.

Apparently, Theta* is the most promising solution for path planning. As a matter of fact, some other graph search algorithms were not considered here, as this chapter would provide a general overview on the main concepts converging in development of these path-planning methods. By the way all the algorithms described have the common drawback of missing any kind of vehicle kinematic constraints in the path generation. The algorithm presented in the following chapter (Kinematic A*) has been developed to bridge this gap and open investigations in this direction.

3.5 Tridimensional path planning with A* and Theta*

The application of A* and Theta* to 3D path planning for mini and micro UAVs was extensively investigated (De Filippis et al., 2010, 2011). The A*-basic algorithm was improved and applied to tri-dimensional path planning on highlands and urban environments. Then this algorithm has been compared with Theta* for the same applications in order to investigate merits and drawbacks of these solutions.

Here is reported the comparison between a path planned with A* with the same one planned with Theta* in order to show the improvements introduced adopting the last algorithm. **Figure 2** is the tri-dimensional view of the two paths implemented for this example.

Map characteristics:

- Cells number: 9990000.
- Δlat: 1 m.
- Δlong: 1 m.
- ΔZ: 1 m.
- Environment matrix dimensions: 300 x 300 x 111 (lat x long x Z).

Path 1		
	A*	Theta*
Path length	386.5 m	372.4 m
Computation time	3.1 s	3.6 s
Number of heading changes	327	6
Number of altitude changes	0	0
Number of waypoints	327	6

Table 1. Example parameters.

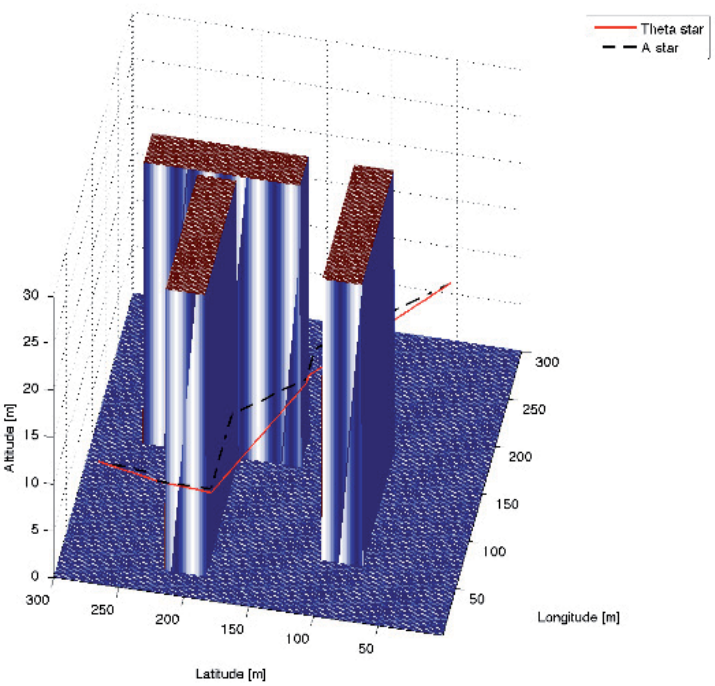


Fig. 2. Comparison between Theta* and A* (3D view).

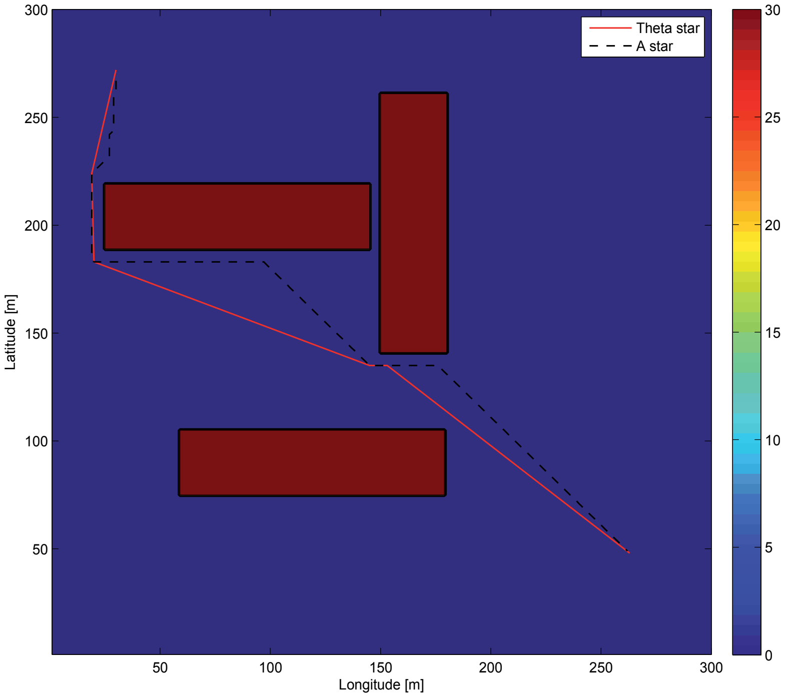


Fig. 3. Comparison between Theta* and A* (Longitude-Latitude plane).

Table 1 collects the map parameters and the algorithm performances while **Figure 3** is the longitude-latitude path view. The paths are planned without altitude changes so the last picture is sufficient to depict differences between them. The path obtained with Theta* is slightly shorter then the A* one, but huge difference is in the number of waypoints composing the path and in the amount of heading changes. These parameters testify the previous statements identifying in Theta* an interesting algorithm among the classical graph search solutions here mentioned.

4. Kinematic A*

The main drawback of applying classical graph search algorithms to path planning problems resides in the lack of correlation between the path and the vehicle kinematic constraints. In this section, a new path-planning algorithm (Kinematic A*) is presented, implementing the graph search logics to generate feasible paths and introducing basic vehicle characteristics to drive the search.

Kinematic A* (KA*) includes a simple kinematic model of the vehicle to evaluate the moving cost between the waypoints of the path in a tridimensional environment. Movements are constrained with the minimum turning radius and the maximum rate of climb. Furthermore, separation from obstacles is imposed, defining a volume along the path free from obstacles (tube-type boundaries), as inside these limits the navigation of the vehicle is assumed to be safe.

The main structure of the algorithm will be presented in this section, together with the most important subroutines composing the path planner.

4.1 From cells to state variables

Classical graph search algorithms solve a discrete optimization problem linking the cost function evaluation to the distance between cells. These cells discretize the motion space representing the discrete state of the system. The states space is finite and discrete containing the positions of the cell centres. The optimization problem requires finding the sequence of states minimizing the total covered distance between the starting and the target cell.

Kinematic A* introduces a vehicle model to generate the states and evaluate the cost function. Each state is made of the model variables and is discrete because the command space is made of discrete variables. So the optimization problem is transformed in finding the discrete sequence of optimal commands generating the minimum path between the starting and the target state.

In the following sections then the concept of cells or nodes of the graph, representing the discrete set of states defining the optimization problem is substituted with the concept of states of the vehicle model and the optimization problem is reformulated.

4.2 The kinematic model

In the following description S is the state of the aircraft at the current position. S is the vector of the model state variables. This simple model is used to generate the possible movements from a given state to the next, i.e. the evolution of S from the current condition to the next.

The model is a set of four differential equations describing the aircraft motion in Ground reference frame (G frame). This is not the typical Nort-East-Down (NED) frame used to write navigation equations in aeronautics. The Ground frame is typical of ground robotic applications that inspired this work. The G-frame origin is placed in the aircraft center of mass. The X and Y axes are aligned with the longitude and latitude directions respectively. Then the Z axis points up completing the frame.

In the G frame distances are measured in meters and two control angles (χ and γ) act as gains on rate of turn and rate of climb along the path:

- χ is the angle between the X axis and the projection of the speed vector (V) on the X-Y plane, the variation of this angle is connected with the rate of turn.
- γ is the angle between the speed vector and its projection on the X-Z plane (see Figure 4), this angle controls the rate of climb.

The model is obtained considering the aircraft flying at constant speed and the Body frame (B frame) aligned with the Wind frame (W frame). The rate of turn is assumed bounded with the minimum turn radius and the rate of climb with the maximum climb angle.

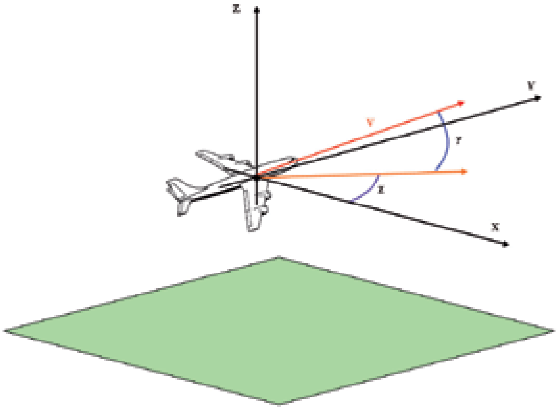


Fig. 4. The Ground Reference frame (G frame).

The speed vector is constant and aligned with the X_B axis. Using the Euler transformation matrix from the body to the ground frame the speed components in G frame are obtained. Combining these differential equations with the turning-rate the aircraft model becomes:

$$\begin{cases} \dot{X} = V\cos(\chi)\cos(\gamma_{\max}\cdot w) \\ \dot{Y} = V\sin(\chi)\cos(\gamma_{\max}\cdot w) \\ \dot{Z} = V\sin(\gamma_{\max}\cdot w) \\ \dot{\chi} = \dfrac{V}{R_{\min}}\cdot u \end{cases} \qquad \begin{matrix} |u| \le 1 \\ |w| \le 1 \end{matrix} \tag{4}$$

where:

X,Y,Z = aircraft positions vector P on the ground frame [m].
V = aircraft speed [m/s].

R_{min} =maximum turning radius [m].
χ = turning angle.
γ_{max} = maximum climbing angle.
u,w = command parameters.

To generate the set of possible movements discrete command values (u_i and w_i) are chosen and the system of equations [4] is integrated in time with the initial conditions given by the current state S:

$$\begin{cases} X(0) = X_s & \\ Y(0) = Y_s & u_i = [\ -1 \quad -0.5 \quad 0 \quad 0.5 \quad 1] \\ Z(0) = Z_s & w_i = [\ -1 \quad -0.5 \quad 0 \quad 0.5 \quad 1] \\ \chi(0) = \chi_s & \end{cases} \tag{5}$$

If the command values are constant along the integration time (Δt), the equations in [4] become:

$$\begin{cases} X_i = X_s + \left(\dfrac{R_{\min}}{u_i}\right) \cdot \cos(\gamma_{\max} \cdot w_i) \cdot \left[\sin\left(\chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t \right) - \sin(\chi_s) \right] \\ Y_i = Y_s - \left(\dfrac{R_{\min}}{u_i}\right) \cdot \cos(\gamma_{\max} \cdot w_i) \cdot \left[\cos\left(\chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t \right) - \cos(\chi_s) \right] \\ Z_i = Z_s + V \cdot \sin(\gamma_{\max} \cdot w_i) \cdot \Delta t \\ \chi_i = \chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t \end{cases} \tag{6}$$

providing the evolution of S for each controls space. On Figure 5 25 trajectories are represented. They are obtained combining the two vectors *u* and *w* presented in [5] and substituting each command couple (5 u_i values x 5 w_i values) in [6]. For each couple the system of equation is integrated over the time step with initial conditions and parameters equal to:

P_s = [0 0 0 0],
V = 25 [m/s],
R_{min} = 120 [m],
γ_{max} = 4 [deg],
Δt = 8 [s].

Once Δt, aircraft speed, minimum turning radius and maximum climbing angle are chosen according with the aircraft kinematic constraints the equations in [6] can be solved at each cycle for the current state and the algorithm can generate the set of possible movements looking for the optimal path.

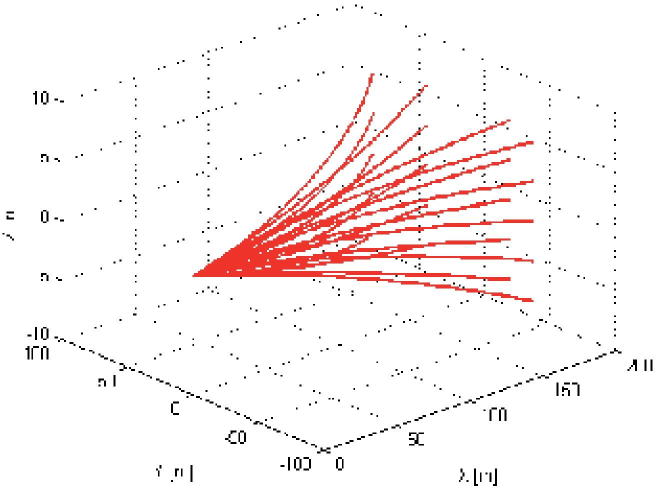


Fig. 5. Sequences of states for a time horizon of 8 seconds.

4.3 The kinematic model with wind

The kinematic model can be improved taking into account the wind effect on the states evolution. Summing to the aircraft speed the wind components in G frame and assuming these components constant on Δt the system of equations [6] become:

$$\begin{cases} X_i = X_s + \left(\dfrac{R_{\min}}{u_i}\right) \cdot \cos(\gamma_{\max} \cdot w_i) \cdot \left[\sin\left(\chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t\right) - \sin(\chi_s)\right] + W_x \cdot \Delta t \\ Y_i = Y_s - \left(\dfrac{R_{\min}}{u_i}\right) \cdot \cos(\gamma_{\max} \cdot w_i) \cdot \left[\cos\left(\chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t\right) - \cos(\chi_s)\right] + W_y \cdot \Delta t \\ Z_i = Z_s + V \cdot \sin(\gamma_{\max} \cdot w_i) \cdot \Delta t + W_z \cdot \Delta t \\ \chi_i = \chi_s + \dfrac{V}{R_{\min}} \cdot u_i \cdot \Delta t \end{cases} \tag{7}$$

where:

$[W_x\ W_y\ W_z]$ = wind speed components in G frame [m/s].

In Figure 6 and Figure 7 a state evolution with wind is compared with the same without wind. The state and parameters used as an example are:

P_s = [0 0 0 0],
V = 25 [m/s],
R_{min} = 120 [m],
γ_{min} = 4 [deg],
Δt = 8 [s].

$W_x = 5$ [m/s]
$W_y = 5$ [m/s]
$W_z = 0$ [m/s]

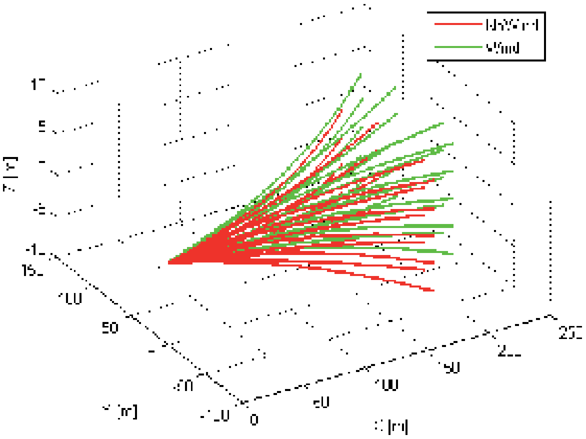


Fig. 6. Comparison of states evolution (3D view) with and without wind (W_x=5 m/s, W_y=5 m/s, W_z=0).

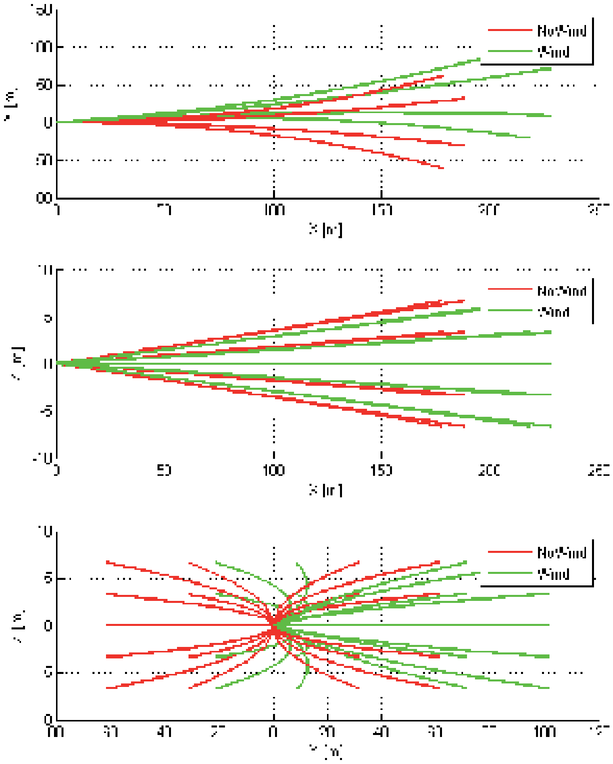


Fig. 7. Comparison of states evolution (2D views) with and without wind (W_x=5 m/s, W_y=5 m/s, W_z=0).

These pictures show how wind affects the evolution of a given state and in turn the effect on the set of possible movements from the current state to the next.

4.4 The problem formulation

The functional J minimized through the optimization process is made up of the costs F_{ij} of each state composing the path. The minimum of J is found summing up the smaller cost F_{ij} of each state. F_{ij} is made of two terms related respectively with the states and the commands. At each step the algorithm generates the set of movements from the current state (shown in Figure 5). Then it evaluates F_{ij} for each new state and chooses the one with the smaller value. The global optimization problem is finding:

$$\min(J) = \sum_{S_0}^{S_t} \min(F_{ij}) = \sum_{S_0}^{S_t} \min(\overline{H}_{ji}^T \cdot \overline{\alpha} \cdot \overline{H}_{ij} + \overline{G}_{ij}^T \cdot \overline{\beta} \cdot \overline{G}_{ij}) \tag{8}$$

The H and G vectors take into account respectively the error on the states and the amount of command due to reach a new state. The matrices α and β are diagonal matrices of gains on the states and on the commands:

$$\begin{gathered} \overline{H}_i = \begin{bmatrix} X_t - X_i \\ Y_t - Y_i \\ Z_t - Z_i \end{bmatrix} = \begin{bmatrix} \Delta X_i \\ \Delta Y_i \\ \Delta Z_i \end{bmatrix} \\ \overline{G}_i = \begin{bmatrix} u_i \\ w_i \end{bmatrix} \end{gathered} \tag{9}$$

$$\begin{gathered} \overline{\alpha} = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{bmatrix} \\ \overline{\beta} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \end{gathered} \tag{10}$$

The H vector is the distance between the new state $[X_i\ Y_i\ Z_i]'$ and the target one $[X_t\ Y_t\ Z_t]'$. On the other hand the G vector evaluates the amount of command needed to reach this new state from the current one. Then choosing the smaller value of F the algorithm selects a new state that reduces the distance from the target minimizing the commands. The gain matrices are used to weight the state variables and the commands in order to tune their importance in F.

To complete the problem formulation, the states in J must be included in the state space respecting the differential equations given in [4] and the initial conditions given in [5]. Then the commands must be chosen in the command space given in [5] in order to minimize the functional J. The state space is constrained by the map limits, the obstacles and the separation requirements and will be described in the following section.

4.5 The state space

If the command space of the problem solved with KA* is bounded by the kinematic constraints and is discretized according with the optimization requirements, the states are bounded only on the X and Y sets (longitude and latitude coordinates) because of constraints on the Z state and on the X and Y states themselves:

- The map bounds: these bounds affect the X-Y sets because points outside the map limits can not be accepted as new states.
- The ground obstacles: these constraints bound the X-Y sets if the Z component of the new state is lower then the ground altitude at the same X-Y coordinates, so the relative new state must be rejected.
- The separation constraints: the new state not only has to have a Z component higher then the ground one, but has also to respect the horizontal and vertical separations from the obstacles. The X-Y sets are bounded because states too close to the obstacles must be rejected.

Figure 8 shows the horizontal (HZ_1 and HZ_2) and vertical (VZ_1) separation constraints imposed on the path from the current state S to the next state I. These constraints guarantee the flight safety along the path because possible tracking errors of the guidance system are acceptable and safe inside the boundaries imposed by the separation constraints.

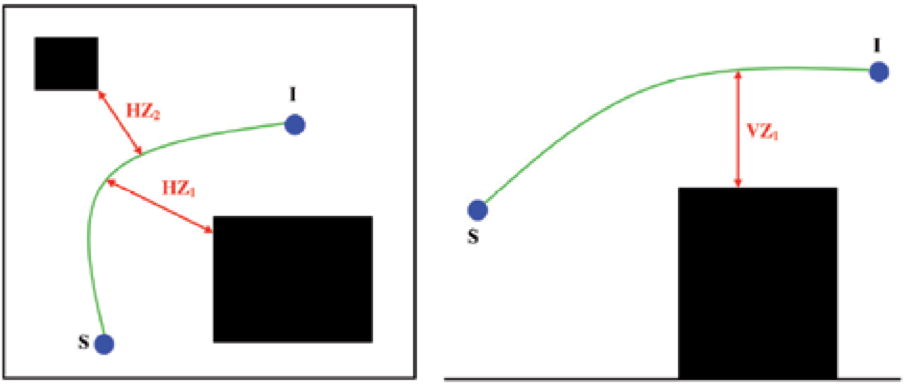


Fig. 8. Horizontal and Vertical separation from an obstacle.

4.6 Algorithm description

- Initialize the control variables and parameters.
- Evaluate the F cost for the initial state
- Add this state to the open list.
- Searching cycle (this cycle breaks when the target state is reached or the open list is empty):
  - Check that the open list is not empty
    - True: go on
    - False: cycle break

- Open list search for the state with the lower F value
- Check that this state is not the target one
  - True: go on
  - False: cycle break
- Add this state to the closed list
- Cancel this state from the open list
- New states evaluation cycle (this cycle breaks when each new state has been evaluated)
  - New states generation through the model equations
  - Check inclusion of the new state inside the state space
    - True: go on
    - False: jump the state
  - Check inclusion of this state in the closed list
    - True: jump the state
    - False: go on
  - Check inclusion of this state in the open list
    - True:
      - Evaluate the F cost for the state
      - Check if the new cost is lower then the previous one:
        - True: substitute this new cost and the new current state to the open list
        - False: jump the state
    - False:
      - Evaluate the F cost for the state
      - Add the state to the open list
  - End of the new state evaluation cycle
- End of the searching cycle.

5. Results

The following chapter collects two tests with the obstacles placed on the map in such a way to force KA* toward its limits. The new algorithm is compared with A* in order to show the improvements introduced. The paths are generated with and without wind to show its effects on the path and to compare the results. These tests then give the opportunity to show limitations of this new technique in order to stimulate future developments.

The first path is on a map with four obstacles symmetrically placed. They are close to the four corners of a square area and the aircraft is forced to slalom between them. The starting and target points are placed respectively at the bottom-left and top-right corner with different altitudes in order to force the algorithm to plan a descent meeting the four obstacles along the flight. Finding the path for this test is easier then finding it for the next one. The algorithm is able to follow the minimum of the cost function without analyzing too many states and it converges rapidly.

The map of the second path has just one wide obstacle placed in the middle. The obstacle is placed slightly closer to the right border of the map in order to obstruct the path to the aircraft that is supposed to move from the bottom-left to the top-right corner. The path search for this test is harder then the previous one because the algorithm has to analyze many states to find the optimum path. Following a monotonic decrease of the cost function along the path search is impossible for this case. The obstacle in the middle forces the aircraft far from the target point. The aircraft has to go around the obstacle to reach the target; this induces a cost increase to move from a state to the next that makes the optimization harder.

The component of wind introduced to implement the following tests is considered constant in time and space on the whole map. This approach clearly does not mean to solve the problems due to wind disturbances in the path optimization. This is a complex and hard problem due to wind model complexity, effects of the wind on the aircraft performances and dynamics, turbulent components effects, etc. Face properly this problem requires specific studies and techniques, but it is useful to introduce this simple study at this level in order to show potential developments of this path planning technique for future applications.

Then the aircraft parameters chosen to implement the tests must be motivated. The small area of the map, induced to chose accordingly the aircraft parameters needed for the model. The reference vehicle is a mini UAV with reduced cruise speed, turning radius and climbing performances but agile enough to perform the required paths. Particularly speed is chosen so that the trajectories needed to avoid the obstacles would be feasible and the turning radius is calculated considering coordinated turns:

$$R_{\min} = \sqrt{\frac{V^4}{g^2 \cdot \left(\left(\frac{1}{\cos(\varphi_{\max})} \right)^2 - 1 \right)}} \tag{11}$$

where:

R_{min} = minimum turning radius.
V = aircraft cruise speed.
g = gravitational acceleration.
φ_{max} = maximum bank angle.

Finally for each test a table collecting all the data is reported. All the reported paths are obtained with the MATLAB version 7.11.0 (R2010b), running on MacBook Pro with Intel Core 2 Duo (2 X 2.53 GHz), 4 Gb RAM and MAC OS X 10.5.8. The table contains:

- map dimensions,
- obstacles dimensions,
- obstacles center position,
- starting point
- target point,
- aircraft parameters,
- optimization parameters,
- obstacles separation parameters,
- wind speed,
- computation time,

- path length,
- number of waypoints.

5.1 Four obstacles

Figure 9 shows the obstacles position on the map and the three paths (KA* with wind, KA* without wind, A*) in tridimensional view.

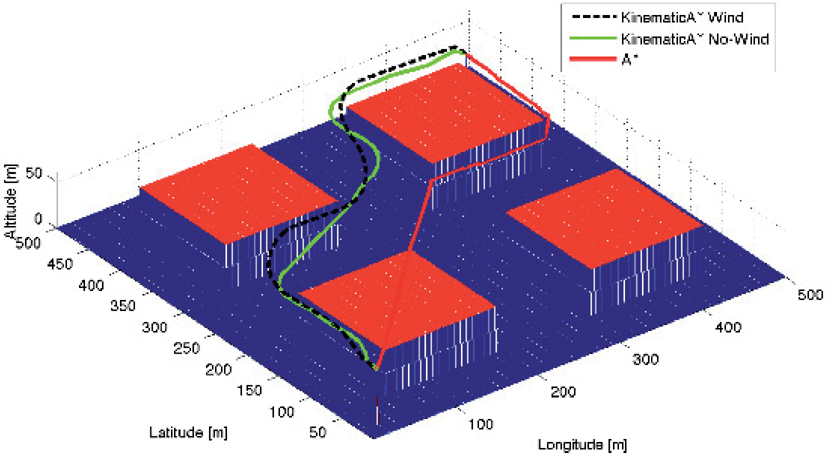


Fig. 9. Four obstacles test (3D view).

Table 2 collects the numerical data used to implement this test. The time step between two states is set to 2 seconds in order to have a sufficiently discretized path without increasing too much the computation time. Horizontal and vertical obstacles separations are set to 15 m and 10 m respectively. This should guarantee sufficient safety without limiting the aircraft agility between the obstacles.

The constant wind along the *Y* ground direction has 5 m/s intensity. This value is sufficiently high and it affects deeply the path as the computation time, the number of waypoints and the path shape testify. The wind pushes the aircraft toward the target, reducing the computation time with respect to the case without wind.

In order to compare the KA* performances with the A* one, it can be noticed that the computation time for the two algorithms is almost the same for the path with wind, but it is increased without wind. This is due to the reduced speed of the aircraft without wind and to the higher number of possible movements from one state to the next. With wind the feasible movements between states are strongly reduced because of the wind disturbance. Flying at lower speed the algorithm is forced to analyze much more states and for each position much more possible movements are feasible. Then the optimization process takes more time. Finally shall be noticed how KA* generates a path with a really small number of waypoints with respect to A*. This permits to obtain more handy waypoints lists without need of post processing, ready to be loaded on the flight control system.

Map dimension			
	X	500	[m]
	Y	500	[m]
	Z	80	[m]
	ΔXY	1	[m]
	ΔXZ	1	[m]
Obstacles dimension			
	X	250	[m]
	Y	125	[m]
	Z	50	[m]
Obstacles-center position (1)			
	X	125	[m]
	Y	125	[m]
Obstacles-center position (2)			
	X	125	[m]
	Y	375	[m]
Obstacles-center position (3)			
	X	375	[m]
	Y	375	[m]
Obstacles-center position (4)			
	X	375	[m]
	Y	125	[m]
Starting point			
	X	20	[m]
	Y	20	[m]
	Z	60	[m]
Target point			
	X	480	[m]
	Y	480	[m]
	Z	30	[m]
Aircraft parameters			
	Speed	10	[m/s]
	Min turning radius	25	[m]
	Max climbing angle	4	[deg]

Optimization parameters			
	Time step	2	[s]
	α	10	
	β	1	
Obstacles separation			
	Horizontal	15	[m]
	Vertical	10	[m]
Wind Speed			
	X	0	[m/s]
	Y	5	[m/s]
	Z	0	[m/s]
Computation time			
	KA* with Wind	2.1827	[s]
	KA* without Wind	9.9657	[s]
	A*	2.3674	[s]
Path length			
	KA* with Wind	772	[m]
	KA* without Wind	769	[m]
	A*	740	[m]
WayPoints			
	KA* with Wind	34	
	KA* without Wind	40	
	A*	591	

Table 2. Four-obstacles test parameters.

In **Figure 10** the paths on the Longitude-Latitude plane are presented. The path obtained with A* pass over the bottom-left obstacle and very close to the top-right one. Planning sharp heading changes to reach the target. This is typical of classical graph search algorithms that do not take into account the vehicle kinematic constraints. The path obtained with KA* on the other hand is smooth and obstacles separation constraints is evident. Comparing the path with wind with the one without wind between the obstacles on the left is evident the disturbance induced by the wind that pushes the path closer to the top-left obstacle.

In **Figure 11** on the *X*-axis is plotted the distance covered from start to target point and on the *Y*-axis the aircraft altitude. Again A* plans sharp altitude changes and particularly sharp descends to reach the target. These changes are unfeasible with real aircrafts. As a matter of fact in general the A* path requires deep post processing and waypoints reallocation to make the path flyable. Analyzing in detail though the algorithm plans a path passing over the bottom-left obstacle and then descending close to the top-right one. Being this descent unfeasible for the aircraft a complete re-planning is needed to reallocate the waypoints sequence. This is one of many cases evidencing that classical graph search algorithms used for tridimensional path planning can generate unfeasible paths because of the strong longitudinal constraints of aircrafts and they need high intrusive post processing algorithms to modify the waypoint sequence.

In **Figure 12** the time history of the turning rate (connected with the *u* command) is plotted. The comparison between the command sequence with and without wind puts in evidence that the path needs more aggressive commands to compensate disturbances introduced from wind, but the average value remains limited thanks to the *G* value in the cost function that takes care of the amount of command needed to perform the path. On the other hand in **Figure 13** the climbing angle (due to the *w* command) is plotted. Also in this case the average amount of command is limited. Limiting turning rates and climbing angles required to follow the path is important. The main path-planning task is to generate a trajectory driving the aircraft from start to target in safe conditions. If tracking the path planned requires aggressive maneuvers, the aircraft performances will be completely absorbed by this task. However in many cases tracking the path is just a low-level task prerogative to accomplish with the high-level mission task (i.e in a save and rescue mission tracking the path could be one of the tasks together with many others. As an example it could be required also to avoid collision with dynamic obstacles along the flight, to deploy the payload and collect data, to interact with other aircrafts involved in the mission). If the aircraft must exploit its best performances to track the path it will not be able to accomplish also with the other mission tasks and this is not acceptable.

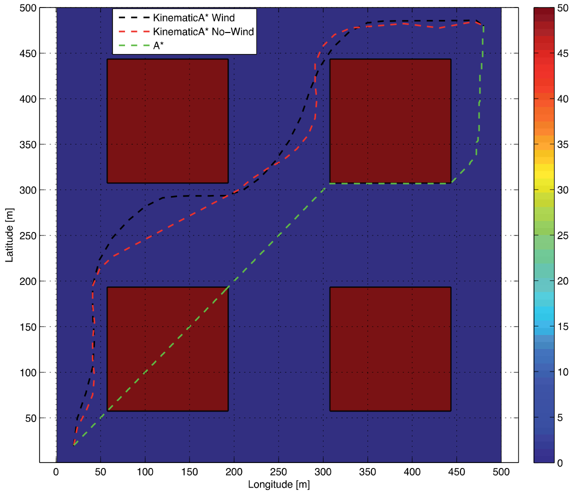


Fig. 10. Longitude-Latitude view (four obstacles test).

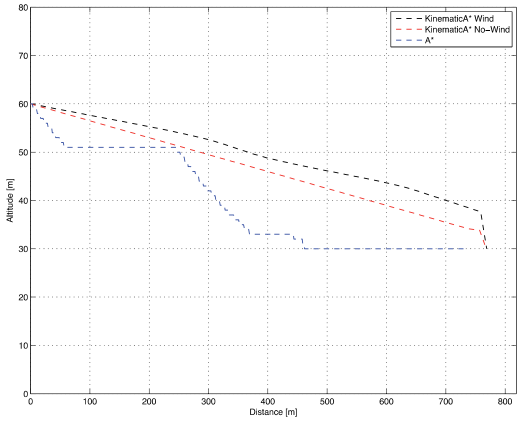


Fig. 11. Distance-Altitude view (four obstacles test).

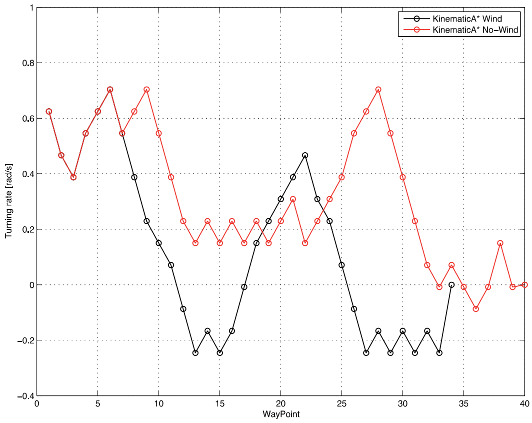


Fig. 12. Turning rate (four obstacles test).

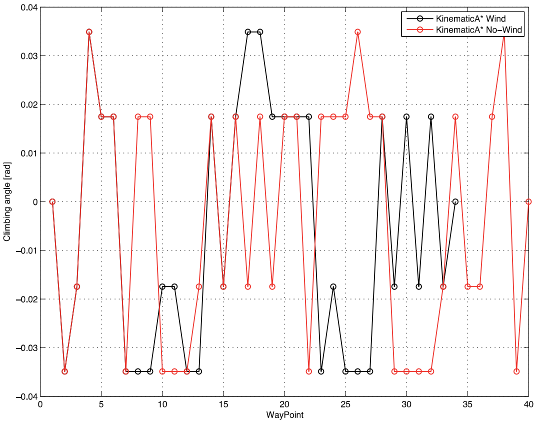


Fig. 13. Climbing angle (four obstacles test).

5.2 One obstacle

Figure 14 is the tridimensional view of the three paths (KA* with wind, KA* without wind, A*) generated with this test. The picture shows that the obstacle obstructs almost completely the path to the aircraft on the right, leaving just a small aisle to reach the target.

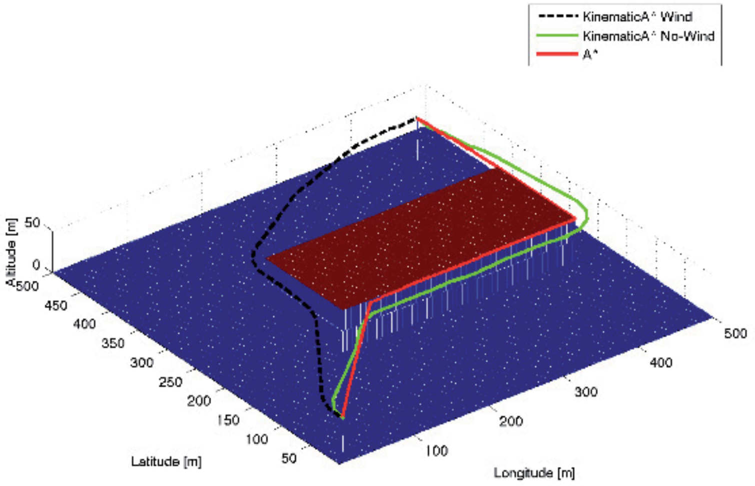


Fig. 14. One obstacle test (3D view).

All the data collected in **Table 3** are almost the same of the previous test. The environment, the aircraft, starting and target point, obstacles separation and optimization parameters do not change. Just the number and distribution of the obstacles is changed, together with the wind speed. The last parameter is changed to investigate the effects of diagonal wind on the path. The wind intensity is reduced to avoid reaching conditions too harsh for the flight. Two [m/s] of wind along the *X* and *Y* ground axes are introduced and the effects on the path are evidenced by the search performances.

The computation time between the path with and without wind is strongly different. As in the previous case wind forces the aircraft to move faster with respect to the ground, but the big difference between the paths now is due to the different path followed to reach the target. As shown in **Figure 14** the path with wind goes to the left of the obstacle and reaches the target directly. This is due to the negative wind speed along the *X*-axis that opposes the tendency of the aircraft to go straight from start to target (as the first part of the path without wind shows). The aircraft is pushed to the left forcing it to find a different path to reach the goal. In this way the computation time is strongly reduced because crossed the obstacle the aircraft can go straight to the target.

On the other hand KA* with wind and A* look for a way to reach the goal crossing the obstacle to the right. This is due to the *H* component in the cost function that drives the

Map dimension			
	X	500	[m]
	Y	500	[m]
	Z	80	[m]
	ΔXY	1	[m]
	ΔXZ	1	[m]
Obstacles dimension			
	X	300	[m]
	Y	125	[m]
	Z	50	[m]
Obstacles-center position			
	X	300	[m]
	Y	250	[m]
Starting point			
	X	20	[m]
	Y	20	[m]
	Z	40	[m]
Target point			
	X	450	[m]
	Y	450	[m]
	Z	50	[m]
Aircraft parameters			
	Speed	10	[m/s]
	Min turning radius	25	[m]
	Max climbing angle	4	[deg]
Optimization parameters			
	Time step	2	[s]
	α	10	
	β	1	
Obstacles separation			
	Horizontal	15	[m]
	Vertical	10	[m]
Wind			
	X	-2	[m/s]
	Y	2	[m/s]

	Z	0	[m/s]
Computation time			
	KA* with Wind	0.4854	[s]
	KA* without Wind	32.476	[s]
	A*	6.5682	[s]
Path length			
	KA* with Wind	689	[m]
	KA* without Wind	796	[m]
	A*	776	[m]
WayPoints			
	KA* with Wind	37	
	KA* without Wind	41	
	A*	698	

Table 3. One-obstacle test parameters.

search along the diagonal between the start and the target point. In this way the algorithms look for the optimal path following the diagonal up to meeting the obstacle. Then the search continues choosing to turn on the right because in that direction the *F*-value is decreasing. Because of this process the computation time is higher and also the covered distance is more then the one with wind.

Analyzing this behavior an important limit of this optimization technique comes out: greedy algorithms become slow when the optimum search does not provide a continuing monotonic decrease of the cost function. Because of this tendency this first version of KA* must be improved in order to accelerate the convergence to the optimal solution in cases where the continuing descent to the minim is not guaranteed.

Figure 15 shows on the Longitude-Latitude plane the different paths planned in this test. In this case, the post processing phase for the A* path would be less intrusive because of the slight altitude variation and of the few heading changes, but the 90 degrees heading change on the right of the obstacle is clearly unfeasible. This is evident comparing the turning radius planned by KA* with the sharp angle planned with A*. Some of these sharp heading changes can be easily corrected with a smoothing algorithm in post processing, but some of them can require a complete waypoints reallocation (as shown in the previous example). The important advantage of KA* is to generate a feasible path respecting the basic aircraft kinematic constraints with low computation workload.

Figure 16 again has on the *X*-axis the distance covered from start to target and on the *Y*-axis the aircraft altitude. For this test the altitude changes are smother then the one in the previous test, but here also it is possible to see the stepwise approach to climb of the A* path compared with the smooth climbing maneuver planned with KA*. About altitude variations the relation with the integration time step must be mentioned. Because of the slower behavior of an aircraft to altitude variations with respect to heading changes, when KA* is used on environments requiring strong altitude variations to avoid the obstacles, the integration time must be increased. The integration time provides to the algorithm the capability to forecast the possible movements from the current state to the next. Then, if the aircraft has to climb to avoid an obstacle flying above it, a longer integration horizon is needed in order to plan in time the climbing maneuver reducing the computation time.

In **Figure 17** the turn rate (related to the *u* command) is plotted. In this case the two command sequences cannot be compared because of the different paths followed. Anyway it is possible to see the strong turning rate imposed to the aircraft reaching the bottom-right corner of the obstacle. Reaching that corner the aircraft has to turn in order to go toward the target respecting the obstacle separation constrains, than strong heading changes are needed. In order to limit turning radiuses and climbing angles along the path and generate smooth and flyable trajectories for the aircraft the *u* and *w* command vectors provided to the model to generate the possible movements are limited to half of the maximum turning rate and climbing angle. Finally **Figure 18** shows the climb angle (related to the *w* command) time history as for the previous test. Here the climb angle is always small for both the paths and the aircraft climbs slowly to the target altitude.

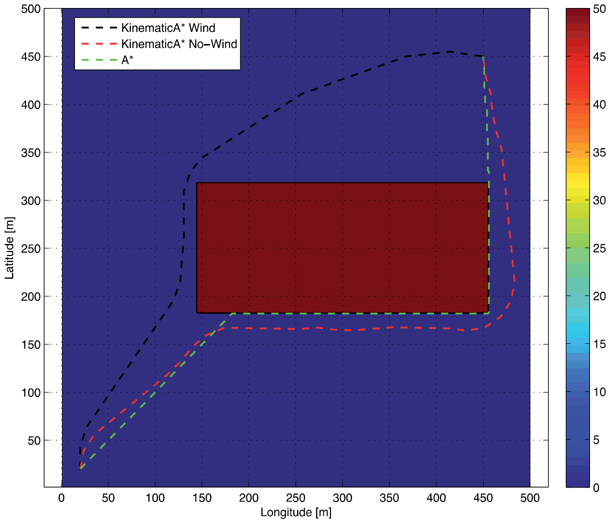


Fig. 15. Longitude-Latitude view (one obstacle test).

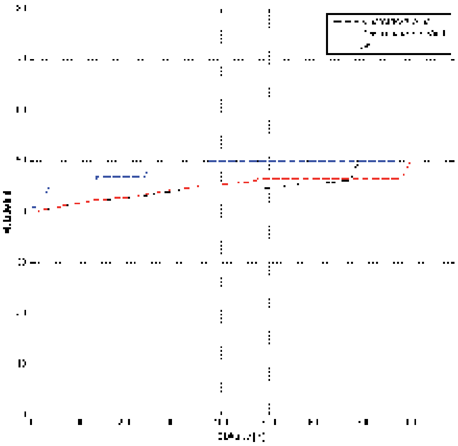


Fig. 16. Distance-Altitude view (one obstacle test).

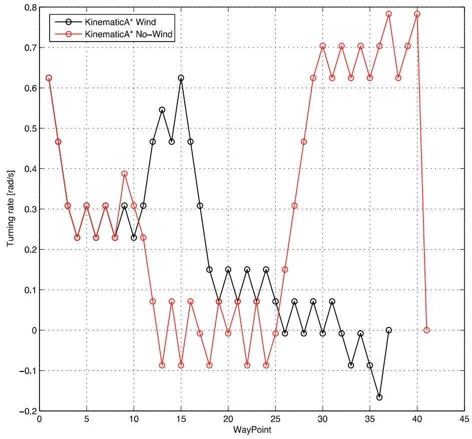


Fig. 17. Turn rate (one obstacle test).

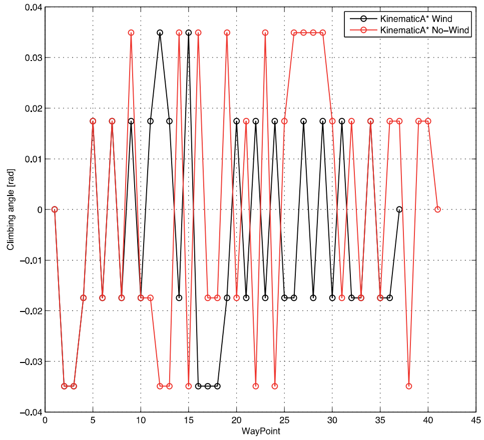


Fig. 18. Climb angle (one obstacle test).

6. Conclusions and future work

Kinematic A* has been developed to fill the relation gap between the aircraft kinematic constraints and the logics used in graph search approaches to find the optimal path. The simple aircraft model generates the state transitions in the state space and drives the search toward feasible directions. This approach has been tested on several cases. The algorithm generates feasible paths respecting limits on vehicle turning rates and climbing angles but the tests evidence also some issues that have to be investigated to improve the algorithm.

As a matter of fact the following conclusion can be taken:

- The goal to obtain paths respecting the basic aircraft kinematic constraints has been reached. KA* generates smooth and safe paths respecting the imposed constraints. No sharp heading changes or strong altitude variations typical of classical graph search paths are shown. Tests put in evidence that some paths obtained with classical graph search algorithms cannot be adapted in post processing with waypoints reallocation, to reach the aircraft kinematic constraints. Some heading changes or altitude variations affect deeply the whole path and a complete re-planning is needed when these unfeasible trajectories are included in the full path. This point addresses the development of an algorithm that like KA* may be able to match the graph search logics with the aircraft kinematic constraints.
- Obstacles separation represents an important improvement with respect to other graph search solutions. In classic graph search formulations obstacles separation was implemented just imposing to skip a given amount of cells around the obstacles whether the algorithm should try to expand them. In KA* the obstacle separation is more elegant; the algorithm skips the states with positions too close to the obstacles modifying accordingly the full planned path.
- Introducing the model to expand the states and perform the graph search is the fundamental novelty introduced with KA*, but is paid with an algorithm increased complexity. A longer computation time was expected but several tests demonstrated just slight time increases to obtain the solution. This is important to save the merit of low computation effort characteristic of graph search algorithms.
- Tests show another important KA* merit: the lower waypoints number on the path generated by KA* with respect to A*. KA* algorithm naturally generates just the amount of waypoints needed to reach the goal and because of this the waypoints filtering and reallocation needed for A* can be skipped.
- In spite of the simplified wind model, the effects of this disturbance are hardly relevant also for the preliminary tests reported in this paper. Wind modifies the state space and forces the algorithm to obtain solutions very different from the one without wind. Because of this further and deeper investigations are required to better understand this problem and improve accordingly the implementation.
- Analyzing heading changes and altitude variations needed to follow the KA* paths it is evident the strong effect that wind has on the path following performances. Paths with wind require harder heading and altitude variations pushing the aircraft to reach its limits in order to follow the path. This aspect shall be taken into account in the following work in order to study deeper the problem and find modifications on the cost function that can improve the algorithm performances.

- The second test puts in evidence an issue still present in KA*: the optimality of the paths planned. The wind effect on the model drives the states expansion along directions that are not taken into account otherwise. This way a shorter and computationally lighter solution is found. This means that the algorithm search must be improved modifying the cost function in order to investigate possible optimality proofs for the generated path. However this is a hard task because the graph search logics in it self makes optimality proof possible just for limited and simple problem formulations.
- Another issue outlined by the tests that must be analyzed in the following work is the exponential increase of computational time when the algorithm cannot follow a monotonic cost function decrease along the states expansion. Part of the problem is due to the graph structure and needs a deep state space analysis to be improved. On the other side, the cost function then needs to be modified, investigating solutions able to drive differently the graph search.
- Finally tests show that the different aircraft behavior on longitudinal with respect to lateral-directional plane affects largely the model time step selection. Particularly, longer time steps are needed when strong altitude variations are required to cross the obstacles. As a consequence, the time step must be tuned in order to improve the algorithm performances according with the test cases.

7. References

Bellman, R. (1958). On a Routing Problem. *Quarterly of Applied Mathematics*, Vol. 16, No 1, pp. 87–90.

Bertuccelli, L.F. & How, J.P. (2005). Robust UAV Search for Environmentas with Imprecise Probability Maps. *Proceedings of IEEE Conference of Decision and Control*, Seville, Dec 2005.

Boissonnat, J.D., Cèrèzo, A. & Leblond, J. (1994). Shortest Paths of Bounded Curvature in the Plane. *Journal of Intelligent and Robotic Systems*, Vol. 11, No. 1-2.

Carroll, D.L. (1996). Chemical Laser Modeling With Genetic Algoritms. *AIAA Journal*, Vol. 34, No.2, pp.338-346.

Chandler, P.R., Rasmussen, S. & Patcher, M. (2000). UAV Cooperative Path Planning. *Proceedings of AIAA Guidance, Navigation and Control Conference*, Denver, USA.

Chitsaz, H., LaValle, S.M. (2007). Time-optimal Paths for a Dubins airplane. *Proceedings of IEEE Conference on Decision and Control*, New Orleans, USA.

De Filippis, L., Guglieri, G., Quagliotti, F. (2009). Flight Analysis and Design for Mini-UAVs. *Proceedings of XX AIDAA Congress*, Milano, Italy.

De Filippis, L., Guglieri, G., Quagliotti, F. (2010). A minimum risk approach for path planning of UAVs. *Journal of Intelligent and Robotic Systems*, Springer, pp. 203-222.

De Filippis, L., Guglieri, G. & Quagliotti, F. (2011). Path Planning strategies for UAVs in 3D environments. *Journal of Intelligent and Robotic Systems*, Springer, pp. 1-18.

Dijkstra, E. W. (1959). A note to two problems in connexion with graphs. *Numerische Mathematik*, Vol:1, pp. 269–271.

Dogan, A. (2003). Probabilistic path planning for UAVs. *proceeding of 2nd AIAA Unmanned Unlimited Systems, Technologies, and Operations - Aerospace, Land, and Sea Conference and Workshop & Exhibition*, San Diego, California, September 15-18.

Dubins, L.E. (1957). On Curves of Minimal Length With a Constraint on Average Curvature and with Prescribed Initial and Terminal Positions and Tangents. *American Journal of Mathematics*, No. 79.

Ferguson, D. & Stentz, A. (2006). Using interpolation to improve path planning: The Field D* algorithm. *Journal of Field Robotics*, 23(2), 79-101.

Floyd, R. W. (1962). Algorithm 97: Shortest Path. *Communications of the ACM*, 5(6), pp. 345.

Ford, L. R. Jr. & Fulkerson, D. R. (1962). *Flows in Networks*. Princeton University Press.

Hart, P., Nilsson, N. & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, SCC-4(2), pp. 100-107.

Horner, D., P. & Healey, A., J. (2004). Use of artificial potential fields for UAV guidance and optimization of WLAN communications. *Autonomous Underwater Vehicles*, 2004 IEEE/OES, vol., no., pp. 88- 95, 17-18.

Jun, M. & D'Andrea, R. (2002). Path Planning for Unmanned Aerial Vehicles in Uncertain and Adversarial Environments. *Models, Applications and Algorithms*, Kluwer Academic Press.

Koenig, S. & Likhachev, M. (2001). Incremental A*. *Proceedings of the Natural Information Processing Systems*.

Koenig, S. & Likhachev, M. (2002). D* Lite. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 476-483.

LaValle, S. M. (2006). *Planning Algorithms.* Cambridge University Press.

Ma, X. & Castanon, D.A. (2006). Receding Horizon Planning for Dubins Traveling Salesman Problems. *Proceedings of IEEE Conference on Decision and Control*, San Diego, USA.

Miele, A. & Pritchard, R.E. (1969). Gradient Methods in Control Theory. Part 2 – Sequential Gradient-Restoration Algorithm. *Aero-Astronautics Report* ,n° 62, Rice University.

Miele, A. (1970). Gradient Methods in Control Theory. Part 6 – Combined Gradient-Restoration Algorithm. *Aero-Astronautics Report*, n° 74, Rice University.

Nash, A., Daniel, K., Koenig, S. & Felner, A. (2007). Theta*: Any-angle path planning on grids. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1177-1183.

Nikolos, I.K., Tsourveloudis, N.C. & Valavanis, K.P. (2003). Evolutionary Algorithm Based Offline/Online Path Planner for UAV Navigation. *IEEE Transactions on Systems, Man and Cybernetics - part B: Cybernetics*, Vol. 33, No. 6.

Papaefthymiou, M. & Rodriguez., J. (1991). Implementing Parallel Shortest-Paths Algorithms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*.

Pfeiffer, B., Batta, R., Klamroth, K. & Nagi, R. (2008). Path Planning for UAVs in the Presence of Threat Zones Using Probabilistic Modelling. *In: Handbook of Military Industrial Engineering*, Taylor and Francis, USA.

Reeds, J.A & Shepp, L.A. (1990). Optimal Path for a Car That Goes Both Forwards and Backwards. *Pacific Journal of Mathematics*, Vol. 145, No. 2.

Stentz, A. (1993). Optimal and efficient path planning for unknown and dynamic environments. *Carnegie Mellon Robotics Institute Technical Report*, CMU-RI-TR-93-20.

Stentz, A. (1995). The focussed D* algorithm for real-time replanning. *Proceedings of the International Joint Conference on Artificial Intelligence*, pp.1652-1659.

Sussmann, H.J. & Tang, W. (1991). Shortest Paths for the Reeds-Shepp Car: a Worked Out Example of the Use of Geometric Technique in Nonlinear Optimal Control. *Report SYCON-91-10*, Rutgers University.

Warshall, S. (1962). A theorem on Boolean matrices. *Journal of the ACM*, 9(1), pp. 11–12.

Waydo, S. & Murray, R.M. (2003). Vehicle Motion Planning Using Stream Functions. P*roceedings of 2003 IEEE International Conference on Robotics and Automation*, September.

9

GNSS Carrier Phase-Based Attitude Determination

Gabriele Giorgi[1] and Peter J. G. Teunissen[2,3]
[1] *Technische Universität München*
[2] *Curtin University of Technology*
[3] *Delft University of Technology*
[1]*Germany*
[2]*Australia*
[3]*The Netherlands*

1. Introduction

The GNSS (Global Navigation Satellite Systems) are a valid aid in support of the aeronautic science. GNSS technology has been successfully implemented in aircraft design, in order to provide accurate position, velocity and heading estimations. Although it does not yet comply with aviation integrity requirements, GNSS-based aircraft navigation is one of the alternative means to traditional dead-reckoning systems. It can provide fast, accurate, and driftless positioning solutions. Additionally, ground-based GNSS receivers may be employed to aid navigation in critical applications, such as precision approaches and landings.

One of the main issues in airborne navigation is the determination of the aircraft attitude, i.e., the orientation of the aircraft with respect to a defined reference system. Many sensors and technologies are available to estimate the attitude of a aircraft, but there is a growing interest in GNSS-based attitude determination (AD), often integrated at various levels of tightness to other types of sensors, typically Inertial Measurements Units (IMU). Although the accuracy of a stand-alone GNSS attitude system might not be comparable with the one obtainable with other modern attitude sensors, a GNSS-based system presents several advantages. It is inherently driftless, a GNSS receiver has low power consumption, it requires minor maintenance, and it is not as expensive as other high-precision systems, such as laser gyroscopes.

GNSS-based AD employs a number of antennas rigidly mounted on the aircraft's structure, as depicted in Figure 1. The orientation of each of the baselines formed between the antennas is determined by computing their relative positions. The use of GNSS carrier phase signals enables very precise range measurements, which can then be related to angular estimations. However, carrier phase measurements are affected by unknown integer ambiguities, since only their fractional part is measured by the receiver. The process of reconstructing the number of whole cycles from a set of measurements affected by errors goes under the name of ambiguity resolution (AR). Only after these ambiguities are correctly resolved to their correct integer values, will reliable baseline measurements and attitude estimations become available. This chapter focuses on novel AR and AD methods. Recent advances in GNSS-based attitude determination have demonstrated that the two problems can be formulated in an integrated manner, i.e., aircraft attitude and the phase ambiguities can be considered as the unknown parameters of a common ambiguity-attitude estimation method. In this integrated approach, the AR and AD problems are solved together by means of the theory of Constrained Integer Least-Squares (C-ILS). This theory extends the well-known least-squares theory (LS), by having geometrical constraints as well as integer constraints imposed on parameter subsets. The novel AR-AD estimation problem is discussed and its various properties are analyzed. The method's complexity is addressed by presenting new numerical algorithms that largely reduce the required processing load. The main objective of this chapter is to provide evidence that:

- GNSS carrier-phase based attitude determination is a viable alternative to existing attitude sensors
- Employing the new ambiguity-attitude estimation method enhances ambiguity resolution performance
- The new method can be implemented such that it is suitable for real-time applications

The structure of this contribution is as follows. Section 2 gives the observation and stochastic model which cast the set of GNSS observations, with special focus on the derivation of the GNSS-based attitude model. Section 3 reviews the most common attitude parameterization and estimation methods, mainly focusing on those widely used in aviation applications. Section 4 introduces a new ambiguity-attitude estimation method, which enhances the existing approach for attitude determination using GNSS signals. Section 5 presents flight-test results, which provide practical evidence of the novel method's performance. Finally, section 6 draws several conclusions.

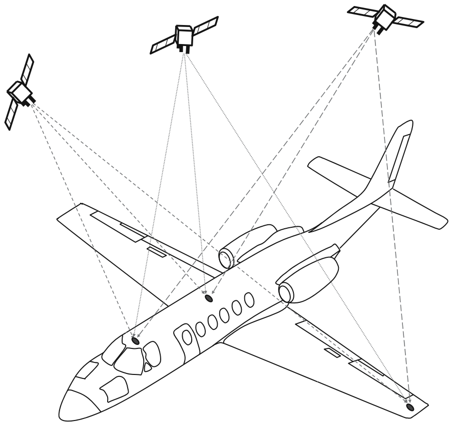


Fig. 1. GNSS data collected on multiple antennas installed on the fuselage and wings allow the estimation of an aircraft's orientation (attitude).

2. The GNSS-based attitude model

A GNSS receiver works by tracking satellites in view and storing the data received. Each GNSS satellite broadcasts a coded message with information about its orbit, the time of transmission, and few other parameters necessary for the correct processing at receiver side (Misra & Enge, 2001). By collecting signals from three or more satellites a GNSS receiver determines its own position with a triangulation procedure, exploiting the knowledge about both the satellites positions and the slant distance (range) by each satellite in view. The range measurements are obtained by detecting the time of arrival of the signal, from which the range can be inferred. This measurement is affected by several error sources: the satellite and receiver clocks are not perfectly synchronized; the signal travels through the atmosphere, which causes delays; the direct signal may be affected by unwanted reflections (multipath) that cannot be perfectly eliminated by careful antenna design. If not properly modeled, each of these effects will limit the achievable GNSS accuracy. The observed pseudorange or code observable is therefore modeled as

$$\begin{aligned} P^s_{r,f}(t) &= \rho^s_r(t, t-\tau^s_r) + I^s_{r,f} + T^s_r + dm^s_{r,f} + c\,[dt_r(t) - dt^s(t-\tau^s_r)] \\ &+ \quad c\left[d_{r,f}(t) + d^s_f(t-\tau^s_r)\right] + \varepsilon^s_{P,r,f} \end{aligned} \tag{1}$$

where the superscript s indicates the satellite and the subscripts r and f indicate the receiver and the frequency, respectively. The different terms are:

P	code observation [m]
τ	signal travel time [s]
ρ	geometrical distance between receiver and satellite [m]
I , T	ionospheric and tropospheric delays [m]
dm	multipath error [m]
c	speed of light : 299 792 458 [$\frac{\text{m}}{\text{s}}$]
dt	clock errors [s]
d	instrumental delays [s]
ε_P	remaining unmodeled code errors [m]

The magnitude of errors involved in these observations - decimeter or meter level - would not allow high-precision applications, such AD, which require cm- or mm-level accuracy in the final positioning product. Therefore, another set of observations is considered: the phase of the tracked signal, modeled as

$$\begin{aligned} \Phi^s_{r,f}(t) &= \rho^s_r(t, t-\tau^s_r) - I^s_{r,f} + T^s_r + \delta m^s_{r,f} + c\,[dt_r(t) - dt^s(t-\tau^s_r)] \\ &+ \quad c\left[\delta_{r,f}(t) + \delta^s_f(t-\tau^s_r)\right] + \lambda_f[\varphi^s_{r,f}(t_0) - \varphi^s_f(t_0)] + \lambda_f z^s_{r,f} + \varepsilon^s_{\Phi,r,f} \end{aligned} \tag{2}$$

with φ the phase of the generated carrier signal (original or replica) in cycles, t_0 the time of reference for phase synchronization, and λ_f the wavelength of frequency f. The phase reading is characterized by different atmospheric delays (the ionosphere causes an anticipation of phase instead of a delay), different instrumental biases (indicated with δ), different multipath and an additional bias which is represented by the unknown number of whole cycles that cannot be detected by the tracking loop, since only the fractional part is measured. These are the integer ambiguities z. In case of GNSS, the precision of the phase measurements far exceeds the one of code observations: typically the phase observable is two orders of magnitude more accurate than the code measurement.

The many sources of error in (1) and (2) can be mitigated in relative positioning models. First, we form the so-called single difference (SD) code and carrier phase observations by taking the differences between observations simultaneously collected at two antennas tracking the same satellite:

$$
\begin{aligned}
P^s_{r_2,f}(t) - P^s_{r_1,f}(t) = P^s_{r_{12},f} = \quad & \rho^s_{r_{12},f} + I^s_{r_{12},f} + T^s_{r_{12}} + dm^s_{r_{12},f} + cdt_{r_{12}} + cd_{r_{12},f} + \varepsilon^s_{P,r_{12},f} \\
\Phi^s_{r_2,f}(t) - \Phi^s_{r_1,f}(t) = \Phi^s_{r_{12},f} = \; & \rho^s_{r_{12},f} - I^s_{r_{12},f} + T^s_{r_{12}} + \delta m^s_{r_{12},f} + cdt_{r_{12}} + c\delta_{r_{12},f} + \lambda_f \varphi^s_{r_{12},f}(t_0) \\
& + \lambda_f z^s_{r_{12},f} + \varepsilon^s_{\Phi,r_{12},f}
\end{aligned}
\tag{3}
$$

where subscript r_{12} indicates the difference between two antennas: $\square_{r_{12}} = \square_{r_2} - \square_{r_1}$. The phase value $\varphi^s_f(t_0)$, relative to the common satellite, is eliminated. The instrumental delays and clock errors of the satellite are usually considered constant over short time spans, since the travel time difference with respect to any two points on the Earth surface is small (Teunissen & Kleusberg, 1998).

The terms $cdt_{r_{12}}$, $cd_{r_{12},f}$ and $\delta_{r_{12},f}$ refer to the relative clock errors and relative instrumental delays between the two receivers. A perfect synchronization between receivers implies the cancellation of the clock biases, and a correct calibration would reduce the impact of instrumental delays. In the case of a single receiver connected to two antennas, these two sources of relative error could cancel out with a proper calibration.

The receiver clock errors and hardware delays in the single difference equations (3) are common for all the satellites tracked at the same frequency. Therefore these terms can be eliminated by forming a double difference (DD) combination, obtained by subtracting two SD measurements from two different satellites:

$$
\begin{aligned}
P^{s_{12}}_{r_{12},f} &= \rho^{s_{12}}_{r_{12},f} + I^{s_{12}}_{r_{12},f} + T^{s_{12}}_{r_{12}} + dm^{s_{12}}_{r_{12},f} + \varepsilon^{s_{12}}_{P,r_{12},f} \\
\Phi^{s_{12}}_{r_{12},f} &= \rho^{s_{12}}_{r_{12},f} - I^{s_{12}}_{r_{12},f} + T^{s_{12}}_{r_{12}} + \delta m^{s_{12}}_{r_{12},f} + \lambda_f z^{s_{12}}_{r_{12},f} + \varepsilon^{s_{12}}_{\Phi,r_{12},f}
\end{aligned}
\tag{4}
$$

It has been assumed that the real-valued initial phase of the receiver replica does not vary for different tracked GNSS satellites.

The differential atmospheric delays depend on the distance between antennas. For sufficiently short baselines - typically shorter than a kilometer - the signals received by the antennas have traveled approximately the same path, thus the atmospheric delays becomes highly correlated. The differencing operation makes these errors negligible with respect to the measurement white noise for the baselines typically employed in AD applications, which rarely exceeds a few hundred meters.

Note that the relation between observations and baseline coordinates is nonlinear, since these are contained in the range term

$$
\rho^s_r = \| r^s(t_r - \tau^s_r) - r_r(t_r) \| \tag{5}
$$

with r^s and r_r the satellite and receiver antenna position vectors, respectively. By assuming the atmospheric delays negligible and applying the Taylor expansion to expression (4) one obtains the linearized relations

$$\begin{aligned} \triangle P^{s_{12}}_{r_{12},f} &= -(u^{s_{12}}_r)^T \triangle r_{r_{12}} + \varepsilon^{s_{12}}_{P,r_{12},f} \\ \triangle \Phi^{s_{12}}_{r_{12},f} &= -(u^{s_{12}}_r)^T \triangle r_{r_{12}} + \lambda_f z^{s_{12}}_{r_{12},f} + \varepsilon^{s_{12}}_{\Phi,r_{12},f} \end{aligned} \tag{6}$$

where the observables are now 'observed minus computed' terms, and the unknowns are expressed as increments with respect to a computed approximate value. $\triangle r_{r_{12}}$ is the baseline vector - the difference between the absolute antennas positions - whereas $u^{s_{12}}_r = u^{s_2}_r - u^{s_1}_r$ is the difference between unit line-of-sight vectors of different satellites. Also note that the multipath terms have been lumped into the remaining unmodeled errors $\varepsilon^{s_{12}}_{P,r_{12},f}$ and $\varepsilon^{s_{12}}_{\Phi,r_{12},f}$.

Consider now two antennas simultaneously tracking the same $m+1$ satellites at N frequencies. The vector of DD observations of type (6) are cast in the linear(ized) functional model (Teunissen & Kleusberg, 1998)

$$y = Az + Gb + \varepsilon \quad ; \quad z \in \mathbb{Z}^{mN} \, , \quad b \in \mathbb{R}^3 \tag{7}$$

with y the $2mN$-vector of code and carrier phase observations, z the unknown integer-valued ambiguities and b the vector of real-valued baseline coordinates. A and G are the design matrices

$$A = \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} \otimes I_m \qquad G = e_N \otimes \begin{bmatrix} U \\ U \end{bmatrix} \tag{8}$$

with Λ the diagonal matrix of N carrier wavelengths and U the $m \times 3$ matrix of DD unit line-of-sight vectors. Symbol $\otimes$ denotes the Kronecker product (Van Loan, 2000).

Model (7) describes the linear relationship between GNSS observables and the parameters of the two antennas. However, a single baseline is generally not sufficient to estimate the full orientation of an aircraft with respect to a given reference frame. At least three non-aligned antennas are necessary to guarantee that each rotation of the aircraft can be tracked unambiguously. It is straightforward to generalize the model formulation (7) to cast n DD baseline observations, obtained with $n+1$ GNSS antennas (Teunissen, 2007a):

$$Y = AZ + GB + \Xi \quad ; \quad Z \in \mathbb{Z}^{mN \times n} \, , \quad B \in \mathbb{R}^{3 \times n} \tag{9}$$

This formulation is obtained by casting the observations at each baseline in the columns of the $2mN \times n$ matrix Y. Consequently, $Z = [z_1, \ldots, z_n]$ is the matrix whose n columns are the integer ambiguity mN-vectors, and $B = [b_1, \ldots, b_n]$ is the $3 \times n$ matrix that contains the n real-valued baseline vectors. We exploited here once again the short baseline hypothesis: the same matrix of line-of-sight vectors U is used for all baselines.

Besides describing the functional relationship between observables and unknowns, a proper modeling should also capture the observation noise, i.e., the measurement error. The error is relative to the receiver, to the satellite, to the frequency and to the type of observations (code or phase). The variance-covariance (v-c) matrix of a vector of DD observations y collected at baseline b will be denoted as $D(y) = Q_{yy}$, with $D(\cdot)$ the dispersion operator. For the multibaseline model (9), the description of measurement errors requires a further step: the observations are cast into a $2mNn$ vector by applying the *vec* operator, which stacks the columns of a matrix. The v-c matrix Q_{YY} that characterizes the error statistic of $vec(Y)$ is

$$D(vec(Y)) = Q_{YY} \tag{10}$$

A simple expression for Q_{YY} is obtained by assuming that each of the baselines is characterized by the same v-c matrix Q_{yy}:

$$D(vec(Y)) = Q_{YY} = P_n \otimes Q_{yy} \tag{11}$$

with P_n the $n \times n$ matrix that takes care of the correlation which is introduced by having a common antenna:

$$P_n = \frac{1}{2}\left[I_n + e_n e_n^T\right] = \begin{bmatrix} 1 & 0.5 \cdots & 0.5 \\ 0.5 & 1 & \vdots \\ \vdots & \ddots & 0.5 \\ 0.5 \ldots & 0.5 & 1 \end{bmatrix} \tag{12}$$

Expressions (9) and (11) define the *GNSS multibaseline model* that we use in this contribution as the foundation of our GNSS-based attitude estimation theory.

With the available code and phase observations it is possible to estimate the set of baseline coordinates. These can then be used to provide the aircraft attitude, but *only* when a further condition is realized: the positions of the antennas installed aboard the given aircraft are known, rigid and do not change over time (or, if change occurs, it is perfectly known and predictable). This is so because it is necessary to have a one-to-one relationship between aircraft attitude and baselines attitude. As an example, consider two antennas mounted on the two extremities of a flexible mast: it is not possible to separate the rotations of the mast from its deformations by only observing the variations of the mutual position between the two antennas.

The rigidity assumption is formalized in the following way. Consider two orthonormal frames, defined by the basis $\{u_1, u_2, u_3\}$ and $\{u_1', u_2', u_3'\}$. Let us assume that the second frame is integrally fixed with the aircraft. An arbitrary vector x can be equivalently described by using either reference system:

$$\begin{aligned} x &= (x^T u_1)\, u_1 + (x^T u_2)\, u_2 + (x^T u_3)\, u_3 \\ x' &= \left(x^T u_1'\right) u_1' + \left(x^T u_2'\right) u_2' + \left(x^T u_3'\right) u_3' \end{aligned} \tag{13}$$

The relation between the components of vectors x and x' is completely defined by the mutual orientation of the two reference systems. The linear transformation $x = Rx'$ allows for a one-to-one relationship. Matrix R, hereafter referred to as *rotation matrix* or *attitude matrix*, belongs to the class of orthonormal matrices $\mathbb{O}$: its column vectors r_i are normal and their product null: $r_i^T r_j = \delta_{i,j}$, with $\delta_{i,j}$ the Kronecker's delta ($\delta_{i,j} = 1$ if $i = j$, 0 otherwise). These constraints are necessary for the admissibility of transformation $x = Rx'$. In absence of deformations, the scalar product between any two vectors should be invariant with respect to the transformation:

$$x'^T y' = x^T R^T R y = x^T y \tag{14}$$

whereas the vectorial product is invariant under rotations about the axis defined by $x' \times y'$:

$$x' \times y' = (Rx) \times (Ry) = |R|\, R\, (x \times y) \tag{15}$$

These conditions are fulfilled for orthonormal rotation matrices with determinant equal to one.

Model (9) can then be reformulated by means of the linear transformation $B = RF$, where F is used to cast the set of known local baseline coordinates and R is the orthonormal ($R^T R = I_q$) matrix that rotates B into F. The complete GNSS attitude model reads then (Teunissen, 2007a; 2011)

$$\begin{aligned} & Y = AZ + GRF + \Xi; \\ & Z \in \mathbb{Z}^{mN \times n}, \\ & R \in \mathbb{O}^{3 \times q} \\ & D(vec(Y)) = Q_{YY} \end{aligned} \tag{16}$$

Parameter q is introduced in order to make model (16) of general applicability. The n baselines may be aligned or coplanar, impeding the estimation of a full 3×3 matrix R. Therefore, q defines the span of matrix F. For baseline sets formed by aligning $n + 1$ antennas we set $q = 1$, whereas configurations of coplanar antennas are defined by $q = 2$. With four or more non-coplanar antennas, $q = 3$.

The GNSS attitude model (16) is a nonlinear model. Although the relation between observables and unknowns remain linear, the orthonormal constraint is of a nonlinear nature, and profoundly affects the estimation process. This is investigated in section 4. First, the following section gives an overview of common attitude parameterization and estimation methods.

3. Attitude parameterization and estimation

The orthonormality of R ($R^T R = I_q$) imposes $\frac{q(q+1)}{2}$ constraints on its components r_{ij}. The full matrix R can then be parameterized with a properly chosen set of variables, whose number can be as little as two (if $q = 1$) or three (if $q \leq 2$). To this purpose, several representations may be used, and few are briefly reviewed in the following.

From a set of code and phase observations cast as in (16), the problem of extracting the components of the attitude representation involves, as shown in section 4, a nonlinear least squares problem. Its formulation and solution are the second topic discussed in this section.

3.1 Attitude parameterization

Several attitude parameterizations are available in the literature, see e.g., Shuster (1993) and references therein. The most common parameterizations are briefly reviewed in the following.

3.1.1 Direction cosine matrix

The transformation between two basis of orthonormal frames reads

$$\{u'_1, u'_2, u'_3\} = R\{u_1, u_2, u_3\} \quad \Longrightarrow \quad u'_i = \sum_{j=1}^{3} r_{ij} u_j \tag{17}$$

with r_{ij} the entries of R. The scalar product between any two unit vectors of the two frames is

$$u_i'^T u_j = \sum_{k=1}^{3} r_{ik}\left(u_k^T u_j\right) = r_{ij} = \cos\left(\widehat{u_i' u_j}\right) \tag{18}$$

Hence, the attitude matrix can be expressed by nine direction cosines, i.e., the nine cosines of the angles formed by the three unit vectors of the first frame and the three unit vectors of the second frame:

$$R = \begin{bmatrix} u_1'^T u_1 & u_1'^T u_2 & u_1'^T u_3 \\ u_2'^T u_1 & u_2'^T u_2 & u_2'^T u_3 \\ u_3'^T u_1 & u_3'^T u_2 & u_3'^T u_3 \end{bmatrix} \tag{19}$$

This representation fully defines the mutual orientation of the two frames, by using a set of nine parameters (see Figure 2). Each configuration can be described without incurring any singularity, at the cost of having a larger number of parameters than other representations.

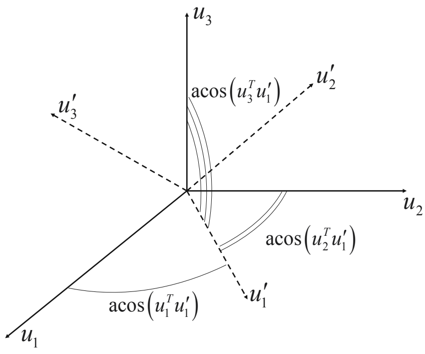

u_3
$\mathrm{acos}\left(u_3^T u_1'\right)$
u_2'
u_3'
u_2
$\mathrm{acos}\left(u_2^T u_1'\right)$
u_1'
$\mathrm{acos}\left(u_1^T u_1'\right)$
u_1

Fig. 2. The main axis u_1' is completely defined by the knowledge of the three direction cosines $u_1^T u_1'$, $u_2^T u_1'$ and $u_3^T u_1'$.

3.1.2 Euler angles

Consider counterclockwise rotations about one of the main axis of a frame $\{u_1, u_2, u_3\}$. Then the rotation matrix R is obtained through one of the following expressions:

$$R\left(u_1, \phi\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & C_\phi & S_\phi \\ 0 & -S_\phi & C_\phi \end{bmatrix}$$

$$R\left(u_2, \phi\right) = \begin{bmatrix} C_\phi & 0 & -S_\phi \\ 0 & 1 & 0 \\ S_\phi & 0 & C_\phi \end{bmatrix} \tag{20}$$

$$R\left(u_3, \phi\right) = \begin{bmatrix} C_\phi & S_\phi & 0 \\ -S_\phi & C_\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Any arbitrary rotation can always be decomposed as a combination of three consecutive rotations about the main axis u_1, u_2 or u_3, represented by one of the relations in (20). Figure 3 shows the example of a 321 rotation: the first rotation is about the third main axis u_3 with magnitude ψ, the second is about the (new) second main axis u_2' with magnitude θ, the last about the (new) first main axis u_1'' with magnitude ϕ. The rotation matrix that defines the transformation between the frames $\{u_1, u_2, u_3\}$ and $\{u_1''', u_2''', u_3'''\}$ is built as

$$\{u_1, u_2, u_3\} \overset{\psi}{\underset{R(u_3,\psi)}{\Longrightarrow}} \{u_1', u_2', u_3'\} \overset{\theta}{\underset{R(u_2,\theta)}{\Longrightarrow}} \{u_1'', u_2'', u_3''\} \overset{\phi}{\underset{R(u_1,\phi)}{\Longrightarrow}} \{u_1''', u_2''', u_3'''\} \tag{21}$$

Therefore, $R_{321}(\psi, \theta, \phi) = R(u_1, \phi) R(u_2, \theta) R(u_3, \psi)$. Twelve combinations of rotations are possible, whose choice depends on the application. As an example, the sequence 321 is commonly used to describe the orientation of an aircraft, where the angles ψ, θ, ϕ are named heading, elevation and bank, respectively.

It is easy to see that the Euler angles representation is not unique: e.g, the combination 321 is equivalently expressed as $R_{321}(\psi, \theta, \phi)$ or $R_{321}(\psi + \pi, \pi - \theta, \phi + \pi)$. This ambiguity is usually avoided by imposing $-90° < \theta \leq 90°$. The main advantage of the Euler angles representation is its straightforward physical interpretation, of importance for human-machine interfaces. The disadvantage lies in fact that the construction of the attitude matrix requires the evaluation of trigonometric functions, of higher computational load than other parameterizations. Also, the derivatives of the components of the rotation matrix are nonlinear (trigonometric), and affected by singularities.

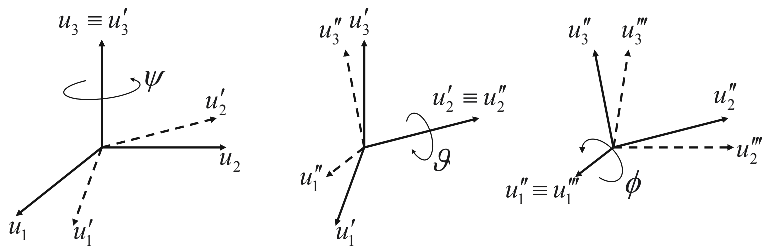


Fig. 3. The three consecutive rotations that rotate the frame $\{u_1, u_2, u_3\}$ into the frame $\{u_1''', u_2''', u_3'''\}$. The first one is about the main axis u_3 and magnitude ψ, the second is about the main axis u_2' and magnitude θ and the third is about the main axis u_1'' and magnitude ϕ.

3.1.3 Quaternions

A quaternion is an order-4 vector whose components can be used to define the mutual rotation between reference systems:

$$\bar{q} = (q_1, q_2, q_3, q_4)^T \tag{22}$$

q_4 is named the scalar (real) component of the quaternion, whereas $(q_1, q_2, q_3)^T$ forms the imaginary (or vectorial) part. The components of a quaternion must respect the constraint $\bar{q}^T \bar{q} = 1$. Physically, the four components of $\bar{q}$ define the magnitude and axis of the rotation necessary to rotate one reference system into the other, see Figure 4. The attitude matrix R is parameterized in terms of quaternions as

$$R(\bar{q}) = R(q, q_4) = \left(q_4^2 - \|q\|^2\right) I_3 + 2qq^T + 2q_4\left[q^+\right]$$

$$= \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix} \tag{23}$$

with

$$\left[q^+\right] = \begin{bmatrix} 0 & q_3 & -q_2 \\ -q_3 & 0 & q_1 \\ q_2 & -q_1 & 0 \end{bmatrix} \tag{24}$$

This parameterization is non ambiguous and it does not involve any trigonometric function, so that the computational burden is lower than with other representations. The quaternion representation is of common use in attitude estimation and control applications, since it guarantees high numerical robustness.

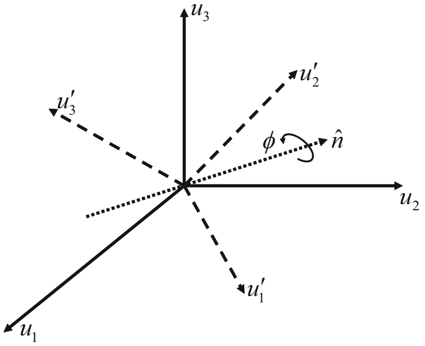


Fig. 4. The frame $\{u_1', u_2', u_3'\}$ can be rotated to equal the orientation of frame $\{u_1, u_2, u_3\}$ by means of a single rotation of magnitude ϕ about axis $\hat{n}$. The four components of a quaternion are proportional to the entries of the normal vector $\hat{n}$ and to the magnitude ϕ.

3.2 Attitude estimation

As it will be shown in the next sections, the least-squares solution of model (16), requires the solution of a constrained least-squares problem of the type:

$$\check{R} = \arg \min_{R \in \mathbb{O}^{3\times q}} \left\|vec\left(\hat{R} - R\right)\right\|_Q^2 \tag{25}$$

with $\|\cdot\|_Q^2 = (\cdot)^T Q^{-1}(\cdot)$. The shape of Q drives the choice of the solution technique to be adopted for solving (25).

If Q is a diagonal matrix, problem (25) becomes an Orthogonal Procustes Problem (OPP), see Schonemann (1966). This class of constrained least-squares problem have been thoroughly analyzed, and fast algorithms have been devised to quickly extract the minimizer $\check{R}$, see (Davenport, 1968; Shuster & Oh, 1981). Various fast methods for the solution of an OPP have been introduced - and widely used in practice - based on the Singular Value Decomposition

(SVD) or the EIGenvalues decomposition (EIG), such as the QUaternion ESTimator (QUEST) (Shuster, 1978; Shuster & Oh, 1981), the Fast Optimal Attitude Matrix (FOAM) (Markley & Landis, 1993), the EStimator of the Optimal Quaternion (ESOQ) (Mortari, 1997) or the Second ESOQ (ESOQ2) (Mortari, 2000) algorithms, which have been extensively compared in Markley & Mortari (1999; 2000) and Cheng & Shuster (2007).

For nondiagonal matrices Q, the extraction of the orthonormal attitude matrix $\check{R}$ has to be performed through nonlinear estimation techniques. A first numerical scheme for the solution of (25) is derived by applying the Lagrangian multipliers method. The Lagrangian function is

$$L(R) = vec\left(\hat{R} - R\right)^T Q^{-1} vec\left(\hat{R} - R\right) - \text{tr}\left[[\lambda]_q \left[R^T R - I_q\right]\right] \tag{26}$$

with $[\lambda]_q$ the q by q matrix of Lagrangian multipliers:

$$[\lambda]_1 = \lambda \quad ; \quad [\lambda]_2 = \begin{bmatrix} \lambda_1 & \frac{1}{2}\lambda_3 \\ \frac{1}{2}\lambda_3 & \lambda_2 \end{bmatrix} \quad ; \quad [\lambda]_3 = \begin{bmatrix} \lambda_1 & \frac{1}{2}\lambda_4 & \frac{1}{2}\lambda_5 \\ \frac{1}{2}\lambda_4 & \lambda_2 & \frac{1}{2}\lambda_6 \\ \frac{1}{2}\lambda_5 & \frac{1}{2}\lambda_6 & \lambda_3 \end{bmatrix} \tag{27}$$

The last term of (26) gives the $\frac{q(q+1)}{2}$ constraining functions that follows from the orthonormality of R: q constraints are given by the normality (unit length) of the columns of R, whereas $\frac{q(q-1)}{2}$ constraints are given by the orthogonality of the columns of R.

The gradient of the Lagrangian function (26), together with the $\frac{q(q+1)}{2}$ constraining functions, defines the nonlinear system to be solved:

$$\begin{cases} \frac{1}{2}\nabla L(R) = \left[Q^{-1} - [\lambda]_q \otimes I_3\right] vec\left(R\right) - Q^{-1} vec\left(\hat{R}\right) = 0 \\ vec\left(R^T R - I_q\right) = 0 \end{cases} \tag{28}$$

Due to the symmetry of matrix $\left[R^T R - I_q\right]$, only its upper (or lower) triangular part has to be considered in (28). The Newton-Raphson method can then be applied to iteratively converge to the sought orthonormal matrix of rotations.

This method is computationally heavier than other iterative schemes, since it requires the explicit computation of larger-sized matrices than other methods given in the following.

A second viable solution scheme is obtained by re-parameterizing the attitude matrix with the vector of Euler angles $\mu = (\psi, \theta, \phi)^T$. Following the reparameterization, matrix $R\left(\mu\right)$ implicitly fulfills the constraint $R^T R = I_q$, and problem (25) is rewritten as

$$\check{\mu} = \arg \min_{\mu \in \mathbb{R}^3} \|h(\mu)\|_I^2 \tag{29}$$

with $h(\mu) = Q^{-\frac{1}{2}} vec\left(\hat{R} - R(\mu)\right)$. The nonlinear least-squares problem (29) is solved by applying iterative methods, e.g., the Newton method. This approach (Euler angles parameterization) works with a minimal set of unknowns - the Euler angles - and it can quickly converge to the sought minimizer if an accurate initial guess is used. The disadvantage is that trigonometric functions have to be evaluated, increasing the computational load.

A third viable approach is devised by employing the quaternions parameterization of R and to solve for (25):

$$\check{\bar{q}} = \arg \min_{\bar{q} \in \mathbb{R}^4, \|\bar{q}\|=1} \left\| vec\left(\hat{R} - R(\bar{q})\right) \right\|_Q^2 \tag{30}$$

The orthonormality of R is guaranteed by the normality of the quaternion: this introduces a single constraint in the minimization problem (30). A Lagrangian function is formed as

$$L'(\bar{q}) = vec\left(\hat{R} - R(\bar{q})\right)^T Q^{-1} vec\left(\hat{R} - R(\bar{q})\right) - \lambda\left(\bar{q}^T\bar{q} - 1\right) \tag{31}$$

and the (nonlinear) system to be solved is

$$\begin{cases} \frac{1}{2}\nabla L'(R(\bar{q})) = J_{R(\bar{q})}^T Q^{-1} vec\left(\hat{R} - R(\bar{q})\right) - \lambda\bar{q} = 0 \\ \bar{q}^T\bar{q} - 1 = 0 \end{cases} \tag{32}$$

with $J_{R(\bar{q})}$ the Jacobian of $vec(R(\bar{q}))$.

The three iterative solutions given above rigorously solve for problem (25), but are generally slower than the methods available for diagonal Q matrices (SVD, EIG, QUEST, FOAM, ESOQ, and ESOQ2). Figure 5a illustrates the mean number of floating-point operations for different attitude estimation methods, per number of baselines employed. 10^4 samples $\hat{R}$ have been generated via Monte Carlo simulations for a given fully-populated Q matrix. The gray bars span between the maximum and minimum numbers obtained for each algorithm. The off-diagonal elements of Q are disregarded when applying the SVD, EIG, QUEST, FOAM, ESOQ, and ESOQ2 methods. These techniques outperform each iterative method: the number of required floating-point operations is generally two to three orders of magnitude lower. Among the iterative methods, the Lagrangian multiplier technique generally requires the highest number of operations, making it the least efficient method, while the Euler angle method and the Quaternion parameterization provide better overall results. Figure 5b shows the corresponding mean, maximum and minimum computational times marked during the simulations. The Lagrangian parameterization method generally takes the longest time to converge, whereas the quaternion and Euler angle methods show better results. Note that higher number of floating operations does not directly translate into longer computational times, because modern processor architectures efficiently operate by means of multi-threading and parallel processing.

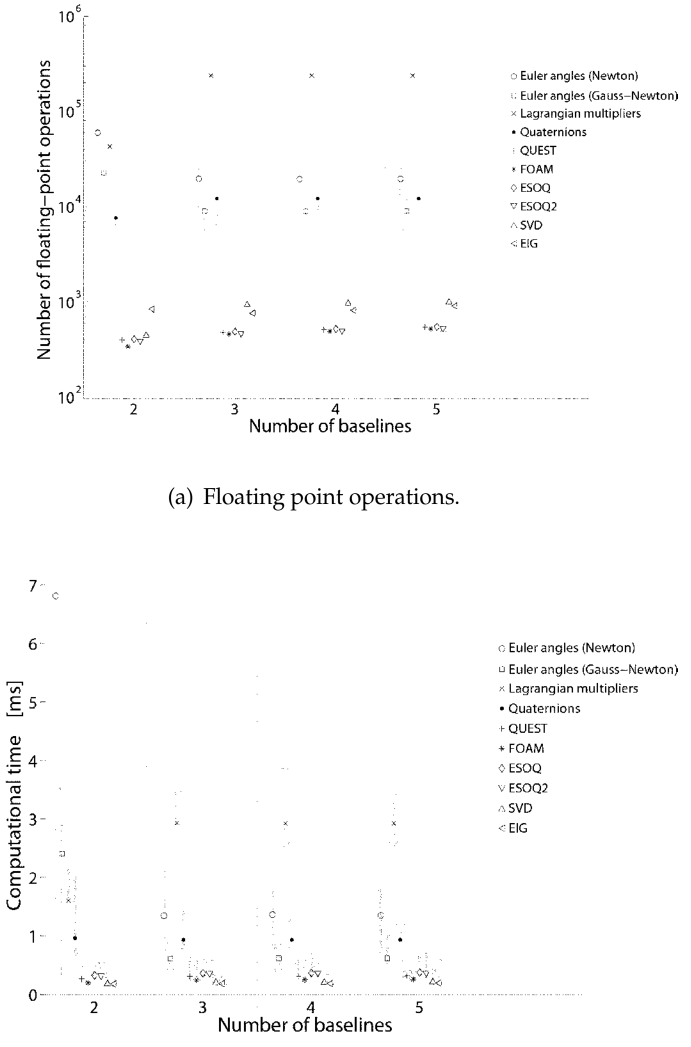
4. Reliable attitude-ambiguity estimation methods

This section reviews the solution of the GNSS attitude model (16). This can be presented by addressing two consecutive steps: float estimation and ambiguity resolution.

4.1 Float ambiguity-attitude solution

We indicate with*float* the solution of (16) obtained by disregarding the whole set of constraints, i.e., the integerness of Z and the orthonormality of R:

$$\left\{\hat{Z}, \hat{R}\right\} = \arg \min_{Z \in \mathbb{R}^{mN \times n},\, R \in \mathbb{R}^{3 \times q}} \left\| vec\left(Y - AZ - GRF\right) \right\|_{Q_{YY}}^2 \tag{33}$$



(a) Floating point operations.

(b) Computational times.

Fig. 5. Mean, maximum and minimum numbers of floating-point operations (left) and computational times (right) per number of baseline, for each of the attitude estimation method analyzed.

This float solution follows from solving the system of normal equations

$$\begin{aligned} M\begin{pmatrix} vec(\hat{R}) \\ vec(\hat{Z}) \end{pmatrix} &= \begin{bmatrix} FP_n^{-1} \otimes G^T Q_{yy}^{-1} \\ P_n^{-1} \otimes A^T Q_{yy}^{-1} \end{bmatrix} vec(Y) \\ M &= \begin{bmatrix} FP_n^{-1}F^T \otimes G^T Q_{yy}^{-1} G & FP_n^{-1} \otimes G^T Q_{yy}^{-1} A \\ P_n^{-1}F^T \otimes A^T Q_{yy}^{-1} G & P_n^{-1} \otimes A^T Q_{yy}^{-1} A \end{bmatrix} \end{aligned} \tag{34}$$

where the v-c matrix Q_{YY} is written as in (11). Inversion of the normal matrix M gives the v-c matrix of the float estimators $\hat{R}$ and $\hat{Z}$:

$$\begin{bmatrix} Q_{\hat{R}\hat{R}} & Q_{\hat{R}\hat{Z}} \\ Q_{\hat{Z}\hat{R}} & Q_{\hat{Z}\hat{Z}} \end{bmatrix} = M^{-1} \tag{35}$$

The float estimators are explicitly derived as

$$\begin{aligned} \hat{R} &= \left[\overline{G}^T Q_{yy}^{-1} \overline{G}\right]^{-1} \overline{G}^T Q_{yy}^{-1} Y P_n^{-1} F^T \left[F P_n^{-1} F^T\right]^{-1} \\ \hat{Z} &= \left[A^T Q_{yy}^{-1} A\right]^{-1} A^T Q_{yy}^{-1} \left[Y - G\hat{R}F\right] \end{aligned} \tag{36}$$

with $\overline{G} = \left[I - A\left[A^T Q_{yy}^{-1} A\right]^{-1} A^T Q_{yy}^{-1}\right] G$. Next to the above float solution, we can also define the following *conditional* float solution for the attitude matrix:

$$\hat{R}(Z) = \arg \min_{R \in \mathbb{R}^{3\times q}} \| vec\,(Y - AZ - GRF) \|^2_{Q_{YY}} \tag{37}$$

In this case the ambiguity matrix is assumed completely known. The solution $\hat{R}(Z)$ can be computed form the float solutions $\hat{R}$ and $\hat{Z}$ as:

$$vec(\hat{R}(Z)) = vec(\hat{R}) - Q_{\hat{R}\hat{Z}} Q_{\hat{Z}\hat{Z}}^{-1} vec(\hat{Z} - Z) \tag{38}$$

Application of the variance propagation law gives

$$Q_{\hat{R}(Z)\hat{R}(Z)} = Q_{\hat{R}\hat{R}} - Q_{\hat{R}\hat{Z}} Q_{\hat{Z}\hat{Z}}^{-1} Q_{\hat{Z}\hat{R}} = \left[F P_n^{-1} F^T\right]^{-1} \otimes \left[G^T Q_{yy}^{-1} G\right]^{-1} \tag{39}$$

There is a very large difference in the precision of the float solution $\hat{R}$ and the precision of the conditional float solution $\hat{R}(Z)$. This can be demonstrated by comparing expression (39) with $Q_{\hat{R}\hat{R}}$ in (35), whose relation with the design matrices can be made explicit as

$$Q_{\hat{R}\hat{R}} = \left[F P_n^{-1} F^T\right]^{-1} \otimes \left[\overline{G}^T Q_{yy}^{-1} \overline{G}\right]^{-1} \tag{40}$$

Matrix $\left[G^T Q_{yy}^{-1} G\right]^{-1}$ is characterized by much smaller entries than $\left[\overline{G}^T Q_{yy}^{-1} \overline{G}\right]^{-1}$. This is demonstrated as follows. Matrices A, G and Q_{yy} may be partitioned as

$$A = \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} \otimes I_m \qquad G = e_N \otimes \begin{bmatrix} U \\ U \end{bmatrix} \qquad Q_{yy} = I_N \otimes \begin{bmatrix} \sigma_P^2 Q & 0 \\ 0 & \sigma_\Phi^2 Q \end{bmatrix} \tag{41}$$

where we assumed, for simplicity, the same code and phase standard deviations for each observation, independent from the combination of satellites, receivers and frequency. Λ is the diagonal matrix of carrier wavelengths, whereas Q is the matrix that introduces correlation due to the DD operation.

It follows that

$$\begin{aligned} \left[\overline{G}^T Q_{yy}^{-1} \overline{G}\right]^{-1} &= \tfrac{\sigma_P^2}{N} \left[U^T Q^{-1} U\right]^{-1} \\ \left[G^T Q_{yy}^{-1} G\right]^{-1} &= \tfrac{1}{N} \frac{\sigma_\Phi^2}{\frac{\sigma_\Phi^2}{\sigma_P^2} + 1} \left[U^T Q^{-1} U\right]^{-1} \approx \tfrac{\sigma_\Phi^2}{N} \left[U^T Q^{-1} U\right]^{-1} \end{aligned} \tag{42}$$

The ratio between the entries of matrix $Q_{\hat{R}\hat{R}}$ and $Q_{\hat{R}(Z)\hat{R}(Z)}$ is then proportional to the ratio $\frac{\sigma_\Phi^2}{\sigma_P^2}$. In GNSS applications, this phase-code variance ratio is in the order of 10^{-4}. This clearly demonstrates the importance of ambiguity resolution: if we can integer-estimate Z with sufficiently high probability, then the attitude matrix R can be estimated with a precision that is comparable with the high precision of $\hat{R}(Z)$.

4.2 Ambiguity resolution

The second step consists of the resolution of the carrier phase integer ambiguities. The solution of model (16) is obtained through the following C-ILS minimization problem:

$$\{\check{Z}, \check{R}\} = \arg \min_{Z \in \mathbb{Z}^{mN \times n},\, R \in \mathbb{O}^{3 \times q}} \| vec\,(Y - AZ - GRF) \|^2_{Q_{YY}} \tag{43}$$

Both sets of constraints are now imposed: the matrix of ambiguities $\check{Z}$ is integer valued and the matrix $\check{R}$ belongs to the class of $3 \times q$ orthonormal matrices $\mathbb{O}^{3 \times q}$. The C-ILS solution $\check{Z}$ can be computed from the float solutions as (Teunissen & Kleusberg, 1998):

$$\check{Z} = \arg \min_{Z \in \mathbb{Z}^{mN \times n}} \underbrace{\left(\| vec(\hat{Z} - Z) \|^2_{Q_{\hat{Z}\hat{Z}}} + \| vec(\hat{R}(Z) - \check{R}(Z)) \|^2_{Q_{\hat{R}(Z)\hat{R}(Z)}} \right)}_{C(Z)} \tag{44}$$

with

$$\check{R}(Z) = \arg \min_{R \in \mathbb{O}^{3 \times q}} \| vec(\hat{R}(Z) - R) \|^2_{Q_{\hat{R}(Z)\hat{R}(Z)}} \tag{45}$$

The cost function $C(Z)$ is the sum of two terms. The first weighs the distance between a candidate integer matrix Z and the float solution $\hat{Z}$, weighted by the v-c matrix $Q_{\hat{Z}\hat{Z}}$. The second weighs the distance between the conditional (on the candidate Z) attitude matrix $\hat{R}(Z)$ and the orthonormal matrix $\check{R}(Z)$ that follows from the solution of (45). Therefore, the computation of cost function $C(Z)$ also involves a term that weighs the distance of the conditional attitude matrix from its orthogonal projection. This second term greatly aids the search for the correct ambiguities: integer candidates Z that produce matrices $\hat{R}(Z)$ too far from their orthonormal projection contribute to a much higher value of the cost function.

Since the minimization problem (44) is not solvable analytically due to the integer nature of the parameter involved, an extensive search in a subset of the space of integer matrices $\mathbb{Z}^{mN \times n}$ has to be performed. The definition of an efficient and fast solution scheme for problem (44) is not a trivial task. In order to highlight the intricacies of such formulation, we first give an approximate solution, obtained by neglecting the orthonormal constraint.

4.2.1 The LAMBDA method

Consider first the integer minimization problem (44) without the orthonormality constraint on R. Then the second term of $C(Z)$ reduces to zero and the integer minimization problem becomes

$$\check{Z}^u = \arg \min_{Z \in \mathbb{Z}^{mN \times n}} \| vec(\hat{Z} - Z) \|^2_{Q_{\hat{Z}\hat{Z}}} \tag{46}$$

This is the usual approach of doing GNSS integer ambiguity resolution. Due to the absence of the orthonormality constraint on R one may expect lower success rates, i.e., lower probability of identifying the correct ambiguity matrix Z. However, the ILS problem (46) is of lower complexity than (44) and a very fast implementation of it is available: the LAMBDA (Least-squares AMBiguity Decorrelation Adjustment) (Teunissen, 1995) method, see, e.g., Boon & Ambrosius (1997); Cox & Brading (2000); Huang et al. (2009); Ji et al. (2007); Kroes et al. (2005). It consists of two steps, namely decorrelation and search.

The integer minimizer has to be extensively searched within a subset of the whole space of integers:

$$\Omega^{u}\left(\chi^{2}\right) = \{Z \in \mathbb{Z}^{mN\times n} \mid \|vec(\hat{Z}-Z)\|^{2}_{Q_{\hat{Z}\hat{Z}}} \leq \chi^{2}\} \tag{47}$$

Ω^{u} is the so-called search space, a region of the space of integer matrices that contains only those candidates Z for which the squared norm (46) is bounded by the value χ^{2}. This can be set by choosing an integer matrix Z_c and taking $\chi^{2} = \|vec(\hat{Z}-Z_c)\|^{2}_{Q_{\hat{Z}\hat{Z}}}$. Rounding the float solution, $Z_c = [\hat{Z}]$, is an option, as well as bootstrapping an integer matrix, as in Teunissen (2000; 2007b).

Searching for the integer minimizer in Ω^{u} proves inefficient due to the weight matrix $Q_{\hat{Z}\hat{Z}}$. Geometrically, the search space defines a hyperellipsoid centered in $\hat{Z}$ and whose shape and orientation are driven by the entries of matrix $Q_{\hat{Z}\hat{Z}}$. The difficulty of the search lies in the fact that the search space is highly elongated, as detailed in Teunissen & Kleusberg (1998). The reason is that the ambiguities are highly correlated. While the set wherein the independent ambiguities (e.g., three ambiguities for a single baseline scenario) can be chosen is rather large, the set of admissible values for the remaining ambiguities is very small. This causes major halting problems during the search, since many times the selected subset of independent ambiguities does not yield admissible integer matrix candidates. This issue is tackled and solved in the LAMBDA method with a decorrelation step. The decorrelation of matrix $Q_{\hat{Z}\hat{Z}}$ is achieved by an admissible transformation matrix T. In order to preserve the integerness, such matrix has to fulfill the following two conditions: T as well as its inverse T^{-1} need to have integer entries. The matrix of transformed ambiguities Z' and corresponding v-c matrix are then obtained as

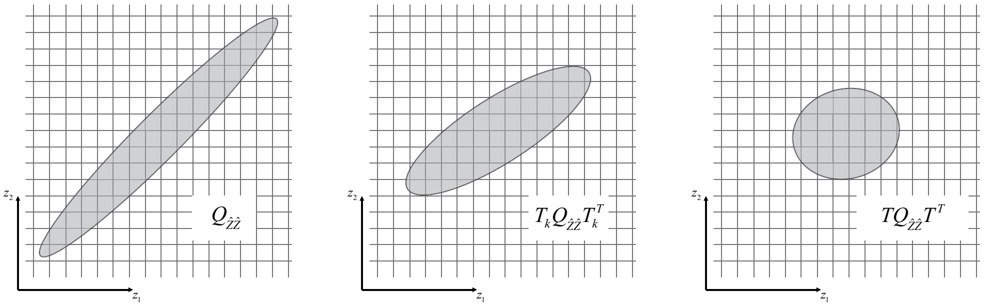
$$Z' = TZ \quad ; \quad Q_{\hat{Z}'\hat{Z}'} = TQ_{\hat{Z}\hat{Z}}T^{T} \tag{48}$$

The decorrelation procedure is described in Teunissen & Kleusberg (1998). The v-c matrix is iteratively decorrelated by a sequence of admissible transformations T_i, until matrix

$$Q_{\hat{Z}'\hat{Z}'} = \left(\prod_{i} T_i\right) Q_{\hat{Z}\hat{Z}} \left(\prod_{i} T_i\right)^{T} = TQ_{\hat{Z}\hat{Z}}T^{T} \tag{49}$$

cannot be further decorrelated. Note that due to the integer conditions on T, a full decorrelation cannot generally be achieved. Figure 6 shows three steps of the decorrelation process for a two-dimensional example. Figure 6a shows the original (elongated) ellipse associated to $Q_{\hat{Z}\hat{Z}}$, Figure 6b shows an intermediate decorrelation step, and Figure 6c shows the final decorrelated search space.

After the decorrelation step, the actual search is performed by operating the LDL^{T} factorization of matrix $Q_{\hat{Z}'\hat{Z}'}$, so that the quadratic form in (46) can be written as a



(a) Original ellipse, defined by $Q_{\hat{Z}\hat{Z}}$.

(b) Intermediate decorrelated ellipse, defined by $T_k Q_{\hat{Z}\hat{Z}} T_k^T$.

(c) Final decorrelated ellipse, defined by $T Q_{\hat{Z}\hat{Z}} T^T$.

Fig. 6. Initial, intermediate and decorrelated search space defined by the (transformed) v-c matrix of the ambiguities.

summation:

$$\left\|vec(\hat{Z}' - Z')\right\|^2_{Q_{\hat{Z}'\hat{Z}'}} = \left\|vec(\hat{Z}' - Z')\right\|^2_{LDL^T} = \sum_{i=1}^{mNn} \frac{\left(\hat{z}'_{i|I} - z'_i\right)^2}{\sigma^2_{i|I}} \leq \chi^2 \tag{50}$$

where the scalars $\hat{z}'_{i|I}$ and $\sigma^2_{i|I}$ are the conditional float ambiguity estimator and its corresponding conditional variance, respectively. These are conditioned to the previous $I = 1, \ldots, i-1$ values, and directly follow from the entries of matrices L and D. More details on the way the search is actually performed can be found in de Jonge & Tiberius (1996).

Due to the decorrelation step, the extensive search for the integer minimizer $\check{Z}^u$ is performed quickly and efficiently, making the LAMBDA method perfectly suitable for real-time applications.

4.2.2 The MC-LAMBDA method

The MC-LAMBDA method is an extension of the LAMBDA method that applies to the geometrically-constrained problem (44). The MC-LAMBDA method shares the same working principle of the LAMBDA method: first the search space is decorrelated, then the search for the integer minimizer is performed. However, an extensive search within a (decorrelated) search space is generally not efficient as it is with the LAMBDA method, as explained in the following.

The search space is now defined as

$$\Omega^c\left(\chi^2\right) = \{Z \in \mathbb{Z}^{mN \times n} \mid C(Z) \leq \chi^2\} \tag{51}$$

The cost function $C(Z)$ takes, for the same candidate Z, much larger values than the first quadratic term in (44), due to the matrix $Q_{\hat{R}(Z)\hat{R}(Z)}$, whose inverse has entries two orders of magnitude larger than the entries of $Q^{-1}_{\hat{Z}\hat{Z}}$ (Giorgi et al., 2011; Teunissen, 2007a). For this reason it is not trivial to set a proper value of χ^2, since the cost function $C(Z)$ is highly sensitive to the choice of Z (Giorgi, 2011; Giorgi et al., 2011). This problem becomes more marked for weaker models (single frequency, low number of satellites tracked, high noise levels). Obviously, larger values of χ^2 imply longer computational times due to the larger number of candidates to be evaluated. Also, the constrained least-squares problem (45) has to be solved for each of the integer candidates in $\Omega^C\left(\chi^2\right)$, thereby further increasing the computational load.

The aforementioned issues are solved with a novel numerical efficient search scheme for the solution of (44). This is achieved by employing easier-to-evaluate bounding functions and introducing new search algorithms.

First, consider two functions, $C_1(Z)$ and $C_2(Z)$, that satisfy the following inequalities:

$$C_1(Z) \leq C(Z) \leq C_2(Z) \tag{52}$$

These functions provide a lower and an upper bound for the cost function $C(Z)$. The choice for these bounding functions is driven by two requirements: their evaluation should be less time consuming than the evaluation of $C(Z)$, and each bound should be sufficiently tight. Several alternatives have been studied in (Giorgi, 2011; Giorgi et al., 2012; Nadarajah et al., 2011; Teunissen, 2007a;c), based on

- the eigenvalues of matrix $Q^{-1}_{\hat{R}(Z)\hat{R}(Z)}$
- the analytical solution of Wahba's problem (Wahba, 1965)
- a tighter geometrical bound based on Procustes problem (Schonemann, 1966)
- a QR factorization (Gram-Schmidt process)

For example, the first method listed exploits the inequalities $\xi_m \|\cdot\|^2_I \leq \|\cdot\|^2_Q \leq \xi_M \|\cdot\|^2_I$, with ξ_m and ξ_M the smallest and largest eigenvalues of $Q^{-1}_{\hat{R}(Z)\hat{R}(Z)}$, respectively. After some manipulation, the two bounding functions read

$$\begin{aligned} C_1(Z) &= \left\|vec(\hat{Z}-Z)\right\|^2_{Q_{\hat{Z}\hat{Z}}} + \xi_m \sum_{i=1}^{q} \left(\|\hat{r}_i(Z)\| - 1\right)^2 \\ C_2(Z) &= \left\|vec(\hat{Z}-Z)\right\|^2_{Q_{\hat{Z}\hat{Z}}} + \xi_M \sum_{i=1}^{q} \left(\|\hat{r}_i(Z)\| + 1\right)^2 \end{aligned} \tag{53}$$

where $\hat{r}_i(Z)$ are the column vectors of $\hat{R}(Z)$.

Two efficient search methods have been developed to reduce the computational burden associated to an extensive search. Independently from the bounding functions used, these novel search schemes allow for a quick minimization of $C(Z)$.

Consider first the lower bound $C_1(Z)$. The search space associated to $C_1(Z)$ is

$$\Omega_1\left(\chi^2\right) = \{Z \in \mathbb{Z}^{mN\times n} \mid C_1(Z) \leq \chi^2\} \supset \Omega^C\left(\chi^2\right) \tag{54}$$

Obviously, the search space $\Omega^C\left(\chi^2\right)$ is contained within $\Omega_1\left(\chi^2\right)$. One may proceed, for example, by choosing $\chi^2 = \left\|vec(\hat{Z}-Z')\right\|^2_{Q_{\hat{Z}\hat{Z}}}$ with Z' a given integer matrix (both rounding the float solution and bootstrapping an integer matrix are viable choices). Then, we can enumerate all the integers matrices contained in $\Omega_1\left(\chi^2\right)$ and compute $C(Z)$ for each candidate (if any, since set $\Omega_1\left(\chi^2\right)$ may also turn out empty), in order to also evaluate $\Omega^c\left(\chi^2\right)$. If this set turns out non-empty, then one has simply to extract the minimizer $\check{Z}$ by sorting the integer matrices according to the values of $C(Z)$. However, there is no guarantee that $\Omega^c\left(\chi^2\right)$ is non-empty. If the search space $\Omega^c\left(\chi^2\right)$ is empty, the size of $\Omega_1\left(\chi^2\right)$ is increased and the process repeated iteratively until the minimizer $\check{Z}$ is found. This search scheme, illustrated with the flow chart in Figure 7, is named *Expansion* approach, since the size of the search space is iteratively 'expanded'.

An alternative approach is devised by considering the upper bound $C_2(Z)$. Its search space is

$$\Omega_2\left(\chi^2\right) = \{Z \in \mathbb{Z}^{mN \times n} \mid C_2(Z) \leq \chi^2\} \subset \Omega^c\left(\chi^2\right) \tag{55}$$

which is contained in the set $\Omega^c\left(\chi^2\right)$. Consider the following iterative procedure. First, the scalar χ^2 is set such that it guarantees the non-emptiness of $\Omega_2\left(\chi^2\right)$, and therefore $\Omega^c\left(\chi^2\right)$ is non-empty either. This can be done by choosing $\chi^2 = C_2(Z')$ for an integer matrix Z', which can be the rounded float solution, a bootstrapped solution, or an integer matrix obtained by other means (see for further options Giorgi et al. (2008)). Then, the search proceeds by looking for an integer candidate in the set $\Omega_2\left(\chi^2\right)$, aiming to find a matrix Z_1 that provides a smaller value for the upper bound $C_2(Z_1) = \chi_1^2 < \chi^2$. When it is found, the set is shrunk to $\Omega_2\left(\chi_1^2\right)$ and the search continues by looking for another integer candidate Z_2 capable of reducing the value $C_2(Z_2) = \chi_2^2 < \chi_1^2$. This process is repeated until the minimizer of $C_2(Z)$, say $\check{Z}_2$, is found. Since this may differ from the minimizer of $C(Z)$, the search space $\Omega^c\left(\overline{\chi}^2\right)$, with $\overline{\chi}^2 = C_2(\check{Z}_2)$, is evaluated and the sought-for integer minimizer $\check{Z}$ extracted. This iterative search scheme is named *Search and Shrink* approach, and it is detailed in the flow chart of Figure 8.

Both the *Expansion* and the *Search and Shrink* approaches implement the search for integer minimizer (44) in a fast and efficient way, such that the algorithm can be used for real-time applications.

The MC-LAMBDA method achieves very high success rates. The success rate is defined as the probability of providing the correct set of integer ambiguities. The inclusion of geometrical constraints, which follow from the a priori knowledge of the antennas relative positions aboard the aircraft, largely aids the ambiguity resolution process, allowing for higher success rates in weaker models, such as with the single-frequency and/or high measurement noise scenarios. These performance improvements associated to the MC-LAMBDA method with respect to classical methods (such as the LAMBDA) are analyzed in the following section with actual data collected during two different flights tests.

5. Flight test results

The performance of the MC-LAMBDA method is analyzed with data collected on two flight-tests performed with a Cessna Citation jet aircraft. The aircraft attitude is extracted from unaided, single-epoch, single-frequency ($N = 1$) GNSS observations, in order to demonstrate the method capabilities in the most challenging scenario, i.e., stand-alone, high observation noise and low measurements redundancy. Also, single-epoch performance is extremely important for dynamic platforms, where a quick recovery from changes of tracked satellites, cycle slips and losses of lock is necessary to avoid undesired loss of guidance. The

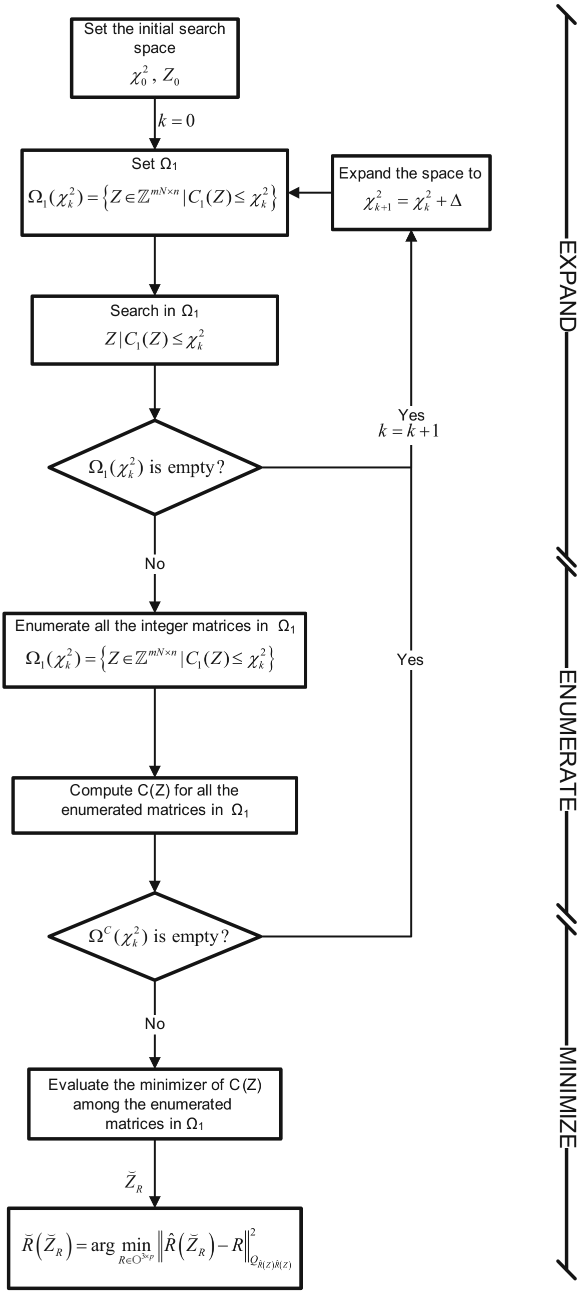

Set the initial search space
χ_0^2, Z_0
$k=0$
Set Ω1
$\Omega_1(\chi_k^2)=\{Z\in\mathbb{Z}^{mN\times n}\,|\,C_1(Z)\leq\chi_k^2\}$
Expand the space to
$\chi_{k+1}^2=\chi_k^2+\Delta$
Search in Ω1
$Z\,|\,C_1(Z)\leq\chi_k^2$
Yes
$k=k+1$
$\Omega_1(\chi_k^2)$ is empty?
No
Enumerate all the integer matrices in Ω1
$\Omega_1(\chi_k^2)=\{Z\in\mathbb{Z}^{mN\times n}\,|\,C_1(Z)\leq\chi_k^2\}$
Yes
Compute C(Z) for all the enumerated matrices in Ω1
$\Omega^C(\chi_k^2)$ is empty?
No
Evaluate the minimizer of C(Z) among the enumerated matrices in Ω1
$\breve{Z}_R$
$\breve{R}\left(\breve{Z}_R\right)=\arg\min_{R\in O^{3\times p}}\left\|\hat{R}\left(\breve{Z}_R\right)-R\right\|^2_{Q_{\hat{R}(Z)\hat{R}(Z)}}$
EXPAND
ENUMERATE
MINIMIZE

Fig. 7. The *Expansion* approach: flow chart.

single-frequency case is of interest for many aerospace applications, where limits on weight and power consumption must often be respected.

In both tests the same receiver (Septentrio PolaRx2@) was connected to three antennas, placed on the middle of the fuselage, on the wing and on the nose (see Figure 9). In the first

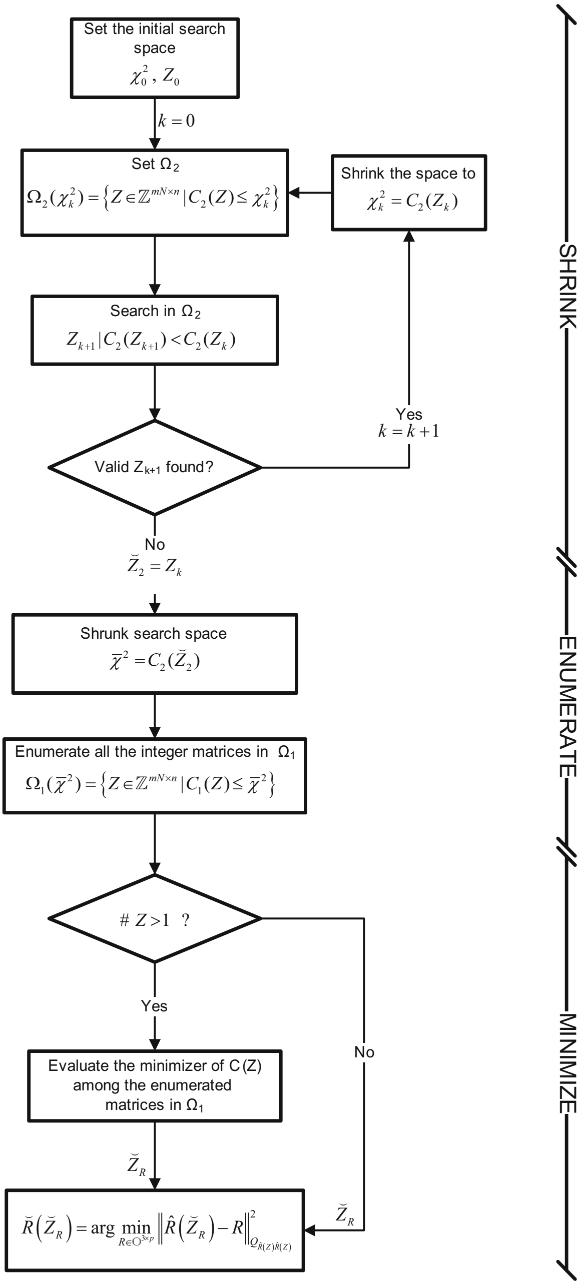

Set the initial search space
χ_0^2 , Z_0
$k=0$
Set Ω2
$\Omega_2(\chi_k^2)=\{Z\in\mathbb{Z}^{mN\times n}|C_2(Z)\le\chi_k^2\}$
Shrink the space to
$\chi_k^2=C_2(Z_k)$
Search in Ω2
$Z_{k+1}|C_2(Z_{k+1})<C_2(Z_k)$
Yes
$k=k+1$
Valid Z_{k+1} found?
No
$\breve{Z}_2=Z_k$
Shrunk search space
$\bar{\chi}^2=C_2(\breve{Z}_2)$
Enumerate all the integer matrices in Ω1
$\Omega_1(\bar{\chi}^2)=\{Z\in\mathbb{Z}^{mN\times n}|C_1(Z)\le\bar{\chi}^2\}$
# Z>1 ?
Yes
No
Evaluate the minimizer of C(Z) among the enumerated matrices in Ω1
$\breve{Z}_R$
$\breve{Z}_R$
$\breve{R}(\breve{Z}_R)=\arg\min_{R\in\mathbb{O}^{3\times p}}\|\hat{R}(\breve{Z}_R)-R\|^2_{Q_{\hat{R}(Z)\hat{R}(Z)}}$
SHRINK
ENUMERATE
MINIMIZE

Fig. 8. The *Search and Shrink* approach: flow chart.

test analyzed ($T-I$) the nose antenna was placed on the extremity of a boom, whereas in second test ($T-II$) it was directly placed on the aircraft body. The two tests largely differ by the flight dynamic. Test $T-I$ was conducted with aggressive maneuvering and few zero-gravity parabolas, whereas $T-II$ was performed as part of a gravimetry campaign, with very few smooth maneuvers, as shown in Figure 10. During test $T-II$, the aircraft

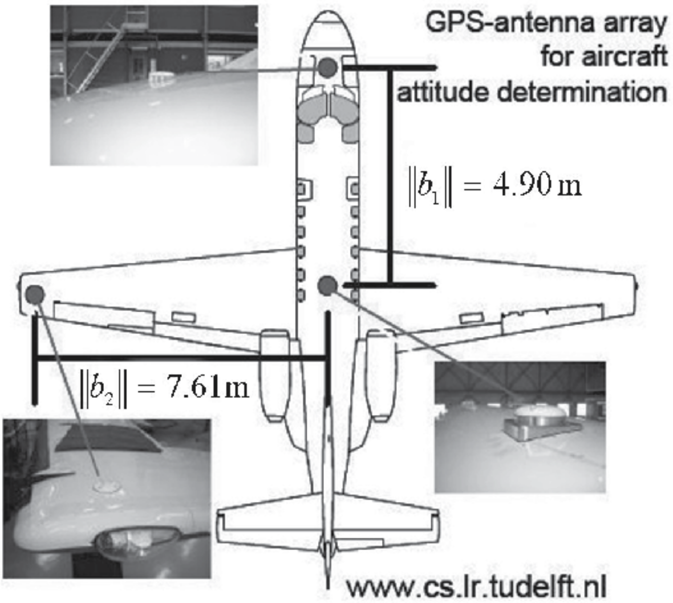


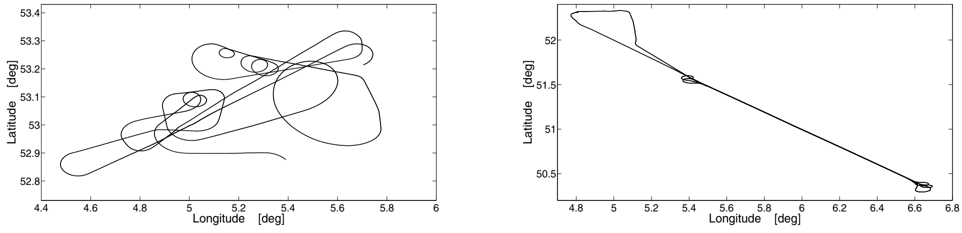
Fig. 9. The antennas set-up onboard the Cessna Citation II.

was also equipped with an Inertial Navigation System (INS), whose output is used to test the GNSS-based attitude estimation accuracy. Figure 11 reports the number of tracked satellites for the duration of the two tests. The PDOP (Precision Dilution of Precision) is also shown.

The matrix of local body-frame baseline coordinates for the two tests are

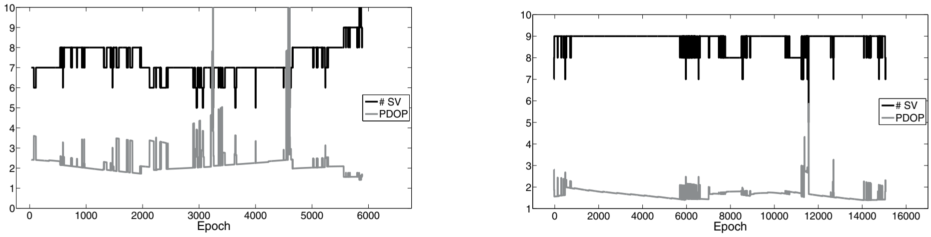
$$F_{T-I} = \begin{bmatrix} 5.45 & -0.34 \\ 0 & 7.60 \end{bmatrix} \quad [\text{m}] \qquad F_{T-II} = \begin{bmatrix} 4.90 & -0.39 \\ 0 & 7.60 \end{bmatrix} \quad [\text{m}] \tag{56}$$

The receiver collected GPS-L1 data for about 6000 epochs (zero cut-off angle 1Hz sampling), between 11:42 and 13:20 UTC, 2nd June 2005 on the first test, and 15000 epochs (zero cut-off angle, 1 Hz sampling), between 11:00 and 14:23 UTC, 1st November 2007 on the second test.



(a) $T-I$ (b) $T-II$

Fig. 10. Ground traces of the two test flights.



(a) $T - I$ (b) $T - II$

Fig. 11. Number of satellites tracked and corresponding PDOP values.

5.1 Instantaneous ambiguity resolution

The success rate marked by the LAMBDA and MC-LAMBDA methods applied to both flight tests is reported in Table 1 (Giorgi et al., 2011). The single-epoch performance of the unconstrained method are rather unsatisfactory. The correct set of integer ambiguities is resolved only for 5.8% of time in test $T - I$ and in 24.7% of time in test $T - II$. The difference is due to the higher number of satellites tracked in the second test.

$T - I$		$T - II$	
LAMBDA	MC-LAMBDA	LAMBDA	MC-LAMBDA
5.8	81.5	24.7	88.1

Table 1. $T - I$ and $T - II$ tests: unaided single-epoch, single-frequency success rate (%) for the LAMBDA and the MC-LAMBDA methods, two-baseline processing.

Instead, application of the MC-LAMBDA method yields a strong performance improvement. The constrained method is capable of providing the correct integer solution for more than 80% of the epochs in test $T - I$ and more than 88% in test $T - II$.

Both airborne tests confirm the very large improvement that is obtained by strengthening the underlying model with the inclusion of geometrical constraints. It is stressed that all the ambiguity resolution performance reported are obtained by processing the GNSS signals without any a priori information or assumption about the attitude or the aircraft motion. Also, mask angles, elevation-dependent models, dynamic models or any kind of filtering are not applied.

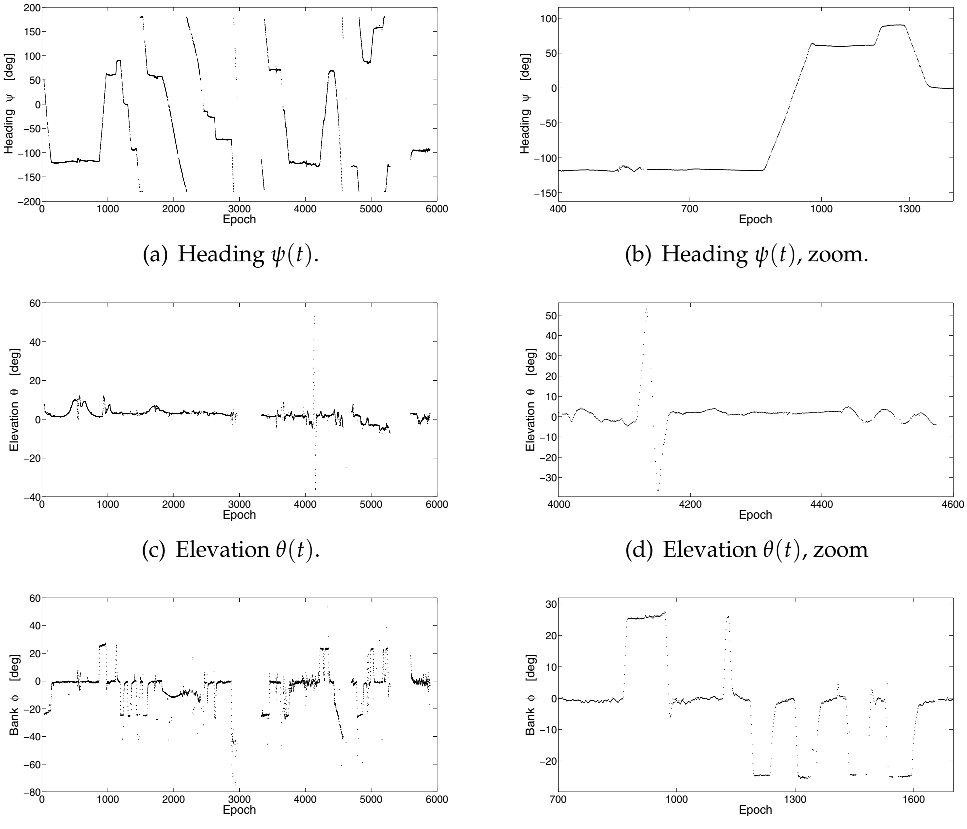
5.2 Attitude determination

Heading	σ_ψ	[deg]	0.07
Elevation	σ_θ	[deg]	0.20
Bank	σ_ϕ	[deg]	0.12

Table 2. $T - II$ test: standard deviations of the differences between GPS and INS attitude angles output.

High single-epoch success rates yield precise epoch-by-epoch attitude solutions for the larger part of the flights duration. The attitude angles based on the correctly fixed integer ambiguities in test $T - I$ are shown in Figure 12. The high dynamics of the flight is evident from the steep variations of the attitude. In particular, Figure 13 shows a zero-gravity maneuver: the aircraft promptly pitched up, gained some altitude, and performed an ample arc to create a virtual absence of gravity on board.

Figure 14 shows the GNSS-based attitude angles for the test $T - II$. The INS solutions are also reported in the figures, in order to provide a comparison between the two systems. Table 2 reports the standard deviations of the differences between the INS and GNSS-based attitude estimations. Taking the precise INS output as benchmark solution, it can be inferred that the accuracy obtained is within the expected range, given the baseline lengths employed. The heading angle is estimated with the highest precision, whereas the elevation estimation is characterized by the highest noise levels. This is due to the relative geometry of the antennas and to the fact that the vertical components of the GNSS-based baseline estimations are inherently less accurate that the horizontal components. The bank angle is estimated with higher precision than the elevation angle, being driven by the longer baseline $Body - Wing$.



(a) Heading $\psi(t)$.

(b) Heading $\psi(t)$, zoom.

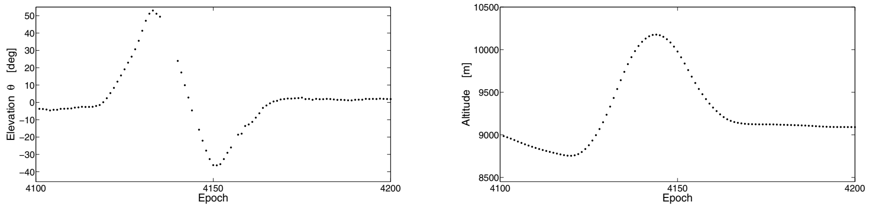
(c) Elevation $\theta(t)$.

(d) Elevation $\theta(t)$, zoom

(e) Bank $\phi(t)$.

(f) Bank $\phi(t)$, zoom

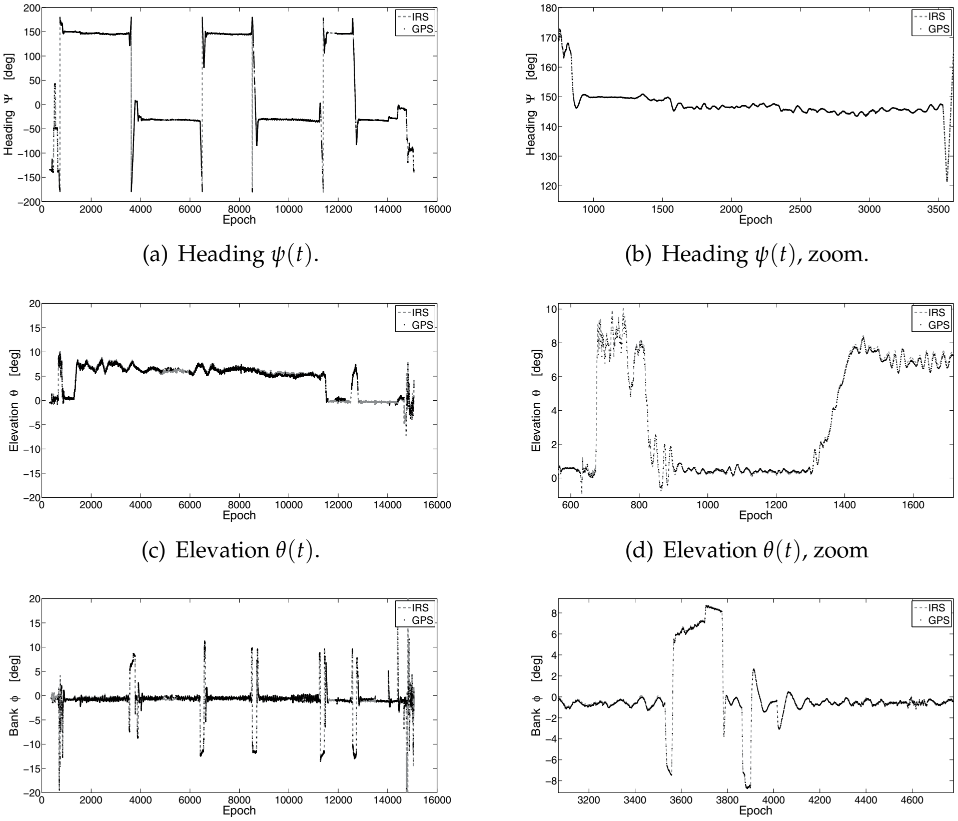
Fig. 12. $T - I$ test: time series of the three attitude angles as estimated via GNSS. On the right, a closer look at the estimates.



(a) Elevation $\theta(t)$.

(b) Altitude profile during the zero-gravity maneuver.

Fig. 13. $T - I$ test: zero-gravity maneuver.



(a) Heading $\psi(t)$.

(b) Heading $\psi(t)$, zoom.

(c) Elevation $\theta(t)$.

(d) Elevation $\theta(t)$, zoom

(e) Bank $\phi(t)$.

(f) Bank $\phi(t)$, zoom

Fig. 14. $T - II$ test: time series of the three attitude angles as estimated via GNSS and provided by the INS. On the right, a closer look at the estimates.

6. Summary and conclusions

Ambiguity resolution can be effectively enhanced by means of a rigorous formulation of the ambiguity-attitude estimation problem. In order to infer the aircraft's orientation from the GNSS antenna positions, each antenna location on the aircraft body has to be precisely known. This geometrical information can be embedded in the ambiguity resolution step, thus strengthening the underlying functional model - i.e., additional information is added to the functional model - and enhancing the whole estimation process. The higher ambiguity resolution performance comes at the cost of an increased computational complexity. In order to overcome the issue, a number of solutions are presented, which allow for fast and reliable solutions without requiring extensive computational loads. A fast implementation of the geometrically constrained problem is obtained by modifying a well-known method for ambiguity resolution: the LAMBDA (Least-squares AMBiguity Decorrelation Adjustment) method. This method is nowadays the standard for carrier-phase based applications, and it is being implemented in a number of receivers employed for high-precision navigation applications. The complexity of the constrained estimation method requires the development of novel strategies to extract the solution in a timely manner. This is achieved by properly modifying the LAMBDA method to address the specific ambiguity-attitude estimation problem: the Multivariate Constrained (MC)-LAMBDA method. Through the use of two novel search schemes the sought-for set of carrier phase ambiguities can be efficiently estimated.

The method is tested on actual data collected on two different flight tests. Each test indicates the feasibility of employing GNSS as attitude sensor, an application that might be increasingly adopted in the aviation industry, either stand-alone for non-critical applications, or in combinations with other sensors for safety-critical applications.

7. Acknowledgment

The second author is the recipient of an Australian Research Council Federation Fellowship (project number FF0883188). This support is gratefully acknowledged.

8. References

Boon, F. & Ambrosius, B. A. C. (1997). Results of Real-Time Applications of the LAMBDA Method in GPS Based Aircraft Landings, *Proceedings KIS97* pp. 339–345.

Cheng, Y. & Shuster, M. D. (2007). Robustness and Accuracy of the QUEST Algorithm, *Advances in the Astronautical Sciences* 127: 41–61.

Cox, D. B. & Brading, J. D. (2000). Integration of LAMBDA Ambiguity Resolution with Kalman Filter for Relative Navigation of Spacecraft, *NAVIGATION* 47(3): 205–210.

Davenport, P. B. (1968). A Vector Approach to the algebra of Rotations with Applications, *NASA Technical Note D-4696, Goddard Space Flight Center* .

de Jonge, P. & Tiberius, C. (1996). The LAMBDA Method for Integer Ambiguity Estimation: Implementation Aspects, *LGR Series 12, Publications of the Delft Geodetic Computing Centre, Delft, The Netherlands* .

Giorgi, G. (2011). GNSS Carrier Phase-based Attitude Determination. Estimation and applications., *PhD dissertation, Delft University of Technology, Delft, The Netherlands* .

Giorgi, G., Teunissen, P. J. G. & Buist, P. J. (2008). A Search and Shrink Approach for the Baseline Constrained LAMBDA: Experimental Results, *Proceedings of the International Symposium on GPS/GNSS 2008. A. Yasuda (Ed.), Tokyo University of Marine Science and Technology* pp. 797–806.

Giorgi, G., Teunissen, P. J. G., Verhagen, S. & Buist, P. J. (2011). Instantaneous Ambiguity Resolution in GNSS-based Attitude Determination Applications: the MC-LAMBDA method, *Journal of Guidance, Control, and Dynamics, to be published* .

Giorgi, G., Teunissen, P. J. G., Verhagen, S. & Buist, P. J. (2012). Integer Ambiguity Resolution with Nonlinear Geometrical Constraints., *N. Sneeuw et al. (eds.), VII Hotine-Marussi Symposium on Mathematical Geodesy, International Association of Geodesy Symposia 137, Springer-Verlag* .

Huang, S. Q., Wang, J. X., Wang, X. Y. & Chen, J. P. (2009). The Application of the LAMBDA Method in the Estimation of the GPS Slant Wet Vapour, *Acta Aeronautica et Astronautica Sinica* 50(1): 60–68.

Ji, S., Chen, W., Zhao, C., Ding, X. & Chen, Y. (2007). Single Epoch Ambiguity Resolution for Galileo with the CAR and LAMBDA Methods, *GPS Solutions* 11(4): 259–268.

Kroes, R., Montenbruck, O., Bertiger, W. & Visser, P. (2005). Precise GRACE Baseline Determination Using GPS, *GPS Solutions* 9(1): 21–31.

Markley, F. L. & Landis, F. (1993). Attitude Determination Using Vector Observations: a Fast Optimal Matrix Algorithm, *The Journal of the Astronautical Sciences* 41(2): 261–280.

Markley, F. L. & Mortari, D. (1999). How to Estimate Attitude from Vector Observations, *Presented at AAS/AIAA Astrodynamics Specialist Conference, Paper 99-427* .

Markley, F. L. & Mortari, D. (2000). Quaternion Attitude Estimation Using Vector Observations, *The Journal of the Astronautical Sciences* 48(2-3): 359–380.

Misra, P. & Enge, P. (2001). *Global Positioning System: Signals, Measurements, and Performance*, 2nd edn, Ganga-Jamuna Press, Lincoln MA.

Mortari, D. (1997). ESOQ: A Closed-form Solution to the Wahba Problem, *The Journal of the Astronautical Sciences* 45(2): 195–204.

Mortari, D. (2000). Second Estimator of the Optimal Quaternion, *Journal of Guidance, Control, and Dynamics* 23(5): 885–888.

Nadarajah, N., Teunissen, P. J. G. & Giorgi, G. (2011). Instantaneous GNSS Attitude Determination for Remote Sensing Platforms, *Presented at the XXV International Union of Geodesy and Geophysics General Assembly (IUGG), Melbourne, Australia* .

Schonemann, P. H. (1966). A Generalized Solution of the Orthogonal Procrustes Problem, *Psychometrika* 31(1): 1–10.

Shuster, M. D. (1978). Approximate Algorithms for Fast Optimal attitude Computation, *Proceedings of the AIAA Guidance and Control conference, Palo Alto, CA, US* pp. 88–95.

Shuster, M. D. (1993). A Survey of Attitude Representations , *The Journal of the Astronautical Sciences* 41(4): 439–517.

Shuster, M. D. & Oh, S. D. (1981). Three-Axis Attitude Determination from Vector Observations, *Journal of Guidance and Control* 4(1): 70–77.

Teunissen, P. J. G. (1995). The Least-Squares Ambiguity Decorrelation Adjustment: a Method for Fast GPS Integer Ambiguity Estimation, *Journal of Geodesy* 70(1-2): 65–82.

Teunissen, P. J. G. (2000). The Success Rate and Precision of GPS Ambiguities, *Journal of Geodesy* 74(3): 321–326.

Teunissen, P. J. G. (2007a). A General Multivariate Formulation of the Multi-Antenna GNSS Attitude Determination Problem, *Artificial Satellites* 42(2): 97–111.

Teunissen, P. J. G. (2007b). Influence of Ambiguity Precision on the Success Rate of GNSS Integer Ambiguity Bootstrapping, *Journal of Geodesy, Springer* 81(5): 351–358.

Teunissen, P. J. G. (2007c). The LAMBDA Method for the GNSS Compass, *Artificial Satellites* 41(3): 89–103.

Teunissen, P. J. G. (2011). A-PPP: Array-aided Precise Point Positioning with Global Navigation Satellite Systems, *IEEE Transactions on Signal Processing (submitted for publication)* pp. 1–12.

Teunissen, P. J. G. & Kleusberg, A. (1998). GPS for Geodesy, *Springer, Berlin Heidelberg New York* .

Van Loan, C. F. (2000). The Ubiquitous Kronecker Product, *Journal of Computational and Applied Mathematics* 123: 85–100.

Wahba, G. (1965). Problem 65-1: A Least Squares Estimate of Spacecraft Attitude, *SIAM Review* 7(3): 384–386.

10

A Variational Approach to the Fuel Optimal Control Problem for UAV Formations

Andrea L'Afflitto and Wassim M. Haddad
Georgia Institute of Technology
USA

1. Introduction

The pivotal role of unmanned aerial vehicles (UAVs) in modern aircraft technology is evidenced by the large number of civil and military applications they are employed in. For example, UAVs successfully serve as platforms carrying payloads aimed at land monitoring (Ramage et al., 2009), wildfire detection and management (Ambrosia & Hinkley, 2008), law enforcement (Haddal & Gertler, 2010), pollution monitoring (Oyekan & Huosheng, 2009), and communication broadcast relay (Majewski, 1999), to name just a few.

A formation of UAVs, defined by a set of vehicles whose states are coupled through a common control law (Scharf et al., 2003b), is often more valuable than a single aircraft because it can accomplish several tasks concurrently. In particular, UAV formations can guarantee higher flexibility and redundancy, as well as increased capability of distributed payloads (Scharf et al., 2003a). For example, an aircraft formation can successfully intercept a vehicle which is faster than its chasers (Jang & Tomlin, 2005). Alternatively, a UAV formation equipped with interferometic synthetic aperture radar (In-SAR) antennas can pursue both along-track and cross-track interferometry, which allow harvesting information that a single radar cannot detect otherwise (Lillesand et al., 2007).

Path planning is one of the main problems when designing missions involving multiple vehicles; a UAV formation typically needs to accomplish diverse tasks while meeting some assigned constraints. For example, a UAV formation may need to intercept given targets while its members maintain an assigned relative attitude. Trajectories should also be optimized with respect to some performance measure capturing minimum time or minimum fuel expenditure. In particular, trajectory optimization is critical for mini and micro UAVs (μUAVs) because they often operate independently from remote human controllers for extended periods of time (Shanmugavel et al., 2010) and also because of limited amount of available energy sources (Plnes & Bohorquez, 2006).

The scope of the present paper is to provide a rigorous and sufficiently broad formulation of the optimal path planning problem for UAV formations, modeled as a system of n 6-degrees of freedom (DoF) rigid bodies subject to a constant gravitational acceleration and aerodynamic forces and moments. Specifically, system trajectories are optimized in terms of control effort, that is, we design a control law that minimizes the forces and moments needed to operate a UAV formation, while meeting all the mission objectives. Minimizing the control effort is equivalent to minimizing the formation‘s fuel consumption in the case of vehicles equipped with conventional fuel-based propulsion systems (Schouwenaars et al., 2006) and is a suitable indicator of the energy consumption for vehicles powered by batteries or other power sources.

In this paper, we derive an optimal control law which is independent of the size of the formation, the system constraints, and the environmental model adopted, and hence, our framework applies to aircraft, spacecraft, autonomous marine vehicles, and robot formations. The direction and magnitude of the optimal control forces and moments is a function of the dynamics of two vectors, namely the translational and rotational primer vectors. In general, finding the dynamics of these two vectors over a given time interval is a demanding task that does not allow for an analytical closed-form solution, and hence, a numerical approach is required. Our main result involves necessary conditions for optimality of the formations' trajectories.

The contents of this paper are as follows. In Section 2, we present notation and definitions of the physical variables needed to formulate the fuel optimization problem. Section 3 gives a problem statement of the UAV path planning optimization problem, whereas Section 4 provides the necessary mathematical background for this problem. Next, in Section 5, we survey the relevant literature and highlight the advantages related to the proposed approach. Section 6 discusses results achieved by applying the theoretical framework developed in Section 4. In Section 7, we present an illustrative numerical example that highlights the efficacy of the proposed approach. Finally, in Section 8, we draw conclusions and highlight future research directions.

2. Notation and definitions

The notation used in this paper is fairly standard. When a word is defined in the text, the concept defined is *italicized* and it should be understood as an "if and only if" statement. Mathematical definitions are introduced by the symbol "$\triangleq$." The symbol $\mathbb{N}$ denotes the set of positive integers, $\mathbb{R}$ denotes the set of real numbers, $\overline{\mathbb{R}}_+$ denotes the set of nonnegative real numbers, $\mathbb{R}^n$ denotes the set of $n \times 1$ column vectors on the field of real numbers, and $\mathbb{R}^{n\times m}$ denotes the set of real $n \times m$ matrices. Both natural and real numbers are denoted by lower case letters, e.g., $j \in \mathbb{N}$ and $a \in \mathbb{R}$, vectors are denoted by bold lower case letters, e.g., $\mathbf{x} \in \mathbb{R}^n$, and matrices are denoted by bold upper case letters, e.g., $\mathbf{A} \in \mathbb{R}^{n\times m}$. Subsets of $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ are denoted by italicized upper case letters, e.g., $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^{n\times m}$. The interior of the set A is denoted by $\mathrm{int}(A)$. The zero vector in $\mathbb{R}^n$ is denoted by $\mathbf{0}_n$, the zero matrix in $\mathbb{R}^{n\times m}$ is denoted by $\mathbf{0}_{n\times m}$, and the identity matrix in $\mathbb{R}^{n\times n}$ is denoted by $\mathbf{I}_n$.

For $\mathbf{x} \in \mathbb{R}^n$ we write $\mathbf{x} \geq\geq \mathbf{0}_n$ (respectively, $\mathbf{x} >> \mathbf{0}_n$) to indicate that every component of $\mathbf{x}$ is nonnegative (respectively, positive). We write $||\cdot||_p$ for the p-norm of a vector and its corresponding equi-induced matrix norm, e.g., $||\mathbf{x}||_p$ and $||\mathbf{A}||_p$. The transpose of a vector or of a matrix is denoted by the superscript $(\cdot)^T$, e.g., $\mathbf{x}^T$ and $\mathbf{A}^T$. The cross product between two vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathbf{a} \wedge \mathbf{b}$. Given $\mathbf{x} \in \mathbb{R}^3$ such that $\mathbf{x} \triangleq [x_1, x_2, x_3]^T$, we define

$$\mathbf{x}^\times \triangleq \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}.$$

The inverse of a square matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$, the transpose of $\mathbf{A}^{-1}$ is denoted by $\mathbf{A}^{-\mathrm{T}}$, the determinant of $\mathbf{A}$ is denoted by $\det(\mathbf{A})$, the diagonal of $\mathbf{A}$ is denoted by $\mathrm{diag}(\mathbf{A})$, and the nullspace of a matrix $\mathbf{A}$ is denoted by $\mathcal{N}(\mathbf{A})$.

Functions are always introduced by specifying their domain and codomain, e.g., $\mathbf{h} : A_1 \times A_2 \to B$. The arguments of a function will not be indicated in the text unless necessary, e.g., $\mathbf{h}(\mathbf{x}, \mathbf{y})$ is simply denoted by $\mathbf{h}$. If a function is dependent on some unspecified variables, then its arguments will be replaced by dots, e.g., $\mathbf{h}(\cdot, \cdot)$. The same convention is used for functionals; however, their arguments are embraced by square brackets, i.e., $\mathrm{J}\left[\mathbf{x}, \mathbf{y}\right]$.

The first derivative with respect to time of a differentiable function $\mathbf{q} : [t_1, t_2] \to \mathbb{R}^{\mathrm{n}}$ is denoted by the a dot on top of the function, e.g., $\dot{\mathbf{q}}(t)$. Given $\mathbf{g} : A \to \mathbb{R}^{\mathrm{m}}$, where $A \subset \mathbb{R}^{\mathrm{n}}$ is an open set, we say that $\mathbf{g}(\cdot)$ *is of class* $\mathcal{C}^{\mathrm{k}}$, that is, $\mathbf{g}(\cdot) \in \mathcal{C}^{\mathrm{k}}(A)$, if $\mathbf{g}(\cdot)$ is continuous on A with k-continuous derivatives. If $\mathbf{g}(\cdot) \in \mathcal{C}^1(A)$, then $\mathbf{g}(\cdot)$ is *continuously differentiable*.

Throughout the paper we use two types of mathematical statements, namely, existential and universal statements. An existential statement has the form: "there exist $\mathbf{x} \in A$ such that condition Φ is satisfied." A universal statement has the form: "condition Φ is satisfied for all $\mathbf{x} \in A$." For universal statements we often omit the words "for all" and write: "condition Φ holds, $\mathbf{x} \in A$."

Time is the only independent variable used in this paper and is denoted by t. In this paper, $t \in [t_1, t_2]$, where $[t_1, t_2] \subset \mathbb{R}$ is a fixed time interval and is a priori assigned. A generic member of a formation of $\mathrm{n} \in \mathbb{N}$ UAVs is identified by the subscript i and, hence, $\mathrm{i} = 1, ..., \mathrm{n}$. We define $\mathbf{r}_{\mathrm{i}} : [t_1, t_2] \to \mathbb{R}^3$ as the *position vector* of the center of mass of the i-th vehicle in a given inertial reference frame, $\boldsymbol{\sigma}_{\mathrm{i}} : [t_1, t_2] \to \mathbb{R}^3$ as the *attitude vector* of the i-th vehicle in modified rodrigues parameters (MRPs) (Shuster, 1993), and $\mathbf{x}_{\mathrm{i}} \triangleq [\mathbf{r}_{\mathrm{i}}^{\mathrm{T}}, \boldsymbol{\sigma}_{\mathrm{i}}^{\mathrm{T}}]^{\mathrm{T}}$ as the *state vector* of the i-th vehicle. The *system's configuration* at time t is defined by $\left[\mathbf{x}_1^{\mathrm{T}}(t), ..., \mathbf{x}_{\mathrm{n}}^{\mathrm{T}}(t)\right]^{\mathrm{T}}$.

The vector $\mathbf{v}_{\mathrm{i}} : [t_1, t_2] \to \mathbb{R}^3$ denotes the *velocity* of the center of mass of the i-th vehicle, $\boldsymbol{\omega}_{\mathrm{i}} : [t_1, t_2] \to \mathbb{R}^3$ denotes the *angular velocity* of the i-th vehicle in a principal body reference frame, and $\tilde{\mathbf{x}}_{\mathrm{i}} \triangleq \left[\mathbf{r}_{\mathrm{i}}^{\mathrm{T}}, \mathbf{v}_{\mathrm{i}}^{\mathrm{T}}, \boldsymbol{\sigma}_{\mathrm{i}}^{\mathrm{T}}, \boldsymbol{\omega}_{\mathrm{i}}^{\mathrm{T}}\right]^{\mathrm{T}}$ is the *augmented state vector* of the i-th vehicle. For all $t \in [t_1, t_2]$, $\mathbf{r}_{\mathrm{i}}(t) = \int_{t_1}^{t} \mathbf{v}_{\mathrm{i}}(\tau)\,\mathrm{d}\tau$ and $\dot{\boldsymbol{\sigma}}_{\mathrm{i}}(t) = \mathbf{R}_{\mathrm{rod}}(\boldsymbol{\sigma}_{\mathrm{i}}(t))\boldsymbol{\omega}_{\mathrm{i}}(t)$, where $\mathbf{R}_{\mathrm{rod}}(\boldsymbol{\sigma}_{\mathrm{i}}(t)) \triangleq \frac{1}{4}(1 - \boldsymbol{\sigma}_{\mathrm{i}}^{\mathrm{T}}(t)\boldsymbol{\sigma}_{\mathrm{i}}(t))\mathbf{I}_3 + \frac{1}{2}\boldsymbol{\sigma}_{\mathrm{i}}^{\times}(t) + \frac{1}{2}\boldsymbol{\sigma}_{\mathrm{i}}(t)\boldsymbol{\sigma}_{\mathrm{i}}^{\mathrm{T}}(t)$ (Neimark & Fufaev, 1972; Shuster, 1993). We assume $\left[\mathbf{x}_1^{\mathrm{T}}(t), ..., \mathbf{x}_{\mathrm{n}}^{\mathrm{T}}(t)\right]^{\mathrm{T}} \in D_{\mathrm{rel}} \subseteq \mathbb{R}^{6\mathrm{n}}$ and $\left[\tilde{\mathbf{x}}_1^{\mathrm{T}}(t), ..., \tilde{\mathbf{x}}_{\mathrm{n}}^{\mathrm{T}}(t)\right]^{\mathrm{T}} \in D_{\mathrm{abs}} \subseteq \mathbb{R}^{12\mathrm{n}}$, $t \in [t_1, t_2]$.

We define $\mathbf{u}_{\mathrm{i,tran}} : [t_1, t_2] \to \Gamma_{\mathrm{i,tran}}$ (respectively, $\mathbf{u}_{\mathrm{i,rot}} : [t_1, t_2] \to \Gamma_{\mathrm{i,rot}}$) as the *translational acceleration* (respectively, the *rotational acceleration*) provided by the control system of the i-th vehicle in the formation, e.g., $\mathbf{u}_{\mathrm{i,tran}}$ is the acceleration provided by the propulsion system and $\mathbf{u}_{\mathrm{i,rot}}$ is the acceleration provided by the ailerons. The vector $\mathbf{u}_{\mathrm{i,tran}}$ (respectively, $\mathbf{u}_{\mathrm{i,rot}}$) is also referred to as the *i-th translational control vector* (respectively, the *i-th rotational control vector*). For a given set of real constants $\rho_{\mathrm{i},1}$, $\rho_{\mathrm{i},2}$, $\rho_{\mathrm{i},3}$, and $\rho_{\mathrm{i},4}$ such that $0 \le \rho_{\mathrm{i},1} < \rho_{\mathrm{i},2}$ and $0 \le \rho_{\mathrm{i},3} < \rho_{\mathrm{i},4}$, $\Gamma_{\mathrm{i,tran}}$ and $\Gamma_{\mathrm{i,rot}}$ are defined as

$$\Gamma_{\mathrm{i,tran}} \triangleq \left\{\mathbf{a} \in \mathbb{R}^3 : \rho_{\mathrm{i},1} \le ||\mathbf{a}||_2 \le \rho_{\mathrm{i},2}\right\} \cup \{\mathbf{0}_3\},$$

$$\Gamma_{\mathrm{i,rot}} \triangleq \left\{\mathbf{a} \in \mathbb{R}^3 : \rho_{\mathrm{i},3} \le ||\mathbf{a}||_2 \le \rho_{\mathrm{i},4}\right\} \cup \{\mathbf{0}_3\}.$$

Finally, for a given set $\Gamma \subset \mathbb{R}^{p}$, $\mathbf{u} : [t_1, t_2] \to \Gamma$ is an *admissible control in* Γ if *i*) $\mathbf{u}(\cdot)$ is continuous at the endpoints of $[t_1, t_2]$, *ii*) $\mathbf{u}(\cdot)$ is continuous for all $t \in (t_1, t_2)$ with the exception of a finite number of times t at which $\mathbf{u}(\cdot)$ may have discontinuities of the first kind, and *iii*) $\mathbf{u}(\tau) = \lim_{t \to \tau^-} \mathbf{u}(t)$, where $\tau \in [t_1, t_2]$ is a point of discontinuity of first kind for $\mathbf{u}(t)$ (Pontryagin et al., 1962). We assume that $\mathbf{u}_{\rm i,tran}$ (respectively, $\mathbf{u}_{\rm i,rot}$) is an admissible control in $\Gamma_{\rm i,tran}$ (respectively, $\Gamma_{\rm i,rot}$) for each $\rm i \in \{1, \ldots, n\}$.

3. Problem statement

3.1 Fuel consumption performance functional

A measure of the effort needed to control the i-th formation vehicle is given by the *performance functional*

$$\mathrm{J}\left[\mathbf{u}_{\rm i}(\cdot)\right] \triangleq \int_{t_1}^{t_2} ||\mathbf{u}_{\rm i}(t)||_2 \, \mathrm{d}t, \tag{1}$$

where $\mathbf{u}_{\rm i}(t) \triangleq [\mathbf{u}_{\rm i,tran}^{\rm T}(t), \mathrm{c}\mathbf{u}_{\rm i,rot}^{\rm T}(t)]^{\rm T}$ and c is a real constant with units of distance. Without loss of generality we assume that $|\mathrm{c}| = 1$. The performance functional $\int_{t_1}^{t_2} ||\mathbf{u}_{\rm i,tran}(t)||_2 \, \mathrm{d}t$ represents a measure of the fuel consumed over the time interval $[t_1, t_2]$ (Schouwenaars et al., 2006). Path planning for UAV formations is sometimes addressed by minimizing the more conservative performance functional $\int_{t_1}^{t_2} ||\mathbf{u}_{\rm i,tran}(t)||_1 \, \mathrm{d}t$ (Blackmore, 2008). It is important to note that $||\mathbf{u}_{\rm i,rot}(t)||_2$ is much smaller than $||\mathbf{u}_{\rm i,tran}(t)||_2$ for conventional aircraft and, hence, its contribution to the performance functional (1) is negligible. However, this assumption does not hold for the case of μUAVs (Bataillé et al., 2009).

The control effort for the entire formation can be captured by the performance measure

$$\mathrm{J}_{\rm formation}\left[\tilde{\mathbf{u}}(\cdot)\right] \triangleq \sum_{\rm i=1}^{\rm n} \mu_{\rm i} \mathrm{J}\left[\mathbf{u}_{\rm i}(\cdot)\right], \tag{2}$$

where $\tilde{\mathbf{u}}(t) \triangleq [\mathbf{u}_1^{\rm T}(t), ..., \mathbf{u}_{\rm n}^{\rm T}(t)]^{\rm T}$ and $\mu_{\rm i} \in [0, 1]$, with $\sum_{\rm i=1}^{\rm n} \mu_{\rm i} = 1$, which represents the relative importance of minimizing the control effort of the i-th vehicle with respect to the others.

3.2 Aircraft dynamic equations

Aircraft are subject to external forces and moments from the environment. Specifically, an aerial vehicle is subject to gravitational forces, aerodynamic forces, and aerodynamic moments. Accelerations induced by external forces and external moments acting on a formation vehicle are denoted by $\mathbf{a} : \mathbb{R}^{12} \to \mathbb{R}^{3}$ and $\mathbf{m} : \mathbb{R}^{12} \to \mathbb{R}^{3}$, respectively, where $\mathbf{a}\left(\widetilde{\mathbf{x}}_{\rm i}\right), \mathbf{m}\left(\widetilde{\mathbf{x}}_{\rm i}\right) \in \mathcal{C}^1\left(\mathbb{R}^{12}\right)$.

The unconstrained dynamic equations for the i-th vehicle are given by (Greenwood, 2003)

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathbf{x}}_{\rm i}(t) = \begin{bmatrix} \mathbf{v}_{\rm i}(t) \\ \mathbf{a}\left(\widetilde{\mathbf{x}}_{\rm i}(t)\right) \\ \mathbf{R}_{\rm rod}(\boldsymbol{\sigma}_{\rm i}(t))\boldsymbol{\omega}_{\rm i}(t) \\ -\mathbf{I}_{\rm in,i}^{-1}\boldsymbol{\omega}_{\rm i}^{\times}\left(\boldsymbol{\omega}_{\rm i}(t)\right)\mathbf{I}_{\rm in,i}\boldsymbol{\omega}_{\rm i}(t) + \widetilde{\boldsymbol{\omega}}_{\rm i}\left(\widetilde{\mathbf{x}}_{\rm i}(t)\right) \end{bmatrix} + \begin{bmatrix} \mathbf{0}_3 \\ \mathbf{u}_{\rm i,tran}(t) \\ \mathbf{0}_3 \\ \mathbf{u}_{\rm i,rot}(t) \end{bmatrix}, \tag{3}$$

where $\mathbf{I}_{\mathrm{in,i}}$ is the inertia matrix of the i-th vehicle in a principal body reference frame and $\widetilde{\boldsymbol{\omega}}_{\mathrm{i}}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}(t)\right) \triangleq \mathbf{I}_{\mathrm{in,i}}^{-1}\mathbf{m}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}(t)\right)$, $t \in [t_1, t_2]$. The boundary conditions for (3) are given by the endpoint constraints discussed in Section 3.3.

3.3 Formation constraints

Given $D_1 \subset \mathbb{R}^{\mathrm{p}}$ and $D_2 \subset \mathbb{R}^{\mathrm{m}}$, the function $\mathbf{S} : D_1 \to D_2$ is a *continuously differentiable manifold* if $\mathbf{S}(\mathbf{y}) = 0$, $\mathrm{m} < \mathrm{p}$, $\mathbf{S}(\mathbf{y}) \in \mathcal{C}^1(D_1)$, and rank $\frac{\partial \mathbf{S}(\mathbf{y})}{\partial \mathbf{y}} = \mathrm{m}$ (Pontryagin et al., 1962). Let $\mathbf{S}_1 : D_{\mathrm{abs}} \to \mathbb{R}^{\mathrm{n}_1}$ and $\mathbf{S}_2 : D_{\mathrm{abs}} \to \mathbb{R}^{\mathrm{n}_2}$ be two continuously differentiable manifolds, and define the *endpoint constraints*

$$\begin{aligned} \mathbf{S}_1\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}(t_1), ..., \widetilde{\mathbf{x}}_{\mathrm{n}}^{\mathrm{T}}(t_1)\right]^{\mathrm{T}}\right) &= \mathbf{0}_{\mathrm{r}_1}, \\ \mathbf{S}_2\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}(t_2), ..., \widetilde{\mathbf{x}}_{\mathrm{n}}^{\mathrm{T}}(t_2)\right]^{\mathrm{T}}\right) &= \mathbf{0}_{\mathrm{r}_2}. \end{aligned} \tag{4}$$

Endpoint constraints partly impose the formation's configuration at times t_1 and t_2, and hence, can model point-to-point or rendezvous maneuvers.

State inequality constraints are given by

$$\mathbf{f}_{\mathrm{ineq}}(\mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)) \leq\leq \mathbf{0}_{\mathrm{r}_3}, \tag{5}$$

where $\mathbf{f}_{\mathrm{ineq}} : D_{\mathrm{rel}} \to \mathbb{R}^{\mathrm{n}_3}$ and $\mathbf{f}_{\mathrm{ineq}}(\mathbf{x}_1, ..., \mathbf{x}_{\mathrm{n}}) \in \mathcal{C}^3(\mathrm{int}\,(D_{\mathrm{rel}}))$. *State equality constraints* are given by

$$\mathbf{f}_{\mathrm{eq}}(t, \mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)) = \mathbf{0}_{\mathrm{r}_4}, \tag{6}$$

where $\mathbf{f}_{\mathrm{eq}} : [t_1, t_2] \times D_{\mathrm{rel}} \to \mathbb{R}^{\mathrm{n}_4}$ and $\mathbf{f}_{\mathrm{eq}}(t, \mathbf{x}_1, ..., \mathbf{x}_{\mathrm{n}}) \in \mathcal{C}^2((t_1, t_2) \times \mathrm{int}\,(D_{\mathrm{rel}}))$. Here we assume that the constraints are *compatible*, that is, for all $t \in [t_1, t_2]$ there exists at least one set of 2n admissible controls $\{\mathbf{u}_{1,\mathrm{tran}}(t), ..., \mathbf{u}_{\mathrm{n,tran}}(t); \mathbf{u}_{1,\mathrm{rot}}(t), ..., \mathbf{u}_{\mathrm{n,rot}}(t)\}$ that satisfies (3) – (6).

State constraints given in terms of $\widetilde{\mathbf{x}}_1(t)$,..., $\widetilde{\mathbf{x}}_{\mathrm{n}}(t)$ that can be reduced to the form given by (5) and (6) are called *holonomic* constraints. In particular, for $\mathrm{n} = 2$ and $t \in [t_1, t_2]$, the constraint $\mathbf{v}_1(t) = \mathbf{v}_2(t)$ is holonomic since it can be rewritten as $\mathbf{r}_1(t) + \mathbf{r}_1(t_1) = \mathbf{r}_2(t) + \mathbf{r}_2(t_1)$, $t \in [t_1, t_2]$. It is important to note that the constraint $\boldsymbol{\omega}_1(t) \leq\leq \boldsymbol{\omega}_2(t)$, $t \in [t_1, t_2]$, is nonholonomic since $\boldsymbol{\sigma}_{\mathrm{i}}(t) \neq \int_{t_1}^{t} \boldsymbol{\omega}_{\mathrm{i}}(\tau)\,\mathrm{d}\tau + \boldsymbol{\sigma}_{\mathrm{i}}(t_1)$, $t \in [t_1, t_2]$ and $\mathrm{i} = 1, 2$ (Greenwood, 2003).

State constraints can model collision avoidance, keeping the formation far from no-fly zones, or the requirement of pointing payloads toward the same target. It is obvious that (6) is a special case of (5); however, as noted in Section 4.2, this distinction is useful in reducing computational complexity.

3.4 Path planning optimization problem

For all $\mathrm{i} = 1, ..., \mathrm{n}$ and $t \in [t_1, t_2]$ find the control vectors $\mathbf{u}_{\mathrm{i,tran}}(t)$ and $\mathbf{u}_{\mathrm{i,rot}}(t)$ among all admissible controls in $\Gamma_{\mathrm{i,tran}}$ and $\Gamma_{\mathrm{i,tran}}$ such that the performance measure (2) is minimized and $\widetilde{\mathbf{x}}_{\mathrm{i}}(t)$ satisfies (3) – (6).

4. Mathematical background

4.1 Slack variables

Inequality constraints (5) can be reduced to equality constraints by introducing $\mathbf{s} : [t_1, t_2] \to \mathbb{R}^{\mathrm{n}_3}$ such that $\mathbf{s}(t) \in \mathcal{C}^2(t_1, t_2)$ and $\mathbf{f}_{\mathrm{ineq}}(t, \mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)) + \frac{1}{2}\mathrm{diag}\left(\mathbf{s}\mathbf{s}^{\mathrm{T}}\right) = \mathbf{0}_{\mathrm{r}_3}$. The components of $\mathbf{s}$ are called *slack variables*. Thus, (5) can be rewritten as (Valentine, 1937)

$$\widetilde{\mathbf{f}}_{\mathrm{ineq}}(\mathbf{s}(t), \mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)) = \mathbf{0}_{\mathrm{r}_3}, \tag{7}$$

where $\widetilde{\mathbf{f}}_{\mathrm{ineq}}\left(\mathbf{s}(t), \mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)\right) \triangleq \mathbf{f}_{\mathrm{ineq}}(\mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t)) + \frac{1}{2}\mathrm{diag}\left(\mathbf{s}\mathbf{s}^{\mathrm{T}}\right)$.

4.2 Lagrange coordinates

The following theorem is needed for the main results of this paper.

Theorem 4.1. *(Pars, 1965) Let $D_{\mathrm{q}} \subseteq \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4}$ be an open connected set and let $\mathbf{q} : [t_1, t_2] \times \mathbb{R}^{\mathrm{n}_3} \times D_{\mathrm{rel}} \to D_q$ be such that $\mathbf{q}\left(t, \mathbf{s}(t), \mathbf{x}_1(t), \dots, \mathbf{x}_{\mathrm{n}}(t)\right) \in \mathcal{C}^2\left((t_1, t_2) \times \mathbb{R}^{\mathrm{n}_3} \times \mathrm{int}\left(\mathrm{D}_{\mathrm{rel}}\right)\right)$. Assume that*

$$\det\left(\frac{\partial\left[\widetilde{\mathbf{f}}^{\mathrm{T}}_{\mathrm{ineq}}\left(\mathbf{s}, \mathbf{x}_1, ..., \mathbf{x}_{\mathrm{n}}\right)\mathbf{f}^{\mathrm{T}}_{\mathrm{eq}}\left(t, \mathbf{s}, \mathbf{x}_1, ..., \mathbf{x}_{\mathrm{n}}\right)\mathbf{q}^{\mathrm{T}}\left(t, \mathbf{s},\, \mathbf{x}_1,\, ...,\, \mathbf{x}_{\mathrm{n}}\right)\right]^{\mathrm{T}}}{\partial\left[\mathbf{s}^{\mathrm{T}},\, \mathbf{x}_1^{\mathrm{T}},\, ...,\, \mathbf{x}_{\mathrm{n}}^{\mathrm{T}}\right]^{\mathrm{T}}}\right) \neq 0 \tag{8}$$

for all $(t, \mathbf{s}, \mathbf{x}_1 ..., \mathbf{x}_{\mathrm{n}}) \in \mathcal{I} \times \Delta$, where $\mathcal{I} \subset (t_1, t_2)$ and $\Delta \subset \mathbb{R}^{\mathrm{n}_3} \times D_{\mathrm{rel}}$ are open connected sets. Then $\mathbf{q}$ can be rewritten as a function of t, that is, $\mathbf{q} : \mathcal{I} \to \mathrm{D}_{\mathrm{q}}$, and $\mathbf{s}$, $\mathbf{x}_1$, ...,$\mathbf{x}_{\mathrm{n}}$, $\widetilde{\mathbf{x}}_1$,..., $\widetilde{\mathbf{x}}_{\mathrm{n}}$ can be rewritten as unique functions of t and $\mathbf{q}$, that is, $\mathbf{s} : \mathcal{I} \times \mathrm{D}_{\mathrm{q}} \to \mathbb{R}^{\mathrm{n}_3}$, $\mathbf{x}_{\mathrm{i}} : \mathcal{I} \times D_{\mathrm{q}} \to \mathbb{R}^6$, and $\widetilde{\mathbf{x}}_{\mathrm{i}} : \mathcal{I} \times D_{\mathrm{q}} \to \mathbb{R}^{12}$ for all $\mathrm{i} = 1, ..., \mathrm{n}$ and $(t, \mathbf{s}, \mathbf{x}_1 ..., \mathbf{x}_{\mathrm{n}}) \in \mathcal{I} \times \Delta$. Furthermore, the components of $\mathbf{q}$ are independent and uniquely characterize the system's configuration.

Under the hypothesis of Theorem 4.1, the components of $\mathbf{q}(t)$ are called *Lagrange coordinates*. As will be shown in Section 4.3, the key advantage of using Lagrange coordinates is that the constraints (5) – (7) are automatically accounted for when rewriting the formation's dynamic equations in terms of t and $\mathbf{q}(t)$ (Pars, 1965). In this paper, we assume that $\mathbf{s}, \mathbf{x}_1, \dots, \mathbf{x}_{\mathrm{n}}, \widetilde{\mathbf{x}}_1, \dots, \widetilde{\mathbf{x}}_{\mathrm{n}}$ are explicit functions of $\mathbf{q}$ only and not t, which occurs in most practical applications (Pars, 1965). In practice, given constraints in the form of (6) and (7), $\mathbf{q}$ is chosen such that Theorem 4.1 holds. As will be further discussed in Section 4.3, we select $\mathbf{q}\left(t, \mathbf{s}(t), \mathbf{x}_1(t), \dots, \mathbf{x}_{\mathrm{n}}(t)\right)$ as an explicit function of $\left(\mathbf{s}(t), \mathbf{x}_1(t), \dots, \mathbf{x}_{\mathrm{n}}(t)\right)$.

Given $\mathbf{q}\left(t, \mathbf{s}(t), \mathbf{x}_1(t), \dots, \mathbf{x}_{\mathrm{n}}(t)\right)$, $\dot{\mathbf{q}}$ is a function of $\mathbf{s}(t)$, $\mathbf{r}_{\mathrm{i}}(t)$, $\boldsymbol{\sigma}_{\mathrm{i}}(t)$, $i = 1, ..., n$, and their first time derivatives. In practice, however, we measure $\boldsymbol{\omega}_{\mathrm{i}}(t)$ rather than $\boldsymbol{\sigma}_{\mathrm{i}}(t)$, and hence, if the assumptions of Theorem 4.1 hold, we define the *kinematic equation*

$$\mathbf{q}_{\mathrm{dot}}(t) \triangleq \boldsymbol{\Psi}\left(\mathbf{q}(t)\right)\dot{\mathbf{q}}(t) + \boldsymbol{\psi}\left(\mathbf{q}(t)\right), \tag{9}$$

where $\boldsymbol{\omega}_{\mathrm{i}}(t)$, $\mathrm{i} = 1, 2, \dots, \mathrm{n}$, explicitly appears in $\mathbf{q}_{\mathrm{dot}}(t)$, $\boldsymbol{\Psi} : D_{\mathrm{q}} \to \mathbb{R}^{(6\mathrm{n}-\mathrm{n}_4)\times(6\mathrm{n}-\mathrm{n}_4)}$ is an invertible continuously differentiable matrix function, and $\boldsymbol{\psi} : D_{\mathrm{q}} \to \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4}$ is continuously differentiable. Consequently, $\mathbf{s}$, $\mathbf{x}_1$, ..., $\mathbf{x}_{\mathrm{n}}$, $\widetilde{\mathbf{x}}_1$, ..., $\widetilde{\mathbf{x}}_{\mathrm{n}}$ can be rewritten as unique functions of $\mathbf{q}$ and $\mathbf{q}_{\mathrm{dot}}$, that is, $\mathbf{s} : D_{\mathrm{q}} \times \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \to \mathbb{R}^{\mathrm{n}_3}$, $\mathbf{x}_{\mathrm{i}} : D_{\mathrm{q}} \times \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \to \mathbb{R}^6$, and $\widetilde{\mathbf{x}}_{\mathrm{i}} : D_{\mathrm{q}} \times \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \to \mathbb{R}^{12}$, $(t, \mathbf{s}, \mathbf{x}_1 ..., \mathbf{x}_{\mathrm{n}}) \in \mathcal{I} \times \Delta$ (Greenwood, 2003). Here, we assume that $\mathbf{q}_{\mathrm{dot}}$ satisfies (23) below.

In the following we assume that the path planning optimization problem can be solved over the time interval $\left[t_1^*, t_2^*\right] \supset [t_1, t_2]$ and that the given set of Lagrange coordinates can be defined on the open connected set $\widetilde{\mathcal{I}}$, where $[t_1, t_2] \subset \widetilde{\mathcal{I}} \subset \left(t_1^*, t_2^*\right)$. Thus, (4) can be rewritten as

$$\mathbf{S}_1\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}\left(\mathbf{q}(t_1), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_1))\right), ..., \widetilde{\mathbf{x}}_{\mathrm{n}}^{\mathrm{T}}\left(\mathbf{q}(t_1), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_1))\right)\right]^{\mathrm{T}}\right) = \mathbf{0}_{\mathrm{r}_1}, \tag{10}$$

$$\mathbf{S}_2\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}\left(\mathbf{q}(t_2), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_2))\right), ..., \widetilde{\mathbf{x}}_{\mathrm{n}}^{\mathrm{T}}\left(\mathbf{q}(t_2), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_2))\right)\right]^{\mathrm{T}}\right) = \mathbf{0}_{\mathrm{r}_2}. \tag{11}$$

Example 4.1. Consider a UAV formation with two vehicles so that $\mathrm{n} = 2$. Assume that

$$\mathbf{f}_{\mathrm{ineq}}\left(\mathbf{x}_1(t), \mathbf{x}_2(t)\right) = \begin{bmatrix} ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 - \mathrm{r}_{\max} \\ \mathrm{r}_{\min} - ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 \end{bmatrix} \leq\leq \mathbf{0}_2, \tag{12}$$

$$\mathbf{f}_{\mathrm{eq}}\left(t, \mathbf{x}_1(t), \mathbf{x}_2(t)\right) = \boldsymbol{\sigma}_1(t) - \boldsymbol{\sigma}_2(t) = \mathbf{0}_3, \tag{13}$$

$$\mathbf{S}_1\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}(t_1)\, \widetilde{\mathbf{x}}_2^{\mathrm{T}}(t_1)\right]^{\mathrm{T}}\right) = \begin{bmatrix} ||\mathbf{r}_1(t_1) - \mathbf{r}_2(t_1)||_2^2 - \left(\frac{\mathrm{r}_{\max} + \mathrm{r}_{\min}}{2}\right) \\ \boldsymbol{\sigma}_1(t_1) - \boldsymbol{\sigma}_2(t_1) \end{bmatrix} = \mathbf{0}_4, \tag{14}$$

$$\mathbf{S}_2\left(\left[\widetilde{\mathbf{x}}_1^{\mathrm{T}}(t_2)\, \widetilde{\mathbf{x}}_2^{\mathrm{T}}(t_2)\right]^{\mathrm{T}}\right) = \begin{bmatrix} ||\mathbf{r}_1(t_2) - \mathbf{r}_2(t_2)||_2^2 - \frac{2(\mathrm{r}_{\max} - \mathrm{r}_{\min})}{3} \\ \boldsymbol{\sigma}_1(t_2) - \boldsymbol{\sigma}_2(t_2) \end{bmatrix} = \mathbf{0}_4, \tag{15}$$

where $\mathrm{r}_{\min}$ and $\mathrm{r}_{\max}$ are real constants such that $0 < \mathrm{r}_{\min} < \mathrm{r}_{\max}$. Equation (12) ensures that $\mathrm{r}_{\min} \leq ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 \leq \mathrm{r}_{\max}$ and (13) ensures that both vehicles always have the same attitude: $D_{\mathrm{rel}} = \left\{\left[\mathbf{x}_1^{\mathrm{T}}(t)\, \mathbf{x}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}} : \mathrm{r}_{\min} \leq ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 \leq \mathrm{r}_{\max}, \boldsymbol{\sigma}_1(t) = \boldsymbol{\sigma}_2(t), t \in [t_1, t_2]\right\}$.

Introducing the slack variables $\mathrm{s}_1 : [t_1, t_2] \to \mathbb{R}$ and $\mathrm{s}_2 : [t_1, t_2] \to \mathbb{R}$, (12) becomes

$$\widetilde{\mathbf{f}}_{\mathrm{ineq}}(\mathbf{s}(t), \mathbf{x}_1(t), \mathbf{x}_2(t)) = \begin{bmatrix} ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 - \mathrm{r}_{\max} + \frac{1}{2}\mathrm{s}_1^2(t) \\ \mathrm{r}_{\min} - ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 + \frac{1}{2}\mathrm{s}_2^2(t) \end{bmatrix} = \mathbf{0}_2. \tag{16}$$

As noted in Section 3.3, the equality constraint (13) can be embedded into (12) to give

$$\widetilde{\mathbf{f}}_{\mathrm{ineq}}(\mathbf{s}(t), \mathbf{x}_1(t), \mathbf{x}_2(t)) = \begin{bmatrix} ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 - \mathrm{r}_{\max} + \frac{1}{2}\mathrm{s}_1^2(t) \\ \mathrm{r}_{\min} - ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 + \frac{1}{2}\mathrm{s}_2^2(t) \\ \boldsymbol{\sigma}_1(t) - \boldsymbol{\sigma}_2(t) + \frac{1}{2}\mathrm{diag}(\mathbf{s}_3\mathbf{s}_3^{\mathrm{T}}) \\ \boldsymbol{\sigma}_2(t) - \boldsymbol{\sigma}_1(t) + \frac{1}{2}\mathrm{diag}(\mathbf{s}_4\mathbf{s}_4^{\mathrm{T}}) \end{bmatrix} = \mathbf{0}_8,$$

where $\mathbf{s}_{\mathrm{j}} : [t_1, t_2] \to \mathbb{R}^3$, $\mathrm{j} = 3, 4$. Note that in this case, the dimension of $\widetilde{\mathbf{f}}_{\mathrm{ineq}}$ is increased since six additional slack variables have been introduced, which increases computational complexity.

Next, define $\mathrm{r}_{\mathrm{i,j}} : [t_1, t_2] \to \mathbb{R}$ (respectively, $\sigma_{\mathrm{i,j}} : [t_1, t_2] \to \mathbb{R}$) as the j-th component of $\mathbf{r}_{\mathrm{i}}(t)$ (respectively, $\boldsymbol{\sigma}_{\mathrm{i}}(t)$). If $\mathbf{q}(t) = \left[s_1(t), s_2(t), \mathbf{r}_1^{\mathrm{T}}(t), \boldsymbol{\sigma}_1^{\mathrm{T}}(t), r_{2,1}(t)\right]^{\mathrm{T}}$, then (8) gives

$$\det\left(\frac{\partial\left[\widetilde{\mathbf{f}}_{\mathrm{ineq}}^{\mathrm{T}}\left(\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2\right), \mathbf{f}_{\mathrm{eq}}^{\mathrm{T}}\left(t, \mathbf{s}, \mathbf{x}_1, \mathbf{x}_2\right), \mathbf{q}^{\mathrm{T}}\left(t, \mathbf{s}, \mathbf{x}_1, \mathbf{x}_2\right)\right]^{\mathrm{T}}}{\partial\left[\mathbf{s}^{\mathrm{T}}, \mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}\right]^{\mathrm{T}}}\right) = 0.$$

Thus, by Theorem 4.1, the components of $\mathbf{q}$ are not Lagrange coordinates.

Alternatively, if $\mathbf{q}(t) = \left[\mathrm{s}_1(t),\ \mathrm{r}_{1,1}(t),\ \mathrm{r}_{1,2}(t),\ \boldsymbol{\sigma}_1^{\mathrm{T}}(t),\ \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$, then

$$\det\left(\frac{\partial\left[\tilde{\mathbf{f}}^{\mathrm{T}}_{\mathrm{ineq}}\left(\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right),\ \mathbf{f}^{\mathrm{T}}_{\mathrm{eq}}\left(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right),\ \mathbf{q}^{\mathrm{T}}\left(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right)\right]^{\mathrm{T}}}{\partial\left[\mathbf{s}^{\mathrm{T}},\ \mathbf{x}_1^{\mathrm{T}},\ \mathbf{x}_2^{\mathrm{T}}\right]^{\mathrm{T}}}\right) = -2\mathrm{s}_2(t)\left(\mathrm{r}_{1,3}(t)-\mathrm{r}_{2,3}(t)\right),$$

for all $(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2) \in (t_1,t_2)\times\mathbb{R}^2\times\operatorname{int}(D_{\mathrm{rel}})$ such that $\mathrm{r}_{1,3}(t) \neq \mathrm{r}_{2,3}(t)$, and hence, the components of $\mathbf{q}$ are suitable Lagrange coordinates if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t)-\mathbf{r}_2(t)||_2^2$ and $\mathrm{r}_{1,3}(t) \neq \mathrm{r}_{2,3}(t)$. In this case, (9) gives

$$\mathbf{q}_{\mathrm{dot}}(t) = \begin{bmatrix} \dot{s}_1(t) \\ \mathrm{v}_{1,1}(t) \\ \mathrm{v}_{1,2}(t) \\ \boldsymbol{\omega}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{R}^{-1}_{\mathrm{rod}}(\boldsymbol{\sigma}_1(t)) & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \dot{s}_1(t) \\ \mathrm{v}_{1,1}(t) \\ \mathrm{v}_{1,2}(t) \\ \dot{\boldsymbol{\sigma}}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix}, \tag{17}$$

where $\mathrm{v}_{1,\mathrm{j}} : [t_1,t_2] \to \mathbb{R}$ is the j-th component of $\mathbf{v}_1(t)$.

A more suitable choice of Lagrange coordinates is given by $\mathbf{q}(t) = \left[\mathbf{x}_1^{\mathrm{T}}(t),\ \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$ since

$$\det\left(\frac{\partial\left[\tilde{\mathbf{f}}^{\mathrm{T}}_{\mathrm{ineq}}\left(\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right),\ \mathbf{f}^{\mathrm{T}}_{\mathrm{eq}}\left(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right),\ \mathbf{q}^{\mathrm{T}}\left(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2\right)\right]^{\mathrm{T}}}{\partial\left[\mathbf{s}^{\mathrm{T}}\ \mathbf{x}_1^{\mathrm{T}},\ \mathbf{x}_2^{\mathrm{T}}\right]^{\mathrm{T}}}\right) = s_1(t)s_2(t)$$

for all $(t,\mathbf{s},\mathbf{x}_1,\mathbf{x}_2) \in (t_1,t_2)\times\mathbb{R}^2\times\operatorname{int}(D_{\mathrm{rel}})$, and hence, the components of $\mathbf{q}$ are suitable Lagrange coordinates if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t)-\mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$. In this case, (9) gives

$$\mathbf{q}_{\mathrm{dot}}(t) = \begin{bmatrix} \mathbf{v}_1(t) \\ \boldsymbol{\omega}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{R}^{-1}_{\mathrm{rod}}(\boldsymbol{\sigma}_1(t)) & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1(t) \\ \dot{\boldsymbol{\sigma}}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix}. \tag{18}$$

Since we use this example throughout the paper, we define $\mathbf{q}_{\mathrm{dot},1} \triangleq [\mathbf{v}_1^{\mathrm{T}},\ \boldsymbol{\omega}_1^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{q}_{\mathrm{dot},2} \triangleq \mathbf{v}_2$, and

$$\boldsymbol{\Psi}_1(\mathbf{x}_1(t)) \triangleq \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{R}^{-1}_{\mathrm{rod}}(\boldsymbol{\sigma}_1(t)) \end{bmatrix}.$$

Finally, note that if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t)-\mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$ for $t \in \left(t_1^*, t_2^*\right) \supset [t_1,t_2]$, then (14) and (15) reduce to

$$||\mathbf{r}_1(t_1)-\mathbf{r}_2(t_1)||_2^2 - \left(\frac{\mathrm{r}_{\max}+\mathrm{r}_{\min}}{2}\right) = 0, \tag{19}$$

$$||\mathbf{r}_1(t_2)-\mathbf{r}_2(t_2)||_2^2 - \frac{2\left(\mathrm{r}_{\max}-\mathrm{r}_{\min}\right)}{3} = 0. \tag{20}$$

4.3 Constrained formation dynamic equations

The formation's kinetic energy is given by *König's theorem* (Pars, 1965) and for our problem takes the form

$$\begin{aligned} \mathrm{k}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right) &= \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{m_i}\mathbf{v}_{\mathrm{i}}^{\mathrm{T}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\mathbf{v}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right) \\ &\quad + \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{\omega}_{\mathrm{i}}^{\mathrm{T}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\mathbf{I}_{\mathrm{in,i}}\boldsymbol{\omega}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right), \end{aligned} \tag{21}$$

where $\mathrm{m_i}$ is the mass of the i-th vehicle, which is assumed to be constant. The dynamic equations of the constrained formation can be written in terms of Lagrange coordinates by applying the *Boltzmann-Hammel equation* (Greenwood, 2003) to give

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathrm{k}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial \mathbf{q}_{\mathrm{dot}}}\right) &= \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{m_i}\mathbf{v}_{\mathrm{i}}^{\mathrm{T}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{v}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial \mathbf{q}_{\mathrm{dot}}} \\ &\quad + \sum_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{\omega}_{\mathrm{i}}^{\mathrm{T}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\mathbf{I}_{\mathrm{in,i}}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \boldsymbol{\omega}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial \mathbf{q}_{\mathrm{dot}}} \\ &\quad + \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathbf{a}\left(\tilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)+\mathbf{u}_{\mathrm{i,tran}}(t)\right)\frac{\partial \mathbf{v}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial \mathbf{q}_{\mathrm{dot}}} \\ &\quad + \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathbf{m}\left(\tilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)+\mathbf{u}_{\mathrm{i,rot}}(t)\right)\frac{\partial \boldsymbol{\omega}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial \mathbf{q}_{\mathrm{dot}}}. \end{aligned} \tag{22}$$

Equations (10) and (11) are the boundary conditions for (22). It is important to note that the dynamic equation (22) is written in terms of Lagrange coordinates, and hence, accounts for (5) and (6).

Analytical optimization techniques such as Pontryagin's minimum principle, Bellman's theorem, and calculus of variations require the dynamic equations to be written as a first-order ordinary differential equation in explicit form. Therefore, using the hypothesis on $\mathbf{q}_{\mathrm{dot}}$, the second-order ordinary differential equation (22) needs to be written in a first-order form

$$\dot{\mathbf{q}}_{\mathrm{dot}}(t) = \mathbf{f}_{\mathrm{dyn}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t),\tilde{\mathbf{u}}(t)), \tag{23}$$

where $\tilde{\mathbf{u}}(t) \triangleq [\mathbf{u}_1^{\mathrm{T}}(t),\ldots,\mathbf{u}_{\mathrm{n}}^{\mathrm{T}}(t)]^{\mathrm{T}}$ and $\mathbf{f}_{\mathrm{dyn}} : D_{\mathrm{q}} \times \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \times \mathbb{R}^{12\mathrm{n}} \to \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4}$. In order to isolate the contribution of $\tilde{\mathbf{u}}$ in (24), we define $\hat{\mathbf{f}}_{\mathrm{dyn}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)) \triangleq \mathbf{f}_{\mathrm{dyn}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t),\tilde{\mathbf{u}}(t)) - \mathbf{u}_{\mathrm{i,tran}}(t)\frac{\partial \mathbf{v}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial \mathbf{q}_{\mathrm{dot}}} - \mathbf{u}_{\mathrm{i,rot}}(t)\frac{\partial \boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial \mathbf{q}_{\mathrm{dot}}}$.

Equation (22) or, equivalently, (23) gives a set of $6\mathrm{n}-\mathrm{n}_4$ equations in $2(6\mathrm{n}-\mathrm{n}_4)$ unknowns, which are $\mathbf{q}$ and $\mathbf{q}_{\mathrm{dot}}$. Thus, (22) needs to be solved together with (9) (Greenwood, 2003) to give

$$\begin{bmatrix} \mathbf{q}_{\mathrm{dot}}(t) \\ \dot{\mathbf{q}}_{\mathrm{dot}}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Upsilon}\left(\mathbf{q}(t)\right)\dot{\mathbf{q}}(t) + \boldsymbol{\psi}\left(\mathbf{q}(t)\right) \\ \mathbf{f}_{\mathrm{dyn}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t),\tilde{\mathbf{u}}(t)) \end{bmatrix}. \tag{24}$$

From (21) it follows that the formation's kinetic energy k is not an explicit function of $\mathbf{s}$, and hence, if $\mathbf{q}$ is chosen as an explicit function of p components of $\mathbf{s}(t) \in \mathbb{R}^{\mathrm{n}_3}$, then p of the $6\mathrm{n} - \mathrm{n}_4$ equations in (22) cannot be straightforwardly recast in the explicit form given by (23). In this case, assume, without loss of generality, that $\mathbf{q}$ explicitly depends on the first p components of $\mathbf{s}$ and substitute the corresponding p equations in (24) with

$$\mathrm{s}_{\mathrm{j}}(t)\ddot{\mathrm{s}}_{\mathrm{j}}(t) = -\dot{\mathrm{s}}_{\mathrm{j}}^{2}(t) - \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathrm{f}_{\mathrm{ineq,j}}(\mathbf{x}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right), \ldots, \mathbf{x}_{\mathrm{n}}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)), \tag{25}$$

which is obtained by differentiating (7). In this case, the boundary conditions are given by

$$\widetilde{\mathbf{f}}_{\mathrm{ineq}}(\mathbf{s}\left(\mathbf{q}(t_1), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_1))\right), \mathbf{x}_1\left(\mathbf{q}(t_1), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_1))\right), ..., \mathbf{x}_{\mathrm{n}}\left(\mathbf{q}(t_1), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_1))\right)) = \mathbf{0}_{\mathrm{r}_3}, \tag{26}$$

$$\widetilde{\mathbf{f}}_{\mathrm{ineq}}(\mathbf{s}\left(\mathbf{q}(t_2), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_2))\right), \mathbf{x}_1\left(\mathbf{q}(t_2), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_2))\right), ..., \mathbf{x}_{\mathrm{n}}\left(\mathbf{q}(t_2), \mathbf{q}_{\mathrm{dot}}(\mathbf{q}(t_2))\right)) = \mathbf{0}_{\mathrm{r}_3}, \tag{27}$$

where $\mathrm{f}_{\mathrm{ineq,j}} : \mathbb{R}^{\mathrm{n}_3} \times D_{\mathrm{rel}} \to \mathbb{R}$ is the j-th component of $\mathbf{f}_{\mathrm{ineq}}(\mathbf{s}(t), \mathbf{x}_1(t), ..., \mathbf{x}_{\mathrm{n}}(t))$ (Jacobson & Lele, 1969), for $\mathrm{j} = 1, \ldots, \mathrm{p}$. If $s_{\mathrm{j}}(t^*) = 0$ for some $t^* \in [t_1, t_2]$, then (25) can be replaced by

$$3\dot{\mathrm{s}}_{\mathrm{j}}(t)\ddot{\mathrm{s}}_{\mathrm{j}}(t) + \mathrm{s}_{\mathrm{j}}(t)\frac{\mathrm{d}^3\mathrm{s}_{\mathrm{j}}(t)}{\mathrm{d}t^3} = -\frac{d^3}{dt^3}f_{\mathrm{ineq,j}}(\mathbf{x}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right), \ldots, \mathbf{x}_{\mathrm{n}}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)), \tag{28}$$

where $\mathbf{s} \in \mathcal{C}^3(t_1, t_2)$. In general, (7) must be differentiated so that $\ddot{s}_{\mathrm{j}}(t)$, or one of its higher-order derivatives, explicitly appears and is multiplied by a term that is non-zero for all $t \in [t_1, t_2]$. In this case, the differentiability assumptions on $\mathbf{s}$ and $\mathbf{f}_{\mathrm{ineq}}$ must be modified accordingly.

Example 4.2. Consider Example 4.1 with $\mathbf{q}(t) = \left[\mathrm{s}_1(t),\ \mathrm{r}_{1,1}(t),\ \mathrm{r}_{1,2}(t),\ \boldsymbol{\sigma}_1^{\mathrm{T}}(t),\ \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$. In this case, the formation's kinetic energy is given by

$$\begin{aligned}\mathrm{k}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right) &= \frac{1}{2}\mathrm{m}_1\mathbf{v}_1^{\mathrm{T}}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\mathbf{v}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\\ &\quad + \frac{1}{2}\mathrm{m}_2\mathbf{v}_2^{\mathrm{T}}(t)\mathbf{v}_2(t) + \frac{1}{2}\sum_{\mathrm{i}=1}^{2}\boldsymbol{\omega}_1^{\mathrm{T}}(t)\mathbf{I}_{\mathrm{in,i}}\boldsymbol{\omega}_1(t).\end{aligned}$$

The dynamic equations can now be found by applying (22) and accounting for (17) giving

$$\mathrm{v}_{1,\mathrm{j}}(t) = \frac{\mathrm{d}\mathrm{r}_{1,\mathrm{j}}(t)}{\mathrm{d}t}, \quad \boldsymbol{\omega}_1(t) = \mathbf{R}_{\mathrm{rod}}^{-1}(\boldsymbol{\sigma}_1(t))\dot{\boldsymbol{\sigma}}_1(t), \quad \mathbf{v}_2(t) = \frac{\mathrm{d}\mathbf{r}_2(t)}{\mathrm{d}t}, \tag{29}$$

$$\mathrm{m}_1\frac{\mathrm{d}\mathrm{v}_{1,1}(t)}{\mathrm{d}t} = \mathrm{m}_1\mathrm{a}_1\left(\widetilde{\mathbf{x}}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathrm{m}_1\mathrm{u}_{1,\mathrm{tran},1}(t), \tag{30}$$

$$\mathrm{m}_1\frac{\mathrm{d}\mathrm{v}_{1,2}(t)}{\mathrm{d}t} = \mathrm{m}_1\mathrm{a}_2\left(\widetilde{\mathbf{x}}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathrm{m}_1\mathrm{u}_{1,\mathrm{tran},2}(t), \tag{31}$$

$$\mathbf{I}_{\mathrm{in},1}\frac{\mathrm{d}\omega_1(t)}{\mathrm{d}t} = -\boldsymbol{\omega}_1^{\times}\left(\boldsymbol{\omega}_1(t)\right)\mathbf{I}_{\mathrm{in},1}\boldsymbol{\omega}_1(t) + \mathbf{m}\left(\widetilde{\mathrm{x}}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t, \mathbf{q}(t))\right)\right) + \mathbf{I}_{\mathrm{in},1}\mathbf{u}_{1,rot}(t), \tag{32}$$

$$\mathrm{m}_2\frac{\mathrm{d}\mathbf{v}_2(t)}{\mathrm{d}t} = \mathrm{m}_2\mathbf{a}\left(\widetilde{\mathrm{x}}_2\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathrm{m}_2\mathbf{u}_{2,\mathrm{tran}}(t), \tag{33}$$

where, for $\mathrm{j} = 1, 2$, $\mathrm{u}_{1,\mathrm{tran,j}} : [t_1, t_2] \to \mathbb{R}$ (respectively, $\mathrm{a}_\mathrm{j} : \mathbb{R}^{12} \to \mathbb{R}$) is the j-th component of $\mathbf{u}_{1,\mathrm{tran}}(t)$ (respectively, $\mathbf{a}\left(\widetilde{\mathbf{x}}_1\left(t, \mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t, \mathbf{q}(t))\right)\right)$). Instead of deducing the dynamics of $\mathrm{s}_1(t)$ from

$$\mathrm{m}_1 \frac{\mathrm{d}\mathrm{v}_{1,3}(t)}{\mathrm{d}t} = \mathrm{m}_1 \mathrm{a}_3\left(\widetilde{\mathbf{x}}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathrm{m}_1 \mathrm{u}_{1,\mathrm{tran},3}(t),$$

we use (16) and (25) to obtain

$$\begin{aligned}\dot{\mathrm{s}}_1^2(t) + \mathrm{s}_1(t)\ddot{\mathrm{s}}_1(t) = &- 2||\mathbf{v}_1(t) - \mathbf{v}_2(t)||_2^2 \\ &+ (\mathbf{r}_1(t) - \mathbf{r}_2(t))^{\mathrm{T}}\left(\mathbf{a}\left(\widetilde{\mathbf{x}}_1\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathbf{u}_{1,\mathrm{tran}}(t)\right) \\ &- (\mathbf{r}_1(t) - \mathbf{r}_2(t))^{\mathrm{T}}\left(\mathbf{a}\left(\widetilde{\mathbf{x}}_2\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right)\right) + \mathbf{u}_{2,\mathrm{tran}}(t)\right), \end{aligned} \tag{34}$$

which can be solved for $\mathrm{s}_1(t)$ if $||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$, $t \in [t_1, t_2]$. In this case , the boundary conditions to (34) are given by

$$\begin{bmatrix} ||\mathbf{r}_1(t_1) - \mathbf{r}_2(t_1)||_2^2 - \mathrm{r}_{\max} + \frac{1}{2}s_1^2(t_1) \\ \mathrm{r}_{\min} - ||\mathbf{r}_1(t_1) - \mathbf{r}_2(t_1)||_2^2 + \frac{1}{2}s_2^2(t_1) \end{bmatrix} = \mathbf{0}_2,$$

$$\begin{bmatrix} ||\mathbf{r}_1(t_2) - \mathbf{r}_2(t_2)||_2^2 - \mathrm{r}_{\max} + \frac{1}{2}s_1^2(t_2) \\ \mathrm{r}_{\min} - ||\mathbf{r}_1(t_2) - \mathbf{r}_2(t_2)||_2^2 + \frac{1}{2}s_2^2(t_2) \end{bmatrix} = \mathbf{0}_2.$$

If, alternatively, $\mathbf{q}(t) = \left[\mathbf{x}_1^{\mathrm{T}}(t), \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$, then the formation's kinetic energy is given by

$$\mathrm{k}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t)\right) = \frac{1}{2}\sum_{\mathrm{i}=1}^{2} \mathrm{m}_\mathrm{i} \mathbf{v}_\mathrm{i}^{\mathrm{T}}(t)\mathbf{v}_\mathrm{i}(t) + \frac{1}{2}\sum_{\mathrm{i}=1}^{2} \boldsymbol{\omega}_1^{\mathrm{T}}(t)\mathbf{I}_{\mathrm{in,i}}\boldsymbol{\omega}_1(t) \tag{35}$$

and the dynamic equations, obtained by applying (22) and (18), are given by

$$\mathbf{v}_1(t) = \frac{\mathrm{d}\mathbf{r}_1(t)}{\mathrm{d}t}, \quad \boldsymbol{\omega}_1(t) = \mathbf{R}_{\mathrm{rod}}^{-1}(\boldsymbol{\sigma}_1(t))\dot{\boldsymbol{\sigma}}_1(t), \quad \mathbf{v}_2(t) = \frac{\mathrm{d}\mathbf{r}_2(t)}{\mathrm{d}t}, \tag{36}$$

$$\mathrm{m}_1 \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}_1(t) = \mathrm{m}_1 \mathbf{a}\left(\widetilde{\mathbf{x}}_1(t)\right) + m_1 \mathbf{u}_{1,\mathrm{tran}}(t), \tag{37}$$

$$\mathbf{I}_{\mathrm{in},1} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\omega}_1(t) = -\boldsymbol{\omega}_1^{\times}\left(\boldsymbol{\omega}_1(t)\right)\mathbf{I}_{\mathrm{in},1}\boldsymbol{\omega}_1(t) + \mathbf{m}\left(\widetilde{\mathbf{x}}_1(t)\right) + \mathbf{I}_{\mathrm{in},1}\mathbf{u}_{1,rot}(t), \tag{38}$$

$$\mathrm{m}_2 \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}_2(t) = \mathrm{m}_2 \mathbf{a}\left(\left[\mathbf{r}_2^{\mathrm{T}}(t), \mathbf{v}_2^{\mathrm{T}}(t), \boldsymbol{\sigma}_1^{\mathrm{T}}(t), \boldsymbol{\omega}_1^{\mathrm{T}}(t)\right]^{\mathrm{T}}\right) + \mathrm{m}_2 \mathbf{u}_{2,\mathrm{tran}}(t). \tag{39}$$

The Lagrange coordinates chosen imply that the first vehicle can be considered as unconstrained, that is, subject to (3), (14), and (15) only, and therefore, the dynamic equations (36) – (38) can be directly deduced from (3). Similarly, the translational dynamics of the second vehicle can be considered as unconstrained. Thus, (39) can be directly obtained from (3). Recall from Example 4.1 that the components of $\mathbf{q}$ are suitable Lagrange coordinates if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$, whereas (29) – (33) hold if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$ and $\mathrm{r}_{1,3}(t) \neq \mathrm{r}_{2,3}(t)$. Thus, $\mathbf{q}(t) = \left[\mathbf{x}_1^{\mathrm{T}}(t), \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$ is a more convenient choice of Lagrange coordinates than $\mathbf{q}(t) = \left[\mathrm{s}_1(t), \mathrm{r}_{1,1}(t), \mathrm{r}_{1,2}(t), \boldsymbol{\sigma}_1^{\mathrm{T}}(t), \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$.

This example will be further elaborated on in Section 6 for $\mathbf{q}(t) = \left[\mathbf{x}_1^{\mathrm{T}}(t),\, \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$, and hence, for notational convenience define $\mathbf{f}_{\mathrm{dyn},2}\left(\tilde{\mathbf{x}}_2(t), \mathbf{u}_{2,\mathrm{tran}}(t)\right) \triangleq \mathbf{a}\left(\tilde{\mathbf{x}}_2(t)\right) + \mathbf{u}_{2,\mathrm{tran}}(t)$ and

$$\mathbf{f}_{\mathrm{dyn},1}(\mathbf{x}_1, \mathbf{q}_{\mathrm{dot},1}(\mathbf{x}_1), \mathbf{u}_1) \triangleq \begin{bmatrix} \mathbf{a}\left(\tilde{\mathbf{x}}_1(t)\right) + \mathbf{u}_{1,\mathrm{tran}}(t) \\ -\mathbf{I}_{\mathrm{in},1}^{-1}\boldsymbol{\omega}_1^{\times}\left(\boldsymbol{\omega}_1(t)\right)\mathbf{I}_{\mathrm{in},1}\boldsymbol{\omega}_1(t) + \tilde{\boldsymbol{\omega}}_{\mathrm{i}}\left(\tilde{\mathbf{x}}_1(t)\right) + \mathbf{u}_{1,\mathrm{rot}}(t) \end{bmatrix}.$$

4.4 Path planning optimization problem revisited

The trajectory optimization problem defined in Section 3.4 can be reformulated as follows. For all $\mathrm{i} = 1, ..., \mathrm{n}$ and $t \in [t_1, t_2]$, find $\mathbf{u}_{\mathrm{i,tran}}(t)$ (respectively, $\mathbf{u}_{\mathrm{i,rot}}(t)$) among all admissible controls in $\Gamma_{\mathrm{i,tran}}$ (respectively, $\Gamma_{\mathrm{i,rot}}$) such that the performance measure (2) is minimized and $\mathbf{q}(t)$ satisfies (24), (10), and (11).

By comparing this problem statement to the problem statement given in Section 3.4, it is clear that (5) and (6) are not explicitly accounted for in the above reformulation of the optimization problem. Hence, the constrained optimization problem has been reduced to an unconstrained optimization problem by the introduction of slack variables and Lagrange coordinates.

4.5 Transversality condition

Let $\mathbf{S} : D_1 \to D_2$, where $D_1 \subset \mathbb{R}^{\mathrm{p}}$ and $D_2 \subset \mathbb{R}^{\mathrm{m}}$, be a a continuously differentiable manifold and let the *manifold tangent* to $\mathbf{S}$ at $\mathbf{y}_0$ be given by

$$\left.\frac{\partial \mathbf{S}(\mathbf{y})}{\partial \mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}_0} (\mathbf{y} - \mathbf{y}_0) = \mathbf{0}_{\mathrm{m}}. \tag{40}$$

Every vector $\mathbf{v} \in \mathbb{R}^{\mathrm{p}}$ that is normal to the manifold tangent to $\mathbf{S}$ at $\mathbf{y}_0$, that is, $\mathbf{v}^{\mathrm{T}}\mathbf{y} = 0$ for all $\mathbf{y} \in \mathbb{R}^{\mathrm{p}}$ such that (40) holds, is said to verify the *transversality condition* for $\mathbf{S}$ at $\mathbf{y}_0$.

4.6 Pontryagin's minimum principle

Assume that a set of Lagrange coordinates has been found and that the formation's dynamic equations can be written in the form given by (24). Define the *costate vectors* $\boldsymbol{\lambda}_{\mathrm{dot}} : [t_1, t_2] \to \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4}$ and $\boldsymbol{\lambda}_{\mathrm{dyn}} : [t_1, t_2] \to \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4}$ so that the *costate equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot}}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn}}(t) \end{bmatrix} = -\left(\frac{\partial}{\partial[\mathbf{q}^{\mathrm{T}},\, \mathbf{q}_{\mathrm{dot}}^{\mathrm{T}}]^{\mathrm{T}}}\begin{bmatrix} \boldsymbol{\Psi}\left(\mathbf{q}(t)\right)\dot{\mathbf{q}}(t) + \boldsymbol{\psi}\left(\mathbf{q}(t)\right) \\ \mathbf{f}_{\mathrm{dyn}}(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t)) \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot}}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn}}(t) \end{bmatrix} \tag{41}$$

holds. The boundary conditions for (41) are given in Theorem 4.2 below. Given $\lambda_0 \in \mathbb{R}$, define the *Hamiltonian function*

$$\begin{aligned} \mathfrak{h}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}_{\mathrm{dyn}}(t), \boldsymbol{\lambda}_{\mathrm{dot}}(t)\right) \triangleq{}& \lambda_0 \sum_{\mathrm{i}=1}^{\mathrm{n}} \mu_{\mathrm{i}}||\mathbf{u}_{\mathrm{i}}(t)||_2 + \boldsymbol{\lambda}_{\mathrm{dot}}^{\mathrm{T}}(t)\mathbf{q}_{\mathrm{dot}}(t) \\ &+ \boldsymbol{\lambda}_{\mathrm{dyn}}^{\mathrm{T}}(t)\mathbf{f}_{\mathrm{dyn}}(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t)). \end{aligned} \tag{42}$$

Finally, define

$$\mathfrak{m}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \boldsymbol{\lambda}_{\mathrm{dyn}}(t), \boldsymbol{\lambda}_{\mathrm{dot}}(t)\right) \triangleq \min_{\tilde{\mathbf{u}} \in \prod_{\mathrm{i}=1}^{\mathrm{n}} (\Gamma_{\mathrm{i,tran}} \times \Gamma_{\mathrm{i,rot}})} \mathfrak{h}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}_{\mathrm{dyn}}(t), \boldsymbol{\lambda}_{\mathrm{dot}}(t)\right). \tag{43}$$

The following theorem is known as the *Pontryagin minimum principle*. For details on this theorem and its numerous applications to optimal control, see Pontryagin et al. (1962).

Theorem 4.2. *(Pontryagin et al., 1962) For all* i = 1, ..., n, *let* $\mathbf{u}^*_{\mathrm{i,tran}}(t)$ *and* $\mathbf{u}^*_{\mathrm{i,rot}}(t)$, $t \in [t_1, t_2]$, *be admissible controls in* $\Gamma_{\mathrm{i,tran}}$ *and* $\Gamma_{\mathrm{i,rot}}$, *respectively, such that* $\mathbf{q}^*(t)$ *satisfies* (24), (10), *and* (11). *If* $\mathbf{u}^*_{\mathrm{i,tran}}(t)$ *and* $\mathbf{u}^*_{\mathrm{i,rot}}(t)$ *solve the trajectory optimization problem stated in Section 4.4, then there exist* $\lambda^*_0 \in \overline{\mathbb{R}}_+$, $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)$, *and* $\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)$ *such that i)* $|\lambda^*_0| + ||\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)||_2 + ||\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)||_2 \neq 0$, $t \in [t_1, t_2]$, *ii)* (41) *holds, iii)* $\mathfrak{h}\left(\mathbf{q}^*(t), \tilde{\mathbf{u}}^*(t), \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right)$ *attains its minimum almost everywhere on* $[t_1, t_2]$ *except on a finite number of points, and iv)* $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t_1)$ *and* $\boldsymbol{\lambda}^*_{\mathrm{dot}}(t_1)$ *(respectively,* $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t_2)$ *and* $\boldsymbol{\lambda}^*_{\mathrm{dot}}(t_2)$*) satisfy the transversality condition for* $\mathbf{S}_1$ *(respectively,* $\mathbf{S}_2$*) at* $\mathbf{q}^*(t_1)$ *(respectively,* $\mathbf{q}^*(t_2)$*).*

Pontryagin minimum principle is a necessary condition for optimality, and hence, it provides *candidate* optimal control vectors. Sufficient conditions for optimality that are currently available in the literature do not apply to the optimization problem discussed herein.

It is worth noting that, instead of introducing the Lagrange coordinates, the equality constraints (7) and (5) can be accounted for by introducing Lagrange multipliers. This approach requires modifying the assigned performance measure and introducing additional costate vectors (Giaquinta & Hildebrandt, 1996; Lee & Markus, 1968). The dynamics of the costate vectors are characterized by ordinary differential equations known as costate equations, which need to be integrated numerically together with the dynamic equations of the state vector. Therefore, the computational complexity of finding optimal trajectories for large formations increases drastically when Lagrange multipliers are employed (L'Afflitto & Sultan, 2010). Alternatively, finding a suitable set of Lagrange coordinates can be a demanding task and in some cases the Lagrange coordinates may not have physical meaning (Pars, 1965); however, this reduces the dimension of the costate equation and consequently reduces the computational complexity.

Finally, we say the optimization problem is *normal* if $\lambda_0 \neq 0$, otherwise the optimization problem is *abnormal*. Normality can be shown by using the *Euler necessary condition*

$$\left.\frac{\partial \mathfrak{h}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}_{\mathrm{dyn}}(t), \boldsymbol{\lambda}_{\mathrm{dot}}(t)\right)}{\partial \tilde{\mathbf{u}}}\right|_{\tilde{\mathbf{u}}=\tilde{\mathbf{u}}^*} = \mathbf{0}^{\mathrm{T}}_{6\mathrm{n}}, \tag{44}$$

where $\tilde{\mathbf{u}}^*(t) \triangleq \left[[\mathbf{u}^{*T}_{1,\mathrm{tran}}(t), \mathbf{u}^{*T}_{1,\mathrm{rot}}(t)]^{\mathrm{T}}, ..., [\mathbf{u}^{*T}_{\mathrm{n,tran}}(t), \mathbf{u}^{*T}_{\mathrm{n,rot}}(t)]^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathrm{int}\,(\prod_{\mathrm{i}=1}^{\mathrm{n}}(\Gamma_{\mathrm{i,tran}} \times \Gamma_{\mathrm{i,rot}}))$. In particular, assume, *ad absurdum*, that $\lambda_0 = 0$. Now, if (41) and (44) imply that $\boldsymbol{\lambda}_{\mathrm{dot}}(t) = \mathbf{0}_{6\mathrm{n}-\mathrm{n}_4}$ and $\boldsymbol{\lambda}_{\mathrm{dyn}}(t) = \mathbf{0}_{6\mathrm{n}-\mathrm{n}_4}$ for some $t \in [t_1, t_2]$, then assertion *i)* of Theorem 4.2 is contradicted. Therefore, $\lambda_0 \neq 0$, and hence, the optimization problem is normal. In this case, we assume without loss of generality that $\lambda_0 = 1$.

5. Analytical and numerical approaches to the optimal path planning problem

Finding minimizers to (2) subject to the constraints (3) – (6) can be formulated as a Lagrange optimization problem (Ewing, 1969), which has been extensively studied both analytically and numerically in the literature. Analytical methods rely on either Lagrange's variational approach using calculus of variations or on the direct approach. In the classical variational approach, candidate minimizers for a given performance functional can be found by applying the Euler necessary condition. In order to find the minimizers, candidate optimal solutions need to be further tested by applying the Clebsh necessary condition, Jacobi necessary condition, Weierstrass necessary condition, as well as the associated sufficient conditions (Ewing, 1969; Giaquinta & Hildebrandt, 1996).

This classical analytical approach is not practical since applying the Euler necessary condition involves solving a differential-algebraic boundary value problem, whose analytical solutions are impossible to find for many practical problems of interest. Moreover, numerical solutions to this boundary value problem are affected by a strong sensitivity to the boundary conditions (Bryson, 1975). Furthermore, verifying the Jacobi necessary condition or the Weierstrass necessary condition can be a dauting task (L'Afflitto & Sultan, 2010).

A variational approach to the optimal path planning problem for a single vehicle, known as *primer vector theory*, was addressed by Lawden (1963). Lawden's problem was formulated using the assumptions that the acceleration vector **a** induced by external forces due to the environment is function of only the position vector, the vehicle is a 3 DoF point mass, and the state and control are only subject to equality constraints (Lawden, 1963). Primer vector theory is successfully employed in spacecraft trajectory optimization (Jamison & Coverstone, 2010), orbit transfers (Petropoulos & Russell, 2008), and optimal rendezvous problems (Zaitri et al., 2010), however, vehicles are often assumed to be point masses subject to only gravitational acceleration. Among the few studies on primer vector theory applied to vehicle formations, it is worth noting the work of Mailhe & Guzman (2004), where the formation initialization problem is addressed. Applications of primer vector theory to 6 DoF single vehicles have been employed to optimize the descent on Mars (Topcu et al., 2007). These studies, however, assume that the spacecraft is subject to a constant gravity acceleration, the control variables are the translational acceleration and the angular rates, and the translational acceleration can be pointed in any direction by rotating the vehicle.

Pontryagin's minimum principle is a variational method that is equivalent to the Weierstrass necessary condition with the advantage of addressing constraints on the control more effectively than applying the classical variational approach. State constraints need to be addressed by applying an optimal switching condition on the costate equation (Pontryagin et al., 1962), which generally increases the complexity of the problem. In the present formulation, the constraints on the formation are addressed by employing Lagrange coordinates, which does not introduce further conditions on the costate vector dynamics.

The direct approach in the calculus of variations, which is more recent than the variational approach, is based on defining a minimizing sequence of control functions $\mathbf{u}_{\mathrm{n}}(t)$ in some set Γ such that $\lim_{\mathrm{n}\to+\infty} \mathbf{u}_{\mathrm{n}}(t) = \mathbf{u}(t)$ is a minimizer of the performance measure $J[\mathbf{u}(\cdot)]$. To this end, the following conditions should be met. *i*) Compactness of Γ, so that a minimizing sequence contains a convergent subsequence, *ii*) closedness of Γ, so that the limit of such a subsequence is contained in Γ, and *iii*) lower semicontinuity of the sequence

$\{\mathbf{u}_n(\cdot)\}_{n=0}^{\infty}$, that is, if $\lim_{n\to+\infty}\mathbf{u}_n(t) = \mathbf{u}(t)$, then $J[\mathbf{u}(t)] \leq \liminf_{n\to+\infty} J[\mathbf{u}_n(t)]$, $\mathbf{u}_n \in \Gamma$. Finally, it is also worth noting that approximate analytical methods can be used to solve the optimal path planning problem such as shape-based approximation methods (Petropoulos & Longuski, 2004), which are generally less effective due to the arbitrary parameterization of the minimizers (Wall, 2008).

Most of the results on the fuel consumption optimization employ numerical methods (Betts, 1998), which can be categorized as indirect or direct. Indirect numerical methods, which mimic the variational approach, suffer from high computational complexity since adjoint variables must be introduced. Alternatively, direct numerical methods are computationally more efficient, however, they require casting the given problem into a parameter optimization problem (Herman & Conway, 1987). Among the numerical methods commonly in use, it is worth mentioning genetic algorithms (Seereram et al., 2000) and particle swarm optimizers (Hassan et al., 2005).

One of the contributions of the present paper is that it extends Lawden's results on primer vector theory to formations of vehicles modeled as 6 DoF rigid bodies subject to generic environmental forces and moments by applying Pontryagin's minimum principle. As in all classical variational methods, Pontryagin's minimum principle is not suitable for numerically computing the optimal trajectory of a formation. However, Pontryagin's minimum principle allows us to draw analytical conclusions since it provides a generalization of the necessary conditions used by Lawden (1963), allows us to formally implement bounded integrable functions as admissible controls, and allows us to account for control constraints. Prussing (2010) and Marec (1979) have used Pontryagin's minimum principle to address primer vector theory using the same assumptions as Lawden (1963). In contrast, the present work provides additional analytical results for generic mission scenarios and complex environmental conditions for which numerical results can be verified. Furthermore, this paper exploits some properties of the costate space and consequently provides further insight into the formation system dynamics problem.

6. Necessary conditions for optimality of UAV formation trajectories

The following propositions are needed to develop the necessary conditions for optimality of the UAV formation problem.

Proposition 6.1. *Consider the performance measure* $J_{\text{formation}}\left[\tilde{\mathbf{u}}(\cdot)\right]$ *given by* (2)*. Then, there exists at least one* $\tilde{\mathbf{u}}^*$ *such that* $J_{\text{formation}}\left[\tilde{\mathbf{u}}^*(\cdot)\right] \leq J_{\text{formation}}\left[\tilde{\mathbf{u}}(\cdot)\right]$ *for all* $\tilde{\mathbf{u}} \in \prod_{i=1}^{n}(\Gamma_{i,\text{tran}} \times \Gamma_{i,\text{rot}})$.

Proof. Since the integrand of the performance measure (1) is a continuous function defined on the compact set $\Gamma_{i,\text{tran}} \times \Gamma_{i,\text{rot}}$, it follows from Weierstrass' theorem that (1) has a global minimizer on $\Gamma_{i,\text{tran}} \times \Gamma_{i,\text{rot}}$. Now, since $\mu_i \in [0,1]$ with $\sum_{i=1}^{n}\mu_i = 1$, the result is immediate. □

Proposition 6.2. *Assume that the hypothesis of Theorem 4.1 hold. If* $\boldsymbol{\lambda}_{\text{dot}}^*(t) \in \mathcal{N}\left(\left.\frac{\partial\boldsymbol{\Psi}(\mathbf{q})}{\partial\mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}^*}\dot{\mathbf{q}}^*(t) + \left.\frac{\partial\psi(\mathbf{q})}{\partial\mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}^*}\right)$*, then the path planning problem is normal.*

Proof. First, note that the Hamiltonian function (42) can be rewritten as

$$\begin{aligned}\mathfrak{h}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t),\tilde{\mathbf{u}}(t),\boldsymbol{\lambda}_{\mathrm{dyn}}(t),\boldsymbol{\lambda}_{\mathrm{dot}}(t)\right) =&\lambda_0\sum_{\mathrm{i}=1}^{\mathrm{n}}\mu_{\mathrm{i}}||\mathbf{u}_{\mathrm{i}}(t)||_2\\ &+\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathbf{u}_{\mathrm{i,tran}}(t)\frac{\partial\mathbf{v}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial\mathbf{q}_{\mathrm{dot}}}\boldsymbol{\lambda}_{\mathrm{dyn}}(t)\\ &+\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathbf{u}_{\mathrm{i,rot}}(t)\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial\mathbf{q}_{\mathrm{dot}}}\boldsymbol{\lambda}_{\mathrm{dyn}}(t)\\ &+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn}}(t)\hat{\mathbf{f}}_{\mathrm{dyn}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t))+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dot}}(t)\mathbf{q}_{\mathrm{dot}}(t).\end{aligned}\tag{45}$$

Furthemore, note that (44) implies that

$$\lambda_0^*\sum_{\mathrm{i}=1}^{\mathrm{n}}\mu_{\mathrm{i}}\frac{\mathbf{u}_{\mathrm{i}}^{*T}(t)}{||\mathbf{u}_{\mathrm{i}}^*(t)||_2}=-\sum_{\mathrm{i}=1}^{\mathrm{n}}\left[\left.\frac{\partial\mathbf{v}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t),\quad\left.\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}\left(\mathbf{q},\mathbf{q}_{\mathrm{dot}}\right)}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)\right],$$

where $\tilde{\mathbf{u}}^*\in\mathit{int}\left(\Pi_{\mathrm{i}=1}^{\mathrm{n}}(\Gamma_{\mathrm{i,tran}}\times\Gamma_{\mathrm{i,rot}})\right)$ and where we use the subscript $(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})$ for $(\mathbf{q},\mathbf{q}_{\mathrm{dot}})=(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})$. Now, assume, *ad absurdum*, that $\lambda_0^*=0$ and note that $\frac{\partial\mathbf{v}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}=\frac{\partial\mathbf{v}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}}\frac{\partial\mathbf{q}}{\partial\mathbf{q}_{\mathrm{dot}}}$ and $\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}=\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}}\frac{\partial\mathbf{q}}{\partial\mathbf{q}_{\mathrm{dot}}}$. Since $\boldsymbol{\Psi}\left(\mathbf{q}\right)$ is diffeomorphic and Theorem 4.1 holds, it follows that $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)=\mathbf{0}_{6\mathrm{n}-\mathrm{n}_4}$. In this case, (41) can be explicitly written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\\ \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)\end{bmatrix}=-\left.\begin{bmatrix}\frac{\partial\boldsymbol{\Psi}(\mathbf{q})}{\partial\mathbf{q}}\dot{\mathbf{q}}(t)+\frac{\partial\boldsymbol{\psi}(\mathbf{q})}{\partial\mathbf{q}} & \mathbf{0}_{(6\mathrm{n}-\mathrm{n}_4)\times(6\mathrm{n}-\mathrm{n}_4)}\\ \frac{\partial\mathbf{f}_{\mathrm{dyn}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}},\tilde{\mathbf{u}})}{\partial\mathbf{q}} & \frac{\partial\mathbf{f}_{\mathrm{dyn}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}},\tilde{\mathbf{u}})}{\partial\mathbf{q}_{\mathrm{dot}}}\end{bmatrix}^{\mathrm{T}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}\begin{bmatrix}\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\\ \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)\end{bmatrix},\tag{46}$$

and hence, $\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)=\mathbf{0}_{6\mathrm{n}-\mathrm{n}_4}$, which contradicts *i*) of Theorem 4.2. □

If follows from Proposition 6.2 that the path planning optimization problem for a constrained formation is abnormal. Example 6.1 below, however, shows that this problem is normal for unconstrained 3 DoF vehicles, which is a well known result in the literature (Lawden, 1963).

Theorem 6.1. *Consider the path planning optimization problem. If* $\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathbf{u}^*_{\mathrm{i,tran}}(t)\left.\frac{\partial\mathbf{v}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}+\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathbf{u}^*_{\mathrm{i,rot}}(t)\left.\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}$ *and* $-\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)$ *are parallel, then the performance measure* (2) *is minimized. Moreover, for all* $\mathrm{i}=1,\ldots,\mathrm{n}$*, the following conditions hold.*

i) *If* $\lambda_0^*\mu_{\mathrm{i}}>\left|\left|\left.\frac{\partial\mathbf{v}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\mathrm{i,tran}}(t)=\mathbf{0}_3$.

ii) *If* $\lambda_0^*\mu_{\mathrm{i}}>\left|\left|\left.\frac{\partial\boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q},\mathbf{q}_{\mathrm{dot}})}{\partial\mathbf{q}_{\mathrm{dot}}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\mathrm{dot}})}\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\mathrm{i,rot}}(t)=\mathbf{0}_3$.

iii) *If* $\lambda_0^*\mu_{\rm i} < \left|\left| \frac{\partial \mathbf{v}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\rm i,tran}(t) = \rho_{\rm i,2}$.

iv) *If* $\lambda_0^*\mu_{\rm i} < \left|\left| \frac{\partial \boldsymbol{\omega}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\rm i,rot}(t) = \rho_{\rm i,4}$.

v) *If* $\lambda_0^*\mu_{\rm i} = \left|\left| \frac{\partial \mathbf{v}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\rm i,tran}(t)$ *is unspecified.*

vi) *If* $\lambda_0^*\mu_{\rm i} = \left|\left| \frac{\partial \boldsymbol{\omega}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$*, then* $\mathbf{u}^*_{\rm i,rot}(t)$ *is unspecified.*

Proof. It follows from (45) that $\mathfrak{h}\left(\mathbf{q}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}_{\rm dyn}(t), \boldsymbol{\lambda}_{\rm dot}(t)\right)$ is minimized if, for all i $= 1,\ldots,$n, $-\frac{\partial \mathbf{v}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\Big|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)$ is parallel to $\mathbf{u}^*_{\rm i,tran}(t)$ and if $-\frac{\partial \boldsymbol{\omega}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\Big|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)$ is parallel to $\mathbf{u}^*_{\rm i,rot}(t)$. Thus, using the triangular inequality, it follows that

$$\mathfrak{h}\left(\mathbf{q}^*(t), \tilde{\mathbf{u}}^*(t), \boldsymbol{\lambda}^*_{\rm dyn}(t), \boldsymbol{\lambda}^*_{\rm dot}(t)\right) - \boldsymbol{\lambda}^{*\rm T}_{\rm dot}(t)\mathbf{q}_{\rm dot}(\mathbf{q}^*(t)) - \boldsymbol{\lambda}^{\rm T}_{\rm dyn}(t)\hat{\mathbf{f}}_{\rm dyn}(\mathbf{q}^*(t), \mathbf{q}_{\rm dot}(\mathbf{q}^*(t)))$$
$$\leq \sum_{\rm i=1}^{\rm n}\left[\left(\lambda_0^*\mu_{\rm i} - \left|\left|\frac{\partial \mathbf{v}_{\rm i}\left(\mathbf{q},\mathbf{q}_{\rm dot}\right)}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2\right)||\mathbf{u}^*_{\rm i,tran}(t)||_2\right]$$
$$+\sum_{\rm i=1}^{\rm n}\left[\left(\lambda_0^*\mu_{\rm i} - \left|\left|\frac{\partial \boldsymbol{\omega}_{\rm i}\left(\mathbf{q},\mathbf{q}_{\rm dot}\right)}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2\right)||\mathbf{u}^*_{\rm i,rot}(t)||_2\right], \tag{47}$$

which proves i) – iv). Next, if $\lambda_0^*\mu_{\rm i} = \left|\left| \frac{\partial \mathbf{v}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$ (respectively, $\lambda_0^*\mu_{\rm i} = \left|\left| \frac{\partial \boldsymbol{\omega}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\right|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)\right|\right|_2$), then Pontryagin's minimum principle does not provide any information about the optimal control, and hence, v) and vi) hold. □

Analogous to Lawden's (Lawden, 1963) primer vector theory, $\frac{\partial \mathbf{v}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\Big|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)$ and $\frac{\partial \boldsymbol{\omega}_{\rm i}(\mathbf{q},\mathbf{q}_{\rm dot})}{\partial \mathbf{q}_{\rm dot}}\Big|_{(\mathbf{q}^*,\mathbf{q}^*_{\rm dot})} \boldsymbol{\lambda}^*_{\rm dyn}(t)$ determine the magnitude and the direction of the control forces, and hence, we denote them as the *translational primer vector* and the *rotational primer vector*, respectively. Moreover, the trajectory given by each of the cases in Theorem 6.1 are called *arcs*. For each i $=$ 1,...,n, the arcs corresponding to i) (respectively, ii)) are called *maximum translational* (respectively, *rotational*) *thrust arcs*. Similarly, arcs corresponding to iii) (respectively, iv)) are called *null translational* (respectively, *rotational*) *thrust arcs*. Finally, arcs corresponding to v) (respectively, vi)) are called *singular translational* (respectively, *rotational*) *thrust arcs*. The optimal translational and rotational control vectors for v) and vi) in Theorem

6.1 need to be deduced by applying the generalized Legendre-Clebsch condition (Giaquinta & Hildebrandt, 1996).

Theorem 6.2. *Consider the path planning optimization problem. Then, there exists* $c^* \in \mathbb{R}$ *such that*

$$\mathfrak{m}\left(\mathbf{q}^*(t), \mathbf{q}^*_{\mathrm{dot}}(t), \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right) = c^*. \tag{48}$$

Proof. It follows from the Weierstrass - Erdmann condition (Giaquinta & Hildebrandt, 1996) that on an optimal trajectory,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{h}\left(\mathbf{q}^*(t), \mathbf{q}^*_{\mathrm{dot}}(t), \tilde{\mathbf{u}}^*(t), \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right) = \frac{\partial}{\partial t}\mathfrak{h}\left(\mathbf{q}(t), \mathbf{q}^*_{\mathrm{dot}}(t), \tilde{\mathbf{u}}^*(t), \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right)$$

holds for all $t \in (t_1, t_2)$. Now, since $\mathfrak{h}$ does not explicitly depend on t, it follows that there exists $\mathrm{c}^* \in \mathbb{R}$ such that $\mathfrak{h}\left(\mathbf{q}^*(t), \mathbf{q}^*_{\mathrm{dot}}(t), \mathbf{0}_{6\mathrm{n}}, \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right) = \mathrm{c}^*$, which proves (48). □

Proposition 6.3. *Consider the costate dynamics given by* (46)*. Then, the dynamics of* $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)$ *are decoupled from the dynamics of* $\boldsymbol{\lambda}^*_{\mathrm{dot}}(t)$*.*

Proof. The result is immediate from the form of (46). □

It follows from Proposition 6.3 that the translational primer vector and the rotational primer vector dynamics are independent of the choice of $\mathbf{q}_{\mathrm{dot}}$. Moreover, in solving for $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(t)$ we need not integrate a system of $2(6\mathrm{n} - \mathrm{n}_4)$ ordinary differential equations as in (41), but rather a system of $(6\mathrm{n} - \mathrm{n}_4)$ ordinary differential equations, which is very advantageous for large formations.

Proposition 6.4. *The translational primer vector and the rotational primer vector are continuously differentiable functions.*

Proof. First, note that $\boldsymbol{\lambda}^*_{\mathrm{dyn}}(\cdot)$ and $\boldsymbol{\lambda}^*_{\mathrm{dot}}(\cdot)$ are continuous with continuous derivatives almost everywhere on $t \in (t_1, t_2)$ except for a finite number of points (Pontryagin et al., 1962). Next, the differentiability assumption on the environmental model for $\mathbf{a}(\cdot)$ and $\mathbf{m}(\cdot)$ implies that the matrix on the right-hand side of (41) is of class $\mathrm{C}^1(\mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \times \mathbb{R}^{6\mathrm{n}-\mathrm{n}_4} \times \mathbb{R}^{12\mathrm{n}})$. Hence, $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\lambda}^*_{\mathrm{dyn}}(\cdot)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\lambda}^*_{\mathrm{dot}}(\cdot)$ are continuous on (t_1, t_2). □

In order to elucidate the translational primer vector and rotational primer vector dynamics for a vehicle formation problem, we focus on specific formation configurations and on a specific environmental model. Hence, in the reminder of the paper we concentrate on the case where $\mathrm{n_v}$ components of $\mathbf{v}_\mathrm{i}$ and n_ω components of $\boldsymbol{\omega}_\mathrm{i}$ are also components of $\mathbf{q}_{\mathrm{dot}}$. A justification for this model is as follows. Assume that the i-th formation vehicle behaves as unconstrained, e.g., the first vehicle in Examples 4.1 and 4.2, or the dynamics of the i-th vehicle can be addressed as partly unconstrained, e.g., the second formation vehicle in the aforementioned examples. In either of these cases, it is natural to choose the unconstrained components of $\mathbf{v}_\mathrm{i}$ and $\boldsymbol{\omega}_\mathrm{i}$ as some of the components of $\mathbf{q}_{\mathrm{dot}}$. This model includes the classical formation configuration known as the *leader-follower* model, whose trajectories are computed as a function of the leader's path (Wang, 1991).

To simplify the environmental model assume that

$$\mathbf{a}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)=\mathbf{a}\left(\left[\mathbf{0}_3^{\mathrm{T}},\,\mathbf{v}_{\mathrm{i}}^{\mathrm{T}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)),\,\mathbf{0}_3^{\mathrm{T}},\,\mathbf{0}_3^{\mathrm{T}}\right]^{\mathrm{T}}\right),\tag{49}$$

$$\widetilde{\boldsymbol{\omega}}_{\mathrm{i}}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)=\widetilde{\boldsymbol{\omega}}_{\mathrm{i}}\left(\left[\mathbf{0}_3^{\mathrm{T}},\,\mathbf{v}_{\mathrm{i}}^{\mathrm{T}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)),\,\mathbf{0}_3^{\mathrm{T}},\,\boldsymbol{\omega}_{\mathrm{i}}^{\mathrm{T}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t))\right]^{\mathrm{T}}\right).\tag{50}$$

For notational convenience, we will refer to (49) and (50) as $\mathbf{a}(\mathbf{v}_{\mathrm{i}}(t))$ and $\widetilde{\boldsymbol{\omega}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}}(t),\boldsymbol{\omega}_{\mathrm{i}}(t))$, respectively. This assumption on the accelerations induced by external forces and external moments is justified by a common environmental model given by (Anderson, 2001)

$$\mathbf{a}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)=\mathbf{g}+||\mathbf{v}_{\mathrm{i}}(t)||_2^2\left(-k_{\mathrm{i,D}}\widehat{\mathbf{v}}_{\mathrm{i}}(t)+k_{\mathrm{i,L}}\widehat{\mathbf{v}}_{\mathrm{i}}^{\mathrm{L}}(t)-k_{\mathrm{i,S}}\widehat{\mathbf{v}}_{\mathrm{i}}^{\mathrm{S}}(t)\right),\tag{51}$$

$$\mathbf{m}\left(\widetilde{\mathbf{x}}_{\mathrm{i}}\left(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t)\right)\right)=||\mathbf{v}_{\mathrm{i}}(t)||_2^2\left(k_{\mathrm{i,R}}\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{R}}(t)+k_{\mathrm{i,P}}\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{P}}(t)+k_{\mathrm{i,Y}}\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{Y}}(t)\right),\tag{52}$$

where $\mathbf{g}$ is the constant gravitational acceleration, $\widehat{\mathbf{v}}_{\mathrm{i}}\triangleq\mathbf{v}_{\mathrm{i}}/||\mathbf{v}_{\mathrm{i}}||_2$, $\widehat{\mathbf{v}}_{\mathrm{i}}^{\mathrm{L}}:[t_1,t_2]\to\mathbb{R}^3$ (respectively, $\widehat{\mathbf{v}}_{\mathrm{i}}^{\mathrm{S}}:[t_1,t_2]\to\mathbb{R}^3$) is the unit vector in the direction of the aerodynamic lift (respectively, in the direction opposite to the aerodynamic side force), $\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{R}}:[t_1,t_2]\to\mathbb{R}^3$ (respectively, $\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{P}}:[t_1,t_2]\to\mathbb{R}^3$ and $\widehat{\boldsymbol{\omega}}_{\mathrm{i}}^{\mathrm{Y}}:[t_1,t_2]\to\mathbb{R}^3$) is the unit vector in the direction of roll (respectively, pitch and yaw), and $k_{\mathrm{i,D}}$, $k_{\mathrm{i,L}}$, $k_{\mathrm{i,S}}$, $k_{\mathrm{i,R}}$, $k_{\mathrm{i,P}}$, and $k_{\mathrm{i,Y}}$, are the drag, lift, side force, roll, pitch, and yaw coefficients, respectively.

Using the above assumptions, it follows from (22) that

$$\dot{\hat{\mathbf{v}}}_{\mathrm{i}}(t)=\hat{\mathbf{a}}(\mathbf{v}_{\mathrm{i}}(t))+\hat{\mathbf{u}}_{\mathrm{i,tran}}(t),\tag{53}$$

$$\dot{\hat{\boldsymbol{\omega}}}_{\mathrm{i}}(t)=\hat{\widetilde{\boldsymbol{\omega}}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}}(t),\boldsymbol{\omega}_{\mathrm{i}}(t))+\hat{\mathbf{u}}_{\mathrm{i,rot}}(t),\tag{54}$$

where $\hat{\mathbf{v}}_{\mathrm{i}}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n_v}}$ (respectively, $\hat{\boldsymbol{\omega}}_{\mathrm{i}}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n}_\omega}$) represents the components of $\mathbf{v}_{\mathrm{i}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t))$ (respectively, $\boldsymbol{\omega}_{\mathrm{i}}(\mathbf{q}(t),\mathbf{q}_{\mathrm{dot}}(t))$) that are also components of $\mathbf{q}_{\mathrm{dot}}(t)$, and $\hat{\mathbf{a}}:\mathbb{R}^3\to\mathbb{R}^{\mathrm{n_v}}$ and $\hat{\mathbf{u}}_{\mathrm{i,tran}}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n_v}}$ (respectively, $\hat{\widetilde{\boldsymbol{\omega}}}_{\mathrm{i}}:\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}^{\mathrm{n}_\omega}$ and $\hat{\mathbf{u}}_{\mathrm{i,rot}}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n}_\omega}$) are the corresponding components of $\mathbf{a}(\mathbf{v}_{\mathrm{i}}(t))$ and $\mathbf{u}_{\mathrm{i,tran}}(t)$ (respectively, $\widetilde{\boldsymbol{\omega}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}}(t),\boldsymbol{\omega}_{\mathrm{i}}(t))$ and $\mathbf{u}_{\mathrm{i,rot}}(t)$).

Next, it follows from (46), (53), and (54) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\mathrm{v}}}^{*}(t)\\ \boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\omega}}^{*}(t)\end{bmatrix}=-\begin{bmatrix}\left(\frac{\partial\hat{\mathbf{a}}(\mathbf{v}_{\mathrm{i}})}{\partial\hat{\mathbf{v}}_{\mathrm{i}}}\right)^{\mathrm{T}} & \left(\frac{\partial\hat{\widetilde{\boldsymbol{\omega}}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}})}{\partial\hat{\mathbf{v}}_{\mathrm{i}}}\right)^{\mathrm{T}}\\ \mathbf{0}_{\mathrm{n}_\omega\times\mathrm{n_v}} & \left(\frac{\partial\hat{\widetilde{\boldsymbol{\omega}}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}})}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right)^{\mathrm{T}}\end{bmatrix}_{(\hat{\mathbf{v}}_{\mathrm{i}}^{*},\hat{\boldsymbol{\omega}}_{\mathrm{i}}^{*})}\begin{bmatrix}\boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\mathrm{v}}}^{*}(t)\\ \boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\omega}(t)}^{*}\end{bmatrix},\tag{55}$$

where $\boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\mathrm{v}}}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n_v}}$ and $\boldsymbol{\lambda}_{\mathrm{dyn,i},\hat{\omega}(t)}:[t_1,t_2]\to\mathbb{R}^{\mathrm{n}_\omega}$ are the $\mathrm{n_v}$ and n_ω components of $\boldsymbol{\lambda}_{\mathrm{dyn,i}}^{*}(t)$ corresponding to the $\mathrm{n_v}$ and n_ω components of $\dot{\mathbf{v}}_{\mathrm{i}}(t)$ and $\dot{\boldsymbol{\omega}}_{\mathrm{i}}(t)$, respectively.

Theorem 6.3. *Assume that* $||\hat{\mathbf{u}}^*_{\mathrm{i,tran}}(t)||_2 = \hat{\rho}_{\mathrm{i,tran}}$, $||\hat{\mathbf{u}}^*_{\mathrm{i,rot}}(t)||_2 = \hat{\rho}_{\mathrm{i,rot}}$, $||\mathbf{u}^*_{\mathrm{i,tran}}(t)||_2 = \rho_{\mathrm{i,tran}}$, $||\mathbf{u}^*_{\mathrm{i,rot}}(t)||_2 = \rho_{\mathrm{i,rot}}$, *where* $\hat{\rho}_{\mathrm{i,tran}}$ *and* $\rho_{\mathrm{i,tran}} \in (\rho_{\mathrm{i},1}, \rho_{\mathrm{i},2})$, $\rho_{\mathrm{i,rot}}$ *and* $\rho_{\mathrm{i,rot}} \in (\rho_{\mathrm{i},3}, \rho_{\mathrm{i},4})$, *and* $\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}})}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}$ *is invertible. Then,*

$$\left|\left|\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]^{-\mathrm{T}}_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}\left(\mathbf{I}_3 + \left[\frac{\partial\hat{\mathbf{a}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}}\right)}{\partial\hat{\mathbf{v}}_{\mathrm{i}}}\right]^T_{\mathbf{v}^*_{\mathrm{i}}}\right)\mathbf{u}^*_{\mathrm{i,tran}}(t)\right|\right|_2 \leq \sqrt{\rho^2_{\mathrm{i,tran}} + \rho^2_{\mathrm{i,rot}}}, \tag{56}$$

$$\left|\left|\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]^{-\mathrm{T}}_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}\dot{\mathbf{u}}^*_{\mathrm{i,rot}}(t)\right|\right|_2 \leq \sqrt{\rho^2_{\mathrm{i,tran}} + \rho^2_{\mathrm{i,rot}}}. \tag{57}$$

Proof. It follows from (44), (42), (53), and (54) that

$$\begin{aligned}\lambda^*_0\mu_{\mathrm{i}}\frac{\hat{\mathbf{u}}^*_{\mathrm{i,tran}}(t)}{||\tilde{\mathbf{u}}^*(t)||_2} &= -\boldsymbol{\lambda}^*_{\mathrm{dyn,i,\hat{v}}}(t),\\ \lambda^*_0\mu_{\mathrm{i}}\frac{\hat{\mathbf{u}}^*_{\mathrm{i,rot}}(t)}{||\tilde{\mathbf{u}}^*(t)||_2} &= -\boldsymbol{\lambda}^*_{\mathrm{rot,i},\hat{\omega}(t)}(t).\end{aligned} \tag{58}$$

Recalling that $\left|\left|\frac{\hat{\mathbf{u}}^*_{\mathrm{i,rot}}(t)}{||\tilde{\mathbf{u}}^*(t)||_2}\right|\right|_2 \leq 1$ and using (55) and (58) we obtain

$$\begin{aligned}&\left|\left|\lambda^*_0\mu_{\mathrm{i}}\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]^{-\mathrm{T}}_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}\frac{||\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)||_2\hat{\mathbf{u}}^*_{\mathrm{i,tran}}(t) + \dot{\tilde{\mathbf{u}}}^{*T}_{\mathrm{i}}(t)\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)\hat{\mathbf{u}}^*_{\mathrm{i,tran}}(t)}{||\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)||^2_2}\right.\\ &\left.\qquad + \lambda^*_0\mu_{\mathrm{i}}\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]^{-\mathrm{T}}_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}\left[\frac{\partial\hat{\mathbf{a}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}}\right)}{\partial\hat{\mathbf{v}}_{\mathrm{i}}}\right]^T_{\mathbf{v}^*_{\mathrm{i}}}\frac{\hat{\mathbf{u}}^*_{\mathrm{i,tran}}(t)}{||\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)||_2}\right|\right|_2 \leq \lambda^*_0\mu_{\mathrm{i}},\\ &\left|\left|\lambda^*_0\mu_{\mathrm{i}}\left[\frac{\partial\hat{\tilde{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial\hat{\boldsymbol{\omega}}_{\mathrm{i}}}\right]^{-\mathrm{T}}_{(\mathbf{v}^*_{\mathrm{i}},\boldsymbol{\omega}^*_{\mathrm{i}})}\frac{||\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)||_2\hat{\mathbf{u}}^*_{\mathrm{i,rot}}(t) + \dot{\tilde{\mathbf{u}}}^{*T}_{\mathrm{i}}(t)\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)\hat{\mathbf{u}}^*_{\mathrm{i,rot}}(t)}{||\tilde{\mathbf{u}}^*_{\mathrm{i}}(t)||^2_2}\right|\right|_2 \leq \lambda^*_0\mu_{\mathrm{i}}.\end{aligned}$$

Now, noting that $\dot{\tilde{\mathbf{u}}}^{*T}_{\mathrm{i}}(t)\tilde{\mathbf{u}}^*_{\mathrm{i}}(t) = 0$, the result follows. □

Since Theorem 6.3 is proven using the Euler necessary condition, it follows that $(\mathbf{u}^*_{\mathrm{i,tran}}, \mathbf{u}^*_{\mathrm{i,rot}}) \in \mathrm{int}(\Gamma_{\mathrm{i,tran}} \times \mathrm{int}(\Gamma_{\mathrm{i,rot}}))$. However, the parameter bounds $\rho_{\mathrm{i,j}}$, $\mathrm{j} = 1, 2, 3, 4$, are imposed by physical and not mathematical considerations, and hence, for practical applications we can assume that there exists $\epsilon > 0$ such that Theorem 6.3 holds for $\rho_{\mathrm{i,tran}} \in (\rho_{\mathrm{i},1} - \epsilon, \rho_{\mathrm{i},2} + \epsilon)$ and $\rho_{\mathrm{i,rot}} \in (\rho_{\mathrm{i},3} - \epsilon, \rho_{\mathrm{i},4} + \epsilon)$. Consequently, for engineering applications we can assume that Theorem 6.3 also holds on arcs of maximum translational and rotational thrust.

Corollary 6.1. *Assume that the hypothesis of Theorem 6.3 hold. If* $\mathrm{n}_\omega = 0$, *then*

$$\left\| \left[\frac{\partial \hat{\mathbf{a}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}}\right)}{\partial \hat{\mathbf{v}}_{\mathrm{i}}} \right]_{\mathbf{v}_{\mathrm{i}}^*}^{-\mathrm{T}} \hat{\mathbf{u}}_{\mathrm{i,tran}}^*(t) \right\|_2 \leq \sqrt{\rho_{\mathrm{i,tran}}^2 + \rho_{\mathrm{i,rot}}^2}. \tag{59}$$

Alternatively, if $\mathrm{n}_{\mathrm{v}} = 0$, *then*

$$\left\| \left[\frac{\partial \hat{\hat{\boldsymbol{\omega}}}_{\mathrm{i}}\left(\mathbf{v}_{\mathrm{i}}, \boldsymbol{\omega}_{\mathrm{i}}\right)}{\partial \hat{\boldsymbol{\omega}}_{\mathrm{i}}} \right]_{\left(\mathbf{v}_{\mathrm{i}}^*, \boldsymbol{\omega}_{\mathrm{i}}^*\right)}^{-\mathrm{T}} \hat{\mathbf{u}}_{\mathrm{i,rot}}^*(t) \right\|_2 \leq \sqrt{\rho_{\mathrm{i,tran}}^2 + \rho_{\mathrm{i,rot}}^2}. \tag{60}$$

Proof. The proof is a direct consequence of Theorem 6.3. □

Example 6.1. Consider the formation of the two vehicles addressed in Examples 4.1 and 4.2, and assume that $\mathbf{q}(t) = \left[\mathbf{x}_1^{\mathrm{T}}(t), \mathbf{r}_2^{\mathrm{T}}(t)\right]^{\mathrm{T}}$. As shown in Example 4.2, if $\mathrm{r}_{\min} < ||\mathbf{r}_1(t) - \mathbf{r}_2(t)||_2^2 < \mathrm{r}_{\max}$, then the first vehicle and the translational dynamics of the second vehicle can be considered unconstrained. Thus, the costate equation (41) can be rewritten as two decoupled ordinary differential equations given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot},1}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn},1}(t) \end{bmatrix} = - \begin{bmatrix} \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \frac{\partial \boldsymbol{R}_{\mathrm{rod}}^{-1}(\boldsymbol{\sigma}_1)}{\partial \boldsymbol{\sigma}_1} \dot{\boldsymbol{\sigma}}_1 \end{bmatrix}^{\mathrm{T}} & \mathbf{0}_{6\times 6} \\ \left(\frac{\partial \mathbf{f}_{\mathrm{dyn},1}(\mathbf{x}_1, \mathbf{q}_{\mathrm{dot},1}(\mathbf{x}_1), \mathbf{u}_1)}{\partial \mathbf{x}_1} \right)^{\mathrm{T}} & \left(\frac{\partial \mathbf{f}_{\mathrm{dyn},1}(\mathbf{x}_1, \mathbf{q}_{\mathrm{dot},1}(\mathbf{q}), \mathbf{u}_1)}{\partial \mathbf{q}_{\mathrm{dot},1}} \right)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot},1}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn},1}(t) \end{bmatrix}, \tag{61}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot},2}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn},2}(t) \end{bmatrix} = - \begin{bmatrix} \mathbf{0}_{3\times 3} & \frac{\partial \mathbf{f}_{\mathrm{dyn},2}(\widetilde{\mathbf{x}}_2(t), \mathbf{u}_{2,\mathrm{tran}}(t))}{\partial \mathbf{r}_2} \\ \mathbf{0}_{3\times 3} & \frac{\partial \mathbf{f}_{\mathrm{dyn},2}(\widetilde{\mathbf{x}}_2(t), \mathbf{u}_{2,\mathrm{tran}}(t))}{\partial \mathbf{v}_2} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\lambda}_{\mathrm{dot},2}(t) \\ \boldsymbol{\lambda}_{\mathrm{dyn},2}(t) \end{bmatrix}, \tag{62}$$

where $\boldsymbol{\lambda}_{\mathrm{dyn}}(t) \triangleq [\boldsymbol{\lambda}_{\mathrm{dyn},1}^{\mathrm{T}}(t)\ \boldsymbol{\lambda}_{\mathrm{dyn},2}^{\mathrm{T}}(t)]^{\mathrm{T}}$, $\boldsymbol{\lambda}_{\mathrm{dyn},1} : [t_1, t_2] \to \mathbb{R}^6$, $\boldsymbol{\lambda}_{\mathrm{dyn},2} : [t_1, t_2] \to \mathbb{R}^3$, $\boldsymbol{\lambda}_{\mathrm{dot}}(t) \triangleq [\boldsymbol{\lambda}_{\mathrm{dot},1}^{\mathrm{T}}(t), \boldsymbol{\lambda}_{\mathrm{dot},2}^{\mathrm{T}}(t)]^{\mathrm{T}}$, $\boldsymbol{\lambda}_{\mathrm{dot},1} : [t_1, t_2] \to \mathbb{R}^6$, and $\boldsymbol{\lambda}_{\mathrm{dot},2} : [t_1, t_2] \to \mathbb{R}^3$.

From (61) and (62) it follows that the path planning optimization problem for the first vehicle is possibly abnormal since we cannot verify a priori whether or not

$$\boldsymbol{\lambda}_{\mathrm{dot},1}^*(t) \in \mathcal{N}\left(\begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \frac{\partial \boldsymbol{R}_{\mathrm{rod}}^{-1}(\boldsymbol{\sigma}_1)}{\partial \boldsymbol{\sigma}_1} \Big| \dot{\boldsymbol{\sigma}}_1(\mathbf{q}(t)) \end{bmatrix}_{\mathbf{q}=\mathbf{q}^*} \right),$$

whereas the path planning optimization problem for the second vehicle is normal since its rotational dynamics are not expressed by (62). Normality for the second formation vehicle can also be proven by rewriting the unconstrained dynamic equations (3) for a 3 DoF vehicle. For details, see L'Afflitto & Sultan (2008).

Using (18) it follows that (45) can be written as

$$\begin{aligned} \mathfrak{h}\left(\mathbf{q}(t), \mathbf{q}_{\mathrm{dot}}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}_{\mathrm{dyn}}(t), \boldsymbol{\lambda}_{\mathrm{dot}}(t)\right) = {} & \mathfrak{h}_1\left(\mathbf{x}_1(t), \mathbf{u}_1(t), \boldsymbol{\lambda}_{\mathrm{dyn},1}(t), \boldsymbol{\lambda}_{\mathrm{dot},1}(t)\right) \\ & + \mathfrak{h}_2\left(\mathbf{x}_2(t), \mathbf{u}_{2,\mathrm{tran}}(t), \boldsymbol{\lambda}_{\mathrm{dyn},2}(t)\right), \end{aligned} \tag{63}$$

where

$$\begin{aligned}\mathfrak{h}_1\left(\mathbf{x}_1(t),\mathbf{u}_1(t),\boldsymbol{\lambda}_{\mathrm{dyn},1}(t),\boldsymbol{\lambda}_{\mathrm{dot},1}(t)\right) &= \lambda_0\mu_1||\mathbf{u}_1(t)||_2+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,1}(t)\mathbf{u}_{1,\mathrm{tran}}(t)\\ &\quad+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,2}(t)\mathbf{u}_{1,\mathrm{rot}}(t)+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,1}(t)\mathbf{a}\left(\widetilde{\mathbf{x}}_1(t)\right)\\ &\quad+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,2}(t)\left(\widetilde{\boldsymbol{\omega}}_1\left(\widetilde{\mathbf{x}}_1(t)\right)-\mathbf{I}^{-1}_{\mathrm{in},1}\boldsymbol{\omega}^{\times}_1\left(\boldsymbol{\omega}_1(t)\right)\mathbf{I}_{\mathrm{in},1}\boldsymbol{\omega}_1(t)\right)\\ &\quad+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dot},1,1}\mathbf{v}_1(t)-\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dot},1,2}\mathbf{R}^{-1}_{\mathrm{rod}}(\boldsymbol{\sigma}_{\mathbf{1}}(\boldsymbol{t}))\dot{\boldsymbol{\sigma}}_1(t), \qquad (64)\\ \mathfrak{h}_2\left(\mathbf{x}_2(t),\mathbf{u}_{2,\mathrm{tran}}(t),\boldsymbol{\lambda}_{\mathrm{dyn},2}(t)\right) &= \mu_2||\mathbf{u}_{2,\mathrm{tran}}(t)||_2+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},2,1}(t)\mathbf{u}_{2,\mathrm{tran}}(t)\\ &\quad+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},2,1}(t)\mathbf{a}\left(\widetilde{\mathbf{x}}_2(t)\right)+\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dot},2,1}\mathbf{v}_2(t), \qquad (65)\end{aligned}$$

where $\boldsymbol{\lambda}_{\mathrm{dyn},1}(t) \triangleq [\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,1}(t),\ \boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,2}(t)]^{\mathrm{T}}$, $\boldsymbol{\lambda}_{\mathrm{dyn},2}(t) \triangleq [\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},2,1}(t),\ \boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},2,2}(t)]^{\mathrm{T}}$, and $\boldsymbol{\lambda}_{\mathrm{dyn,j,k}}$: $[t_1,t_2] \rightarrow \mathbb{R}^3$, j, k = 1, 2. Now, using Theorem 6.3 we can construct a candidate optimal control law. Remarkably, the same candidate optimal control law can be obtained by applying Theorem 6.3 to (64) and (65) independently. The fact that the candidate optimal control law for the the first vehicle can be found independently from the second vehicle is another advantage in employing Lagrange coordinates. The minimization of $\mathfrak{h}_2$ leads to the same candidate optimal control law as given by primer vector theory with the only difference being that the arcs of maximum, null, and singular thrust are not characterized by the sign of $||\boldsymbol{\lambda}^{*}_{\mathrm{dyn},2,1}(t)||_2-1$ as in Lawden's work (Lawden, 1963) but rather by the sign of $||\boldsymbol{\lambda}^{*}_{\mathrm{dyn},2,1}(t)||_2-\mu_2$.

Singular translational thrust arcs for the first vehicle occur when

$$(\lambda_0\mu_1)^2=\boldsymbol{\lambda}^{\mathrm{T}}_{\mathrm{dyn},1,1}(t)\boldsymbol{\lambda}_{\mathrm{dyn},1,1}(t) \qquad (66)$$

and, as shown in Theorem 6.3, $\mathbf{u}^{*}_{2,\mathrm{tran}}$ cannot be found on singular arcs by applying Pontryagin's minimum principle. However, from (44) and (64), we note that $\lambda_0\mu_1\frac{\mathbf{u}^{*}_{1,\mathrm{tran}}(t)}{||\mathbf{u}^{*}_1(t)||_2}=-\boldsymbol{\lambda}^{*}_{\mathrm{dyn},1,1}(t)$, and hence, (66) yields

$$||\mathbf{u}^{*}_1(t)||^2_2=\mathbf{u}^{*\mathrm{T}}_{1,\mathrm{tran}}(t)\mathbf{u}^{*}_{1,\mathrm{tran}}(t). \qquad (67)$$

Thus, on singular translational thrust arcs for the first vehicle $\mathbf{u}^{*}_{1,\mathrm{rot}}(t)=\mathbf{0}_3$. Similarly, it can be shown that $\mathbf{u}^{*}_{1,\mathrm{tran}}(t)=\mathbf{0}_3$ on singular rotational thrust arcs for the first vehicle. Finally, singular arcs for the second vehicle occur when

$$\mu_2^2=\boldsymbol{\lambda}^{*\mathrm{T}}_{\mathrm{dyn},2,1}(t)\boldsymbol{\lambda}^{*}_{\mathrm{dyn},2,1}(t). \qquad (68)$$

From (44) and (65) it follows that $\mu_2\frac{\mathbf{u}^{*}_{2,\mathrm{tran}}(t)}{||\mathbf{u}^{*}_{2,\mathrm{tran}}(t)||_2}=-\boldsymbol{\lambda}^{*}_{\mathrm{dyn},2,1}(t)$, which satisfies (68). Hence, any admissible $\mathbf{u}_{2,\mathrm{tran}}$ can be applied on singular arcs. This was first noted by Lawden (1963).

7. Illustrative numerical example

In this section, we present a numerical example to highlight the efficacy of the framework presented in the paper. In particular, we consider the two vehicles presented in Examples

4.1, 4.2, and 6.1 with masses 0.1kg and inertia matrices $0.40\mathbf{I}_3\,\mathrm{kgm}^4$ flying in an environment modeled by (51) and (52), where $\mathbf{g} = [0, 0, -9.81]^{\mathrm{T}}\frac{\mathrm{m}}{\mathrm{s}^2}$, $k_{\mathrm{i,D}} = 0.20$, $k_{\mathrm{i,L}} = 1.20$, $k_{\mathrm{i,S}} = 0.50$, $k_{\mathrm{i,R}} = 0.30$, $k_{\mathrm{i,P}} = 0.30$, and $k_{\mathrm{i,Y}} = 0.30$, for $\mathrm{i} = 1, 2$. Furthermore, we assume that $t_1 = 0.00$ s, $t_2 = 60.00$ s, $\mathbf{r}_1(t_1) = [0.00, 0.00, 0.00]^{\mathrm{T}}$ m, $\mathbf{r}_1(t_2) = [0.90, -10.00, -1.80]^{\mathrm{T}}$ m, $\boldsymbol{\sigma}_1(t_1) = [0.00, 0.00, 0.00]^{\mathrm{T}}$, and $\boldsymbol{\sigma}_1(t_2) = [0.00, 0.00, 120.00\frac{\pi}{180.00}]^{\mathrm{T}}$. For our simulation we take $\rho_{\mathrm{i},1} = 10.00\frac{\mathrm{m}}{\mathrm{s}^2}$, $\rho_{\mathrm{i},2} = 45.00\frac{\mathrm{m}}{\mathrm{s}^2}$, $\rho_{\mathrm{i},3} = 10.00\frac{1}{\mathrm{s}^2}$, and $\rho_{\mathrm{i},1} = 20.00\frac{1}{\mathrm{s}^2}$, for $\mathrm{i} = 1, 2$. The boundary conditions for the second vehicle are deduced from (14) and (15) by assuming that $\mathrm{r}_{\max} = \frac{21}{25}$m and $\mathrm{r}_{\min} = \frac{33}{50}$m. It can be easily verified that the constraints given by (12) and (13) hold for all $t \in [t_1, t_2]$. Letting $\mu_1 = \mu_2 = \frac{1}{2}$ and applying Theorem 6.1, we obtain the optimal trajectory shown in Figure 1. Figures 2 and 3 show the optimal control as a function of the norm of the translational primer vector and the rotational primer vector, as well as time, respectively. For this example $\mathrm{J}[\mathbf{u}_1(\cdot)] = 10.00\frac{\mathrm{m}}{\mathrm{s}}$ and $\mathrm{J}[\mathbf{u}_2(\cdot)] = 11.60\frac{\mathrm{m}}{\mathrm{s}}$. Since $\mathfrak{m}\left(\mathbf{q}^*(t), \mathbf{q}^*_{\mathrm{dot}}(t), \boldsymbol{\lambda}^*_{\mathrm{dyn}}(t), \boldsymbol{\lambda}^*_{\mathrm{dot}}(t)\right) = 22.30\frac{\mathrm{m}}{\mathrm{s}^2}$, Theorem 6.2 holds. Finally, Figure 4 shows the translational primer vector and the rotational primer vector of the first vehicle as a function of time.

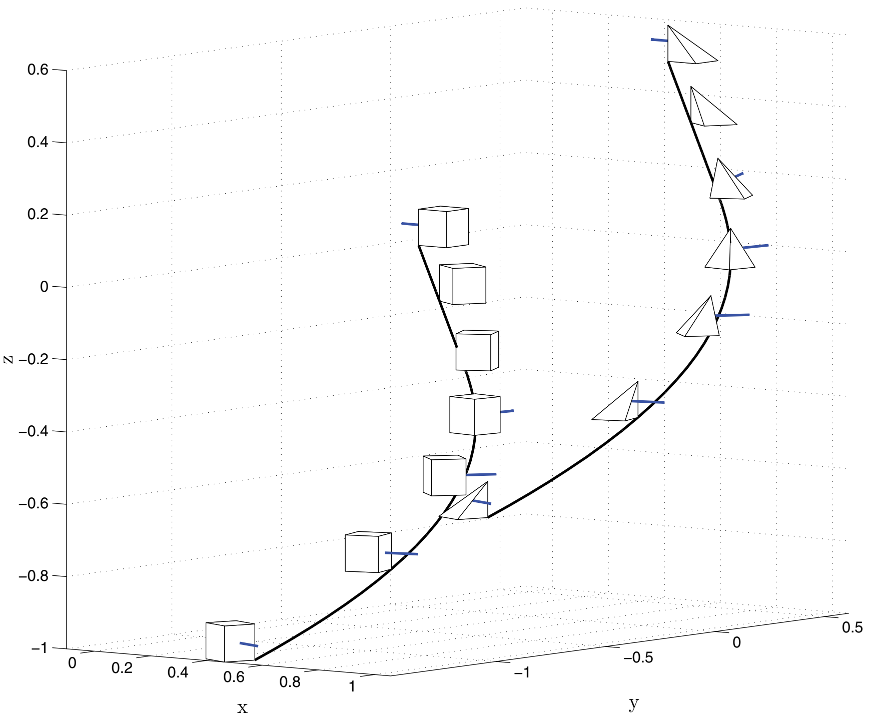


Fig. 1. Optimal trajectories for vehicles 1 and 2. The cube represents the first vehicle and the prism represents the second vehicle.

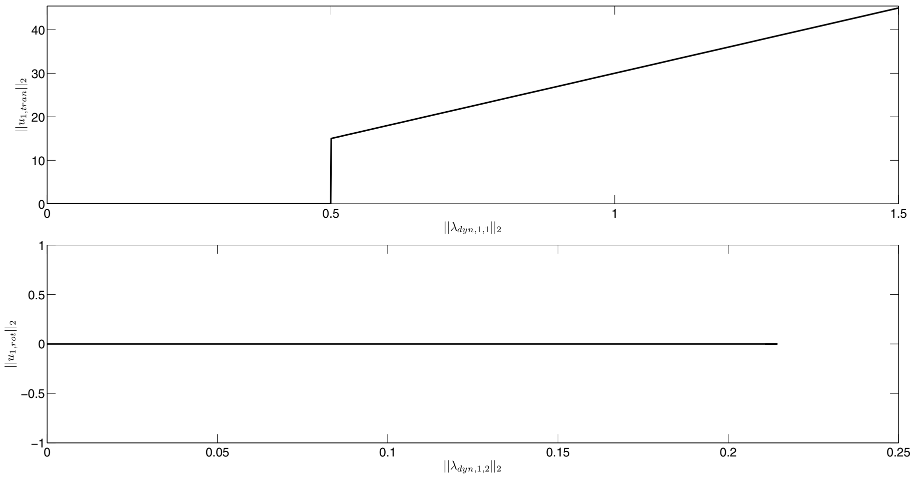


Fig. 2. Optimal control for the first vehicle as function of the norm of the translational primer vector and the rotational primer vector.

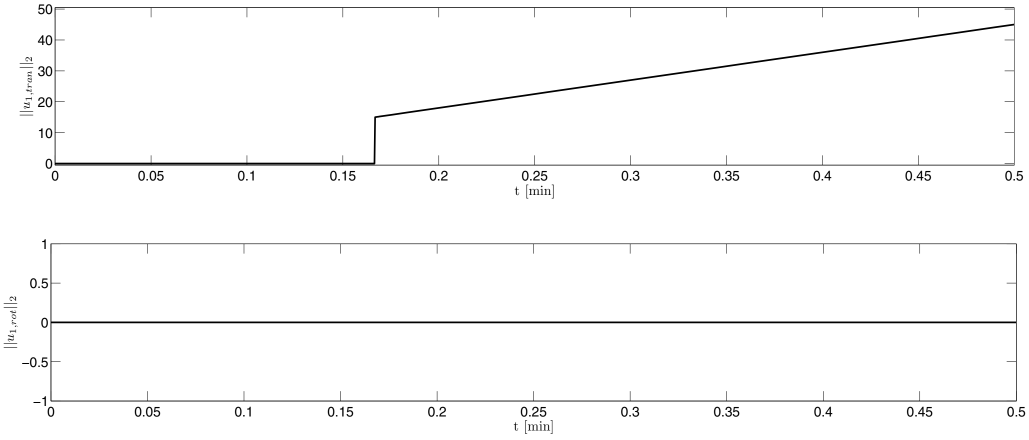


Fig. 3. Optimal control for the first vehicle as function of time.

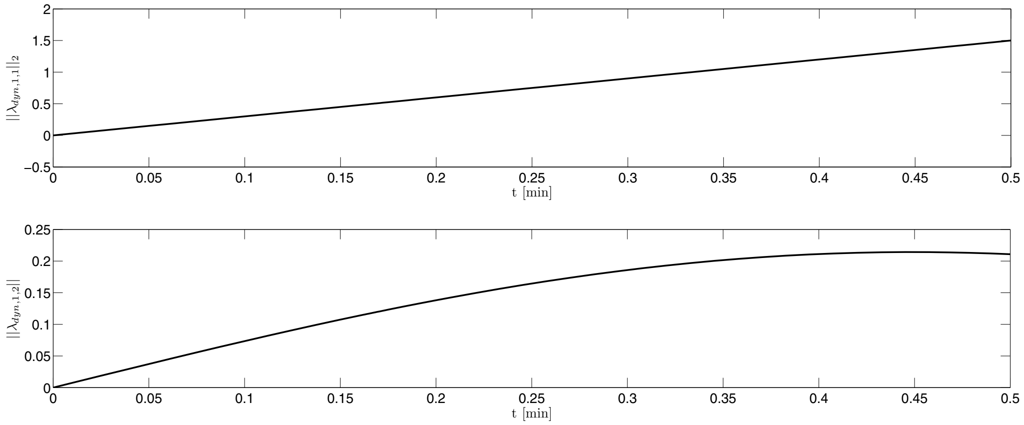


Fig. 4. Translational and rotational primer vector norms as functions of time for the first vehicle.

8. Conclusion and recommendations for future research

In this paper, we addressed the problem of minimizing the control effort needed to operate a formation of n UAVs. Specifically, a candidate optimal control law as well as necessary conditions for optimality that characterize the resulting optimal trajectories are derived and discussed assuming that the formation vehicles are 6 DoF rigid bodies flying in generic environmental conditions and subject to equality and inequality constraints. The results presented extend Lawden's seminal work (Lawden, 1963) and several papers predicated on his work.

An illustrative numerical example involving a formation of two vehicles is provided to illustrate the mathematical path planning optimization framework presented in the paper. Furthermore, we show that our framework is not restricted to UAV formations and can be applied to formations of robots, spacecraft, and underwater vehicles.

The results of the present paper can be further extended in several directions. Specifically, an analytical study of the translational primer vector and the rotational primer vector can be useful in identifying numerous properties of the formation's optimal path. In particular, the translational primer vector and the rotational primer vector can be used to measure the sensitivity of the candidate optimal control law to uncertainties in the dynamical model. In this paper, we provide a generic formulation to the optimal path planning problem in order to address a large number of formation problems. However, specializing our results to a particular formation and a particular environmental model can lead to analytical tools that can be amenable to efficient numerical methods. Additionally, nonholonomic constraints have not been accounted in our framework and can be addressed by modifying Theorem 4.1. Finally, in this paper, we penalize vehicle control effort by tuning the constants $\mu_1,...,\mu_n$ in (2). In many practical applications, however, it is preferable to trade-off the control effort in a formation of vehicles by optimizing over the free parameters $\mu_1,...,\mu_n$.

9. Acknowledgments

The first-named author would like to thank Drs. C. Sultan and E. Cliff at Virginia Polytechnic Institute and State University for several helpful discussions. This research was supported in part by the Domenica Rea d'Onofrio Fellowship Foundation and the Air Force Office of Scientific Research under Grant FA9550-09-1-0429.

10. References

Ambrosia, V. & Hinkley, E. (2008). NASA science serving society: Improving capabilities for fire characterization to effect reduction in disaster losses, *IEEE International Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008.*, Vol. 4, pp. IV –628 –IV –631.

Anderson, J. D. (2001). *Fundamentals of Aerodynamics*, McGraw Hill, New York, NY.

Bataillé, B., Moschetta, J. M., Poinsot, D., Bérard, C. & Piquereau, A. (2009). Development of a VTOL mini UAV for multi-tasking missions, *The Aeronautical Journal* 13: 87–98.

Betts, J. T. (1998). Survey of numerical methods for trajectory optimization, *AIAA Journal of Guidance, Control, and Dynamics* 21: 193–207.

Blackmore, L. (2008). Robust path planning and feedback design under stochastic uncertainty, *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Honolulu, HI.

Bryson, A. E. (1975). *Applied Optimal Control*, Hemisphere, New York, NY.

Ewing, E. G. (1969). *Calculus of Variations with Applications*, Dover Edition, New York, NY.

Giaquinta, M. & Hildebrandt, S. (1996). *Calculus of Variations I*, Springer-Verlag, Berlin, Germany.

Greenwood, T. D. (2003). *Advanced Dynamics*, Cambridge University Press, New York, NY.

Haddal, C. C. & Gertler, J. (2010). Homeland security: Unmanned aerial vehicles and border surveillance, *Technical Report RS21698*, Congressional Research Service, Washington, D.C.

Hassan, R., Cohanim, B. & de Weck, O. (2005). A comparison of particle swarm optimization and the genetic algorithm, *Proceedings of the 46th AIAA Structures, Structural Dynamics and Materials Conference*, AIAA, Breckenridge, CO.

Herman, A. L. & Conway, B. A. (1987). Direct optimization using nonlinear programming and collocation, *AIAA Journal of Guidance, Control, and Dynamics* 10: 338–342.

Jacobson, D. & Lele, M. (1969). A transformation technique for optimal control problems with a state variable inequality constraint, *IEEE Transactions on Automatic Control* 14(5): 457–464.

Jamison, B. R. & Coverstone, V. (2010). Analytical study of the primer vector and orbit transfer switching function, *AIAA Journal of Guidance, Control, and Dynamics* 33: 235–245.

Jang, J. S. & Tomlin, C. J. (2005). Control strategies in multi-player pursuit and evasion game, *Proceeding AIAA Guidance, Navigation, and Control Conference*, AIAA, San Francisco, CA.

L'Afflitto, A. & Sultan, C. (2008). Applications of calculus of variations to aircraft and spacecraft path planning, *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Chicago, IL.

L'Afflitto, A. & Sultan, C. (2010). On calculus of variations in aircraft and spacecraft formation flying path planning, *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Toronto, Canada.

Lawden, D. F. (1963). *Optimal Trajectories for Space Navigation*, Butterworths, London, UK.

Lee, E. B. & Markus, L. (1968). *Foundations of Optimal Control Theory*, Wiley, New York, NY.

Lillesand, T., Kiefer, R. W. & Chipman, J. (2007). *Remote Sensing and Image Interpretation*, Wiley, New York, NY.

Mailhe, L. & Guzman, J. (2004). Initialization and resizing of formation flying using global and local optimization methods, *Proceedings IEEE Aerospace Conference*, Vol. 1, pp. 547–556.

Majewski, S. E. (1999). Naval command and control for future UAVs. MS Thesis, Naval Postgraduate School, Monterey, CA.

Marec, J. P. (1979). *Optimal Space Trajectories*, Elsevier, New York, NY.

Neimark, J. I. & Fufaev, N. A. (1972). *Dynamics of Nonholonimic Systems*, American Mathematical Society, New York, NY.

Oyekan, J. & Huosheng, H. (2009). Toward bacterial swarm for environmental monitoring, *IEEE International Conference on Automation and Logistics*, pp. 399 –404.

Pars, L. A. (1965). *A Treatise on Analytical Dynamics*, Wiley, New York, NY.

Petropoulos, A. E. & Longuski, J. M. (2004). Shape-based algorithm for automated design of low-thrust, gravity-assist trajectories, *AIAA Journal of Guidance, Control, and Dynamics* 32: 95–101.

Petropoulos, A. E. & Russell, R. P. (2008). Low-thrust transfers using primer vector theory and a second-order penalty method, *Proceedings of the AIAA Astrodynamics Specialist Conference*, AIAA, Honolulu, HI.

Plnes, D. & Bohorquez, F. (2006). Challenges facing future micro-air-vehicle development, *AIAA Journal of Aircraft* 43: 290–305.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V. & Mishchenko, E. F. (1962). *The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York, NY.

Prussing, J. E. (2010). Primer vector theory and applications, *in* B. A. Conway (ed.), *Spacecraft Trajectory Optimization*, Cambridge University Press, Chicago, IL, pp. 155–188.

Ramage, J., Avalle, M., Berglund, E., Crovella, L., Frampton, R., Krogmann, U., Ravat, C., Robinson, M., Shulte, A. & Wood, S. (2009). Automation technologies and application considerations for highly integrated mission systems, *Technical Report TR-SCI-118*, North Atlantic Treaty Organisation.

Scharf, D., Hadaegh, F. & Ploen, S. (2003a). A survey of spacecraft formation flying guidance and control (part 1): Guidance, *Proceedings of the American Control Conference*, pp. 1733 – 1739.

Scharf, D., Hadaegh, F. & Ploen, S. (2003b). A survey of spacecraft formation flying guidance and control (part 2): Control, *Proceedings of the American Control Conference*, pp. 1740 – 1748.

Schouwenaars, T., Feron, E. & How, J. (2006). Multi-vehicle path planning for non-line of sight communication, *Proceedings of the American Control Conference*, pp. 5758–5762.

Seereram, S., Li, E., Ravichandran, B., Mehra, R. K., Smith, R. & Beard, R. (2000). Multispacecraft formation initialization using genetic algorithm techniques, *Proceedings of the 23rd Annual AAS Guidance and Control Conference*, AAS, Breckenridge, CO.

Shanmugavel, M., Tsourdos, A. & White, B. (2010). Collision avoidance and path planning of multiple UAVs using flyable paths in 3D, *15th International Conference on Methods and Models in Automation and Robotics*, pp. 218–222.

Shuster, M. D. (1993). Survey of attitude representations, *Journal of the Astronautical Sciences* 11: 439–517.

Topcu, U., Casoliva, J. & Mease, K. D. (2007). Minimum-fuel powered descent for Mars pinpoint landing, *AIAA Journal of Spacecraft and Rockets* 44(2): 324–331.

Valentine, F. A. (1937). The problem of Lagrange with differential inequalities as added side conditions, *in* G. A. Bliss (ed.), *Contributions to the Calculus of Variations*, Chicago University Press, Chicago, IL, pp. 407–448.

Wall, B. J. (2008). Shape-based approximation method for low-thrust trajectory optimization, *Proceedings of the AIAA Astrodynamics Specialist Conference*, AIAA, Honolulu, HI.

Wang, P. K. C. (1991). Navigation strategies for multiple autonomous mobile robots moving in formation, *Journal of Robotic Systems* 8: 177 – 195.

Zaitri, M. K., Arzelier, D. & Louembert, C. (2010). Mixed iterative algorithm for solving optimal impulsive time-fixed rendezvous problem, *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Toronto, Canada.

Measuring and Managing Uncertainty Through Data Fusion for Application to Aircraft Identification System

Peter Pong[1] and Subhash Challa[2]
[1]*Jacobs Australia / University of Melbourne*
[2]*NICTA Victoria Research Laboratory / University of Melbourne*
Australia

1. Introduction

Despite the use of modern Identification Friend Foe (IFF) technology, aircraft recognition remains problematic even though a great deal of research effort has already been invested in this area. In the military context, IFF identification is supposed to be initiated when the interrogator transmits a signal to the aircraft and friendly aircraft are 'supposed' to reply to the signal by transmitting an identification code to the interrogator. Hostile aircraft often become unresponsive to the interrogator because it is either does not have the appropriate transponder or is trying to avoid being identified as an unfriendly aircraft. In the civilian air transport system, the Secondary Surveillance Radar (SSR) allows the location of the civilian aircraft being transmitted (through transponder) to the Air Traffic Controller (ATC). However, in extreme incidents, such as the attacks on the World Trade Center on 11th September 2001, the SSR transponders were manually disabled, which prevented the ATC detecting flight path alternation. To avoid the drawback of the transponder based aircraft identification system, the technique of Non-Cooperative Target Recognition (NCTR) has become a useful technology, because it does not require the participation of friendly aircraft. The NCTR technique relies primarily on the ground based target classification technology. In a typical classification problem, the goal is to develop a classifier that is capable to discriminate targets. This technology shares a great deal of similarity with the modern Electronics Support Measures (ESM) system that often employs as a Radar Warning Receiver (RWR) for modern military aircraft self-protection. Acknowledging the number of successful classifier technologies reported in this area, the goal of this work is not to propose any new algorithm to enhance the classification technology. Instead, a novel method, based on uncertainty measures, is introduced to improve the classification function by employing a data fusion technique. Data fusion applying evidential reasoning framework is a well established technique to fuse diverse sources of information. A number of fusion methods within this formalism were introduced including Dempster-Shafer Theory (DST) Fusion, Dezert Samarandche Fusion (DSmT), and Smets' Transferable Belief Model (TBM) based fusion. However, the impact of fusion on the level of uncertainty within these techniques was not studied in detail. While the use of Shannon entropy with the Bayesian fusion is well understood, the measures of uncertainty within the Dempster-Shafer formalism is not widely regarded. In this paper, an uncertainty based technique is proposed to quantify the evolution of DST fusion. This technique is then utilised to determine the optimal combination of sensor information to achieve the least uncertainty in the context of the aircraft identification problem using sensors operating the NCTR technique.

2. Background

Information fusion is often used as a data-processing technique to integrate uncertain information from multiple sensors. Information often contains uncertainties, which are usually related to physical constrains, detection algorithms and the transmitting channel of the sensors. Whilst the intuitive approaches, such as Dempster-Shafer Fusion (Shafer, 1976), Dezert Samarandche Fusion (DSmT)(Dezert & Smarandache, 2006) and Smets' Transferable Belief Model (TBM) (B.Ristic & P.Smets, 2005) aggregate all available information, these approaches do not always guarantee optimum results. Acknowledging that these techniques have associated measurement costs, the essence is to derive a fusion technique to minimise global uncertainties.

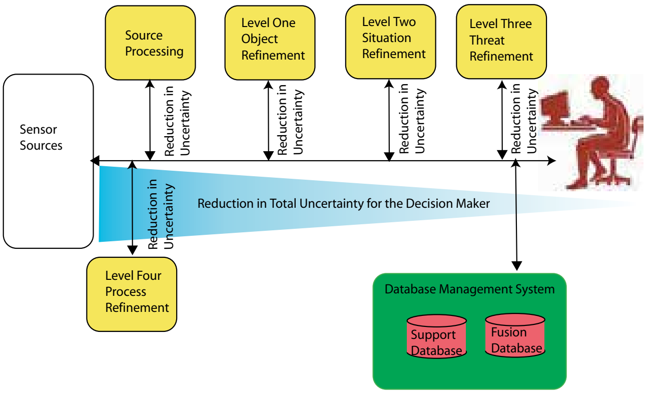

Source Processing
Level One Object Refinement
Level Two Situation Refinement
Level Three Threat Refinement
Sensor Sources
Reduction in Uncertainty
Reduction in Uncertainty
Reduction in Uncertainty
Reduction in Uncertainty
Reduction in Uncertainty
Reduction in Total Uncertainty for the Decision Maker
Level Four Process Refinement
Database Management System
Support Database
Fusion Database

Fig. 1. JDL Model and Uncertainty

In the aerospace community, there is an increasing trend to automate decision processes based on information fusion techniques. As an example, fighter pilots may rely on various forms of data fusion models to assist in assessing the current situations, when uncertain information co-exists at all levels of fusion. Considering the many data fusion models, the Joint Defence Laboratory (JDL) model (Hall & Llinas, 2001) is one the most commonly referred frameworks, which consists of Level 1 Object Assessment, Level 2 Situation Assessment, Level 3 Impact Assessment and Level 4 Process Refinement. The decision maker is supposed to treat the JDL model at 4 independent levels of functions, however, each level of fusion often includes unavoidable uncertainties. That means any aircraft identification system employing real-time situation analysis technology is required to manage uncertainty in the most effective manner. The techniques based on statistical models employed in aircraft tracking were widely acknowledged, but the methods based on uncertainty measures for target identification are not well understood in the aviation community. In recognition of this deficiency, this paper explores a novel aircraft identification technique by leveraging a new uncertainty based fusion concept.

The new concept introduced in this work explores a number of uncertainty measures under the reasoning framework and attempts to introduce a methodology to manage uncertainty variation under the DST based fusion. An example derived from an Aircraft Identification (AI)

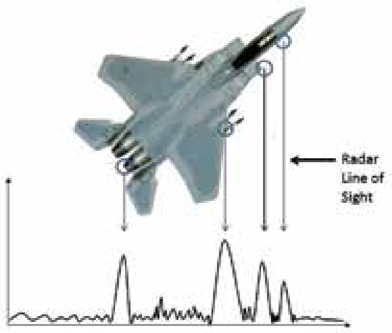


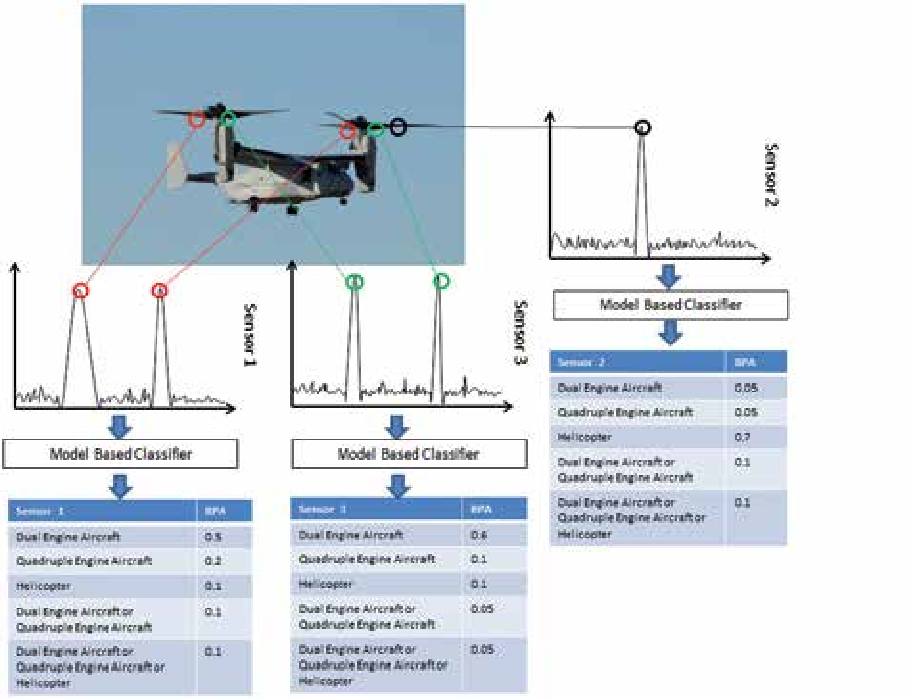
Fig. 2. Example of a radar range profile of a fighter aircraft

system is employed to demonstrate the characteristics of uncertainty variation. In terms of target tracking, significant advancements have been made in the past two decades to improve tracking technology by employing sophisticated data fusion techniques. Some of the earlier works went even further by incorporating Target Identification information, such as IFF data, to improve the overall track quality (Leung & Wu, 2000), (Carson & Peters, 26-30 Oct 1997), (Bastiere, 1997), and (Perlovsky & Schoendorf, 1995). When legitimate statistical information is presented, the techniques employed by tracking and identification using IFF information are relatively mature. However, when conflicting information is presented to the NCTR system, most techniques employed today may find it difficult to discriminate the contradicting information. In this work, we propose a technique based on uncertainty measures to resolve this problem. The employment of uncertainty in recent aviation research was reported in areas, such as air traffic control (Porretta & Ochieng, 2010), navigation (Deng & Liu, 2011) and airport surface movement management (Schuster & Ochieng, 2011), however, all these works essentially model uncertainty based on the target statistical characteristics, such as model based classified illustrated in Figure 2. Instead of treating uncertainty implicitly using their statistical values, the concept proposed in this work treats uncertainty measures directly as input parameters. In this way, we could explicitly quantify the fusion performance to make the best target identification.

3. Sensor selection and decision making

Information fusion is often perceived to produce improved decision. This assumption is generally true when sensor availability is limited, however, one has to question whether fusing all available data guarantee synergy. The focus of this work is on the reduction of uncertainties by expressing the relevant uncertainties in the reasoning system and utilise these measures to achieve the best information fusion strategy. In order to develop an uncertainty based information fusion in the aircraft identification context, the authors argue that the best fusion decision can only be observed when (i) the information fusion could provide the least ambiguous choice, (ii) the result produced by the fusion system induces the least vague answer under the reasoning framework, and (iii) the final recommendation provided by the fusion system has the fewest uncertainties. These three axioms underlying this paper are used to define the best fusion configuration. It is apparent that the goal of uncertainty based fusion is to choose the result with the least uncertainty. A fusion process based on uncertainties has the potential to lead to a biased result. However, it is difficult to neglect a decision based on information fusion when it is the least uncertain, least ambiguous and the most defined answer when compared with other potential solutions.

Figure 3 depicts an illustrative example where an aircraft identification scenario is considered. Assuming a model based classifier is employed to identify three kinds of aircraft types - Dual engines aircraft (D), Quadruple engines aircraft (Q) and Helicopter (H). Also assuming that the sensors produced an "unknown" state in the form of {D, Q, H}, where the decision of the aircraft type is not possible to be classified. Three sensors are utilised in this example to simplify the demonstration, where a classification value based on Basic Probability Assignment (BPA) are given to each of the classification reports with details also summarised in Figure 3.



Sensor 1	BPA
Dual Engine Aircraft	0.5
Quadruple Engine Aircraft	0.2
Helicopter	0.1
Dual Engine Aircraft or Quadruple Engine Aircraft	0.1
Dual Engine Aircraft or Quadruple Engine Aircraft or Helicopter	0.1

Sensor 3	BPA
Dual Engine Aircraft	0.6
Quadruple Engine Aircraft	0.1
Helicopter	0.1
Dual Engine Aircraft or Quadruple Engine Aircraft	0.05
Dual Engine Aircraft or Quadruple Engine Aircraft or Helicopter	0.05

Sensor 2	BPA
Dual Engine Aircraft	0.05
Quadruple Engine Aircraft	0.05
Helicopter	0.7
Dual Engine Aircraft or Quadruple Engine Aircraft	0.1
Dual Engine Aircraft or Quadruple Engine Aircraft or Helicopter	0.1

Fig. 3. Multi-sensor aircraft classification

If the identification process performed by each sensor is independent, information provided by Sensor 2 is clearly contradicting with Sensor 1 and Sensor 3. The errors can be induced by the incorrect scatter angle, or simply estimated by an inaccurate model. Based on the axioms discussed, it is observed that fusing Sensor 1 and Sensor 2, or Sensor 2 and Sensor 3 under DST (which be discussed in the next section) will not produce a pronounced result to identify the aircraft type. The result of the fusion is illustrated in Table 1, where only the combination of Sensor 1 and Sensor 3 could provide an unambiguous fusion result. This example highlights the criticality of uncertainty measures in relation to the standard DST fusion process. Section 5 and Section 6 of this paper provide an empirical uncertainty measures analysis in the

	Sensor 1&2	Sensor 1&3	Sensors 2&3
D,Q,H	0.026	0.022	0.0356
D,Q	0.0779	0.037	0.0595
H	0.3506	0.7704	0.3810
Q	0.1558	0.1185	0.0833
D	0.3896	0.0519	0.4405

Table 1. Sensor fusion example with contradicted information

reasoning framework, and provides an insight into how this method can be applied in an aircraft identification capability.

4. Evidential reasoning framework

The notion of Basic Probability Assignment (BPA) (Shafer, 1976) is defined with respect to a finite universe of propositions or frame of discernment, Ω. The sum of the probabilities assigned to all subsets of Ω and all propositions which support Ω must be in unity, as such BPA is a function from the set of subsets, 2^{Ω}, of Ω to the unit interval $[0, 1]$. In accordance with the convention proposed by Shafer (Shafer, 1976):

$$m(\emptyset) = 0 \tag{1}$$

and

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{2}$$

The *subset* A of Ω such that $m(A) > 0$ is called a *focal element* of m, and $\emptyset$ is the empty set. Whilst the summation of BPA must be unity, it is not manditory for the BPA of a proposition A and its negation $\overline{A}$ sum to unity.

4.1 Belief and plausibility measures

The idea of linking belief with evidential measures was first discussed by Shafer, and the idea of Belief function in reference to the BPA is defined as,

Definition 1. *Bel:* $2^{\Omega} \rightarrow [0, 1]$ *is a belief function over* Ω *if it satisfies:*

- $Bel(\emptyset) = 0$
- $Bel(\Omega) = 1$
- *for every integer* $n > 0$ *and collection of subsets* $A_1, ..., A_n$ *of* Ω

$$Bel(A_1 \cup ... \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap ... \cap A_n)$$

BPA gives a measure of support that is assigned exactly to the focal elements of a given frame of discernment. In order to aggregate the total belief in a subset A, the extent to which all the available evidence supports A, one needs to sum together the BPAs of all the subsets of A for a belief measurement.

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad \forall A \subseteq \Omega \tag{3}$$

The remaining evidence may not necessarily support the negation $\overline{A}$. In fact some of them may be assigned to propositions which are not disjointed from A, and hence, could be plausibly transferred directly to A for further information. Shafer called this the plausibility of A:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega \tag{4}$$

4.2 Dempster-Shafer fusion under an iterative process

Dempster′s rule of combination forms a new body of evidence with which the focal elements are all non-empty intersections $X \cap Y$. Given any $S \subseteq U$ there are many pairs $X, Y \subseteq U$ such that $X \cap Y = S$ and so the total weight of agreement assignable to the focal subset $X \cap Y$ is $\sum_{X \cap Y = S} m(X)m'(Y)$. Once normalising the agreement with the "non-conflicting values" $(1 - K)$, Dempster′s rule of combination for imprecise evidence becomes,

$$(m * m')(S) = \frac{1}{1-K} \sum_{X \cap Y = S} m(X)m'(Y) \tag{5}$$

for all $\emptyset \neq S \subseteq U$. The *conflict* between two bodies of evidence m, m' is the total weight of contradiction between the events of m and the events of m':

$$K(m, m') = \sum_{X \cap Y = \emptyset} m(X)m'(Y) \tag{6}$$

The quantity $1 - K$ is the cumulative degree to which the two bodies of evidence do not contradict with each other and is called the *agreement* between m and m'. In general evidential theory, Dampster-Shafer rules, belief functions, plausibility functions and BPA forms a suite of significant tools to construct probabilities through carefully modelled evidence. Through this combination process, two new measurement values - *non-specificity* and *conflict*, are also generated as a by-product. An empirical analysis is presented in Section 5 in conjunction with the theory of Aggregated Uncertainty (AU) and the recently proposed generalised Total Uncertainty (TU) measures.

5. Uncertainty measures within the evidential reasoning framework

While the classical uncertainties are often measured by the Hartley and Shannon functions, the two functions are tailored for different purposes. In order to cater for both uncertainties, evidential based uncertainty measures are adopted. Two types of classical evidential based uncertainties - non-specificity and conflict - are often measured as part of the DST fusion (Harmanec, 1996). In this section, an overview is introduced to the concept of Hartley Uncertainty measures, Aggregrated Uncertainty (AU) measures and Total Uncertainty (TU) measures which was proposed by Klir (Klir, 2006). This analysis covers the context of the DST fusion system and their subsequent implication. A practical example based on aircraft identification applying uncertainty measures as sensor discrimination matrices is discussed in Section 7 to verify our observations.

5.1 Hartley uncertainty

The technique of uncertainty measures was first addressed by Shannon. Under his proposal, the way to quantify uncertainty measures expressed by a probability distribution function p

on a singleton set is in the form of,

$$-c \sum p(x) \log_b p(x) \tag{7}$$

where b and c are positive constants, and $b \neq 1$. While this technique is useful to apply in sensor management system operating under the probabilistic framework, it cannot be used under a finite set condition. An alternative is to employ the legacy Hartley measures (Hartley, n.d.), where it seems to be the only meaningful way to measure uncertainty in the form of,

$$c \log_b \sum_{x \in \Omega} r_A(x) \tag{8}$$

or alternatively

$$c \log_b |A| \tag{9}$$

where A is a finite set and $|A|$ is the cardinality of the finite set. b and c are positive constants, and $b \neq 1$. When uncertainty is measured in *bits*, $c \log_b 2 = 1$. Harley uncertainty measures, H, defined for any basic possibility functions, r_A,

$$H(r_A) = \log_2 |A| \tag{10}$$

On closer examination of (10), $H(r_A)$ is a measure directly related to the specificity of a finite set. In other words, the larger the size of a set, the less specific the measurement becomes. This type of measures was defined as *non-specificity* by Klir (Klir, 2006). In the reasoning framework, Hartley Measures are usually treated as a weighted average of all the focal subsets in the form of BPA function (Klir, 2006).The concept of generalised Harley measures in the context of DST framework is thus defined by the function,

$$GH(m) = \sum_{A \in \Omega} m(A) \log_2 |A| \tag{11}$$

where Ω is the superset of the focal elements.

5.2 Aggregated uncertainty measures

Suppose the goal of information fusion is to reduce global uncertainties, Harmanec (Harmanec, 1996) was the first to explore the concept of uncertainty measures in the DST framework. The idea of AU uncertainty measures was proposed as the optimum uncertainty measures technique under the DST domain, because it is the only way to incorporate the value of non-specificity and conflict simultaneously, which often coexist in the DST framework.

Definition 2. *The measure of the Aggregated Uncertainty contained in Bel, denoted as $AU(Bel)$, is defined by*

$$AU(Bel) = \max\{-\sum_{x \in \Omega} p_x \log_2 p_x\} \tag{12}$$

where the maximum is taken over all $\{p_x\}_{x \in \Omega}$ such that $p_x \in [0,1]$ for all $x \in \Omega$, $\sum_{x \in \Omega} p_x = 1$ and for all $A \subseteq \Omega$, $Bel(A) \leq \sum_{x \in A} p_x$.

Although the AU technique is not an efficient algorithm, it does satisfy all the properties defined as uncertainty measures (Harmanec, 1996), and specifically, the subadditivity/additivity characteristics.

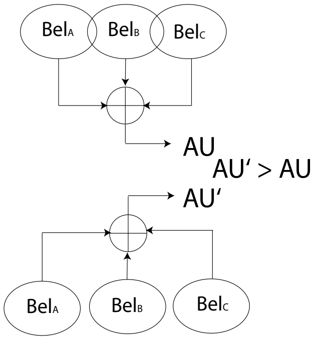


Fig. 4. Additivity and Subadditivity

Subadditivity. If Bel is an arbitrary joint belief function on $X \times Y$ and the associated marginal belief functions are Bel_X and Bel_Y, then

$$AU(Bel) \leq AU(Bel_X) + AU(Bel_Y) \tag{13}$$

Additivity. If Bel is a joint belief function on $X \times Y$, and the marginal belief functions Bel_X and Bel_Y are noninteractive, then

$$AU(Bel) = AU(Bel_X) + AU(Bel_Y) \tag{14}$$

The property of additivity/subadditivity of AU call forth the assumption that uncertainties could be reduced if sensors share common interaction prior the information fusion process occurring. Assuming sensor dependency exists among Bel_A, Bel_B and Bel_C, the characteristics of the resultant uncertainty under an evidential fusion system is illustrated pictorially in Figure 4. The algorithm to compute AU uncertainty was originated by Harmanec (Harmanec, 1996). Under the proposed algorithm, the input is treated in the form of a frame of discernment X, with a belief function Bel on X. This algorithm's computation completes once a finite number of steps have been taken and the output is the correct value of the function $AU(Bel)$, since $\{p_x\}_{x \in X}$ maximises the Shannon entropy within the constraints induced by Bel.

5.3 Total uncertainty measures

The concept of generalised Total Uncertainty (TU) was proposed by Klir (Klir & Smith, 2001) not long after the introduction of AU uncertainty. This measure is defined as a combination of AU uncertainty and Generalised Hartley Measures,

$$TU = \langle GH, GS \rangle \tag{15}$$

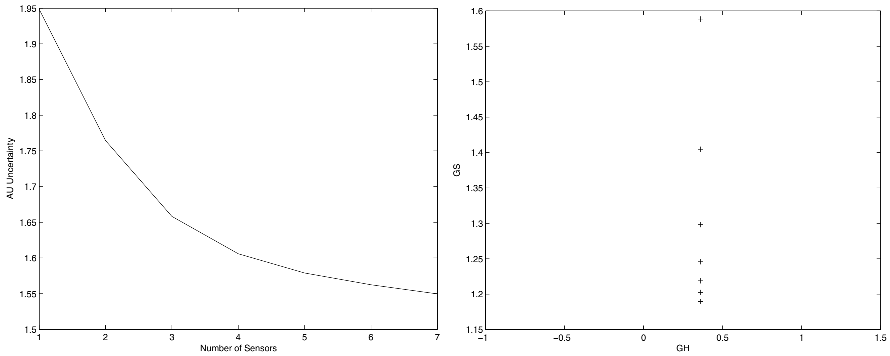
where GH represent the Generalised Hartley measures which was discussed in (11). The factor GS is called Generalised Shannon measurement (Klir, 2006), which is the conflicts measurement with the consideration of evident specificity. In other words, it is $GS = AU - GH$, the Aggregated Uncertainty with the reduction of specificity consideration. One advantage of the disaggregated TU, in comparison with AU, is that it expresses amounts of both types of uncertainty (non-specificity and conflict) explicitly, and consequently, it is highly sensitive to changes in evidence. These new features of uncertainty measures allow one to work with any set of recognised and well-developed theories of uncertainty as a whole, which are commonly seen in any evidential based fusion problem.

Classification	1 Sensor	3 Sensors	7 Sensors
A	0.22	0.3485	0.4125
B	0.25	0.3309	0.3525
C	0.26	0.2845	0.2343
D	0,00	0.0015	0.0001
A,B	0.07	0.0163	0.0004
A,C	0.03	0.005	0.0001
A,D	0.03	0.005	0.0001
B,C	0.015	0.0022	0.0000
B,D	0.005	0.0007	0.0000
C,D	0.01	0.0014	0.0000
A,B,C,D	0.1	0.0042	0.0000

Table 2. Classification Results with DST Fusion

6. Analysis of uncertainty measures under the Dempster Shafer fusion framework

To appreciate the impact of uncertainty variation, an example with a set of arbitrary data is illustrated in Table 2. The data set is exactly the same measurement values, such that an iterative DST fusion can be performed. The results in Table 2 confirmed that sensor information can be refined and appears to have a reduction of ambiguity under an iterative DST fusion process. However, the merit of these results cannot be examined further, unless an acceptable matrices is used to quantify the fusion. To address this point, the results illustrated in Figure 5 a demonstrate how AU uncertainty reduction could quantify the DST fusion process. Whilst the AU uncertainty measure are a useful index to quantify the DST fusion process, it is suggested to be insensitive to small change in evidences (Klir, 2006). Acknowledging the inherited issues with the AU uncertainty measures, this work also examines the concept of employing Total Uncertainty Map (TUM) to evaluate a standard DST Fusion process. Considering TU is an amalgamation of GH and GS, the uncertainty variation becomes significant if it is illustrated in two dimensional space. Figure 5b is an illustration of how a TUM can be used to visualise the recursive DST fusion. To assist the interpretation, the results of GS/GH are also provided in Figure 5a to enhance the illustration. In this case, GS and GH are treated as an unified parameters with the variation under the DST fusion process observed. Due to the equivalent sensor input for the DST fusion, the weighted average of



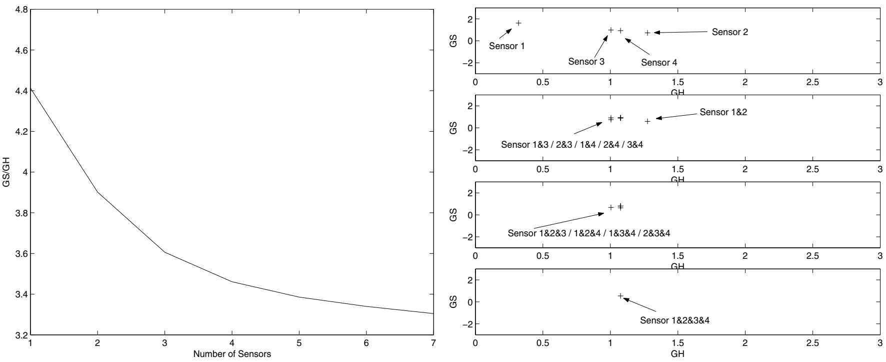
(a) AU Uncertainty Variation Under the DS Fusion (b) TU Map Variation Under the DS Fusion

Fig. 5. Uncertainty Variation

Sensor 1	Sensor 2	Sensor 3	Sensor 4
{ A } = 0.26	{B} = 0.2	{B} = 0.1	{A} = 0.05
{B}= 0.26	{A,B}= 0.1	{C} = 0.1	{B} = 0.05
{C}= 0.26	{A,C} = 0.1	{A,B}=0.16	{D} = 0.2
{A,B}= 0.07	{A,B,C}=0.1	{B,C}=0.14	{A,B} = 0.11
{A,C}= 0.01	{A,C,D}=0.1	{B,D}=0.05	{A,C} = 0.03
{A,D}= 0.01	{B,C,D}=0.3	{A,C}=0.1	{A,D} = 0.03
{B,C}= 0.01		{A,B,C}=0.2	{C,D} = 0.03
{B,D}= 0.01		{B,C,D}=0.15	{B,C,D} = 0.3
{C,D}= 0.01			{A,B,C,D} = 0.2
{A,B,C,D}= 0.1			

Table 3. Random Sensor Input

each focal subset are virtually unchanged, which is why the GH values displayed in Figure 5b remain constant throughout the iterative DST fusion process. Further observation shows, however, that other uncertainty in the form of conflicts are gradually reduced as part of the DST fusion process. To further explore the characteristics of uncertainty variation, four arbitrary sensor data sets are outlined in Table 3. The TU uncertainty is displayed in Figure 6 b. These results are further broken down into four levels and each level represents the number of sensors fused by the DST fusion. Based on the sample results, it is difficult to provide a consolidated uncertainty variation within the DST fusion framework. However, a potential optimisation solution exists when the fusion goal is to present the most specific and least conflicted information to the decision maker. This concept will be covered in Section 7 by leveraging a NCTR based AI example.



(a) GS/GH Variation Under the DS Fusion

(b) TU Map Under the DS Fusion with Random Sensors Input Data

Fig. 6. Extended Uncertainty Variation Modelling

7. NCTR based Aircraft Identification (AI)

This case study utilises an example commonly encountered in a model based classification system. Assuming each NCTR sensor has a potential to produce feature detection of,

$$B = \{E0, E1, E2, ..., E36\}$$

where B is the frame of discernment of the aircraft's type attributes, and this example allows seven model based classifiers to report aircraft type identification. To reduce the

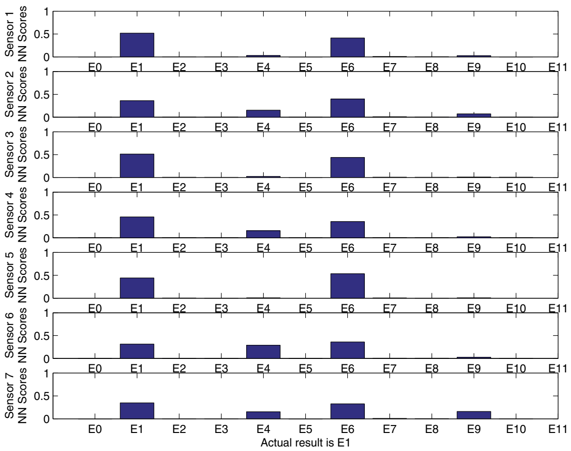


Fig. 7. Model based classifier for aircraft type detection

Sensor1	Sensor2	Sensor3	Sensor4	Sensor5	Sensor6
$m\{e_1\} = 0.5188$	$m\{e_1\} = 0.3617$	$m\{e_1\} = 0.5126$	$m\{e_1\} = 0.4565$	$m\{e_1\} = 0.4414$	$m\{e_1\} = 0.3480$
$m\{e_6\} = 0.4124$	$m\{e_4\} = 0.1540$	$m\{e_6\} = 0.4387$	$m\{e_4\} = 0.1551$	$m\{e_6\} = 0.5342$	$m\{e_4\} = 0.1533$
$m\{\Theta\} = 0.0687$	$m\{e_6\} = 0.3971$	$m\{\Theta\} = 0.0487$	$m\{e_6\} = 0.3546$	$m\{\Theta\} = 0.0244$	$m\{e_6\} = 0.3254$
	$m\{e_9\} = 0.0733$		$m\{\Theta\} = 0.0337$		$m\{e_9\} = 0.1602$
	$m\{\Theta\} = 0.0138$				$m\{\Theta\} = 0.0357$

Table 4. Normalised Aircraft Detection

computational workload this example only employs 12 of the target type signature instead of the potential 37 type of targets, where the results are depicted in Figure 7. The 12 aircraft type signatures selected for this simulation share similar characteristics, and often cause confusion to this particular NCTR platform. The remaining 25 emitter detections are not discarded, but are consolidated as detection CLUTTER. This method is similar to the strategy reported in (Yu & Sycara, 2006), instead this case study treats all aircraft signatures as the total frame of discernment Θ_E, $\{ e_0, e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11} \}$. In terms of the simulation, each emitter signature is considered as $e_i \in E$, where $m(e_i)$ is the normalised confidence level assigned by the post threshold detection process. For instance, the normalised post-detection confidence level with Sensor 2 are $m\{e_1\} = 0.3617$, $m\{e_4\} = 0.1540$, $m\{e_6\} = 0.3971$ and $m\{e_9\} = 0.0733$. To include the non-mutually exclusive aircraft type as CLUTTER, $m\{\Theta_E\} = c(CLUTTER)$, where we assign the confidence of CLUTTER to the set of all possible aircraft types. In this case, the normalised $m\{\Theta_E\}$ based on the pre-detection process is 0.0138.

Upon completion with the BPA preparation, we performed a DST based fusion with a permutation space of 2^7. Figure 8 shows the uncertainty in the form of AU as gradually reduced with the increment of DST fusion. However, the results become less effective when more sensors are fused together. In accordance with the discussions covered in Section 6, the authors believe the optimum approach when conducting an uncertainty based DST fusion cannot rely on one single parameter alone. Depending on the computational workload and the tolerance of conflicts, the uncertainty based fusion process ought to be determined by a TU map, where GS and GH are to be treated separately. The preliminary results based on this concept are illustrated in Figure 9.

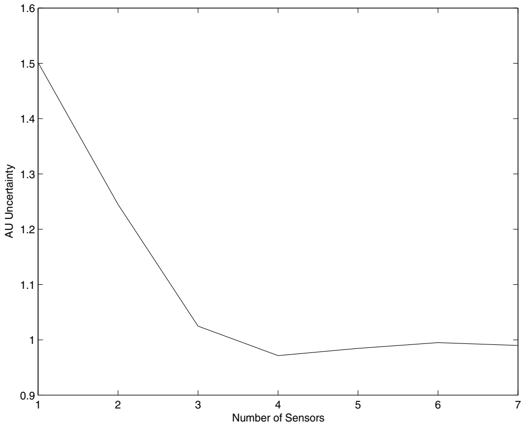
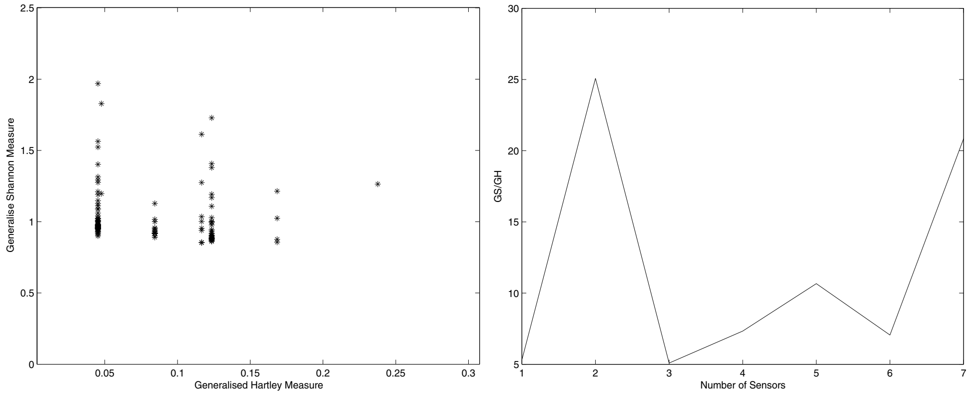


Fig. 8. AU Uncertainty Variation Under the model based classifier DST Fusion



(a) TU Map Variation Under the model based classifier DST Fusion

(b) GS/GH Variation Under the model based classifier DST Fusion

Fig. 9. Uncertainty Variation

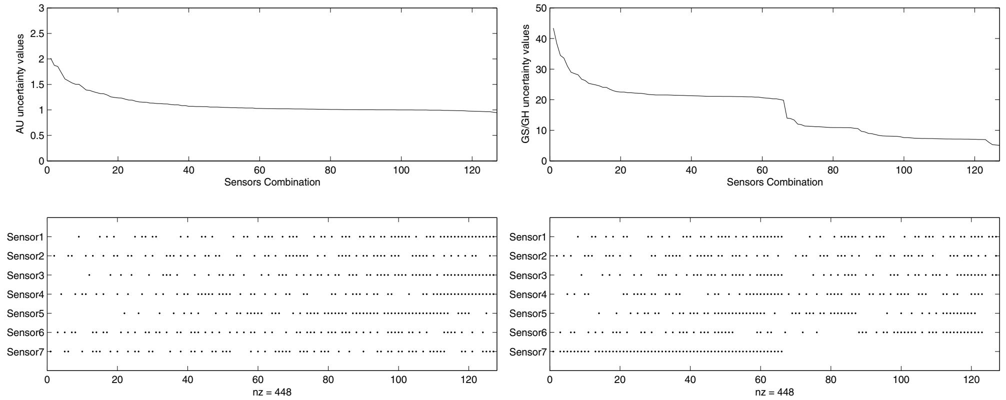
Notwithstanding the treatment of uncertainty in the DST context, Figure 9a outlined a method when adopting the theory of AU uncertainty to search for the least uncertain post-fusion results. For comparison purposes, the results of GS/GH measures are also displayed in Figure 9b. Under such a process, the final result is to be determined by the fusion that produces the minimum AU uncertainty. In this particular example, Sensors 1, 3, 4 and 7 can be selected to participate in the fusion process. Based on the least AU uncertainty, the final BPA for the detected emitters are given below:

$$m\{\Theta_E\} = 0$$

$$m\{e_1\} = 0.5604$$

$$m\{e_6\} = 0.4396$$

With a similar approach, and adopting the GS/GH characteristics, Sensors 1, 2 and 3 are selected to join the fusion process. Based on the least GS/GH uncertainty, the final BPA for



(a) AU Uncertainty measurements in conjunction with sensor fusion

(b) GS/GH Uncertainty measurements in conjunction with sensor fusion

Fig. 10. Uncertainty Variation

the detected emitters are given below, which is equivalent to sensor combination with the least AU uncertainty:

$$m\{\Theta_E\} = 0$$
$$m\{e_1\} = 0.5604$$
$$m\{e_6\} = 0.4396$$

Although the final results obtained from the uncertainty based DST fusion do not yield distinct decisions, the results justify that aircraft type e_1 or aircraft type e_6 are detected.

8. Conclusion

This paper reviews the role of uncertainty measures in the data fusion framework within the context of evidential reasoning. An empirical analysis of the AU and TU uncertainty variations is conducted under the DST fusion framework. A preliminary method to choose sensors based on the uncertainty level is proposed. This technique is illustrated with an aircraft identification problem, when the radar range profile classifier is employed to support an identification system such as NCTR. Since the amount of reflected radar energy is different for different parts of the aircraft, inconsistency often occurs even when the same target is being observed by a number of sensors despite using the same classifier model. It is this inconsistency which makes the uncertainty based fusion technique useful in resolving aircraft identification problems. While the proposed technique can be computationally intensive, the idea underwrites a conservative result with the least measurable uncertainty. This approach essentially yields the potential to evaluate all kinds of reasoning based fusion systems. We have certainly not reached the end of our research effort yet, as the proposed concept only considers primarily the reduction of AU uncertainty. The authors recognise the benefits in further investigation of TUM in conjunction with the theory of optimisation, when a trade-off can be computed based on the classification's precision and accuracy. At the moment, our proposed concept does not take into account the sensor information based on human originated data. It is certainly an exciting future research topic, if the proposed concept is to be extended to cover identification systems where human originate information is employed.

9. References

Bastiere, A. (1997). Fusion methods for mltisensor classification of airborne targets, *Aerospace Science and Technology* 1: 83–94.

B.Ristic & P.Smets (2005). Target classification approach based on the belief function theory, *IEEE Transactions On Aerospace And Electronics Systems* 41(2).

Carson, R. Meyer, M. & Peters, D. (26-30 Oct 1997). Fusion of iff and radar data, *16th AIAA/IEEE Digital Avionics Systems Conference (DASC)* 1: 5.3–9–15.

Deng, H. Chao, P. & Liu, J. (2011). Entropy flow-aided navigation, *The Journal of Navigation*, Vol. 64, The Royal Institute of Navigation, pp. 109–125.

Dezert, J. & Smarandache, F. (2006). Dsmt: A new paradigm shift for information fusion, *Cogis ′ 06 Conference, Paris* .

Hall, H. & Llinas, J. (2001). *Handbook of Multisensor Data Fusion*, CRC.

Harmanec, D. (1996). *Uncertainty in Dempster-Shafer Theory*, PhD Dissertation, State University of New York.

Hartley, R. (n.d.). Transmission of information, *The Bell System Technical Journal* 7(3): 535–563.

Klir, G. (2006). *Uncertainty and Information, Fundations of Generalised Information theory*, Wiley Interscience.

Klir, G. & Smith, R. (2001). On measuring uncertainty and uncertainty based information: Recent developments, *Annals of Mathematics and Artificial Intelligence* 32: 5–33.

Leung, H. & Wu, J. (2000). Bayesian and dempster-shafer target identification for radar surveillance, *IEEE Transactions on Aerospace and Electronic Systems* 36(2): 432–447.

Perlovsky, L. Chernick, J. & Schoendorf, W. (1995). Multi-sensor atr and identification of friend or foe using mlans, *Neural Networks* 8(7/8).

Porretta, M. Schuster, W. & Ochieng, W. (2010). Strategic conflict detection and resolution using aircraft intent information, *The Journal of Navigation*, Vol. 63, The Royal Institute of Navigation, pp. 61–88.

Schuster, W. & Ochieng, W. (2011). Airport surface movement - critical analysis of navigation system performance requirements, *The Journal of Navigation*, Vol. 64, The Royal Institute of Navigation, pp. 281–294.

Shafer, G. (1976). *A mathematical theory of evidence*, Princeton University Press.

Yu, B. & Sycara, K. (2006). Learning the quality of sensor data in distributed decision fusion, *Proceeding of the 9th International Conference on Information Fusion* .

12

Subjective Factors in Flight Safety

Jozsef Rohacs
Budapest University of Technology and Economics
Hungary

1. Introduction

The central deterministic element of the aircraft conventional control systems is the pilot – operator. Such systems are called as active endogenous subjective systems, because (i) the actively used control inputs (ii) origin from inside elements (pilots) of the system as (iii) results of operators' subjective decisions. The decisions depend on situation awareness, knowledge, practice and skills of pilot-operators. They may make decisions in situations characterized by a lack of information, human robust behaviors and their individual possibilities. These attributes as subjective factors have direct influences on the system characteristics, system quality and safety.

Aircraft control containing human operator in loop can be characterized by subjective analysis and vehicle motion models. The general model of solving the control problems includes the passive (information, energy - like vehicle control system in its physical form) and active (physical, intellectual, psychophysiology, etc. behaviors of subjects - operators) resources. The decision-making is the appropriate selection of the required results leading to the best (effective, safety, etc.) solutions.

This chapter defines the flight safety and investigates aircraft stochastic motion. It shows the disadvantages of the stochastic approximation and discusses, how, the methods of subjective analysis can be applied for the evaluation of flight safety.

The applicability of the developed method of investigation will be demonstrated by analysis of the aircraft controlled landing. The applied equation of motions describes the motion of aircraft in vertical plane, only. The boundary constraints are defined for velocity, trajectory angle and altitude. The subjective factor is the ratio of required and available time to decision on the go-around. The decision depends on the available information and psycho-physiological condition of operator pilots and can be determined by the theory of statistical hypotheses. The endogenous dynamics of the given active system is modeled by a modified Lorenz attractor.

2. Flight safety

2.1 Definitions

Safety is the condition of being safe; freedom from danger, risk, or injury. From the technical point of view, safety is a set of methods, rules, technologies applied to avoid the emergency situation caused by unwanted system uncertainties, errors or failures appearing randomly.

Safety and security are the twin brothers. The difference between them could be defined such as follows:

- *Safety:* avoid emergency situation caused by unwanted system uncertainties, errors or failures appearing randomly.
- *Security:* avoid emergency situations caused by unlawful acts (of unauthorized persons) – threats.

Safety related investigations start as early as the development of the given system. At the definition and preliminary phase of a new system, one should also concentrate some efforts on the (i) potential safety problems, (ii) critical situations, (iii) critical system failures, (iv) and their possible classification, identification. After the risk assessment, the next step is the development of a set of policies and strategies to mitigate those risks. Generally, the safety policies and strategies are based on the synergy of the

- physical safety (characteristics of the applied materials, structural solutions, system architecture that help to overcome safety critical – emergency situations),
- technical safety (dedicated active or passive safety systems including e.g. sensors to enhance situation awareness),
- non-technical safety (such as policy manuals, traffic rules, awareness and mitigation programs).

The safety of any systems can be evaluated by using the risk analysis methods. Risk is the probability that an emergency situation occurs in the future, and which could also be avoided or mitigated, rather than present problems that must be immediately addressed.

2.2 Flight safety metrics

The evaluation of the flight safety is not a simple task. There is no uniformly applicable metrics for the evaluation. Some governments have already published (CASA, 2005; FAA, 2006; Transport, 2007) their opinion and possible methodologies for flight safety measures that are applied by evaluators (Ropp & Dillmann, 2008). The problem is associated with the very complex character of flight safety depending on the developed and applied

- safety plan with management commitment,
- documentation management,
- risk monitoring,
- education and training,
- safety assurance (quality management on safety),
- emergency response plan.

Risk analyses methods defining the probability of emergency situations or risks are very widely used for flight safety evaluation. Metrics of risk is the probability of the given risk as an unwanted danger event. This probability has at least four slightly different interpretations:

- classic - the unwanted event,
- logic - the necessary evil,
- objective - relative frequency,
- subjective - individual explanation of the events.

In practice, the analysis of accident statistics could characterize the flight risks. Such statistics give the evidences for the well-known facts (Rohacs, 1995, 2000; Statistical 2008): (i) the longest part of the flight (with about 50 - 80 % of flight time) is the cruise phase, which only accounts for 5 - 8 % of the total accidents and 6 - 10 % of the total fatal accidents, (ii) the most dangerous phases of flight are the take-off and landing, because during this about 2 % of flight time the 25 - 28 % of fatal accidents are occurring, and (iii) generally nearly 80 % of the accidents are caused by human factors and about 50 % of them are initiated by the pilots.

A good example of using accident statistics is shown in Figure 1. Beside showing the effects of technological development on the reduction of flight risks, it also shows that since 2003, the European fatal accident rate - as fatalities per 10 million flights - has increased, without knowing – so far – the reason causing it.

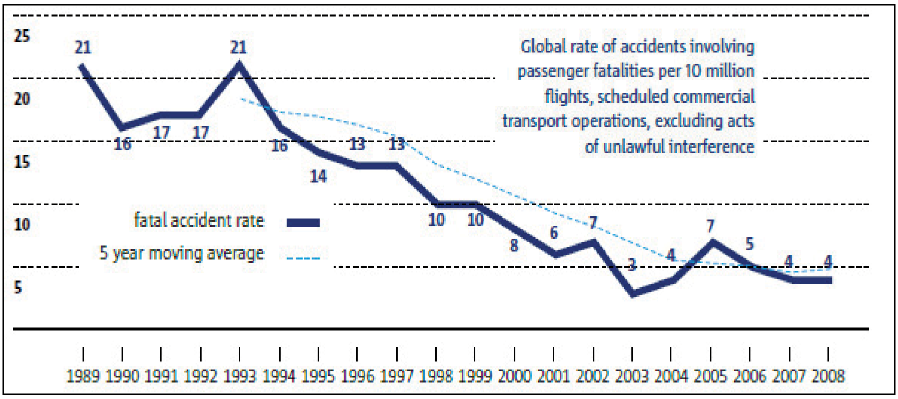


Fig. 1. Characterization of the European accident statistics (EASA, 2008).

The accident statistics could be also used for flight safety analysis in original, or unusual method. While accident statistics demonstrate a considerable higher risk, accident rates for small aircraft, according to the Figure 2., the ratio of all and fatal accidents are nearly the same for airlines and general aviation. This means that the small and larger civilian aircraft are developed, designed, and produced with the same philosophy, at least the same safety approach and ‘structural damping of damage processes’. The flight performances, flight dynamics, load conditions, structural solutions are different for small and larger aircraft, and therefore the accidents rates are also different. However, the risk of hard aftermath, appearing the fatal accident following the accidents are the same.

2.3 Human factors

In 1908, 80 % of licensed pilots were killed in flight accident (Flight, 2000). Since that, the World and the aviation have changed a lot. After 1945, the role of technical factors in causing the accidents (and generally in safe piloting) is continuously decreasing while the role of human factors is increasing.

As it was outlined already, nearly 80 % of accidents are caused by human factors. (Rohacs, 1995, 2000; Statistical, 2008). While, only 4 -7 % of accidents are defined by the "independent investigators" as accident caused by unknown factors. According to Ponomerenko (2000) this figure might be changed when one tries to establish the truth in fatal accidents,

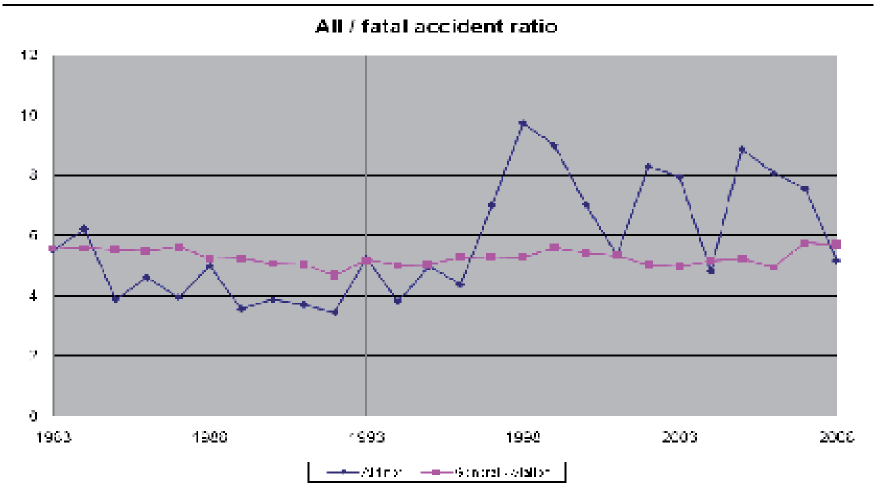


Fig. 2. An original way to compare airliner and GA accident statistics.

especially by taking into account the socio-psychological aspects and use of " 'guilt' and 'guilty' as the 'master key' to unlock the true cause of the accident. Hence, the bias of the investigators often does not represent the interest of the victims, but that of the administrative superstructure. It side steps the legal and socio-psychological estimation of aircrew behavior, and replaces it by formal logic analysis of known rules: permitted/forbidden, man or machine, chance/relationship, violated/not violated, etc."

Accident investigations show that human factors could be divided into three groups depending on their origins.

- Technical factors: disharmony in human - machine interface. Most known cases from this group are called as PIDs (pilot induced oscillations). Some of these factors, like limitations of the control stick forces are included even into the airworthiness requirements.
- Ergonomic factors: a lack of ergonomic information display, guidance control, out-of-cockpit visibility, design of instrument panel, as well as of adequate training [Ponomarenko 2000].
- Subjective factors: un-predictable and non-uniform man's behavior. Making wrong decisions because the lack of knowledge and practice of operators.

The different groups have nearly the same role in accident casualty, equal to 25, 35, 40 %, respectively. Others (Lee, 2003) call the same type of factors as system data problems, human limitation and time related problems.

The first group of human factors, harmonization of the man-machine interface from the technical side of view is well investigated and such type of human factors are taken into account in aircraft development and design processes. Generally, the handling quality or (nowadays) the car free characteristics are the merits and used as main philosophical approaches to solve these types of problems.

The ergonomic factors have been investigated a lot for last 40 - 50 years. The third generation of the fighters had been developed with the use of ergonomics, especially in development of the cockpit, that were radically redesigned for that period. However, the ergonomic investigations had used the governing idea, how to make better for operator. A new approach has developed for last 20 years that investigates the 'ergatic' systems (see for example Pavlov & Chepijenko, 2009) in which the operator (pilot) one of the important (might be most important) element of the systems, and the psycho-physiological behaviors of the operator may play determining roles in operation of system.

The third group of human factors has not investigated on the required level yet. Generally, the key element of human reaction on the situation, especially on the emergency situation is the time. *However, the speed and time of reaction is "... not determined by the amount of processed information, but by the choice of the signal's importance, which is always subjective and affected by individual personality traits" (Ponomarenko 2000)*. In an emergency situation, flight safety does not depend as much on the detailed information on the emergency situation and the size of pilot supporting information, as on the whole picture including space and time, knowledge and practice of pilots and the actual determination of the ethical limits of man's struggle with the arisen situation.

Flight safety could also be analyzed with the prediction of the future air transport characteristics. For example, the NASA initiated zero accident project, (Commercial, 2000; Shin, 2000; White, 2009) leads to the following general conclusion: before introducing the wide-body aircraft, the risk of flight was decreased by a factor of 10, but this cannot be further reduced with the present technical and technological methods (Rohacs, 1998; Shin, 2000). Even so, the number of aircraft and the number of yearly, daily flights are continuously increasing (Fig. 3.); Seeing this, the absolute number of accidents is expected to increase in the future, which might even lead to the vision made by Boeing, in which by 2016/17, each week one large-body aircraft is envisioned to have an accident. "Given the very visible, damaging, and tragic effects of even a single major accident, this number of accidents would clearly have an unacceptable impact upon the public's confidence in the aviation system and impede the anticipated growth of the commercial air-travel market" (Shin, 2000). Therefore, new methods like emergency management might need to be developed and applied to keep the absolute number of accidents on the present level.

Seeing the envisioned rapid development of the future aviation, especially the small aircraft transportation system, the conclusion derived from the zero accident program and use of the subjective analysis in flight safety investigation might be relevant to be kept in mind.

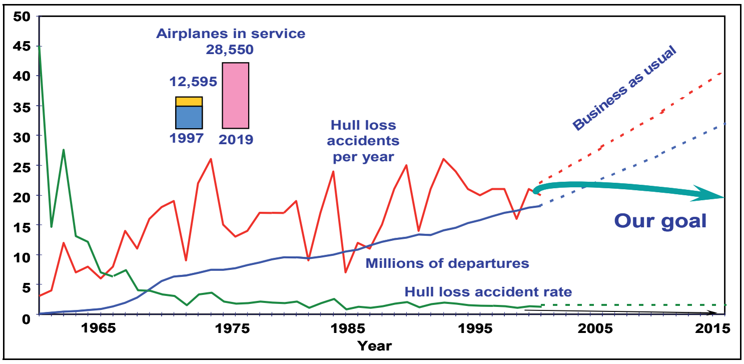


Fig. 3. The NASA zero accident program (Commercial, 2000).

3. Flight safety evaluation

3.1 Technical approach to flight safety evaluation

Technically, flight risks are always initiated by the deviations in the system parameters. Therefore, the investigation of the system parameter uncertainties and anomalies might be applied as a basis to evaluate flight safety. Flight safety is the risk that an emergency situation occurs, when the system parameters (at least one of them) are out of the tolerance zones. In the view of this, flight safety might be characterized by the probability of the deviations (in the structural and operational characteristics) being larger than those predetermined by the airworthiness (safety) requirements (Bezapasnostj 1988, Rohacs & Németh, 1997).

Mathematically, flight operation quality, $\mathbf{Q}_r(t)$, could be given in the following simple form:

$$\mathbf{Q} \equiv \{a_i\} \quad , \quad i = 1,\mathrm{n} \tag{1}$$

where a_i are the parameters defining the attributes of the given aircraft or system. In a more general form, it could be given as:

$$a_i = f\left(a_1, a_2, \ldots a_{i-1}, a_{i+1}, \ldots a_n\right) . \tag{2}$$

In real flight situations, the real quality of operation $\mathbf{Q}_r(t)$ is deviated from the design (nominal) quality $\mathbf{Q}_r^n(t)$:

$$\delta\,\mathbf{Q}_\mathrm{r}(t) = \mathbf{Q}_\mathrm{r}(t) - \mathbf{Q}_\mathrm{r}^\mathrm{n}(t) \quad . \tag{3}$$

For each case, the acceptable level of deviation is maximized by the flight safety threshold (δ_fs),

$$\left|\delta\,\mathbf{Q}_\mathrm{r}(t)\right| \geq \delta_\mathrm{fs} \quad , \tag{4}$$

where $P\left(\left|\delta\,\mathbf{Q}_\mathrm{r}(t)\right| \geq \delta_\mathrm{fs}\right)$ describes the probability of a flight event (flight out of prescribed operational modes).

By summing all the potential flight events, flight safety (P_{fs}) could be given with the following probability:

$$P_\mathrm{fs} = 1 - \sum_{i=1}^{n} R_i(t) P_i(t) \tag{5}$$

where $R_i(t)$ - is the risk of flight accident.

For time period [0, T] the following integral risk can be applied:

$$\tilde{P}_\mathrm{fs} = 1 - \frac{1}{\mathrm{T}} \int_0^\mathrm{T} \left(1\text{-}P_\mathrm{fs}(t)\right) dt \quad . \tag{6}$$

Because $\delta\,\mathbf{Q}_r(t)$ is the random value with probability density, $\rho\,(\delta\,\mathbf{Q}_r(t))$, the flight safety level can be given as:

$$P_{\text{fs}} \equiv P\left(\left|\delta\,\mathbf{Q}_r(t)\right| \le \delta_{\text{fs}}\right) = \int_{-\delta_{\text{fs}}}^{\delta_{\text{fs}}} \rho\,\left(\delta\,\mathbf{Q}_r\right) d\delta\,\mathbf{Q}_r \quad . \tag{7}$$

According to the Tchebyshev inequality

$$P(\left|\delta\,\mathbf{Q}_r(t)\right| \succ \delta_{\text{fs}}) \le D\left(\delta\,\mathbf{Q}_r\right) / \delta_{\text{fs}}^2 \tag{8}$$

the flight safety level takes the form:

$$P_{\text{fs}} \equiv P\left(\left|\delta\,\mathbf{Q}_r(t)\right| \le \delta_{\text{fs}}\right) \ge 1 - D\left(\delta\,\mathrm{Q}_r\right) / \delta_{\text{fs}}^2 \quad , \tag{9}$$

where $D(\delta\,\mathbf{Q}_r)$ is the dispersion of $\delta\,\mathbf{Q}_r$.

Such type of system approach was developed, applied and improved. Generally, once the aircraft is investigated as a dynamic system, the effects of the system anomalies could be given by the following type of probabilities (Rohacs 1986; Rohacs & Nemeth, 1997):

$$P_1\left\{\,\mathbf{y}(t)\Big|_{t_0\ \le\ t\ \le\ t+\tau,\ \mathbf{x}\in\Omega_{\text{x}},\ \mathbf{u}\in\Omega_{\text{u}},\ \mathbf{z}\in\Omega_{\text{z}},\ \mathbf{p}\in\Omega_{\text{p}}}\right\} \quad , \tag{10.a}$$

$$P_2\left\{\,\mathbf{u}(t)\Big|_{t_0\ \le\ t\ \le\ t+\tau,\ \mathbf{x}\in\Omega_{\text{x}},\ \mathbf{y}\in\Omega_{\text{y}},\ \mathbf{z}\in\Omega_{\text{z}},\ \mathbf{p}\in\Omega_{\text{p}}}\right\} \quad , \tag{10.b}$$

where $\mathbf{y} \in R_r$ defines the output (measurable) signal vector (measured vector of operational characteristics) $\mathbf{x} \in R_n$ is the state vector, $\mathbf{u} \in R_m$ gives the input (control) vector, $\mathbf{z} \in R_i$ stands for the vector of environmental characteristics (vector of service conditions), $\mathbf{p} \in R_k$ is the parameter vector characterizing the state of the aircraft, t defines the time, τ provides the elementary time, Ω_{x}, Ω_{y}, Ω_{z}, Ω_{u}, Ω_{p} are the allowed ranges for the given characteristics.

If the joint density function,

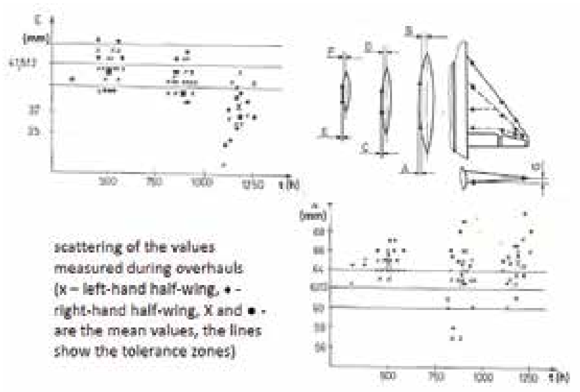
$$f_{\Sigma} = f\left[\mathbf{x}(\mathrm{t}),\mathbf{u}(\mathrm{t}),\mathbf{z}(\mathrm{t}),\mathbf{p}(\mathrm{t}),\mathbf{y}(\mathrm{t})\right] \tag{11}$$

is known, then the recommended characteristics can be calculated as:

$$\mathrm{P}_1\left\{\mathbf{y}(\mathrm{t}) \in \mathbf{\Omega}_{\mathrm{y}}\middle|\ldots\right\} = \frac{\displaystyle\int_{\Omega_i} f_{\Sigma}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{y}}{\displaystyle\int_{-\infty}^{+\infty}\mathrm{d}\mathbf{y}\int_{\Omega_j} f_{\Sigma}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{p}} \qquad \begin{matrix}(i \in \mathbf{x},\mathbf{u},\mathbf{z},\mathbf{p},\mathbf{y})\\ (j \in \mathbf{x},\mathbf{u},\mathbf{z},\mathbf{p})\end{matrix} \quad , \tag{12.a}$$

$$\mathrm{P}_2\left\{\mathbf{u}(\mathrm{t}) \in \mathbf{\Omega}_{\mathrm{u}}\middle|\ldots\right\}. = \frac{\displaystyle\int_{\Omega_i} f_{\Sigma}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{y}}{\displaystyle\int_{-\infty}^{+\infty}\mathrm{d}\mathbf{u}\int_{\Omega_j} f_{\Sigma}\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z}\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{y}} \qquad \begin{matrix}(i \in \mathbf{x},\mathbf{u},\mathbf{z},\mathbf{p},\mathbf{y})\\ (j \in \mathbf{x},\mathbf{z},\mathbf{p},\mathbf{y})\end{matrix} \quad . \tag{12.b}$$

Unfortunately, this method of determining the effects of the system anomalies on the flight safety is often considered to be too complex, while it is found to be reasonable, since the formulas given above could be supported with statistical data collected during aircraft operation. The method of determining the flight risk on the probability approach (as given in (Gudkov & Lesakov, 1968; Howard, 1980)) is envisioned to be too complicated, once it is also desirable to consider the so-called common (failures appearing at the same time due to different reasons) and depending failures or errors. The Figures 4 and 5 show a nice example of using the described method is the investigation changes in geometrical and operational characteristics of aircraft investigated by (Rohacs 1986) and published in several articles, like (Rohacs, 1990).



scattering of the values measured during overhauls (x – left-hand half-wing, • - right-hand half-wing, X and ■ - are the mean values, the lines show the tolerance zones)

Fig. 4. The level book and examples of the measuring data for Mig-21.

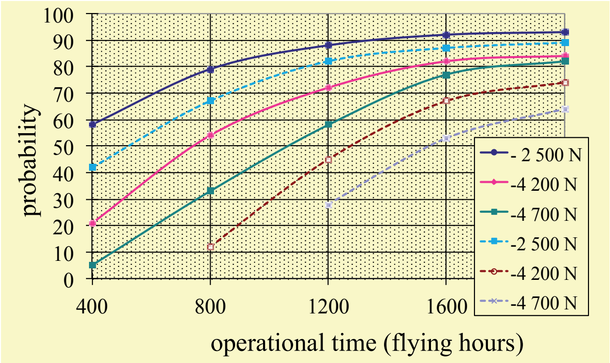


Fig. 5. Probability of lack of generated lift at fighters Míg-21 due the changes in wing geometry during the operation (line - single seat, dot line - double seats aircraft)

3.2 Stochastic model of flight risk

The aircraft's motion is the result of the deterministic control and the stochastic disturbance processes. Such motion might be mathematically given by the following stochastic (random) differential equation, called as diffusion process (Gardiner, 2004):

$$\dot{x} = f(x,t) + \sigma(x,t)\eta(t) \quad , \tag{13}$$

Naturally, this equation might be also given in vector form. The first part of the right side of the equation describes the drift (direction of the changes) of the stochastic process passing through $x(t) = X$ at the moment *t*, while the second part shows the scattering (variance) of the random process. Here $\eta(t)$ is the random disturbance (e.g. air turbulence, or cumulative effects of random load processes, including even extreme loads as hard touchdown, etc.).

Seeing that the future states depend only on the present sate, the equation (13) is in fact a Markov process (Ibe, 2008; Rohacs & Simon, 1989; Tihonov, 1977). Such process can be fully described by its transition probability density function:

$$p(x_2,t_2|X_1,t_1), \qquad (t_2 > t_1) \ , \tag{14}$$

which characterizes the distribution probability of the continuous random process ($x(t)$) at the moment t_2, once it's passing through the $x(t) = X$ at time t_1.

The transition probability density function can be described by the following Fokker - Planck - Kolmogorov equations (Gardiner, 2004):

$$\frac{\partial p\left(x_2,\, t_2|X_1,\, t_1\right)}{\partial t_2} = -\frac{\partial}{\partial x_2}\left[f\left(x_2,\, t_2\right)p\left(x_2,\, t_2|X_1,\, t_1\right)\right] +$$

$$+\frac{1}{2}\frac{\partial^2}{\partial x_2^2}\left[\sigma^2\left(x_2,\, t_2\right)p\left(x_2,\, t_2|X_1,\, t_1\right)\right] \ , \tag{15.a}$$

or

$$\frac{\partial\left(x,\, t\right)}{\partial t} = -\frac{\partial}{\partial x}\left[f\left(x,\, t\right)p\left(x,\, t\right)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2\left(x,\, t\right)p\left(x,\, t\right)\right]. \tag{15.b}$$

Statistic flight mechanics has already worked out several methods for the application of such models. For example, the statistical linearization through the proof of the sensitivity function matrix to the flight mechanic models and generating out the set of equations for the moments of the investigated stochastic process could be used to study the scattering of the process.

Using the equations (15.a), (15.b), which define the Markov process, the following definition could be made:

$$p\left(X_2,\, t_2|X_1,\, t_1\right) = \sum_{X(t)} p\left(X_2,\, t_2|x,\, t\right)p\left(x,\, t|X_1,\, t_1\right), \qquad \left(t_2 \geq t \geq t_1\right) \ , \tag{16}$$

This is called Chapman - Kolmogorov – Smoluchovski equation. It gives the possibility to approximate the investigated non-linear stochastic process with continuous time and state space with a Markov chain with continuous time and discrete state space. This leads us back to the situation chain process.

The space of the motion variables can be divided into several subspaces, called as situations. The motion of the aircraft is in fact a time invariant series of situations. This is the situation dynamics.

Accidents are the results of the situation process, which is assumed to be similar to the one given in the Figure 6. Here, *N* marks the normal, conventional flight, S_1, S_2, S_3 are different states related to the case when the aircraft has one (*F1),* two (*F2*), or three serious system failures (F_3), while *A* shows the accident situation (Rohacs & Nemeth, 1997; Rohacs, 2000).

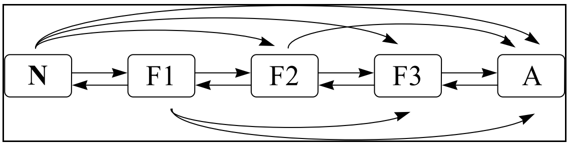


Fig. 6. Simple graph model of aircraft pre-accident process

The Markov chain can be described by the transition probabilities, $\beta_{i,j}$. These variables give the probability of moving the aircraft from a state (situation) S_i to a state S_j. As it is known, this type of process can be approximated by Markov process, under the following conditions:

- the transition from one state into another occurs in a significantly short time,
- the probability of a transfer from one state into another through one or more other states is a limited, and
- the time spent in the states could be approximated by an exponential distribution.

Under the conditions mentioned above, the process could be described with the following model:

$$\dot{\mathbf{P}}(t) = \mathbf{P}_t(t)\mathbf{P}(t) \ , \tag{17}$$

where **P**(t)=[Pi(t)] is a vector of probabilities defining the states S_i (*i*=*N*, *F1*, *F2*, *F3*, *A*).

At this stage, one should give the applicable graph model and estimate the transition probability matrix.

In this simple case, the aircraft's operational process – as a stochastic process with continuous time and discrete states shown in the Fig. 6. – could be approximated by the following Markov model:

$$\dot{\mathbf{P}}(t) = \boldsymbol{\beta}(t)\mathbf{P}(t) \tag{18}$$

where **P**(t)=[P i(t)] is a vector of probabilities that the aircraft is in the states S_i (*i*=*N*, *F1*, *F2*, F_3, *A*), and

$$\beta(t)=[\beta\ i,j] \tag{19}$$

is a time depending transition matrix:

$$\boldsymbol{\beta}(t) = \begin{bmatrix} -\beta_{N,F1}-\beta_{N,F2}-\beta_{N,F3}-\beta_{N,A} & -\beta_{F1,N} & 0 & 0 & 0 \\ \beta_{N,F1} & -\beta_{F1,N}-\beta_{F1,F2}-\beta_{F1,F3}-\beta_{F1,A} & -\beta_{F2,F3} & 0 & 0 \\ \beta_{N,F2} & \beta_{F1,F2} & -\beta_{F2,F1}-\beta_{F2,F3}-\beta_{F2,A} & -\beta_{F3,F2} & 0 \\ \beta_{N,F3} & \beta_{F1,F3} & \beta_{F2,F3} & -\beta_{F3,F2}-\beta_{F2,A} & -\beta_{A,F3} \\ \beta_{N,A} & \beta_{F1,A} & \beta_{F2,A} & \beta_{F3,A} & -\beta_{A,F3} \end{bmatrix}$$

Our theoretical and practical investigations on flight safety showed that the aircraft's operational process is a complicated process. For example, if a pilot reports an in-operating engine, than ATCOs are often to make 40 - 100 times more mistakes relative to normal circumstances. The simplified graph model of flight situations - taking into account such effects - is given in the Figure 7. The advantage of this representation method over the others, could be summarized in the followings. Firstly, this model includes a new state, called state of anomalies (*An*), in which the aircraft does not have any failures or errors, but still, its characteristics are essentially deviating from their nominal values. Secondly, the total amount of states are decomposed or grouped into four subparts (structure, pilot, air traffic control, surroundings).

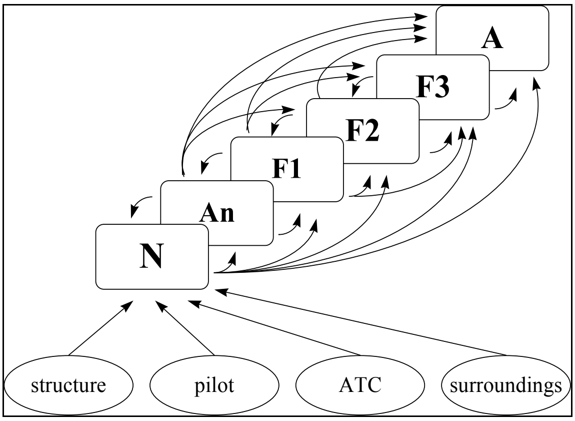


Fig. 7. The suggested general graph model of aircraft.

To simplify the representation of this method, the Figure 7. shows only the nominal state decomposition (Rohacs & Nemeth, 1997; Rohacs, 2000). Even so, the different numbers of failures are further decomposed. States *N* is a prescribed nominal state. States *An* and *F1* might only be initiated by the anomalies or failures in one of the aircraft's flight operation subsystems (e.g. aircraft structure, pilot, ATC, surroundings). On the other hand, the states *F2*, *F3* might be initiated by two or three failures appearing in any combination of the subsystems. For example *F2* may contain mistake of the pilot and ATCO, or two aircraft structural (system) failures.

According to these specific features of the model, the general Markov model should have 43 states. For example in our model, the state number 21, is the state with two failures generated in the structure and one is initiated by the mistake of the pilot. As a consequence, the transfer matrix is composed of 43 x 43 elements, while the elements of the matrix are the linear functions of **P**(t):

$$\beta i,j = \beta i,j,o + \mathbf{K}i,j\ \mathbf{P}(t)\ ; \qquad (20)$$

where, βi,j,o is the initial transfer matrix element, **K**i,j is the vector of coefficients. The vector **K**i,j may contain zero elements, too, if the given state has no influence on transfer process.

The determination of the vector elements **K**i,j, is based on the theory of anomalies, dealing with the calculation of the real deviations, characteristics, and distributions. For example, human error depends on weather, traffic situations, or possible system failures. Naturally, if the aircraft is piloted by pilot with limited skills, then the coefficients would be higher than it is for the conventional small aircraft operations. After the evaluation of different models based on the above discussed Markov and semi-Markov processes, we found that the inadequate initial data and the relatively large number of states makes the semi-Markov process irrelevant for our purposes.

Due to the large number of states, the developed model might be seen too complex. On the other hand, by the analysis of the potential methods to simplify the model, it was found that the suggested approach can be transferred to the model shown in the Figure 7. This is reasonable, since from a flight safety point of view, the most important is the transfer of one state into another, and not the detail how that transfer could be made. Therefore, the transition matrix element, β_{F1}, β_{F2}, describing the transfer from one failure state (*F1*) into the state with two failures (*F2*) can be given in the following form:

$$\beta_{F1,F2} = \frac{\sum_{i,j} \beta_{F1_i,F2_j} P_{F1_i}}{\sum_{k,i} \beta_{An_k,F1_i} P_{An_k}} \quad , \tag{21}$$

where *An* indicates the state with anomalies, and *k, i, j* are indexes defining the states.

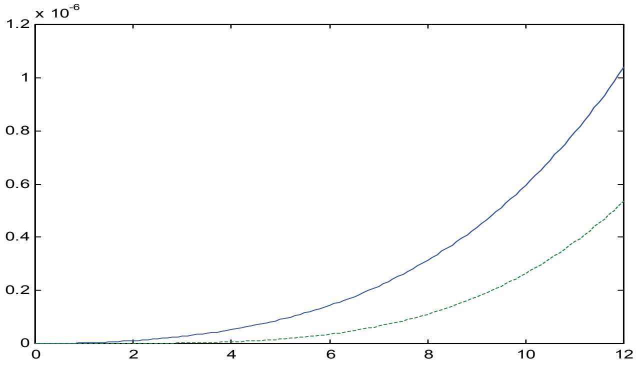


Fig. 8. Flight risk by considering (state An included - solid blue line) or neglecting the effects of anomalies (green dashed line).

As a result, the general model – describing the real interactions between different types of failures, distinguishing common and depending failures – could be reduced to a simple model.

The developed model was used for the analysis of the aircraft control. Some results are shown in Figures 8. and 9.

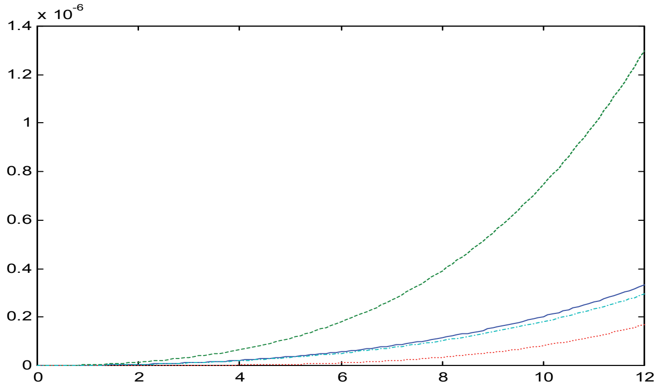


Fig. 9. Probability of appearance of first failures (solid blue line - pilot error (failure), green dashed line - pilot error in case of system anomalies red dashed line - structure failure, blue dashed dot line - structure failure calculated considering the influences of the anomalies).

4. Subjective analysis and flight safety

4.1 Theoretical background

The major determinative element of the aircraft's conventional control systems is the pilot. Such systems are called as ergatic active endogenous systems [Kasyanov 2007], since the systems are actively controlled by solutions initiated by ergates (Greek ἐργάτης ergatēs - worker), human organism (e.g. nervous cells). So the control solution becomes from inside the system, from the operator. Such effects are called often as endogenous feedback or endogenous dynamics (Banos, Lamnabhi-Lagarrigue & Montoya, 2001; Fliens et all, 1999, Nieuwstadt 1997]. Because pilots make their decision upon their situation awareness, knowledge, practice and skills, e.g. on the subjective way, the system would be also subjective. Beside human robust behaviors and individual possibilities, pilots – in certain circumstances – should also make decisions, even if the information for an appropriate reaction is limited.

Safety of active systems is determined by risks initiated by subjects being the central elements of the given system. For example, flight safety is the probability that a flight happens without an accident. Aircraft are moving in the three dimensional space, in function of their aerodynamic characteristics, flight dynamics, environmental stochastic disturbances (e.g. wind, air turbulence) and applied control. Pilots make decision upon their situation awareness. They must define the problem and choose the solution from their resources, which makes human controlled active systems endogenous. Resources are methods or technologies that can be applied to solve the problems (Kasyanov, 2007). These could be classified into the so-called (i) passive (finance, materials, information, energy - like aircraft control system in its physical form) and (ii) active (physical, intellectual, psycho-physiological behaviors, possibilities of subjects) resources. The passive resources are therefore the resources of the system (e.g. air transportation system, ATM, services provided), while the active resources are related to the pilot itself. Based on these, decision making is in fact the process of choosing the right resources that leads to an optimal solution.

Subjects (like pilots) could develop their active resources (or competences) with theoretical studies and practical lessons. However, the ability of choosing and using the right resources is highly depending on (i) the information support, (ii) the available time, (iii) the real knowledge, (iv) the way of thinking, and (v) the skills of the subject. Such decisions are the results of the subjective analysis.

There is insufficient information on the physical, systematic, intellectual, physiological characteristics of the subjective analysis, as well as on the way of thinking, and making decision of subjects-operators like pilots. Only limited information is available on the time effects, possible damping the non-linear oscillations, the long-term memory, which makes the decision system chaotic.

Flight safety can be evaluated by the combination of subjective analysis and aircraft motion models.

At first, the pilot as subject (Σ) must identify and understand the problem or the situation ($S_{i,}$), then from the set of accessible or possible devices, methods and factors (S_p) must choose the disposable resources (R^{disp}) available to solve the identified problems, to finally decide and apply the required resources (R^{req}) (Kasyanov 2007) (Fig.10.). For this task, the pilot applies its active and passive resources. The active resources will define how the passive resources are used:

$$R_{\text{a}}^{\text{req}} = f\left(R_{\text{p}}^{\text{req}}\right) \tag{22}$$

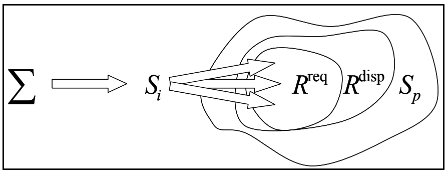


Fig. 10. Pilot decision – action process (endogenous dynamics) in aircraft operation (control) system.

Instead of the function between the resources (22), the literature often uses the velocity of transferring the passive resources into the actives:

$$v_{\text{a}}^{\text{req}} = f_v\left(v_{\text{a}}^{\text{req}}\right)v_{\text{a}}^{\text{req}}, \tag{23}$$

where

$$v_{\text{a}}^{\text{req}} = \frac{dR_{\text{a}}^{\text{req}}}{dt}, \qquad v_{\text{p}}^{\text{req}} = \frac{dR_{\text{p}}^{\text{req}}}{dt}, \tag{24}$$

and in simple cases

$$f_v = \frac{\partial R_{\text{a}}^{\text{req}}}{\partial R_{\text{p}}^{\text{req}}}. \tag{25}$$

It is clear that the operational processes can be given by a series of situations: pilot identifies the situation (S_i,), makes decision, controls (R_a^{req}), which transits the aircraft into the next situation (S_j,). (The situation S_j, is one of the set of possible situations). This is a repeating process (Fig. 11.), in which the transition from one situation into another depends on (i) the evaluation (identification) of the given situation, (ii) the available resources, (iii) the appropriate decision of the pilot, (iv) the correct application of the active resources, (v) the limitation of the resources and (vi) the affecting disturbances.

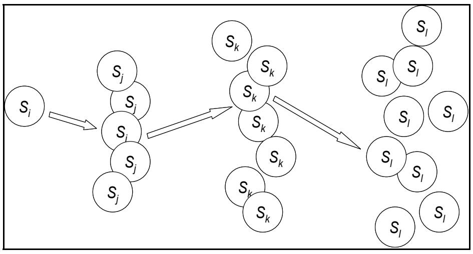


Fig. 11. Situation chain process of aircraft operational process as a result of an active subjective endogenous control.

The situation chain process can be given by the following mathematical formula:

$$c(t):\quad \left(x_0, t_0, \omega\left(t_f \in [t_0, t_0+\tau]\right); R^{disp}(t_0), R^{req}(t_0), \ldots\right), \tag{26}$$

or in a more general approach:

$$c(t):\quad \left(P : \sigma_0(t_0) \rightarrow \sigma_j\left(t_f \in [t_0, t_0+\tau]\right) \in S_f \subset S_a, R^{disp}(t_0), R^{req}(t_0), \ldots\right); \tag{27}$$

where x_0 is the vector of parameters at the initial (actually starting) state at t_0 time; σ gives the state of the system in the given time; τ defines the available time for the transition of the state vector into the set of ω not later than $[t_0, t_0+\tau]$; P are the problems how to transit the system from the initial state into the one of the possible state $S_f \subset S_a$ not later than τ .

During a flight, one flight situation is followed by another. Therefore, the aircraft flight operational process with continuous state space and time can be approximated by the stochastic process with continuous time and discrete state space, flight situations. This means that a flight is a typical situation chain process. (This is a basis for using the stochastic model of flight risk - see 2.2. point.)

4.2 Using the developed model to investigation of the aircraft landing

Final approach and landing are the most dangerous phases of flights. It is even a more significant problem for personal flights, controlled by less-skilled pilots.

The developed method using the subjective analysis to the flight safety evaluation was applied to investigate a landing procedure of a small aircraft.

In this investigation, no side wing, and no lateral motion were considered. By using the trajectory reference system – in which the x axis shows the direction of the wind, z axis is perpendicular to x in the local vertical plane, while centre of the coordinate system is located in the aircraft's centre of gravity – the motion of the aircraft could be given by the motion and the rotation of its center of gravity (Kasyanov 2004):

$$m\frac{dV}{dt} = T(V,z,t) - W\sin\theta - D(V,z,t) \ , \tag{28.a}$$

$$mV\frac{d\theta}{dt} = L(V,z,t) - W\cos\theta \ , \tag{28.b}$$

$$I_y\frac{dq}{dt} = M(\alpha,q,V,z,t) \ . \tag{28.c}$$

Due to the applied control, the trust (*T*), the lift (*L*), the drag (*D*) and the aerodynamic moment (*M*) are all clearly depending on time. The altitude (*z*) has also an influence on the variable above, through the ground effect. Mass (*m*) and therefore the weight (*W*) of the aircraft are assumed to be constant. The aircraft's velocity (*V*) and pitch rate (*q*) describes the motion, while the flight path angle (or descent angle θ) gives the position of the aircraft. The angle of attack (α) is the difference between the pitch attitude, ϑ and flight path angles:

$$\alpha = \upsilon - \theta \ . \tag{29}$$

The pitch rate and the modification of the altitude could be easily given by:

$$q = \frac{d\upsilon}{dt} \ , \tag{30}$$

$$\frac{dH}{dt} = V\sin\theta \ . \tag{31}$$

According to the flight operational manuals and airworthiness requirements, limitations (mi - minimum and ma - maximum) should be applied on the velocity, the descent angle and the decision altitude:

$$V \in \left[V_{mi}^{*}, V_{ma}^{*}\right], \tag{32.a}$$

$$\theta \in \left[\theta_{mi}^{*}, \theta_{ma}^{*}\right], \tag{32.b}$$

$$H \geq H_{Dmi}^{*} \ . \tag{32.c}$$

A simple assumption could be applied: during an approach, pilots should decide whether to land or to make a go-around. For this decision they need time, which is the sum of (i) the time to understand and evaluate the given situation, σ_k, (ii) the time for decision making and (iii) the time to react (covering also the reaction time of the aircraft for the applied decision) (Kasyanov 2007):

$$t^{req} = t_{ue}^{req}\left(\sigma_k\right) + t_{dec}^{req}\left(S_a\right) + t_{react}^{req}\left(\sigma_k, S_a\right) \ . \tag{33}$$

Here σ_k defines all possible situations (e.g. σ_1 might be the situation of landing at first approach without any problems, σ_2 could be related to the situation when the under carriage system could not be opened, σ_3 might stand for a landing on the fuselage, σ_5 for go-around, or σ_5 for a successful landing after second approach).

S_a is the chosen solution from the set of possible solutions. It is clear that all solutions have a limited drawback, such as extra cost, or extra fuel.

The subjective factor of pilots might be introduced with the use of the ratio of the required and disposable resources (Kasyanov 2007):

$$\overline{r}_k = \frac{R^{req}\left(\sigma_k\right)}{R^{disp}\left(\sigma_k\right)} = \overline{t}_k = \frac{t^{req}\left(\sigma_k\right)}{t^{disp}\left(\sigma_k\right)} \ . \tag{34}$$

In this case, an endogenous index can be defined as

$$\varepsilon_k\left(\sigma_k\right) = \frac{\overline{r}_k}{1 - \overline{r}_k} = \frac{t^{req}\left(\sigma_k\right)}{t^{disp}\left(\sigma_k\right) - t^{req}\left(\sigma_k\right)} \quad or \quad \varepsilon_k\left(\sigma_k\right) = \frac{t^{req}\left(\sigma_k\right) + t^{dec}\left(S_a\right)}{t^{disp}\left(\sigma_k\right) + t^{dec} - t^{req}\left(\sigma_k\right)} \ , \tag{35}$$

where $t^{dec}\left(S_a\right)$ is a time required to recognize the set of alternative strategies.

Naturally, we can assume that pilots are able to evaluate the consequences of their decisions, and therefore they can evaluate the risk of the applied solutions. Such evaluation can be defined as the subjective probability of situations: $P\left(\sigma_k\right)$, canonic distribution of which as the distribution of canonic assemble of the preferences is assumed to hold the following form:

$$p\left(\sigma_k\right) = \frac{P^{-\alpha}\left(\sigma_k\right) e^{-\beta\varepsilon_k\left(\sigma_k\right)}}{\sum_{q=1}^{2} P^{-\alpha}\left(\sigma_q\right) e^{-\beta\varepsilon_k\left(\sigma_q\right)}} \ , \tag{36}$$

where $p\left(\sigma_k\right)$ describes the distribution of the best alternatives from a negative point of view.

The time-depending coefficients α and β should be chosen in a way to model the endogenous dynamics, model the subjective psycho physiological personalities of pilots. The qualities of the pilots are depending on different factors including "periodical" incapacity to make decisions that increases while getting closer to the decision time (altitude) of go-around.

The (36) has special features: in case of $\overline{t}_k = \frac{t^{req}\left(\sigma_k\right)}{t^{disp}\left(\sigma_k\right)} \to 0$ preferences are determined by the subjective probability, $P\left(\sigma_k\right)$, only, and in case $\overline{t}_k \to 1$, the preference turn into zero. The (36) comes from the solution of the following function:

$$\Phi_p = -\sum_{k=1}^{N} p(\sigma_k)\ln p(\sigma_k) - \beta\sum_{k=1}^{N} p(\sigma_k)\varepsilon_k(\sigma_k) - \alpha\sum_{k=1}^{N} p(\sigma_k)\ln P(\sigma_k) + \gamma\sum_{k=1}^{N} p(\sigma_k) \ . \tag{37}$$

A special feature of this function is that the structure of the efficiency function includes the logarithm of the subjective probability:

$$\eta_p = -\sum_{k=1}^{N}\left(\alpha \ln P(\sigma_k) + \beta\varepsilon(\sigma_k)\right)p(\sigma_k) . \tag{38}$$

The complexity of decision making could be characterized by the uncertainties or the pilots' incapacity to make decisions, which is increasing while getting closer to the minimum decision altitude, H_{Dmi}^{*}. To make decisions, the pilots must overcome their "entropic barrier", H_p. The rate of incapacity could be defined with the norm of entropy:

$$\bar{H}_p = \frac{H_p}{\ln N} \ . \tag{39}$$

Figure 12. shows a simplified decision making situation at an approach about the go-around [Kasyanov 2004, 2007]. At t_0, x_0, $S_a : (\sigma_1, \sigma_2)$ indicates the set of alternative situations with the distribution of preferences $p(\sigma_1)$ and $p(\sigma_2)$ (where σ_1 indicates the landing and σ_2 defines the go-around).

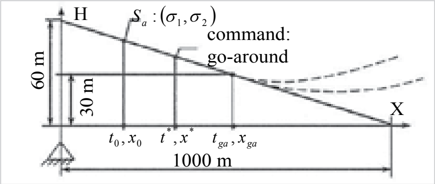


Fig. 12. Final phase of aircraft approach.

The preferences are oscillating, because of the exogenous fluctuation (while decision altitude is getting closer) and the endogenous processes (depending on the uncertainties in the situation awareness and operators (pilots) incapacity to make decisions). If pilots are able to overcome their entropy barrier up to command for go-around (reaching the decision minimum altitude), t^*, x^*, then they could make a decision. Due to this decision, the set of situations, S_a , can be given with the followings:

$$\begin{array}{c} S_a : (\sigma_2); p(\sigma_2) \\ t \prec t^* \\ p(\sigma_1) + p(\sigma_2) = 1 \end{array} \begin{array}{c} \nearrow \\ \\ \searrow \end{array} \begin{array}{c} S_{a1} : (\sigma_2); p(\sigma_2) = 1; p(\sigma_1) = 0 \\ \\ S_{a2} : (\sigma_1); p(\sigma_1) = 1; p(\sigma_2) = 0 \\ t \geq t^* \end{array} \tag{40}$$

If pilots are not able to overcome their entropy barrier before reaching t^*, x^*, the flight situation would become more complex, and therefore the possibility to perform a go-around (case σ_2) might be even out of the possible set of situations.

4.3 Modeling the human way of thinking and decision making

A human as "biomotoric system" uses the information provided by sense organs (sight, hearing, balance, etc.) to determine the motoric actions (Zamora, 2004). From a piloting point of view, balance is the most important from the human sense organs. (As known, pilots are flying upon their "botty" for sensing the aircraft's real spatial position, orientation and motion dynamics (Rohacs, 2006).) The sense of balance (Zamora, 2004) is maintained by a complex interaction of visual inputs (the proprioceptive sensors being affected by gravity and stretch sensors found in muscles, skin, and joints), the inner ear vestibular system, and the central nervous system. Disturbances occurring in any part of the balance system, or even within the brain's integration of inputs, could cause dizziness or unsteadiness.

In addition to this, human has another sensing, kinesthesia (Zamora, 2004) that is the precise awareness of muscle and joint movement that allows us to coordinate our muscles when we walk, talk, and use our hands. It is the sense of kinesthesia that enables us to touch the tip of our nose with our eyes closed or to know which part of the body we should scratch when we itch. This type of sensing is very important in controlling an aircraft and moving in 3D space. (Some scientists believe that future aircraft control system must be operated by thumbs, as the new generation is trained on video-games such as "Game Boy" (Rohacs, 2006).)

The main element of the "human biomotoric system" is the human brain that is the anteriormost part of the central nervous system in humans as well as the primary control center for the peripheral nervous system.

The human brain (Russel, 1979; Davidmann, 1998). is a very complex system based on the net of brain cells called as neurons that specialize in communication. The brain contains circuits of interconnected neurons that pass information between themselves.

The neurons contain the dendrites, cell body and axon. In neurons, information passes from dendrites through the cell body and down the axon (Russel, 1979; Davidmann, 1998).

Principally, transmission of information through the neuron is an electrical process. The passage of a nerve impulse starts at a dendrite, it then travels through the cell body, down the axon to an axon terminal. Axon terminals lie close to the dendrites of neighboring neurons.

From control theory point of view, the most important behavior of human brain is the memory, namely learning, memorizing and remembering (Receiving, Storing and Recalling). Generally, human beings are learning all the time, storing information and then recalling it when it is required (Davidmann, 1998). After the investigation of human thinking, including recognition, information analysis, reasoning, decision support (Rohacs, 2006; 2007) the human way of thinking is found to be have the following behaviors:

- syntactic and semantic processing of the sensed information,
- working on the basis of large net of small and simplified articles (neurons),
- using the complex system oriented approach,
- making parallel thinking and activity,
- learning (synthesis of the new knowledge),
- model-formation and using the models (including verbal models applied in learning processes and complex mathematical representation),
- long-term memory,

- tacit knowledge (took in practice),
- intentional thinking (goal and wish),
- intuition (subconscious thinking),
- creativity (finding the contexts),
- innovativity (making originally new minds, things),
- unexpected values can be appeared,
- jumping from quantity to quality.

Seeing all the features listed above, it is clear that human thinking and decision making is a very complex process, containing some chaotic effects.

There is not enough information on the physical, systematic, intellectual, psychophysiology, etc. characteristics of the subjective analysis, about the way of thinking and making decision of subjects-operators like pilots. Only limited information is available on the time effects, possible damping the non-linear oscillations, long term memory, etc. making the decision system chaotic.

Professor Kasyanov introduced a special chaotic model (Kasyanov, 2007) based on the modified Lorenz attractor (Stogatz, 1994) for modeling the endogenous dynamics of the described process.

$$\begin{aligned} \frac{dX}{dt} &= aY - bZ - hX^2 + f(t); \\ \frac{dY}{dt} &= -Y - XZ + cX - mY^2; \\ \frac{dZ}{dt} &= XY - dZ - nZ^2. \end{aligned} \tag{41}$$

where *a, b, c, d, h, m, n* are the constants while *f* takes into account the disturbance. (In case of *h*=*m*=*n*=0 and *f(t)*=0 the model turns into the classic form of Lorenz attractor.)

Principally, there are no strong arguments explaining the use of Lorenz attractor to model the human way of decision making (human thinking) (Dartnell, 2010; Krakovska, 2009), but the results of application are close to real situations.

4.4 Results of investigations

Professor Kasyanov investigated various model types, and evaluated the model parameters (Kasyanov, 2007). For a medium sized aircraft (weight of aircraft, W = 10^6 N; wing area, S = 100 m^2; wing aspect ratio A = 7; thrust T = 9.4 x 10^4 N; and velocity V = 70 m/sec) with commercial pilots, he recommended to use the following values: a=8; b=8; c=20; d.43; f=0.8; h= 0.065; m=0.065; n=0.065.

Using these parameters, the subjective probabilities might be chosen as $P(\sigma_1) = 0.53$, $P(\sigma_2) = 0.6$ and $\varepsilon_1 = 5.5 + 0.01t$, $\varepsilon_2 = 5.4 + 0.04t$ take into account the decreasing difference in the required and the available time for the decision. The typical results of using the described model are shown in the Figure 13., demonstrating the chaotic character of decision making.

In this example, the figures demonstrate that pilots are unfixed for a period of about 10 sec, during which their preferences (A, B) are changing by sudden oscillations and the H entropy at the beginning is rather high. If the limit for the entropy would be 0.7 (that is still quit high) then decisions could be made in about 10 sec. This means that pilots will not able to do that according to the Figure 12.

If the parameters are set to a=10; b=10; c=35; d=1; f=0; h= 0.065; m=0.065; n=0.065 and $P(\sigma_1)=0.53,\ \ P(\sigma_2)=0.6$, then (see Figure 14) the entropy would quickly decrease and the decision could be made in about 3 sec. According to the ICAO requirements, time $t=t_{ga}-t^*$ (see Figure 12.) should not be less than 3.16 sec. Therefore, if the situation presented in the Figure 12. appears before t_0, x_0, then the right decision could be made.

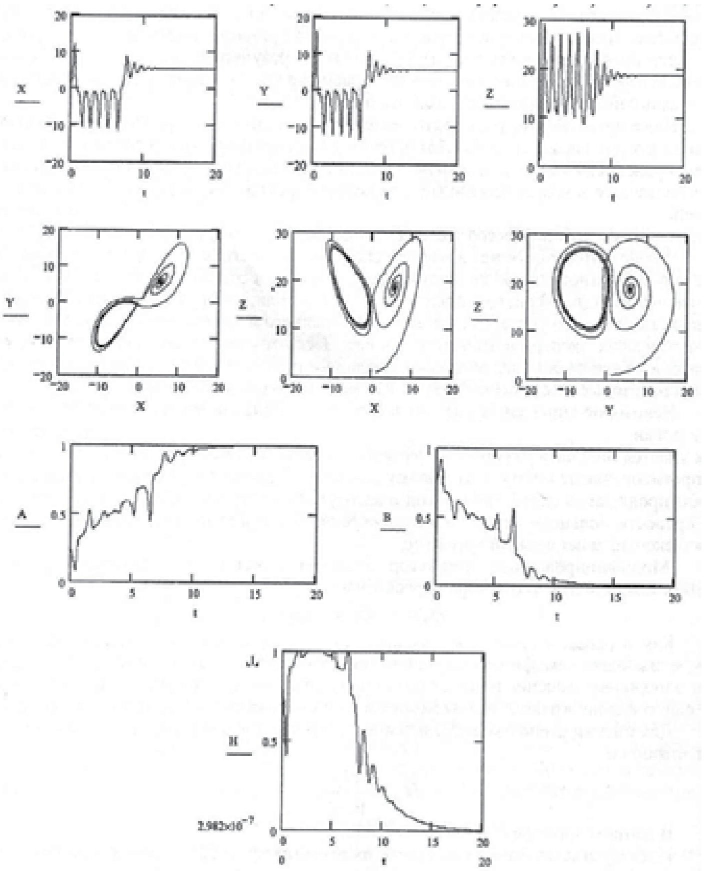


Fig. 13. Results of using the developed model to landing of a medium sized aircraft.

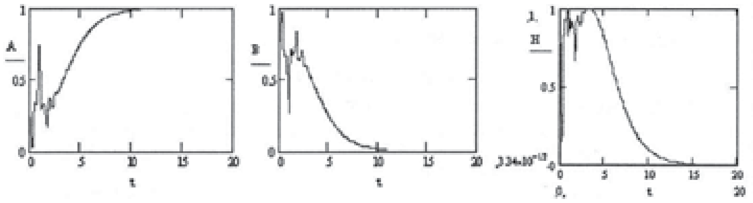


Fig. 14. Results, when the parameters are chosen for well-skilled pilots.

From the results of using the developed model to the landing phase of a small aircraft (such as analyzed in the Hungarian national projects SafeFly: development of the innovative safety technologies for a 4 seats composite aircraft and EU FP7 project PPlane: Personal Plane: Assessment and Validation of Pioneering Concepts for Personal Air Transport Systems, Grant agreement no.233805) several important conclusions had been made (Rohacs et all, 2011; Rohacs & Kasyanov, 2011; Rohacs, 2010).

During the final approach, the common airliner pilots require about three times more time for making decision on go-around than the well practiced colleagues.

Using the developed model and condition defined by Figure 12, the descent velocity of a small aircraft could be determined to about 100 km/h for airliner common pilots, and 75 km/h for those of less-skilled.

In this case, the airport can be designed with a landing distance of less than 600 m (runway about 250 - 300 m) and a protected zone under the approach (to overfly the altitude of 100 m) of about 1500 m. These characteristics enable to place small airports close / closer to the city center.

5. Conclusions

This chapter introduced the subjective analysis methodology into the investigation of the real flight situation, flight safety. The subject, as pilot operator generates his decision on the basis of his subjective situation analysis depending on the available information and his psycho-physiological condition. The subjective factor is the time available for the decision of the given tasks.

After the general discussion on flight safety, its metrics and accident statistics, an original approach was introduced to study the role of human factors in flight safety. The deterministic or stochastic models of flight safety are not included clearly the subjective behaviors of human operators. However, the subjective analysis may open a new vision on the flight safety and may result to improve the aircraft development methods and tools.

The subjective decision making of pilots was modeled by the modified Lorenz attractor that needs further investigation and explanation. The applicability of the developed methodology was applied to study the small aircraft final approach and landing. It demonstrates that the model is suitable to investigate the difference between the well trained and less-skilled pilots. The model helped in the definition of the aircraft and airport characteristics for the personal air transportation system.

This work is connected to the scientific program of the "Development of the innovative safety technologies for a 4 seats composite aircraft - SafeFly" (NKTH-MAG ZRt. OM-000167/2008) supported by Hungarian National Development Office and Personal Plane - PPLANE Projhect supported by EU FPO7 (Contract No - 233805) and the research is supported by the Hungarian National New Széchenyi Plan (TÁMOP-4.2.2/B-10/1-2010-0009)

6. References

Afrazeh A & Bartsch H (2007) Human reliability and flight safety. International Journal of Reliability, Quality and Safety Engineering 14(5): 501–516

Banos, A., Lamnabhi-Lagarrigue, F., Montoya, F. J. (2001) Advances in the Control of Nonlinear Systems, Lecture Notes in Control and Information Sciences 264, Spinger - Verlag London Berlin Heidelberg, 2001,

Bezopastnostj (1988) poletov letateljnüh apparatov (pod red. A.I. Starikova). Transport, Moscow, 1988.

CASA (2005) AC 139-16(0): Developing a Safety Management System at Your Aerodrome, Australian Government – Civil Aviation Safety Authority (CASA) Advisory Circular, 2005.

Commercial (2000) Aviation Safety Team (CAST), Process Overview, http://www.icao.int/fsix/cast/CAST%20Process%20Overview%209-29-03.ppt

Dartnell, L. (2010) Chaos in the brain, (2010), http://plus.maths.org/content/chaos-brain

Davidmann, M. (1998) How the Human Brain Developed and How the Human Mind Works, 1998, 2006, http://www.solbaram.org/articles/humind.html

EASA (2008) Annual Safety Review, EASA 2008.

FAA (2006) Introduction to Safety Management Systems for Air Operators, Federal Aviation Administration Advisory Circular 120-92: Appendix 1, Jun. 22, 2006.

Fliens, M., Levine, J., Martin, P., Rouchen, P. (1999) A Lie-Bäcklund Approach to Equivalence and Flatness of Nonlinear Systems, IEEE Transactions on Automatic Control, Vol. 44, No. 5, MAY 1999, 922 - 937.

Flight (2000) Control design – Best Practice, NATO, RTO-TR-029, AC/323(SCI)TP/23, Neuilly-sur-Seine Cedex, France, 2000.

Gardiner, C. W. (2004) Handbook of Stochastic Methods for Physics, Chemistry and the natural Sciences, Springer Series in Synergetics, Springer-Verlag Berlin Heidenberg New York, 2004.

Gudkov A. I., Lesakov P. S. (1968) Vneisnie nagruzki i prochnostj letateljnih apparatov, Masinostroyeniye Moscow, 1968.

Howard R. W. (1980) Progress in the Use of Automatic Flight Controls in Safety Critical Applications, The Aeronautical Journals, 1980 v. 84. X. No. 837. pp.316-326.

Ibe, O. C. (2008) Markov Process for Stochastic Modeling, Academic press. 2008

Kasyanov, V. A. (2004) Flight modelling (in Russian),), National Aviation University, Kiev, 2004, 400 p.

Kasyanov, V. A. (2007) Subjective analysis (in Russian), National Aviation University, Kiev, 2007, 512. p.

Krakovska A. (2009) Two Decades of Search for Chaos in Brain, MEASUREMENT 2009, Proceedings of the 7th International Conference, Smolenice, Slovakia, pp. 90 - 94.

Lee, C. A. (2003) Human error in aviation (2003) http://www.carrielee.net/pdfs/HumanError.pdf

Pavlov, V.V., Chepijenko, V. I. (2009) Ergaticheskie sistemii upravleniya (Ergatic control systems), gasudarstvennij nauchno-Isledovatjelskij Institute Avitacii, Kiev, http://194.44.242.245:8080/dspace/bitstream/handle/123456789/7645/01-Pavlov.pdf?sequence=1

Ponomarenko, V. (2000) Kingdom in the Sky – Earthly Fetters and Heavenly Freedoms. The Pilot's Approach to the Military Flight Environment, NATO RTO-AG-338 AC/323(HFM)TP/5, July, 2000

Rohacs, J. (1986) Deviation of Aerodynamic Characteristics and Performance Data of Aircraft in the Operational Process. (Ph.D. thesis) KIIGA, Kiev , 1986.

Rohacs, J. (1990) Analysis of Methods for Modeling Real Flight Situations. 17th Congress of the International Council of the Aeronautical Sciences, Stockholm, Sweden, Sept. 9-14. 1990, ICAS Proceedings 1990. pp. 2046-2054.

Rohács, J. (1995) Repülések biztonsága (Safety of Flights) Bólyai János Műszaki Katonai Főiskola (Military Technology High School Named János Bólyai), Budapest, 1995.

Rohács, J. (1998) Revolution in Safety Sciences -- Application of the Micro Devices „Progress in Safety Sciences and Technology" (Edited by Zeng Quingxuan, Wang Liqiong, Xie Xianping, Qian Xinming) Science Press Beijing / New York, 1998, pp. 969 – 973.

Rohacs, J. (2000) Risk Analysis of Systems with System Anomalies and Common Failures „Progress in Safety Sciences and Technology" Vol. II. Part. A. (edited by Li Shengcai, Jing Guoxun, Qian Xinming), Chemical Industry Press, Beijing, 2000, 550–560.

Rohacs, J. (2006) Development of the control based on the biologycal principles, ICAS Congress, Hamburg, 2006 Sept. CD-ROM, ICAS, 2006

Rohacs, J. (2007) Some thoughts about the biological principle based control, Sixth International Conference on Mathematical Probéems and Engineering and Aerospace Sciences (ed. By Sivasundaram, S.), Cambridge Scientific Publisherm 2007, pp. 627-638, ISBN 978-1-904868-56-9

Rohacs, J. (2010) Subjective Aspects of the less-skilled Pilots, Performance, Safety and Well-being in Aviation, Proceedings of the 29th Conference of the European Association for Aviation Psychology, 20-24 September 2010, Budapest, Hungary, (edited by A. Droog, M. Heese), ISBN: 978-90-815253-2-9 pp. 153-159

Rohacs, J., Kasyanov, V. A. (2011) Pilot subjective decisions in aircraft active control system, J. Theor. Appl. Mech., 49, 1, pp. 175-186, 2011

Rohács, J., Németh, M. (1997) Effects of Aircraft Anomalies on Flight Safety „Aviation Safety (Editor: Hans M. Soekkha) VSP, Ultrecht, The Netherland, Tokyo, Japan, 1997, pp. 203–211.

Rohacs, J., Rohacs, D., Jankovics, I., Rozental, S., Hlinka, J, Katrnak, T., Helena, T. (2011) Personal aircraft system improvements Internal report, PPLANE (EU FP 7 projects), Budapest, 2011.

Rohacs, J., Simon I. (1989) Repülőgépek és helikopterek üzemeltetési zsebkönyve (The handbook of airplane and helicopter operation) Müszaki Könyvkiadó, Budapest, 1989.

Ropp, T. D., Dillman, B. G. (2008) Standardized Measures of Safety: Finding Global Common Ground for Safety Metrics, IAJC -IJME Conference, International Conference on Engineering and Technology, 2008, Nashville, TN, US, ENT 203: Topics in Aviation Safety, Paper No. 29.

Russel, P. (1979) The Brain Book, Penguin Group, new York, 1979.

Shin, J. (2000) The NASA Aviation Safety Program: Overview, Nasa, 2000, NASA/TM—2000-209810, http://gltrs.grc.nasa.gov/reports/2000/TM-2000-209810.pdf

Statistical (2008) summary of commercial jet airplane accidents worldwide operations 1959 - 2008, Boeing, http://www.boeing.com/news/techissues/pdf/statsum.pdf

Strogatz, S. (1994) Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering. Perseus Books, Massachusetts, US, 1994.

Tihonov, V.I., Mironov, M.A. (1977) Markovskie processi. Sovetskoe Radio, Moscow, 1977.

Transport (2007) Canada, TP 14343, Implementation Procedures guide for Air Operators and Approved Maintenance Organizations, April, 2007.

White, J.: (2009) Aviation safety program, NASA, (2009) http://www.docstoc.com/docs/798142/NASA-s-Aviation-Safety-Program

Zamora, A. (2004) "Human Sense Organs - The Five Senses." Anatomy and Structure of Human Sense Organs. Scientific Psychic, 2004, 2011, http://www.scientificpsychic.com/workbook/chapter2.htm

Part 3

Aircraft Electrical Systems

13

Power Generation and Distribution System for a More Electric Aircraft - A Review

Ahmed Abdel-Hafez
Shaqra University
Kingdom of Saudi Arabia

1. Introduction

More-Electric Aircraft (MEA) is the future trend in adopting single power type for driving the non-propulsive aircraft systems; i.e. is the electrical power. The MEA is anticipated to achieve numerous advantages such as optimising the aircraft performance and decreasing the operation and maintenance costs. Moreover, MEA reduces the emissions of air pollutant gases from aircrafts, which can contribute in signifcantly solving some of the problems of climate change. However, the MEA puts some challenges on the aircraft electrical system, both in the amount of the required power and the processing and management of this power. This chapter introduces the outline for MEA. It investigates possible topologies for the power system of the aircraft. The different electric power generation options are highlighted; while at the same time assessing the generator topologies. It also includes a general review of the power electronic interfacing circuits. Also, the key design requirements for an interfacing circuit are addressed. Finally, a glance at protection facilities for the aircraft power system is given.

2. More electric aircraft

Recently, the aircraft industry has achieved a tremendous progress both in civil and military sectors (AbdElhafez & Forsyth, 2008,2009; Cronin, 1990; Moir & Seabridge, 2001). For example some current commercial aircraft operate at weights of over 300 000 kg and have the ability to fly up to 16 000 km in non-stop journey at speed of 1000 km/h (AbdElhafez & Forsyth, 2009).

The non-propulsive aircraft systems are typically driven by a combination of different secondary power drives/subsystems such as hydraulic, pneumatic, electrical and mechanical (AbdElhafez & Forsyth, 2008,2009; Jones, 1999; Moir, 1999; Moir & Seabridge, 2001; Quigley, 1993). These powers subsystems are all soured from the aircraft main engine by different methods. For example, mechanical power is extracted from the engine by a driven shaft and distributed to a gearbox to drive lubrication pumps, fuel pumps, hydraulic pumps and electrical generators (AbdElhafez & Forsyth, 2009; Jones, 1999; Moir, 1999; Quigley, 1993). Pneumatic power is obtained by bleeding the compressor to drive turbine motors for the engine's starter subsystem, and wing anti-icing and Environmental Control Systems (ECS), while electrical power and hydraulic power subsystems are distributed throughout the aircraft for driving actuation systems such as flight control actuators, landing gear brakes, utility actuators, avionics, lighting, galleys, commercial loads and weapon systems (AbdElhafez & Forsyth, 2009, Howse, 2003; Jones, 1999; Moir, 1998, 1999; Quigley, 1993).

This combination had always been debated, because these systems had become rather complicated, and their interactions reduce the efficiency of the whole system. For example, a simple leak in pneumatic or hydraulic system jeopardises the journey by grounding the aircraft, and eventually causing inconvenient flight delays. The leak is usually difficult to locate and once located it cannot easily be handled (AbdElhafez & Forsyth, 2009; Cutts, 2002; Hoffman, 1985; Moir, 1998; Pearson, 1998; Rosero, et al, 2007; Weimer, 1993). Furthermore, from manufacturing point of view reducing the cost of ownership, increasing the profit and some anticipated future legislation regarding the climate changes demand radical changes to the entire aircraft, as it is no longer sufficient to optimise the current aircraft sub-systems and components individually to achieve these goals (AbdElhafez & Forsyth, 2009; Andrade, 1992; Cutts, 2002; Clyod, 1997; Emadi & Ehsani, 2000; Hoffman, 1985; Moir, 1998; Pearson, 1998; Ponton, 1998; Rosero, etal, 2007; Weimer, 1993).

The trend is using the electrical power for sourcing and distributing non-propulsive aircraft engine powers. This trend is defined as MEA. The MEA concept is utterly not a new concept, it has been investigated for several decades since W.W. II (Andrade, 1992; Cutts, 2002; Pearson, 1998; Ponton, 1998; Weimer, 1993). However, due to the lack of electric power generation capabilities and prohibitive volume of power conditioning equipments, the focus has been drifted into the conventional power types. Relatively, the recent technology breakthroughs in the field of power electronics systems, fault-tolerant electric machines, electro- hydrostatic actuators, electromechanical actuators, and fault-tolerant electrical power systems have renewed the interest in MEA (AbdElhafez & Forsyth, 2009; Andrade, 1992; Cutts, 2002; Clyod, 1997; Emadi & Ehsani, 2000; Hoffman, 1985; Moir, 1998; Pearson, 1998; Ponton, 1998; Rosero, etal, 2007; Weimer, 1993). A comparison between conventional aircraft subsystems and MEA subsystems is shown in Fig. 1 (AbdElhafez & Forsyth, 2009).

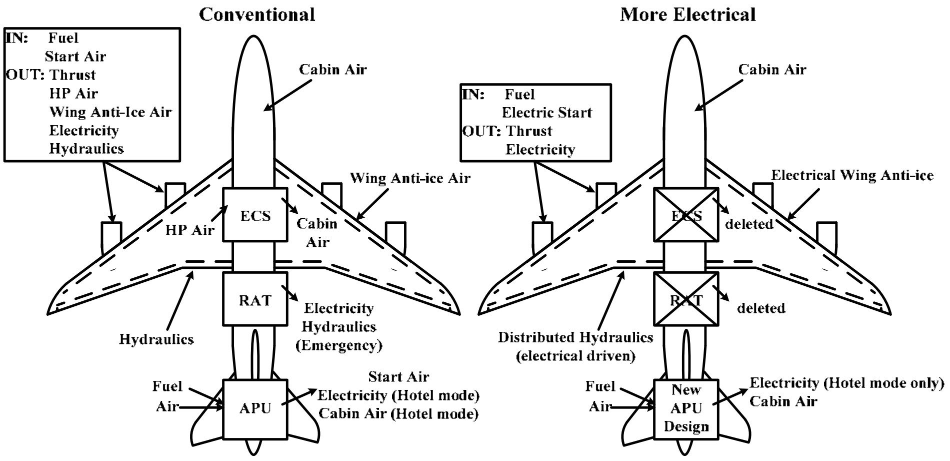


Fig. 1. Comparison between conventional systems aircraft and MEA systems (AbdElhafez & Forsyth, 2009).

The adoption of MEA in the future aircraft both in civil and military sectors will result in tremendous benefits such as:-

1. Removal of hydraulic systems, which are costly, labour-intensive, and susceptible to leakage and contamination problems, improves the aircraft reliability, vulnerability, and reduces complexity, redundancy, weight, installation and running cost (Cutts, 2002; Pearson, 1998; Ponton, 1998; Quigely, 1993; Weimer, 1993).
2. Deployment of electrical starting for the aero-engine through the engine starter/generator scheme eliminates the engine tower shaft and gears, power take-off shaft, accessory gearboxes and reduces engine starting power especially in the cold conditions and aircraft front area (Clyod, 1997; Emadi & Ehsani, 2000; Jones, 1999; Moir & Seabridge, 2001).
3. Utilization of the Advanced Magnetic Bearing (AMB) system, which could be integrated into the internal starter/generator for both the main engine and auxiliary power units, allows for oil-free, gear-free engine area (AbdElhafez & Forsyth, 2009; Andrade & Tenning, 1992a, 1992b; Hoffman et al., 1985; Jones, 1999; Moir & Seabridge, 2001).
4. In MEA, using a fan shaft generator that allowing emergency power extraction under windmill conditions removes the conventional inefficient single-shot ram air turbine, which increases the aircraft's reliability, and survivability under engine-failure conditions (AbdElhafez & Forsyth, 2009; Andrade & Tenning, 1992a, 1992b; Quigley, 1993).
5. Replacement of the engine-bleed system by electric motor-driven pumps reduces the complexity and the installation cost, and improves the efficiency (Jones, 1999).

In general, adopting MEA will revolutionise the aerospace industry completely, and significant improvements in terms of aircraft-empty weight, reconfigureability, fuel consumption, overall cost, maintainability, supportability, and system reliability will be achieved (AbdElhafez & Forsyth, 2009; Clyod, 1997; Cronin, 1990; Emadi & Ehsani, 2000; Hoffman et at., 1985; Moir, 21998, 1999, Weimer, 1993).

On the other hand, the MEA requires more demand on the aircraft electric power system in areas of power generation and handling, reliability, and fault tolerance. These entails innovations in power generation, processing, distribution and management systems (AbdElhafez & Forsyth, 2009; Clyod, 1997; Cronin, 1990; Emadi & Ehsani, 2000; Hoffman et at., 1985; Moir, 21998, 1999).

The proceeding sections briefly discuss a general overview of the electrical power distribution and management, generation and processing systems in MEA.

3. Distribution systems

The power distribution system of the most in-service civil aircrafts is composed of combined of AC and DC topologies. E.g., an AC supply of 115V/400Hz is used to power large loads as such as galleys, while the DC supply of 28V DC is used for avionics, flight control and battery-driven vital services.

Recently there is a trend for using only high voltage DC system for power distribution and management in MEA. A number of factors encouraged this trend (AbdElhafez & Forsyth, 2009; Cross et al., 2002; Hoffman, 1985; Jones, 1999; Glennon, 1998; Maldonado et al., 1996, 1997, 1999; Mallov et al., 2000; Quigely, 1993; Worth, 1990) :

1. Adopting the new generation options as variable frequency,
2. Recent advancements in the areas of interfacing circuits, control techniques and protection systems,
3. The advantages of the high voltage DC distribution system in reducing the weight, the size and the losses, while increasing the levels of the transmitted power.

Some values of the system voltage are presently under research. These values are: 270, 350 and 540V. The exact value, however, is determined by a number of factors such as, the capabilities of DC switchgear, the availability of the components and the risk of corona discharge at high altitude and reduced pressure (Brockschmidt, 1999).

Different topologies were suggested for implementing the distribution system in MEA (Cross et al., 2002; Hoffman, 1985; Glennon, 1998; Maldonado et al., 1996, 1997, 1999; Mallov et al., 2000; Worth, 1990). In the following four main candidates of these topologies are briefly reviewed, as follows :

1. Centralized Electrical Power Distribution System (CEPDS),
2. Semi-Distributed Electrical Power Distribution System (SDEPDS),
3. Advanced Electrical Power Distribution System (AEPDS),
4. Fault-Tolerant Electrical Power Distribution System (FTEPDS).

3.1 Centralized Electrical Power Distribution System (CEPDS)

CEPDS is a point-to-point radial power distribution system as shown in Figure 2. It has only one distribution centre. The generators supply this distribution centre. The electrical power is being processed and fed to the different electrical loads. The distribution centre is normally positioned in the avionics bay, Figure 2, where the voltage regulation is also located. In this system, each load is supplied individually from the power distribution centre (Cross et al., 2002; Worth et al., 1990). CEPDS has a number of advantages, such as :

1. The ease of maintenance, since all equipments are located in one place, i.e. avionics bay.
2. Decoupling between loads; thus the disturbance in a load is not transferred to the others.
3. Fault-tolerance, as the main buses are highly protected.

As stated CEPDS may have significant advantages, however it also has a number of disadvantages, such as:

1. CEPDS suffers from the difficulty of upgrading.
2. The faults in the distribution system affect probably all loads and disable the entire system.
3. CEPDS is cumbersome, expensive and unreliable, as each load has to be wired from the avionics bay.
4. Costly and bulky protection system has to be deployed to protect the distribution system.

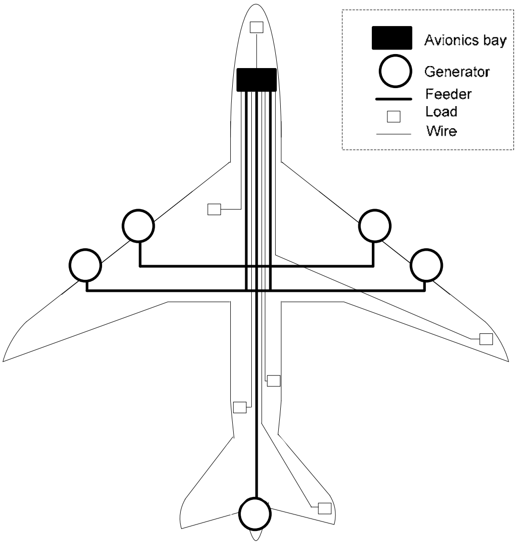


Fig. 2. Centralised Electrical Power Distribution System CEPDS for the MEA (AbdElhafez & Forsyth, 2009).

3.2 Semi-Distributed Electrical Power Distribution System (SDEPDS)

SDEPDS was proposed to overcome the problems of CEPDS (AbdElhafez & Forsyth; 2009; Cross et al., 2002; Hoffman, 1985; Glennon, 1998; Maldonado et al., 1996, 1997, 1999; Mallov et al., 2000; Worth, 1990) . The SDEPDS as shown in Figure 3 has a large number of Power Distribution Centres (PDCs). These centres are scaled versions of PDCs in CEPDS. The PDCs are distributed around the aircraft in such way to optimise the system volume, weight and reliability. They are located, Figure 3, close to load centres.

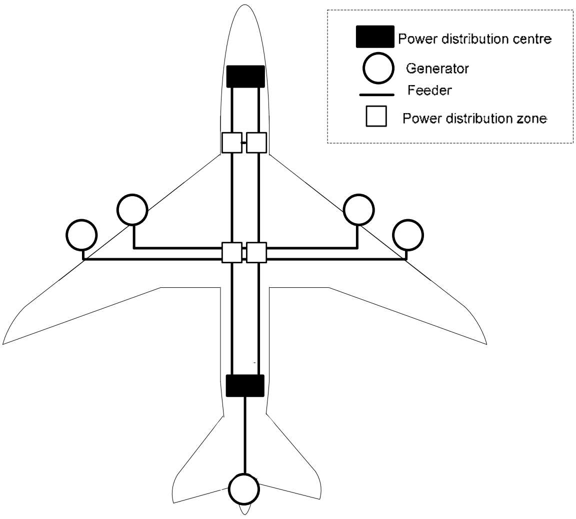


Fig. 3. Semi-Distributed Electrical Power Distribution System SDEPDS for the MEA (AbdElhafez & Forsyth, 2009)

SDEPDS has a number of advantages :

1. Elevated power quality and improved Electromagntic compatibility, due to the position of the distribution centres near to the loads,
2. High efficiency and cost effective, attributed to the deployment of electrical components with small weight/volume in PDCs,
3. Efficient and stable system operation, due to reduced losses/voltage drops across the distribution network.
4. High level of redundancy in primary power distribution path, due to the strategy of increasing and distributing the PDCs,
5. Simplicity and flexibility of upgrading.

On the other hand, the close coupling between the loads in SDEPDS may reduce the reliability, as faults/ disturbances in a load can propagate to nearby loads. Moreover, extra equipments are required to perform the monitoring and control of the distributed PDCs.

3.3 Advanced Electrical Power Distribution System (AEPDS)

AEPDS is a flexible, fault-tolerant system controlled by a redundant microprocessor system. This system is developed to replace the conventionally centralized and semi-distributed systems.

AEPDS as shown in Figure 4, is highly protected. The electrical power from the generators, Auxiliary Power Unit (APU), battery and ground sources is supplied to the primary power distribution, where the Contactor Control Units (CCU) and high power contactors are located. The primary power distribution centre performs a number of tasks: voltage/frequency regulation, damping oscillation and transient and controlling the flow of the reactive power.

The aircraft loads are supplied via the Relay Switching Units (RSU). Each RSU is controlled and monitored by a Remote Terminal (RT) unit. The AEPDS is controlled by either one of the two redundant Electrical load Management Units (ELMU). The ELMU interact and exchange data/control strategies with the RTs through a quad redundant data bus (Mollov et al., 2002; Worth, 1990) .

The AEPDS has improved performance than CEPDS and SDEPDS. This is attributed for the following (Worth, 1990):

1. AEPDS reduces the aircraft life cycle cost, as the system reconfiguration in case of aircraft modification/upgrade can easily be accommodated.
2. AEPDS can detect deviant conditions of current/voltage and provide instantaneous load shut-off.
3. A major reduction in the weight and wiring in the AEPDS is achieved due to the elimination of circuit breaker panels from the flight deck stands.
4. AEPDS is fault-tolerant distribution system.

The AEPDS has the disadvantage of concentrating the distribution and the management of power supplied by the generating units/sources into a single unit; therefore a fault in this unit may interrupt the whole system operation.

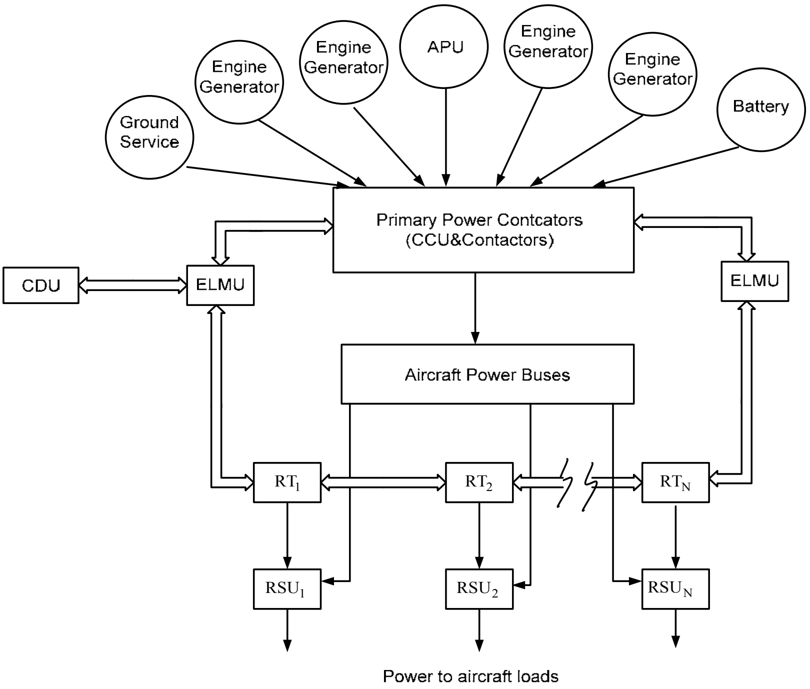


Fig. 4. Advanced Electrical Power Distribution System AEPDS for MEA (AbdElhafez & Forsyth, 2009).

3.4 Faulted-Tolerant Electrical Power Distribution System (FTEPDS)

FTEPDS is adequatly protected. A typical FTEPDS for a two-engine aircraft is shown in Figure 5. The system is composed of two switch matrices, six multi-purpose converter, six generators and different loads. The source and load switch matrices could be implemented by using mechanical or solid-state switches. However, the latter has the advantages of controllability, fast response and high efficiency (Cross et al., 2002; Hoffman et al., 1985; Glennon, 1998; Maldonado et al., 1996, 1997) over the former.

FTEPDS is a mixed distribution system; the AC power from generators and airport grid are connected to source switch matrix, while 270V DC system is interfaced with the converters. The bi-directional power flow in the generators indicates that system allows integral starter/generator operation, where the generator initally acts as a motor to start the jet engine; then it operates as generator to supply the aircraft electrical system. Also 270V DC system has a bi-directional power flow; this is to charge the batteries and other energy storage units during normal flight conditions. However, during faults and disturbances the DC system injects power to stabilize the aircraft distribution system.

FTEPDS enjoys the following advantages:

1. The ability to start the aircraft engine by generator/starter scheme,
2. High redundancy,
3. Fault-tolerant, the ability of the system to continue functioning even under an engine failure,

However FEEPDS has a serious drawback; a fault in source/load switch matrices may interrupt the operation of the entire system.

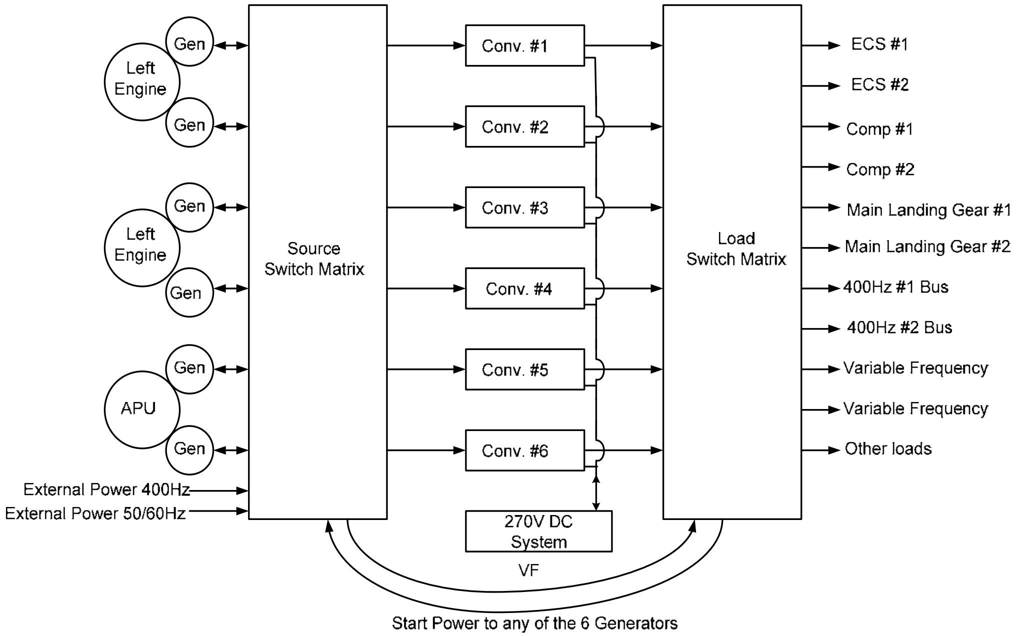


Fig. 5. Fault-tolerant Electrical Power Distribution System FTEPDS for MEA (AbdElhafez & Forsyth, 2009).

4. Electric power generation in MEA

Since its advent, generated electrical power uilization has been rising rapidly. The growth of electrical power generation/application in aircrafts is shown in Figure 6. The quadratic growth is attributed to the increase aircraft system loading such as : galley and In-Flight Entertainment (IFE) systems.

MEA recently is one of the major driving force in electric generation in aircrafts (AbdElhafez & Forsyth, 2009; Andrade, 1992; Bansal et al., 2003, 2005; Howse, 2003; Jones, 1999; Quigely, 1993; Mellor et al., 2005; Moir & Seabridge, 2001; Moir, 1999; Raimondi et al., 2002). Not only are aircraft electrical system power levels growing, but the diversity of the power generation types is increasing as well.

4.1 Schemes of power generation

The various in-service and prospect schemes of electrical power generation are shown in Figure 7 (AbdElhafez & Forsyth, 2009; Cossar, 2004)

Examples of civil/military aircraft and the corresponding generation scheme are given in Table 1.

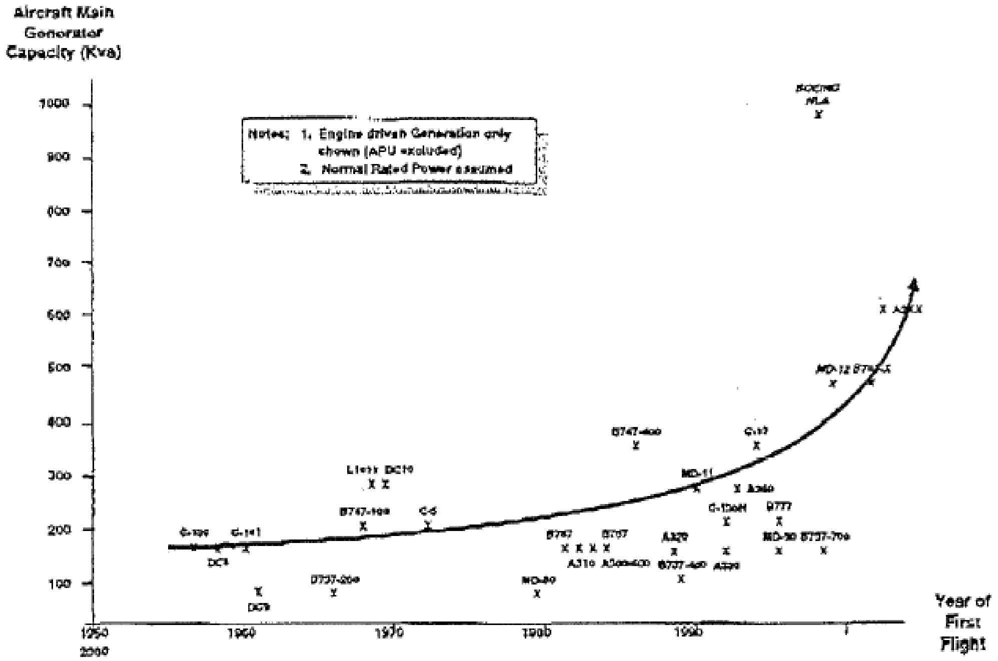


Fig. 6. Growth of generated electrical power in aircraft since the first flight (AbdElhafez & Forsyth, 2009).

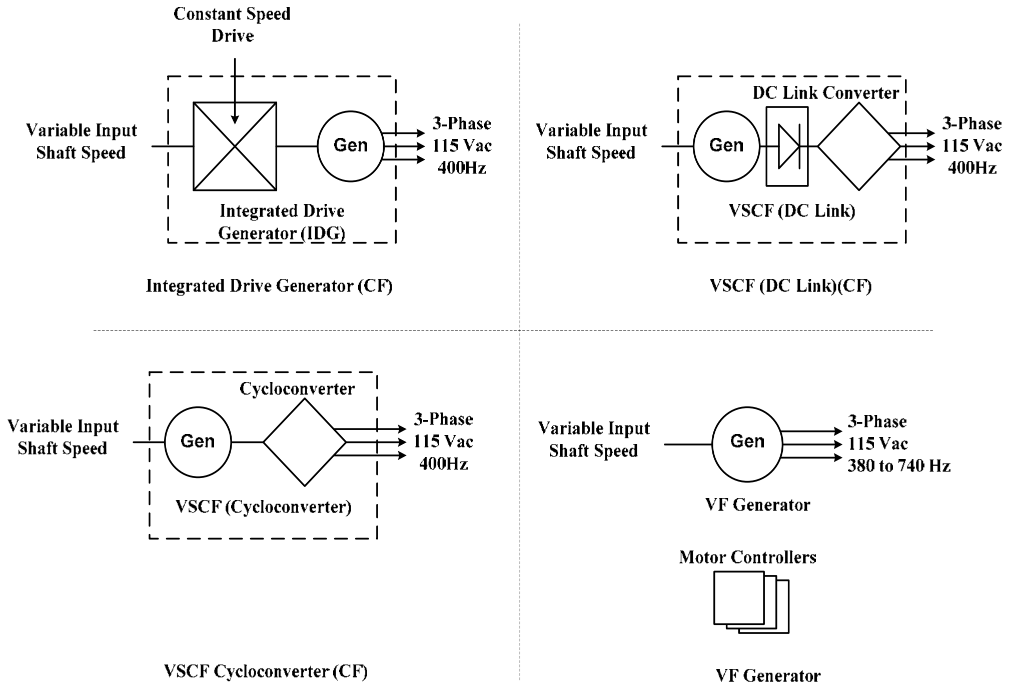


Fig. 7. Aircraft Electrical Power Generation Options (AbdElhafez & Forsyth, 2009).

Generation scheme	Civil		Military	
	Aircraft	Rating (kVA)	Aircraft	Rating (kVA)
CF(IDG)	B777 A340 B373NG MD-12 B747-X B717 B767-400 Do728	2x120 4x90 2x90 4x120 4x120 2x40 2x40 2x40		
VSCF (Cycloconverters)			F-18E/F	2x60/65
VSCF(DC-link)	B77(backup) MD-90	2x20 2x75	C145	2x120
VF	Global Ex Horizon A3xx	4x50 2x20/25 4x150	Boeing JSF	2x50
270VDC			F-22 Raptor X-35A/B/C	2x70 2x50

Table 1. Civil/Military aircraft and electrical power generation techniques (AbdElhafez & Forsyth, 2009).

A brief review of the different generation techniques is given below where the focus is on the merits/demerits of each.

4.1.1 Constant frequency

The constant Frequency (CF), three-phase 115V/400Hz scheme is the most common electric power generation option. This scheme is in-service in most civil aircrafts as shown in Table 1. The CF is alternatively termed Integrated Drive Generator (IDG).

In CF system, the generator is attached to the engine through unreliable and cumbersome mechanical gearbox. This gearbox is essential to ensure that the generator speed is constant irrespective of the engine speed and aircraft status. The frequency f of the generated power is related to generator speed N by,

$$f=\frac{PN}{120} \tag{1}$$

where f is output frequency in cycle/sec(Hz); N is generator speed in revolution per minutes (rpm) and P is the number of magnetic poles. Maintaining generator speed N constant ensures that output frequency remains fixed; however the CF has a number of disadvantages (AbdElhafez & Forsyth, 2009; Cossar, 2004; Howse, 2003; Jones, 1999; Quigely, 1993; Moir, 1999; Raimondi et al., 2002):

1. The interfacing mechanical gear box is unreliable, inefficient and costly, which reduces the overall system efficiency.

2. The system has to be examined for every flight, increasing the operational costs.
3. CF could not allow internal starting for the aero-engine by integral starter/generator scheme.

4.1.2 DC-link system

Variable Speed Constant Frequency (VSCF) DC-link system is now the preferred option for most new military aircraft and some commercial aircraft, Table 1. The generator in this scheme, Figure 7, is attached directly to the engine, thus according to (1) the output frequency will vary with engine speed. The engine speed is subjected to wide variation during the normal course of flight, and so does the frequency; therefore interfacing circuits are required to change the generator output power into usable form.

The output of the generator is supplied to diode rectifiers, which converts the variable frequency AC power into DC form. Then three-phase inverters are used to convert the DC power into three-phase 115V/400Hz AC type. This is the typical form of VSCF DC-link system. However, recently several topologies were reported. These new topologies produce improved performance regarding harmonics, reactive power flow and system stability. Moreover, the range of VSCF DC-link system has been widened due to the recent advancements in field of high power electronic switches. VSCF DC-link option is generally characterised by simplicity and reliability (AbdElhafez & Forsyth, 2009; Hoffman et al., 1985; Ferriera, 1995; Moir, 1999; Quigley, 1993; Olaiya, &. Buchan, 1999; Ying shing & Lin, 1995).

4.1.3 Cycloconverters

Variable Speed Constant Frequency (VSCF) Cycloconverters as shown in Figure 2 convert directly the variable frequency AC input power into AC form with fixed frequency and amplitude, three-phase 115V/400Hz (AbdElhafez & Forsyth, 2009; Cloyd, 1997; Cronin, 1990; Emad & Ehasni, 2000; Howse, 2003, Jones, 1999; Moir & Seabridge, 2001). The output frequency is lower than the input frequency; thus, making it possible for the generator to be attached to the engine with a fixed turns ratio gearbox. In the typical form of cycloconverters, three bidirectional switches interface each generator phase with the corresponding supply phase.

The VSCF cycloconverters are more efficient than CF and VSCF DC-link; however they require sophisticated control. The power generation efficiency of the cycloconverters increases as the power factor decrease, which would be beneficial if this technique is applied to motor loads with significant lagging power factors (AbdElhafez & Forsyth, 2009).

4.1.4 Wild frequency

Variable Frequency (VF), commonly known as wild frequency, is the most recent electric power generation contender. In VF approach, the generator is attached directly to the engine shaft. This method is commonly termed embedded generation (Raimondi et al., 2002). Generator direct allocation in the engine shafts de-rates power take-off shaft and the associated gearbox, which reduce their size and weight and increase the reliability. However, a number of implications will arise, in case of embedding one or more electrical machines within the core of the engine:

1. Accommodation of the embedded generators requires revision of the design of the engine components from their current state, which may change the components structure and probably the profile of the airflow through the engine.
2. The heat loss within the generator places a significant burden on the engine oil cooling system, requiring additional or alternative heat exchange.
3. If the generator rotor is only supported through main engine bearings, the small air gap requirement of the generator may lead to obligatory stiffening of the engine structure. The latter being nessary to ensure that rotor and stator do not come into contact under high acceleration
4. Transmitting high levels of electrical power to and from the core of the engine would require significant alterations in the supporting engine core structure relative to the engine pylon (Raimondi et al., 2002).

In VF, variations in engine speed would manifest directly into the output frequency as shown from (1) and Figure 2. The promising features of VF are the small size, weight, volume, and cost as compared with other aircraft electrical power generation options. Also VF offers a very cost-effective source of power for the galley loads, which consumes a lot of on-board power. However VF may pose significant risk at higher power levels, particularly with high power motor loads. Furthermore, the cost of motor controllers required due to the variation in the supply frequency, need to be taken into consideration when assessing VF (AbdElhafez & Forsyth, 2009; Cronin, 2005; Elbuluk & Kankam, 1997; Hoffman, 1999; ; Moir, 1998, Pearson, 1998; Weimer, 1993).

4.2 Generator topologies

The anticipated increase in electrical power generation requirements on MEA suggests that high power generators should be attached directly to the engine, mounted on the engine shaft and used for the engine start in Integral Starter/Generator (IS/G) scheme . The harsh operating conditions and the high ambient temperatures push most materials close to or even beyond their limits, requiring more innovations in materials, processes and thermal management systems design.

Consequently, Induction, Switched Reluctance, Synchronous and Permanent Magnet machine types (Hoffman et al., 1985; Mollov et al., 2000; Cross, 2002) have been considered for application in MEA due to their robust features.

4.2.1 Induction generator

Induction Generators (IGs) are characterized by their robustness, reduced cost and ability to withstand harsh environment. However, the IG requires complex power electronics and is considered unlikely to have the power density of the other machines (Khatounian et al.,2003; Ying & Lin, 1995; Bansal et. al, 2003, 2005).

4.2.2 Synchronous generator

The current generator technology employed on most commercial and military aircraft is the three-stage wound field synchronous generator (Hoffman, 1985). This machine is reliable and inherently safe; as the field excitation can be removed, de-energising the machine.

Therefore, the rating of the three-stage synchronous generator has increased over the years reaching to 150KVA (Hoffman, 1985) on the Airbus A380. The synchronous machine has the ability to absorb/generate reactive power, which enhances the stability of the aircraft power system. However, this machine requires external DC excitation, which unfortunately decreases the reliability and the efficiency.

4.2.3 Switched reluctance generator

The Switched Reluctance (SR) machine has a very simple robust structure, and can operate over a wide speed range. The three-phase type has a salient rotor similar to salient pole synchronous machine. The stator consists of three phases; each phase is interfaced with the DC supply through two pairs of anti-parallel switch-diode combination. Thus, the SR machine is inherently fault-tolerant. However the machine has the severe disadvantage of producing high acoustic noise and torque ripples (Mitcham & Cullenm, 2002, 2005; Pollock & Chi-Yao, 1997; Trainer & Cullen, 2005; Skvarenina et al., 1996,1997).

4.2.4 Permanent generator

The Permanent Magnet (PM) generator has a number of favourable characteristics (AbdElhafez & Forsyth, 2009; Argile, 2008; Bianchi, 2003; Jack et al., 1996; Pollock & Chi-Yao, 1997; Mecrow et al., 1996; Mitcham & Cullenm, 2002, 2005):

1. Ease of cooling, as the PM generator theoretically has almost zero rotor losses.
2. High efficiency compared to other machine types.
3. High volumetric and gravimetric power density.
4. High pole number with reduced length of stator end windings.
5. Self excitation at all times.

However, conventional PM machines are claimed to have inferior fault tolerance compared with SR machines (Argile, 2008; Mecrow et al., 1996; White, 1996). Conventional PM generators are intolerant to elevated temperatures. Furthermore, PM generators require power converters with high VA rating to cater for a wide speed range of operation (AbdElhafez & Forsyth, 2009; Bianchi, 2003 ;Jack et al., 1996;Mecrow et al., 1996; Mitcham & Cullenm, 2002, 2005). Therefore, a different implementation is mandatory in PM machine technology if they are to be used in aero-engines.

The fault-tolerant PM machines are one solution and offer high levels of redundancy and fault tolerance (Argile, 2008; Ho et al.,1988; Mitcham & Grum, 1998; Mellor, at al.,2005). These machines are designed with a high number of phases, such that the machine can continue to deliver a satisfactory level of torque/power after a fault in one or more phases. Furthermore, each phase has minimal electrical, magnetic, and thermal impact upon the others (Argile, 2008; Jack et al., 1996; Jones & Drager, 1997;Mecrow et al., 1996; Mitcham & Cullenm, 2002, 2005; White, 1996). This is realised by:

1. The number of magnetic poles in the machine being similar to the stator slot number; each phase winding can be placed in a single slot, which is thermally isolated from the other phases (AbdElhafez, 2008; Adefajo, 2008; Jones & Drager, 1997;Mecrow et al., 1996; Mitcham & Cullenm, 2002).
2. The stator coils being wound around alternate teeth, which provides physical and magnetic isolation between the phases (AbdElhafez, 2008; Jones & Drager, 1997).

3. Each phase being attached to a separate single-phase power converter, which achieves the electrical isolation (AbdElhafez, 2008; Adefajo, 2008; Jack et al., 1996; Jones & Drager, 1997;Mecrow et al., 1996; Mitcham & Cullenm, 2002, 2005).
4. The machine synchronous reactance per phase is typically 1.0 p.u., limiting the short-circuit fault current to no greater than the rated phase current (AbdElhafez, 2008; Jack et al., 1996; Jones & Drager, 1997;Mecrow et al., 1996; Mitcham & Cullenm, 2002, 2005).

4.3 Integrated generation

MEA as mentioned, suggests innovative strategies for optimizing the aircraft performance and reducing the installation and operational costs, such as IS/G and emergency power generation schemes.

4.3.1 Integral starter/generator

Commonly, jet engines are externally started by pneumatic power from a ground cart. This reduces the system reliability and increases maintenance and running cost. A move toward internal starting for the engine is adopted in MEA.

The jet engine has two shafts: High Pressure (HP) and Low Pressure (LP) shafts. The main generator is usually attached to the HP shaft . The trend is to use that generator as the prime mover to start the engine. Once the engine is started, the generator returns to its default operation, generator. The prime mover (starter) is powered from the aircraft system, which during this stage is supplied from energy storage devices. ISG scheme has a number of advantages (AbdElhafez & Forsyth, 2009; Ganev, 2006; Elbuluk & Kankam, 1997; Ferreira, 1995; Skvarenina, 1996, 1997) :

1. Improves the aircraft reconfigureability by eliminating the arrangement used previously for ground starting.
2. Allows the adoption of All Electric Aircraft (AEA)
3. Uses AMB system that results in reliable robust and compact engine.
4. Reduces the operational and maintenance cost, which boosts the air traffic industry

Different machine topologies are suggested for IS/G scheme; however the SR and fault-tolerant PM machines are most reliable. These machines do not require external excitation or sophisticated control techniques. Also, they are either inherently or artificially fault-tolerant.

4.3.2 Emergency power generation

The level of the emergency power is expected to grow significantly for future aircrafts, due to rising demands of critical aircraft loads/services. Currently, the emergency power is sourced from generators coupled to a Ram Air Turbine (RAT). This scheme is deployed only under emergency conditions, and suffers from serious drawbacks such as (AbdElhafez et al., 2006a, 2006b, 2008; Adefajo, 2008; Bianchi, 2003) :

1. It is expensive to develop, install and maintain.
2. It is unpopular with the airliners.
3. The integrity of such a ‘one-shot’ system is always subject to some doubt.

The proposal is to utilize the windmill effect of the aero-engine fan, which is driven from the LP shaft, for emergency power generation. While, the fan is normally rotating, the heath of the emergency generation system is continuously monitored and backup power will be immediately available following a main generator failure. Also the stored inertial energy of the engine is significant and could be recovered as another source of emergency power (AbdElhafez & Forsyth, 2008, 2009; Ganev, 2006).

Different machine topologies are competing for LP emergency generators. Trade-off studies were conducted to identify the most suitable machine technology. Due to the difficulty of the location, reliability is paramount and it is clear that a brushless machine format is required. The harsh operating environment particularly extremely high ambient temperatures, pushes many common materials, e.g. permanent magnet materials and insulation materials close to or beyond their operating limits. Consequenclty, cooling or alternative materials and process would be required (AbdElhafez & Forsyth, 2008, 2009; Mitcham &. Grum, 1998) .

Machine efficiency is another crucial issue, since dissipated heat needs to be absorbed by the engine cooling system. Currently, the generator loss is absorbed by the engine oil system and this is in turn mainly cooled by the fuel entering the engine. This restricts the amount of heat that can be dissipated without introducing an alternative cooling method.

Some key requirements, assisting in the choice of LP generator type are list below (AbdElhafez & Forsyth, 2008, 2009):

1. The machine operates only as a generator, drive torque is not allowed.
2. The machine is subject to a harsh operating environmental conditions (specifically high temperature), with limited access for maintenance.
3. Power must be generated over a very wide speed range (approximately 12:1) with an output voltage compatible with the aircraft DC-distribution system voltage 350 V dc.
4. The machine is fault tolerant, such that it continues to run even if there is a fault on one or two phases without significantly degrading the output power.

Also the operating speed range, weight and volume constraints are important parameters that affect the choice of machine type.

Several brushless machine types seem to have the required ruggedness and hence the capability of operation in such environment. These include: IG, SR and PM machines (AbdElhafez & Forsyth, 2008, 2009; Mitcham &. Grum, 1998).

5. Interfacing circuits

There are many occasions within the aircraft industry where it is required to convert the electrical power from one level/form to another level/form, resulting in a wide range of Power Electronics Circuits (PECs) such as AC/DC, DC/DC, DC/AC and matrix converters (AbdElhafez & Forsyth, 2009; Chivite-Zabalza, 2004; Cutts, 2002; Lawless & Clark, 1997; Matheson, &. Karimi, 2002; Moir & Seabridge, 2001; Singh et. al, 2008). There are general requirements, which PEC should satisfy:

1. PEC should have reduced weight and volumetric dimension.
2. PEC should be fault-tolerant, which implies its ability to continue functioning under abnormal conditions without much loss in its output power or degradation of its performance.

3. PEC should be efficient and have the ability for operation in harsh conditions such as high temperature and low maintenance.
4. PEC should emit minimum levels of harmonic and Electromagnetic Interference (EMC).
5. PEC could be easily upgraded and computerized.

Innovation in the area of power electronics components is required to enable realisation of MEA. Wide-Band Gap (WBG) High-Temperature Electronics (THE) is an example of these developments. The devices manufactured from WBG-THE are capable of operating at both higher temperatures (600 ^{0}C) (Reinhardt & Marciniak, 1996) and higher efficiencies compared to Si-based devices (-55 ^{0}C to 125 ^{0}C). A number of advantages are expected to be realized from employing WBG-THE devices (AbdElhafez et al., 2006, 2008, Howse, 2003; Gong et al., 2003; Lawless & Clark, 1997; Matheson, &. Karimi, 2002; Moir & Seabridge, 1998, 2001; Trainer & Cullen, 2005):

1. Eliminating/reducing of ECS required for cooling flight control electronics and other critical PECs
2. Reducing the engine control system weight and volumetric dimension
3. Improving the system reliability by using a distributed processing architecture
4. Optimizing the aircraft system and reducing the installation and running cost
5. Improving system fault-tolerance and redundancy

Another main challenge for PECs in the aircraft is passive electrical component size, as the current components are heavy and bulky, especially for the high power level expected in the MEA. However, the on-going research in the design and fabrication of the passive components for MEA gives some optimistic results. For example, some advanced polymer insulation materials such as Eymyd, L-30N, and Upilex S (AbdElhafez & Forsyth, 2009; Cutts, 2002; Lawless & Clark, 1997; Moir & Seabridge, 2001) have the ability to operate over a wide temperature range (-269 ^{0}C to 300 ^{0}C). Also these materials can withstand the environmental conditions such as humidity, ultraviolet radiation, basic solution and solvent at high altitudes (AbdElhafez & Forsyth, 2009; Lawless & Clark, 1997). The ceramic capacitor is a good example, which offers remarkable advantages in volumetric density compared to other capacitor technology (Lawless & Clark, 1997).

	SSCB	Conventional
Mechanism	The breaker consists of bidirectional switches that allow current flow in both directions. The gating signal of the switches are blocked to inhabit the faulty current	Commonly an isolating air gap is developed in the path of the fault current. A upon disconnection, an arc is created. Depending on the arc distinguishing methodology the breaker is termed.
Response time	Very small	Long
Power rating	Small	Medium to high
Volumetric/weight	Compact/small	Bulky/heavy
Cost	Expensive	Cheap
Functionality	Multi-task, they perform current monitoring and status reporting	They should be instructed to be opened

Table 2. Comparison between SSCB and conventional breakers.

6. Protection system

The distribution system of aircraft is adequatly protected; different types of Circuit Breakers (CBs) are utilized. Thus includes the conventional and power electronics based. The conventional CBs include air, SF6, and oil, while the Solid-State Circuit Breakers (SSCBs) represent the power electronics based breakers (AbdElhafez & Forsyth, 2009; Jones, 1999; Moir & Seabridge, 2001). A comparison between SSCB and a generic conventional CB is given in Table 2 above.

7. References

AbdElhafez, A. (2008). *Active Rectifier Control for Multi-Phase Fault-Tolerant Generators*. PhD desertion, University of Manchester, UK.

AbdEl-Hafez, A.; Cross, A.; Forsyth, A.; Mitcham, A.; Trainer, D.& Cullen, J. (2006). Fault Tolerant Starter-Generator Converter Optimisation ",Patent Application Rolls-Royce, Ed. UK, 2006.

AbdEl-hafez, A.; Cross, A.; Forsyth, A.;Trainer, D.& Cullen, J. (2006). Single-Phase Active Rectifier Selection for Fault Tolerant Machine, *in 3rd IET International Conference on Power Electronics, Machines and Drives, PEMD* 2006, pp. 435-439, April 2006.

AbdElHafez, A & Forsyth, A. J. (2009) . A Review of More-Electric Aircraft, *Proceedings of The 13rd international conference on Aerospace Science and Aviation Technology conference* ,ASAT-13. Cairo, Egypt, May 26-28, 2009.

AbdEl-Hafez, A.; Todd, R.; Forsyth, A.& Long, S. (2008). Single-Phase Controller Design for a Fault Tolerant Permanent Magnet Generator, in *IEEE Vehicle Power and Propulsion Conference, VPPC 2008,* pp. 250-257, September 2008.

Adefajo, O.; Barnes, M.; Smith, A.; Long, S.; Trainer, D.; AbdEl-hafez, A.; & Forsyth, A. (2008). Voltage Control On An Uninhabited Autonomous Vehicle Electrical Distribution System," in *The 4th IET International Conference on Power Electronics, Machines and Drives, PEMD 2008*, pp. 676-680, April 2-4, 2008.

Andrade, L. & Tenning, C.(1992). Design of the Boeing 777 Electric System, *IEEE National Aerospace and Electronics Conference*,.pp.1281 - 1290, May 18-22, 1992.

Andrade, L. & Tenning, C.(1992). Design of Boeing 777 electric system, *IEEE Aerospace and Electronic Systems Magazine*, Vol. 7, (1992) pp. 4-11.

Argile, R.; Mecrow, B.; Atkinson, D.; Jack, A.& Sangha, P. (2008). reliability analysis of fault tolerant drive topologies, *the Proceeding of The 4th IET International Conference on Power Electronics, Machines and Drives,PEMD 2008,* pp 11-15, April 2-4, 2008.

Bansal, R.; Bhatti, T.& Kothari, D.(2003). Bibliography on the application of induction generators in nonconventional energy systems," *IEEE Transaction on Energy Conversion,* Vol. 18, (September 2003), pp. 433-439.

Bansal, R. (2005). Three-phase self-excited induction generators: an overview, *IEEE Transaction on Energy Conversion,* Vol. 20, (20005), pp. 292-299.

Bianchi, N.; Bolognani, S.; Zigliotto, M. & Zordan, M. (2003). Innovative remedial strategies for inverter faults in IPM synchronous motor drives, *IEEE Transaction on Energy Conversion,,* Vol. 18, (June 2003), pp. 306-314.

Brock, A. & Schmidt, T. (1999). Electrical environments in aerospace applications, *in Proceedings of International Conference Electric Machines and Drives, IEMD '99,* pp. 719-721, May 9-12, 1999.

Chivite-Zabalza, F.; Forsyth, A. & Trainer, D. (2004). Analysis and practical evaluation of an 18-pulse rectifier for aerospace applications, *in Second International Conference on Power*

Electronics, Machines and Drives, PEMD 2004, Vol.1, pp. 338-343, March-31 April-2, 2004.

Cloyd, J. (1997). A status of the United States Air Force's More Electric Aircraft initiative," in *Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference, IECEC-97,* Vol.1, pp. 681-686, July-27 August-1 1997.

Cross, M.; Forsyth, A & Mason, G. (2002) . Modelling and simulation strategies for the electric system of large passenger aircraft," *in SAE 2002 conference*, pp. 450-459, 2002.

Cutts, S. (2002). A collaborative approach to the More Electric Aircraft, *in Proceedings* of *International Conference on Power Electronics, Machines and Drives,PEMD 2002*, pp. 223-228, April 16-18, 2002.

Cossar, C.& Sawata, T. (2004). Microprocessor controlled DC power supply for the generator control unit of a future aircraft generator with a wide operating speed range, in *Second International Conference on Power Electronics, Machines and Drives, PEMD 2004*, Vol.2, pp. 458-463, March 31, April-2, 2004.

Cronin, M. (1990). The all-electric aircraft, *IEE Review,* Vol. 36, (September 1990), pp. 309-311.

Elbuluk, M.& Kankam, P. (1997). Potential starter/generator technologies for future aerospace applications, *IEEE Aerospace and Electronic Systems Magazine,* Vol. 12, (May 1997),pp. 24-31.

Emadi, K &. Ehsani, M. (2000). Aircraft power systems: technology, state of the art, and future trends, *IEEE Aerospace and Electronic Systems Magazine,* Vol. 15, (January 2000), pp. 28-32.

Ferreira, C.; Jones, S.; Heglund, W.& Jones, W. (1995). Detailed design of a 30-kW switched reluctance starter/generator system for a gas turbine engine application, *IEEE Transactions on Industry Applications,* Vol. 31, (May/June 1995). pp. 553-561.

Ganev, E. (2006). High-Reactance Permanent Magnet Machine for High-Performance Power Generation Systems" *SAE Power Systems Conference*, pp. 247-253, November, 2006.

Glennon, T. (1998). Fault tolerant generating and distribution system architecture," in *IEE Colloquium on All Electric Aircraft,* (June 1998), pp. 1-4.

Gong, G.; Drofenik, U.& Kolar, J. (2003). 12-pulse rectifier for more electric aircraft applications, *in IEEE International Conference on Industrial Technology, V*ol.2, pp. 1096-1101, December 10-12, 2003.

Hoffman, A; Hansen, A. Beach, R.; Plencner, R.; Dengler, R.; Jefferies, K. & Frye, R. (1985) Advanced secondary power system for transport aircraft, *NASA Technical Paper* 2463, (May 1985) http://ntrs.nasa.gov/archive/nasa/19850020632_1985020632.pdf

Hoffman, A.; Hansen, I.; Beach, R.; Plencner, R.;Dengler, R.; Jefferies, K. & Frye, J. (1985) "Advanced secondary power system for transport aircraft," in *IEE Colloquium on All Electric Aircraft,* (June 1995), pp. 1-4.

Ho, T.; Bayles, R. & Sieger, E. (1988). Aircraft VSCF generator expert system," *IEEE Aerospace and Electronic Systems Magazine,* Vol. 3, (April 1988), pp. 6-13.

Howse, M. (2003). All electric aircraft, *Power Engineer Journal,* Vol. 17, (2003) pp. 35-37.

Jack, A.; Mecrow, B. and Haylock, J. (1996). A comparative study of permanent magnet and switched reluctance motors for high-performance fault-tolerant applications, *IEEE Transactions on Industry Applications,* Vol. 32, (July/August 1996), pp. 889-895.

Jones, R. (1999). The More Electric Aircraft: the past and the future?, *in IEE Colloquium on Electrical Machines and Systems for the More Electric Aircraft,*(November 1999), pp. 1-4.

Jones, S. & Drager, B. (1997). Sensorless switched reluctance starter/generator performance,*IEEE Industry Applications Magazine*, Vol. 3, (1997), pp. 33-38.

Khatounian, F.; Monmasson, E.; Berthereau, F.; Delaleau, E.& Louis, J. (2003). Control of a doubly fed induction generator for aircraft application, *in Proceedings of 29th Annual*

Conference of the IEEE Industrial Electronics Society, IECON '03, Vol.3, (November 2003), pp. 2711-2716.

Lawless, W.& Clark, C. (1997). Energy storage at 77 K in multilayer ceramic capacitors, *IEEE Aerospace and Electronic Systems Magazine, ,* Vol. 12, (August 1997), pp. 32-35.

Maldonado, M. & Korba, G. (1999). Power management and distribution system for a more-electric aircraft (MADMEL)," I*EEE Aerospace and Electronic Systems Magazine,* Vol. 14, (1999), pp. 3-8.

Maldonado, M.; Shah, N.; Cleek, K.; Walia, P. & Korba, G. (1996). Power management and distribution system for a more-electric aircraft (MADMEL)-program status," i*n Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference, IECEC-97*. Vol. 1, pp. 148-153, 1996.

Maldonado, M.; Shah, N.; Cleek, K.; Walia, P. & Korba, G. (1997). "Power Management and Distribution System for a More-Electric Aircraft (MADMEL)-program status," i*n Proceedings of the 33nd Intersociety Energy Conversion Engineering Conference, IECEC-97*. Vol. 1, pp. 274-279. 1997.

Matheson, E.; Karimi, K. (2002). Power Quality Specification Development for More Electric Airplane Architectures, in *SAE International Conference,* Vol. 2, pp. 343-347, 2002.

Mecrow, B.; Jack, A.; Haylock, J. & Coles, J. (1996). Fault-tolerant permanent magnet machine drives, *IEE Electric Power Applications,* Vol. 143, (November 1996), pp. 437-442.

Mellor, P.; Burrow, S.; Sawata, T.& Holme, M. (2005). A wide-speed-range hybrid variable-reluctance/permanent-magnet generator for future embedded aircraft generation systems, *IEEE Transactions on Industry Applications,* Vol. 41, (Marc-April 2005), pp. 551-556.

Mitcham, A. ; Antonopoulos , G. & Cullen, J. (2002). Favourable slot and pole number combinations for fault-tolerant PM machines, *in Proceedings of IEE Electric Power Applications,* Vol. 151, (September 2004), pp. 520-525.

Mitcham, A. & Grum, N. (1998). An integrated LP shaft generator for the more electric aircraft. *in IEE Colloquium on All Electric Aircraft*, (June 1998), pp. 1-9.

Mitcham, A. & Cullen, J. (2005). Permanent Magnet Modular Machines: New design Philosophy, in *Electrical Drive Systems for the More Electric Aircraft one-Day Seminar*, pp. 1-8, 2005.

Mitcham, A. & Cullen, J. (2002). Permanent magnet generator options for the More Electric Aircraft, *in Proceeding of International Conference on Power Electronics, Machines and Drives, PEMD* 2002, pp. 241-245, April 16-18, 2002.

Moir, I. (1999). .More-electric aircraft-system considerations, *in IEE Colloquium on Electrical Machines and Systems for the More Electric Aircraft*, (1999), pp. 1-9.

Moir, I. & Seabridge, A. (2001). *Aircraft systems : mechanical, electrical, and avionics subsystems integration,* London press, London, UK

Moir, I. (1998). .The all-electric aircraft-major challenges," *in IEE Colloquium on All Electric Aircraft*, (June 1998), pp. 1-6.

Mollov, S.; Forsyth, A. & Bailey, M. (2000). System modelling of advanced electric power distribution architecture for large aircraft," *SAE Transaction* (2000), pp. 904-913.

Olaiya, M.& Buchan, N. (1999) "High power variable frequency generator for large civil aircraft," in *IEE Colloquium on Electrical Machines and Systems for the More Electric Aircraft, (*November 1999), pp. 1-4.

Pearson, W. (1998). The more electric/all electric aircraft-a military fast jet perspective, *in IEE Colloquium on All Electric Aircraft* (June 1998), pp. 1-7.

Ponton, A. & at al (1998). "Rolls-Royce Market Outlook 1998-2017," *Rolls-Royce Publication No TS22388* (1998).

Pollock C. & Chi-Yao, W. (1997). Acoustic noise cancellation techniques for switched reluctance drives," *IEEE Transactions on Industry Applications,* Vol. 33, (March/April 1997), pp. 477-484.

Provost, M. (2002). The More Electric Aero-engine: a general overview from an engine manufacturer, *in Proceedings of International Conference on Power Electronics, Machines and Drives, PEMD 2002*, pp. 246-251, April 16-18 , 2002.

Quigley, R. (1993). .More Electric Aircraft, *in Proceedings of Eighth Annual Applied Power Electronics Conference and Exposition, APEC '93*, pp. 906-911, March 1993.

Raimondi, C.; Sawata, T.; Holme, M.; Barton, A.; White, G.; Coles, J.; Mellor, P. & Sidell, N (2002). Aircraft embedded generation systems, in *Proceeding of International Conference on Power Electronics, Machines and Drives,* PEMD 2002, pp. 217-222, April 16-18, 2002.

Reinhardt, K.& Marciniak, M. (1996). Wide-band gap power electronics for the More Electric Aircraft, *in Proceedings of the 31st Intersociety Energy Conversion Engineering Conference, IECEC 96.,* pp. 127 – 132, August 11-16, 1996.

Richter, E. & Ferreira, C. (1995). Performance evaluation of a 250 kW switched reluctance starter generator, *in Thirtieth IAS Annual Meeting IEEE Industry Applications Conference, IAS '95.,* Vol.1, pp. 434-440, October 8-12, 1995.

Rosero, J; Ortega, J.; Aldabas, E. & Romeral, L. (2007) Moving towards a more electric aircraft, *IEEE Aerospace and Electronic Systems Magazine*, Vol. 22, (2007), pp. 3-9.

Shing, Y.& Lin, C. (1995). A prototype induction generator VSCF system for aircraft, *in International IEEE/IAS Conference on Industrial Automation and Control: Emerging Technologies,* pp. 148-155, May 22-27, 1995.

Singh, B.; Gairola, S.; Singh, N.; Chandra, A.& Al-Haddad, K. (2008). Multiples AC-DC Converters for Improving Power Quality: A Review, *IEEE Transactions on Power Electronics,* Vol. 23, (January 2008.), pp. 260-281.

Skvarenina, T.; Pekarek, S.; Wasynczuk, O.; Krause, P.; Thibodeaux, R. & and Weimer, J. (1997), Simulation of a switched reluctance, More Electric Aircraft power system using a graphical user interface, *in Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference, IECEC-97,* 1997, Vol.1, pp. 580-584, July-27 August - 1, 1997.

Skvarenina, T.; Wasynczuk, O.; Krause, P.; Zon, W.; Thibodeaux, R.& Weimer, J. (1996). Simulation and analysis of a switched reluctance generator/More Electric Aircraft power system, *in Proceedings of the 31st Intersociety Energy Conversion Engineering Conference, IECEC 96.,* Vol.1, pp. 143-147, August 11-16, 1996.

Trainer, D. & Cullen, J. (2005). Active Rectifier for Fault Tolerant Machine Application, Derby, internal memorandum, February 24, 2005.

White, R. & Miles, M. (1996). Principles of fault tolerance, *the Proceeding of Eighth Annual Applied Power Electronics Conference and Exposition, APEC '96*, Vol.1, pp. 18-25, 1996.

Weimer, J. (1993). Electrical power technology for the more electric aircraft, *in Proceedings of 12th AIAA/IEEE Digital Avionics Systems Conference, DASC 1993*, pp. 445-450, October 25-28, 1993.

Welchko, B.; Lipo, T.; Jahns, T. & Schulz, S.(2004). Fault tolerant three-phase AC motor drive topologies: a comparison of features, cost, and limitations, *IEEE Transactions on Power Electronics, V*ol. 19, (July 2004), pp. 1108-1116.

Worth, F.; Forker, V.; Cronin, M. (1990). Advanced Electrical System (AES)," *in Aerospace and Electronics Conference,* pp. 400 - 403, 1990.

14

Power Electronics Application for More Electric Aircraft

Mohamad Hussien Taha
Hariri Canadian University
Lebanon

1. Introduction

In the competitive world of airline economics, where low cost carriers are driving dawn profit margins on airline seat miles, techniques for reducing the direct operating costs of aircraft are in great demand. In effort to meet this demand, the aircraft manufacturing industry is placing greater emphasis on the use of technology, which can influence maintenance costs and fuel usage. (Faleiro, 2005)

There is a general move in the aerospace industry to increase the amount of electrically powered equipments on future aircraft. This trend is referred to as the "More Electric Aircraft". It assumes using electrical energy instead of hydraulic, pneumatic and mechanical means to power virtually all aircraft subsystem including flight control actuation, environmental control system and utility function. The concept offers advantages of reduced overall aircraft weight, reduced need for ground support equipment and maintenance and increased reliability (Taha,2007,Wiemer,1999).

Many aircraft power systems are now operating with a variable frequency over a typical range of 360 Hz to 800 Hz.

Distribution voltages for an aircraft system can be classified as:

a. Nominal 115/200 V rms and 230/400 V rms ac, both one phase and three phase, over variable frequency range.
b. Nominal 14, 28 and 42 V DC.
c. High DC voltage which could be suitable for use with an electric actuator (or other) aircraft loads.

This chapter presents studies, analysis and simulation results for a boost and buck converters at variable input frequency using vector control scheme. The design poses significant challenges due to the supply frequency variation and requires many features such as:

1. The supply current to the converter must have a low harmonic contents to minimize its impact on the aircraft variable frequency electrical system.
2. A high input power factor must be achieved to minimize reactive power requirements.
3. Power density must be maximized for minimum size and weight.

2. Boost converter for aircraft application

A three phase boost converter which is shown in fig. 1 with six steps PWM provide DC output and sinusoidal input current with no low frequency harmonic. However the switching frequency harmonics contained in the input currents must be suppressed by the input filter. Referring to fig.1 after the output capacitor has charged up via the diodes to a voltage equals to 1.73V_{pk}, the diodes are all reverse biased. Turning one of the MOSFETs in each of the three phases will cause the inductor current to increase. Assume the input voltage V_a is positive, if S_2 is turned on, the inductor current increases through the diode D_4 or D_6 and the magnetic energy is stored in the inductor. Since the diodes D_1, D_3 and D_5 are reverse biased, the output capacitor C_{dc} provides the power to the load. When S_2 is turned off, the stored energy in the inductor and the AC source are transferred to C_{dc} and the load via the diodes. When the AC voltage is negative, S_1 is turned on and the inductor current increases through the diode D_3 or D_5. The same operation modes are involved for phase B and phase C (Taha., 2008; Habetler.,1993). Fig.2 and fig. 3 show different operating modes.

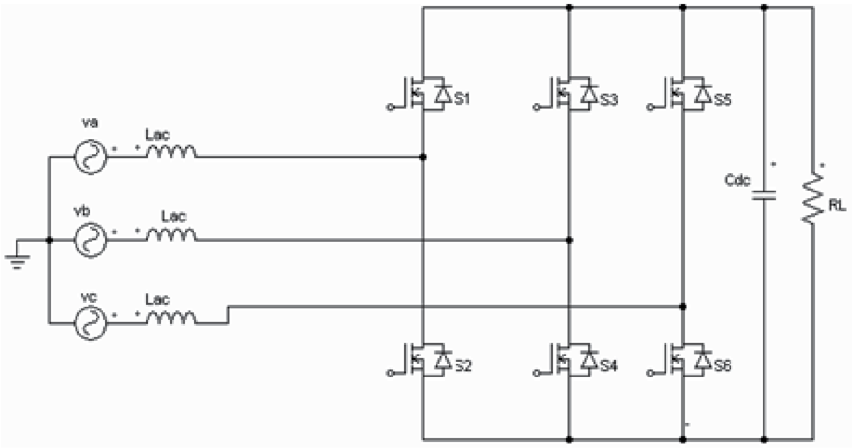


Fig. 1. Boost converter.

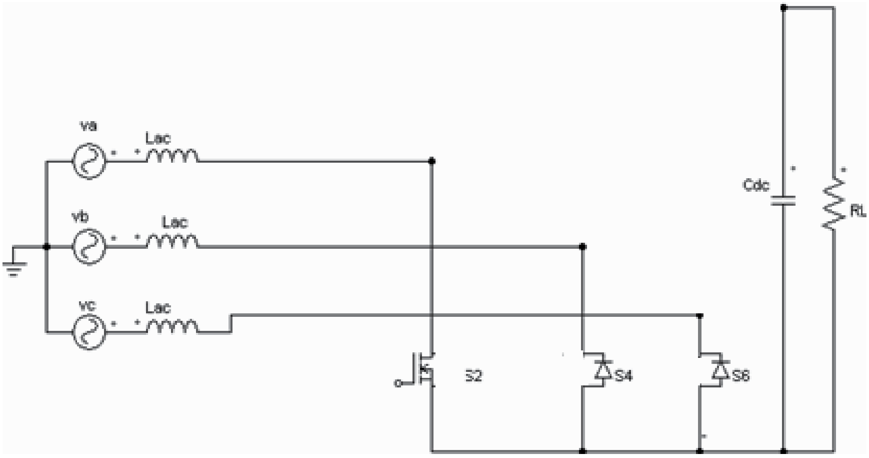


Fig. 2. Boost converter when V_a is positive.

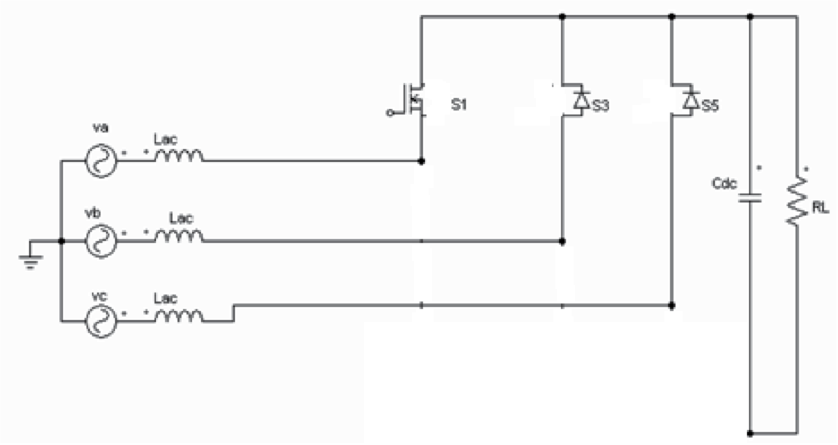


Fig. 3. Boost converter when V_a is negative.

3. Buck converter for aircraft application

The buck 3-phase/dc converter is a controlled current circuit which relies on pulse width modulation of a constant current to achieve low distortion. As shown in fig. 4. The circuit consists of 3 power MOSFETs and 12 diodes, an AC side filter and DC side filter.

The AC side input and DC side output filters are standard second order low pass L-C filters. For the input filter, the carrier frequency has to be considerably higher than the filter resonance frequency in order to avoid resonance effects and ensure carrier attenuation. The Ac side filter is arranged to bypass the commutating energy when the MOSFETs are turning off and to absorb the harmonic for the high frequency switching. At the DC side, the inductor is used to maintain a constant current, this inductor can be relatively small since the ripple frequency will be related to the switching frequency. The magnitude and the phase of the input current can be controlled and hence the power transfer that occurs between the AC and DC sides can also be controlled. (Green et al., 1997).

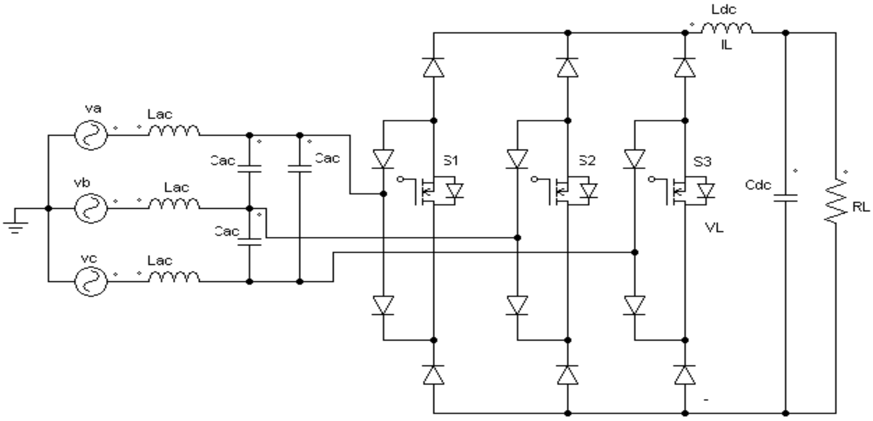


Fig. 4. Buck converter.

The input phase voltages V_a, V_b, V_c and the input currents I_a, I_b, I_c are assumed to be sinusoidal of equal magnitude and symmetrical

$$V_a = V_{pk} \sin(\omega t) \quad I_a = I_{pk} \sin(\omega t + \varphi) \tag{1}$$

$$V_b = V_{pk} \sin(\omega t - 2\pi/3) \quad I_b = I_{pk} \sin(\omega t - 2\pi/3 + \varphi) \tag{2}$$

$$V_c = V_{pk} \sin(\omega t + 2\pi/3) \quad I_c = I_{pk} \sin(\omega t + 2\pi/3 + \varphi) \tag{3}$$

Figure (5) shows 60 degrees of two sine waveforms.

$$T_a = TM \sin(\omega t + \varphi) \tag{4}$$

$$T_b = TM \sin(\omega t + 2\pi/3 + \varphi) \tag{5}$$

The freewheeling time T_f is equal to:

$$T_f = T - T_a - T_b \tag{6}$$

Where: φ is the displacement angle,
Vpk is the peak phase voltage,.
M modulation index.
T is the PWM switching period.

The general operation of the system is as follows: The switching of the devices is divided into six equal intervals of the 360 degrees main cycle. The waveforms repeat a similar pattern at each interval. At any time during the switching interval, only two converter legs are modulated independently and the third leg is always on. There are some time intervals that only one device is only on, thus providing a freewheeling for the DC current since at this time the energy stored on the DC inductor feeds the load.

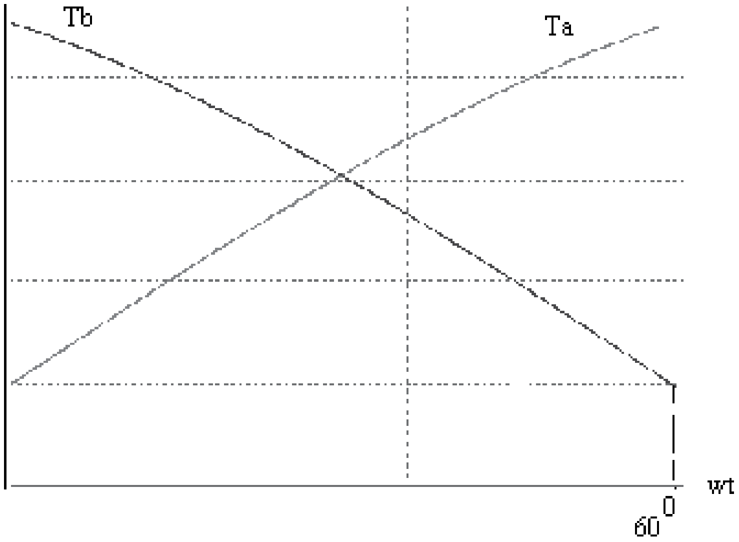


Fig. 5. Two 60° sine waves.

Mode 1

With S_2 on and S_1 is modulated by reference T_a, current flows in phase (a) and phase (b). $I_a>0$ and $I_b<0$, The bridge output voltage V_L is connected to main line supply V_{ab} which opposed by V_{DC} is applied across the inductor. Current I_L increases in the inductor.

$$I_L = I_a = -I_b \tag{7}$$

$$V_L = V_{ab} \tag{8}$$

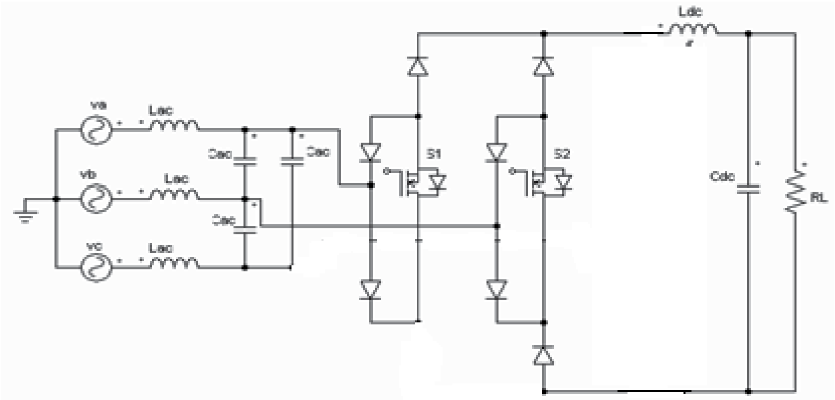


Fig. 6. Mode 1 equivalent circuit.

Mode 2

With S_2 on and S_3 is modulated by T_b, current flows in phase (c) and phase (b). $I_c>0$ and $I_b<0$. Line voltage V_{cb} opposed by V_{DC} is applied across the inductor. Again the inductor current increases.

$$I_L = I_c = -I_b \tag{9}$$

$$V_L = V_{cb} \tag{10}$$

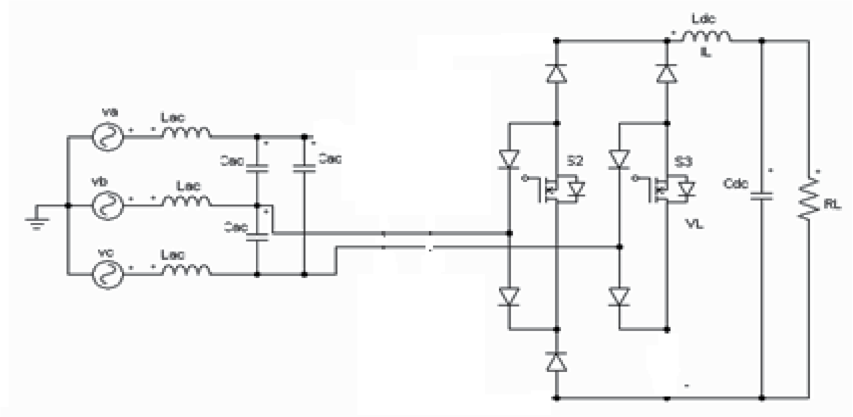


Fig. 7. Mode 2 equivalent circuit.

Mode 3

In this mode, only one MOSFET is on (S_2), the inductor current freewheels and the converter is disconnected from the mains and the DC voltage is zero.

$$V_L = 0 \tag{11}$$

Therefore the average voltage V_L over one switching period T is:

$$V_L = [(V_{ab} \text{ x } T_a) + (V_{cb} \text{ x } T_b)] / T \tag{12}$$

Where:

$$V_{ab} = V_a - V_b = 1.5V_{pk} \sin (\omega t) + 0.866 \, V_{pk} \cos (\omega t) \tag{13}$$

$$V_{cb} = V_c - V_b = 1.73 \, V_{pk} \cos (\omega t) \tag{14}$$

By substituting equations 8, 10, 12 and 13 into equation 12 yeilds:

$$V_L = V_{DC} = [(V_{ab} \text{ x } T_a) + (V_{cb} \text{ x } T_b)] / T = 1.5 \, M \, V_{pk} \cos\varphi \tag{15}$$

By assuming an ideal power converter in which the power losses are negligible, the power nput is then equal to power output, and by assuming $\cos\varphi = 1$, The DC output voltage can be defined as :

$$V_{DC} = 1.5 \, M \, V_{pk} \tag{16}$$

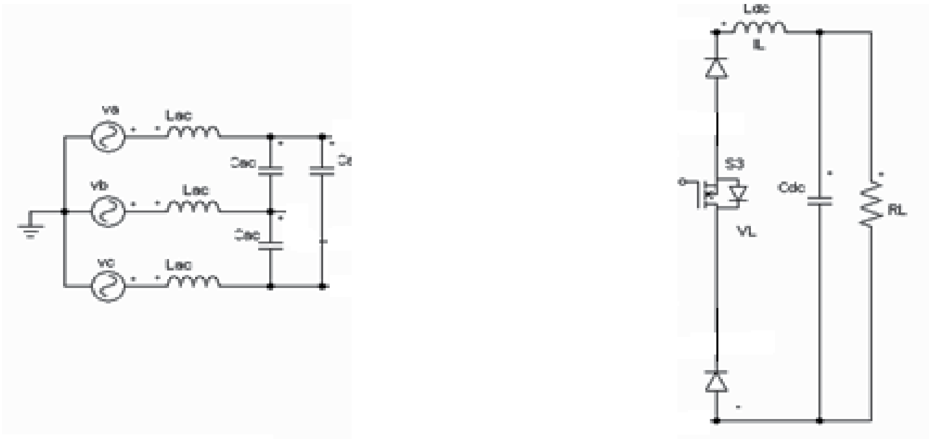


Fig. 8. Mode 3 equivalent circuit.

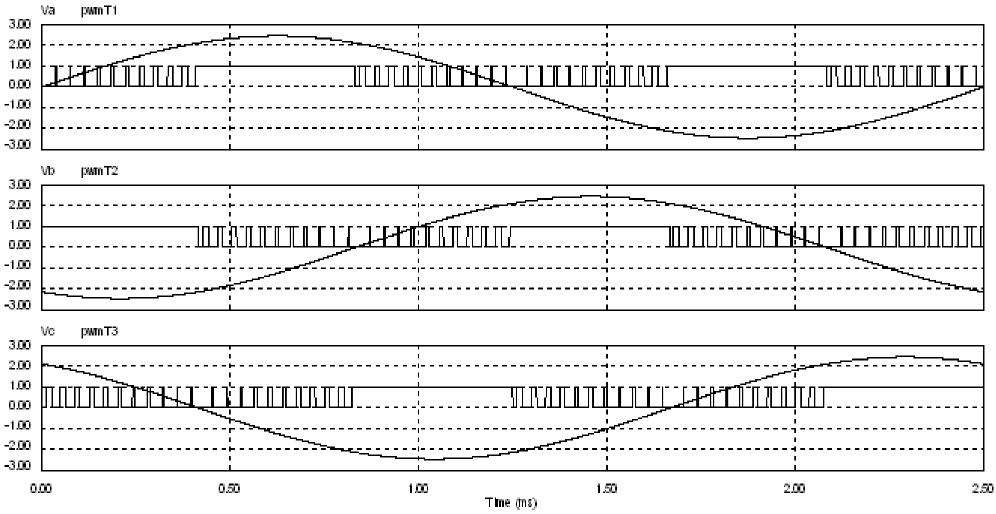


Fig. 9. PWM for the buck converter.

4. Adaptive reactive power control using boost converter

In comparison to the operating frequencies of land based power systems, which is normally 50-60 Hz, the operation of aircraft power systems at these relatively high frequencies can present some technical difficulties. One area of importance is associated with the impedance of the potentially long cables (which may run along part of the wing and a large proportion of the fuselage). These cables connect electrical loads, such as electric actuators for aircraft flight surfaces, to the AC supply, or "point of regulation" (POR). In large modern aircraft, the cables can be in excess of 200 ft and contribute impedance which is dependent on the cable's inductance and resistance. The inductive reactance, X_L, is proportional to the operating frequency of the power system and is given by $X_L = 2\pi fL$, where f is the operating frequency and L is the inductance of the cable and therefore the reactance changes with operating frequency. (Taha M, Trainer R D 2004). As the connected load draws a current, the cable develops a voltage drop due to its impedance which is out of phase with respect to the voltage at the POR and has two detrimental effects:

The voltage at the load is reduced below the regulated voltage at the point of regulation which is usually at the generator output.

The power factor of the load seen at the point of regulation reduces (even for a purely resistive load).

The voltage drop across the cable is clearly disadvantageous. The voltage drop may be tolerated and the connected loads have to be correspondingly down rated for the lower received voltage. Alternatively the voltage drop across the length of the cable is not allowed to exceed a threshold (typically 4 V) and it is necessary to provide cables that are both large and heavy such that their resistance remains low. Clearly space and weight are at a premium in aerospace applications. There can be significant weight saving if smaller, high resistance cables are used, particularly where low duty cycle, pulsed loads like electric actuators are supplied. The detrimental effects of such cables may be offset if the system designer uses the high inductive reactance present at the higher operating frequencies to affect voltage boost and power factor correction.

The simplest type of compensation for this type of problem is to connect a set of 3-phase capacitors (star or delta) at the point of connection of the load, in a similar way to ac motor-start capacitors. The capacitors can be used as a generator of reactive power but the beneficial effects are limited since the capacitive compensation is mainly controlled by the voltage magnitude and system frequency rather than the requirements of the load. Having noted the limitations of connecting shunt capacitors, there may be some applications where this type of compensation is applicable.

There is growing interest in the use of advanced power electronic circuits for aerospace loads, particularly in the motor-drives associated with electric actuators. The main two classes of converters currently being considered are active rectifiers and direct ac-ac frequency changer circuits (e.g. Matrix converters). Both types of converter can be made to operate with leading, lagging or unity power factor by suitable control of the semi-conductor switching elements.

The current view in the aerospace industry appears to be that the operation of these converters should be limited to unity power factor and little (or no) work has been carried out to explore the true system level benefits of variable power factor operation.

Fig.10 shows a basic circuit diagram for an electric actuator load incorporating an advanced power electronic converter with power factor control. It is clear that by controlling the power factor of the converter (shown leading), the effects of cable inductance can be eliminated so that the load as seen from the POR becomes unity power factor. Other operating power factors may be desirable in order to optimize the operation of the overall power system, including the generator loading.

Because the effects are proportional to the load current flowing through the cable and the system frequency, the reactive power compensation provided by the converter also needs to be variable.

The voltage magnitude at the load can be made the same as that at the POR. It could be beneficial in some applications to boost the input voltage by increasing the capacitive compensation provided by the power electronic converter.

The main benefit of using the advanced power electronic converter as a source of reactive power is to reduce (or eliminate) the voltage drop down the connecting cable. This gives us the possibility to use high impedance cables with benefits of reduced conductor diameter and significantly lower weight.

In order to understand the benefits of reactive power control, it is convenient to consider the flow of real and reactive current separately as shown in figure 10. Superposition can then be used to assess the net effect of both forms of current flow.

Therefore:

$$i = i_p + ji_q = I\,(\cos\theta_1 + j\sin\theta_1) \tag{17}$$

Where

$$\theta_1 = \tan^{-1}(i_q/i_p) \tag{18}$$

$$E = V + (i_p + ji_q)R + (i_p + ji_q)jX_L \tag{19}$$

$$E = V + i_p R - i_q X_L + j(i_q R + i_p X_L) \tag{20}$$

$$E = E\ ((\cos\theta_2 + j\sin\theta_2) \tag{21}$$

Where

$$\theta_2 = \tan^{-1}((i_q R + i_p X_L)/\ (V + i_p R - i_q X_L)) \tag{22}$$

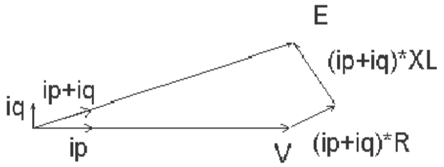
For unity power factor θ_1 should equals θ_2.
Therefore:

$$(i_q/i_p) = (i_q R + i_p X_L)/\ (V + i_p R - i_q X_L) \tag{23}$$

$$i_q = ((V/\ X_L) \pm ((V/\ X_L)^2 - 4\ i_p^2)^{1/2})/2 \tag{24}$$

In a practical system, ip could take the form of a current demand and iq would be a separate reactive current demand that is made to vary as a function of ip and X_L (frequency dependant). R and X_L are cable dependant parameters.

Referring to Fig. 10, the inputs here are system frequency and load current, the output is Q demand, which is an input to the power electronic converter. The parameters of the cable are stored and used within the electronic circuitry to calculate the required compensation for the system under consideration.



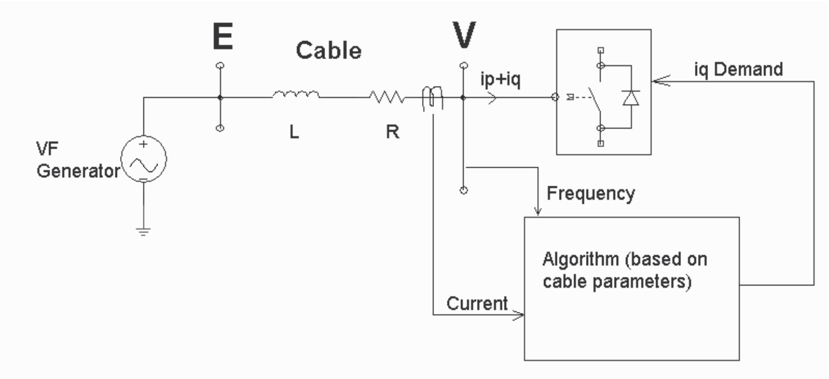


Fig. 10. System performance for reactive power compensation.

5. DQ vector control for the converters

In the DQ vector control strategy the instantaneous 3 phase voltages and currents are transferred to a 2-axis reference frame system which rotates at the angular frequency of the supply. This has the effect of transforming the three phase AC quantities (representing rotating volt and current phasors in the stationary co-ordinate frame) into DC quantities in the synchronously rotating frame (Taha et al.,2002, Taha,. 2008). If the D axis is chosen to be aligned with the voltage phasor, the D and Q axis current components represent the active and reactive components respectively. Fig. 11 shows the schematic of the DQ control scheme implemented in the input converter.

The proposed control scheme consists of two parts:

1. An outer voltage controller.
2. An inner current controller.

The outer voltage controller regulates the DC link voltage. The error signal is used as input for the PI voltage controller this provides a reference to the D current of the inner current controller. The Q current reference is set to zero to give unity power factor. A PI inner current control is used to determine the demand of the stationary DQ voltage values (Taha M & Trainer R D 2004; Kazmierkoski et al., 1991).

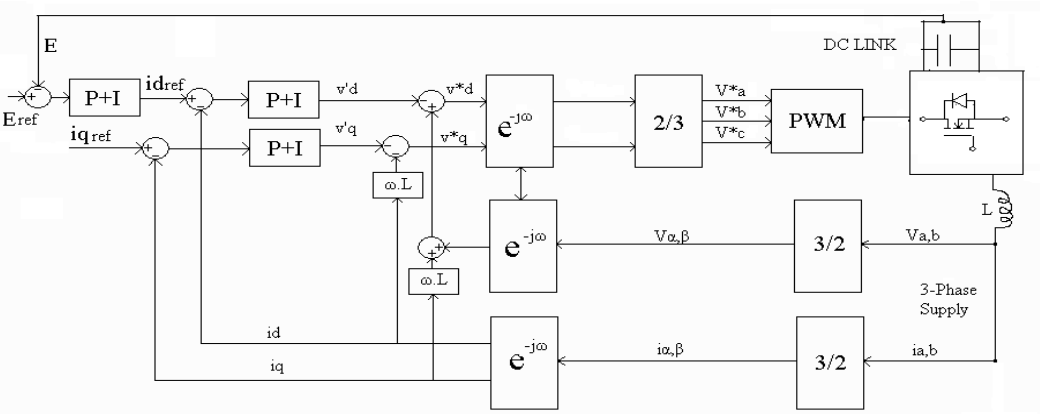


Fig. 11. DQ Control lock diagram.

Each gain in the controller affects the system characteristics differently. Settling time, steady state error and system stability are affected by the amount of the proportional gain. Selecting a large gain attains faster system response, but cost of large overshoot and longer settling time. Application of the integral feedback drives the steady state error to zero. The integral term increases as the sum of the steady state error increases causing the error to eventually be zero. However it can cause overshoot and ringing.

Selection of the two gain constants is critical in providing fast system response with good system characteristics.

The general formulas for DQ transformations are given as follows. We assume that the three-phase source voltages v_a, v_b and v_c are balanced and sinusoidal with an angular frequencyω.

The components of the input voltage phasor along the axes of a stationary orthogonal reference frame (α, β) are given by:

$$v_\alpha = v_a \tag{25}$$

$$v_\beta = \frac{1}{\sqrt{3}}\ (2\, v_b + v_a) \tag{26}$$

The input voltage can then be transformed to a rotating reference frame DQ chosen with the D axis aligned with the voltage phasor. The voltage components are given by:

$$v_d = v_\alpha \cos \omega t\text{-}\ v_\beta .\ \sin \omega t \tag{27}$$

$$v_q = v_\alpha \sin \omega t + v_\beta \cos \omega t \tag{28}$$

The same transformations are applied to the phase currents.

$$i_d = i_\alpha \cos \omega t - i_\beta .\ \sin \omega t \tag{29}$$

$$i_q = i_\alpha \sin \omega t + i_\beta \cos \omega t \tag{30}$$

Let v_{a1}, v_{b1} and v_{c1} be the fundamental voltages per phase at the input of the converter.

$$v_a = Ri_a + L\ di_a/dt + v_{a1} \tag{31}$$

$$v_b = Ri_b + L\ di_b/dt + v_{b1} \tag{32}$$

$$v_c = Ri_c + L\ di_c/dt + v_{c1} \tag{33}$$

where L is the value of input line inductance and R is its resistance of the inductor.

Taking the steady state DQ transformation for the inductor, the input voltage to the converter in the DQ reference frame is given by:

$$v_d = Ri_d + L.di_d/dt - \omega Li_q + v_{d1} \tag{34}$$

$$v_q = Ri_q + L.di_q/dt + \omega Li_d + v_{q1} \tag{35}$$

The active and reactive powers are given by:

$$P = v_d.i_d + v_q.i_q \tag{36}$$

$$Q = v_d.i_q - v_q.i_d \tag{37}$$

Inverse DQ transformations then need to be applied to provide the three phase modulating waves (v_{aref}, v_{bref} and v_{cref}) for the PWM generation.

The main advantages of the DQ control are :

1. Direct control the active and reactive power.
2. Fast dynamics of current control loops.

The PWM generator based on a regular asymmetric PWM strategy.

Voltage Control

The DC side may be modelled by a capacitor C, representing the smoothing capacitors, and a resistor R, representing the load. This is shown in Figure 12.

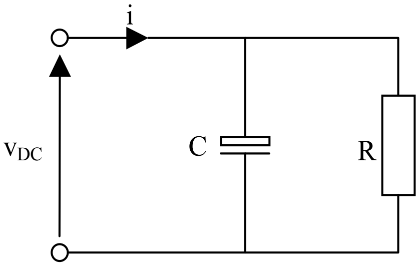


Fig. 12. Schematic of dc voltage link.

The linearised model for the DC side is given by the open-loop transfer function relating the DC link voltage to the supply current:

$$G(s) = \frac{v_{DC}(s)}{i(s)} = \frac{R}{1 + RCs} \tag{38}$$

Applying the PI controller illustrated, i(s) is given by:

$$i(s) = \left(K_p + \frac{K_i}{s}\right)\left(v_{REF} - v_{DC}\right) \tag{39}$$

Thus, the closed-loop transfer function is given by:

$$\frac{v_{DC}(s)}{v_{REF}(s)} = \frac{\left(K_p s + K_i\right)/C}{s^2 + \frac{s\left(1 + RK_p\right)}{RC} + \frac{K_i}{C}} \tag{40}$$

To give a damped response, the poles of the system should be placed along the real axis in the s-domain, i.e. at $s = -\omega_1$ and $s = -\omega_2$, giving the transfer function:

$$\frac{v_{DC}(s)}{v_{REF}(s)} = \frac{\left(K_p s + K_i\right)/C}{s^2 + s\left(\omega_1 + \omega_2\right) + \omega_1 \omega_2} \tag{41}$$

By equating the coefficients of the denominators of the above equations, the proportional and integral gains are:

$$K_p = C\left(\omega_1 + \omega_2\right) - 1/R \tag{42}$$

$$K_i = C\omega_1\omega_2 \tag{43}$$

The zero of the transfer function is where:

$$\left(K_p s + K_i\right)\big/C = \left(\left(C\left(\omega_1 + \omega_2\right) - 1/R\right)s + C\omega_1\omega_2\right)\big/C = 0 \tag{44}$$

Therefore:

$$s = \frac{-C\omega_1\omega_2}{C\left(\omega_1 + \omega_2\right) - 1/R} \tag{45}$$

An approach to the controller design is to locate this zero to coincide with one of the poles, say at $s = -\omega_1$, so as to cancel its effect. This gives:

$$\omega_1 = 1/RC \tag{46}$$

The second pole can then be placed at any desired location, to give the desired bandwidth. This gives the proportional and integral gains:

$$K_p = C\omega_2 \tag{47}$$

$$K_i = \omega_2/R \tag{48}$$

Current Control for boost converter

In this case the system is the line from the generator to the input converter, which may be modelled by an inductor in series with a resistor. The generator e.m.f. is assumed to have no dynamic effect, and so is represented as a short circuit. The system schematic is shown in Fig. 13.

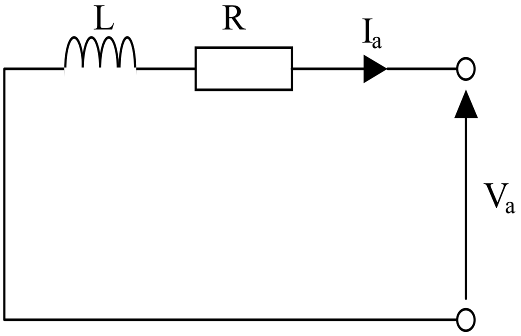


Fig. 13. Schematic of current control for the boost converter.

The phase current is given by:

$$i_a = -\frac{v_a}{R} - \frac{L}{R}\frac{di_a}{dt} \tag{49}$$

The open loop transfer function relating the phase current to the phase voltage is, therefore:

$$\frac{i_a}{v_a} = -\frac{1}{R + Ls} \tag{50}$$

From this simple transfer function, it would appear that the PI controller proposed would suffice, driving the steady-state current error to zero and allowing the behaviour and bandwidth, (i.e. the positions of the poles, of the closed loop system) to be fully determined by choosing the proportional and integral gains. The D-axis and Q-axis currents are compared to their respective demanded values and the error is applied to individual PI controllers to give voltage demands referred to the D-axis and Q-axis. With the feed-forward and dc-coupling terms, the transfer functions of the systems being controlled are:

$$\frac{i_d(s)}{v_d'(s)} = \frac{1}{Ls + R} \tag{51}$$

$$\frac{i_q(s)}{v_q'(s)} = \frac{1}{Ls + R} \tag{52}$$

Again, these are first-order equations and similar to the voltage control loop, the PI controllers will drive the steady-state error to zero and enable the behaviour and bandwidth of the closed-loop system to be determined by placing the poles appropriately.

Current Control for buck converter

The idea of controlling the current of the AC side LC filter has been proposed as a way of suppressing the excitation of the resonance of this filter. In steady state and in the absence of distortion there are no current components to excite the resonance because the resonant frequency will have been chosen to fall between the fundamental and the switching frequency. During the transient, the resonance of the filter can be damped by choosing the characteristics impedance to match the resistance and the inductance.

In this case the system is the line from the generator to the input converter, which may be modeled by an inductor in series with a resistor and capacitor.. The system schematic is shown in Figure 14.

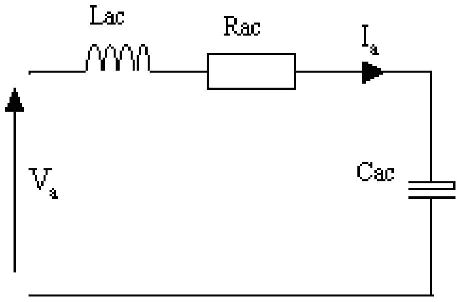


Fig. 14. Schematic of current control for the buck converter.

The open loop transfer function relating the phase current to the phase voltage is, therefore:

$$G(s) = \frac{1}{s^2\, L_{ac}C_{ac} + sR_{ac}C_{ac}\ + 1} \tag{53}$$

From this transfer function it would be appear that the PI controller proposed would suffice, driving the steady state error to zero and allowing the behaviour and bandwidth (position of the poles) of closed loop system to be fully determined by choosing the proportional and integral gains. The procedure for this is the same as that described above for voltage control.

6. Hardware design

All of the converters components had to be selected so that normal service maintenance would ensure the retention of their specified characteristics through the full range of operational and environmental conditions likely to be encountered through the life of the aircraft, or support facility, in which they are installed (Taha M 1999).

6.1 Capacitors

The choice of capacitors is very important for aerospace industry. Wet aluminum electrolytic capacitors are not suitable due to their limited operating temperature range and hence limited life. Equivalent series resistance is also a problem for these and other types of electrolytic capacitor and therefore alternative technologies, such as ceramic or plastic, are recommended.

Ceramic capacitors have a good lifetime, low series resistance and they work in high temperature conditions. On the other hand for a rating of a few hundred volts this type of capacitor has a very small value per unit volume and are only available in units of up to 20uF. The size and weight for this converter are very important. Therefore care was taken to choose the optimal value of the DC capacitor

6.2 Magnetic components

Another important factor is the design of the magnetic components. In order to achieve a small air gap, minimum winding turns, minimum eddy current losses and small inductor size, the inductor should be designed to operate at the maximum possible flux density. Also, care should be taken to ensure that the filter inductors do not reach a saturated state during the overload condition. As the cores saturate, the inductance falls and the THD rises.

7. Simulation results for boost and buck converter

The power conversion in the boost or buck converter is exclusively performed in switched mode. Operation in the switch mode ensures that the efficiency of the power conversion is high. The switching losses of the devices increase with the switching frequency and this should preferably be high in order have small THD therefore choosing the switching frequency poses significant challenges due to:

1. Supply frequency Variation (360 to 800 Hz).

For the boost converter the simulation carried out with a fixed switching frequency. However, for the buck converter. one of the method could be used is a variable switching frequency which depend on the input frequency. Trade off between the values of the filters and the switching frequency have been studies, in order to maintain the THD within the required value at different input frequency. Another method is to use the same switching frequency for different input frequency, here the highest input frequency should be considered.

The parameter values used for the simulation are shown in table 1. Fig. 15 to fig. 18 show, input AC voltage and current and Dc output voltage.

Boost converter	Buck converter
RMS phase voltage = 115 V	RMS phase voltage = 115 V
DC voltage setting = 400 V	DC voltage setting = 42 V
AC Input Filter = L_{ac}= 100uH	AC Input Filter = L_{ac}= 150uH; C_{ac} = 1μF
Dc Output filter C_{dc} = 50μF	Dc Output filter = L_{dc} = 1mH ; C_{dc} = 50μF
Load = 10 Ω	Load = 0.5 Ω
Switching freq for 800 input freq =20000 Hz	Switching freq for 800 input freq =33600 Hz
Switching freq for 360 input freq =20000 Hz	Switching freq for 360 input freq =23760 Hz

Table 1. Simulation parameters.

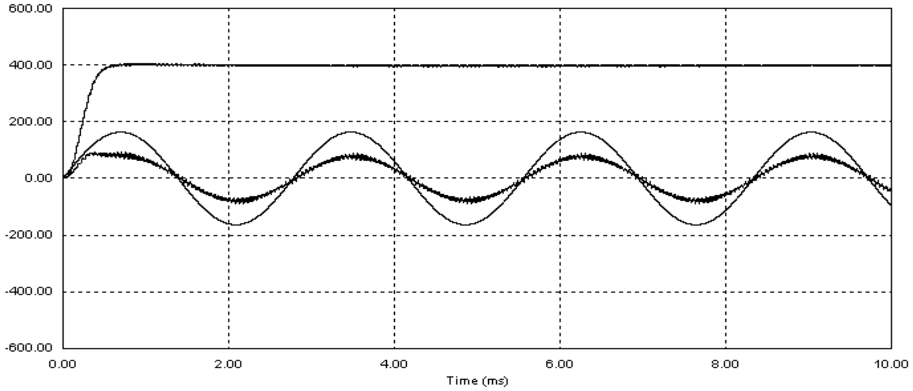


Fig. 15. Boost converter simulation results at 360 Hz input frequency.

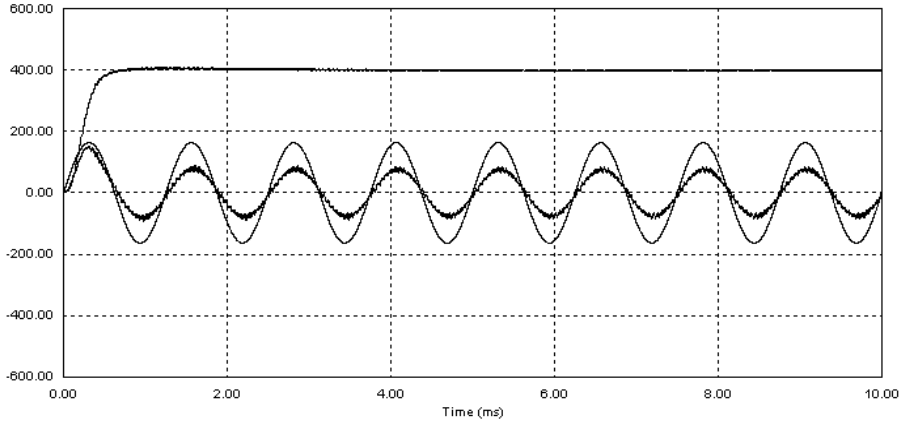


Fig. 16. Boost converter simulation results at 800 Hz input frequency.

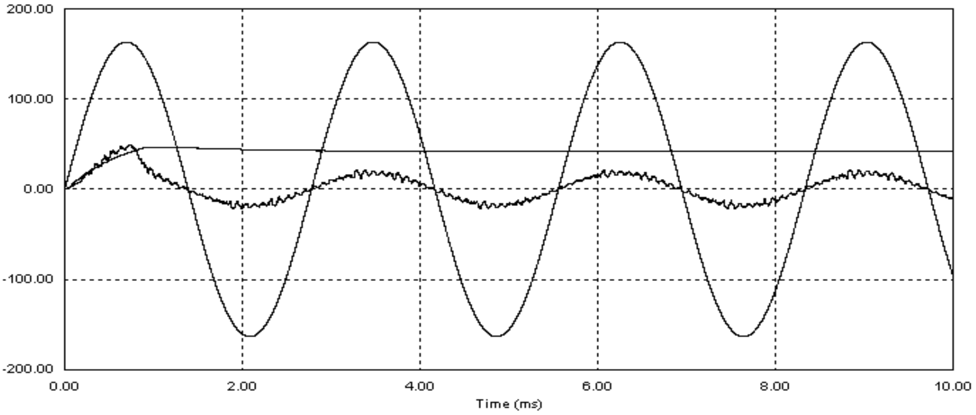


Fig. 17. Buck converter simulation results at 360 Hz input frequency.

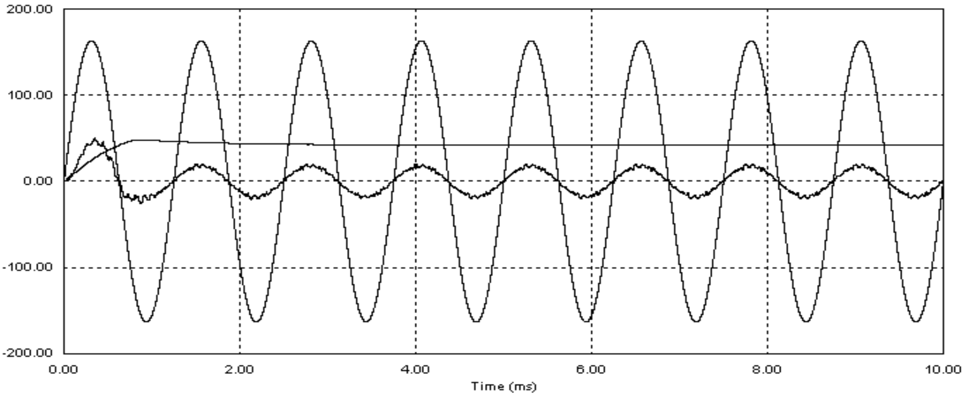
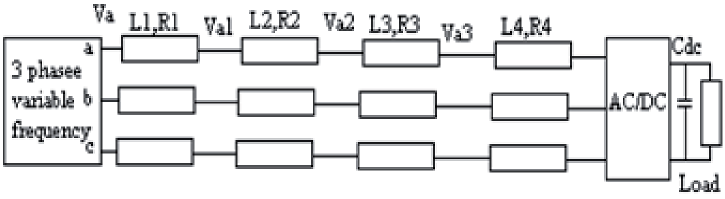


Fig. 18. Buck converter simulation results at 800 Hz input frequency.

8. Simulation results for the adaptive power control

Simulation has been done at 16 kW approximately. Fig. 19 shows the "per phase" parameter values used. Fig. 20 and fig. 21 show results for 360 Hz, the voltage drops to 107.5 V at the input filter of the converter. To compensate for the voltage drop across the cable, q (reactive demand) has been set this gave leading power factor. Fig. 21 shows that the voltage Va3 increase to 111.3 V at 0.9 PF.

Fig. 22 and Fig. 23 show results for 800 Hz, the voltage drops to 106 V at the input filter of the converter. Fig. 23 shows that the voltage Va3 increase to 1113 V at 0.9 PF.



V_{a1} is the point of regulation voltage
V_{a3} is the point of connection of the load
L_1 = 25uH,R_1= 0.015 ohms is the generator inductance and resistance.
L_2 = 10uH,R_2= 0.01 ohms is the cable inductance and resistance from the generator to contactor.
L_3 = 20uH,R_2= 0.1 ohms is the cable inductance and resistance from the contactor to the load.
L_4 = 100uH,R_4= 0.1 ohms is the inductance and resistance of the load converter input filter.

Fig. 19. Single phase parameters for the adaptive power control.

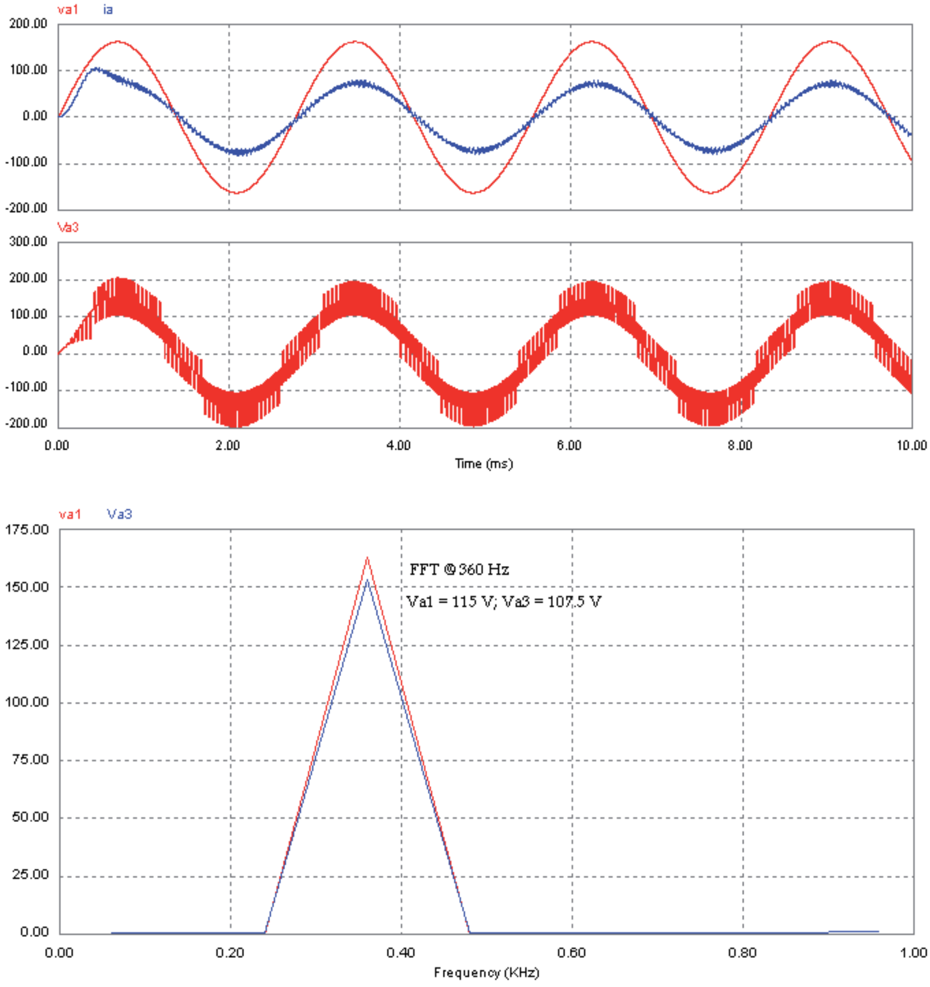


Fig. 20. Results at 360 Hz input frequency for unity power factor.

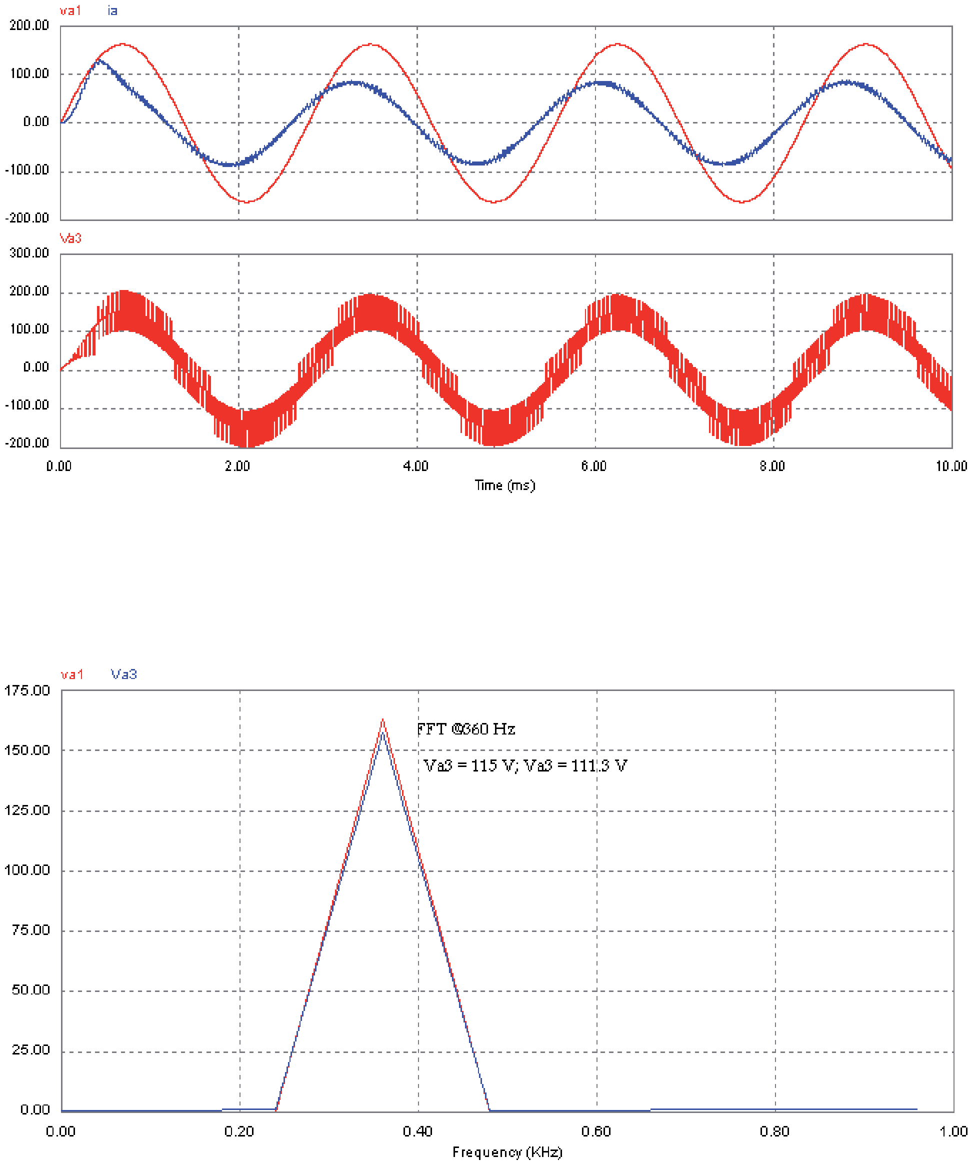


Fig. 21. Results at 360 Hz input frequency for 0.9 power factor.

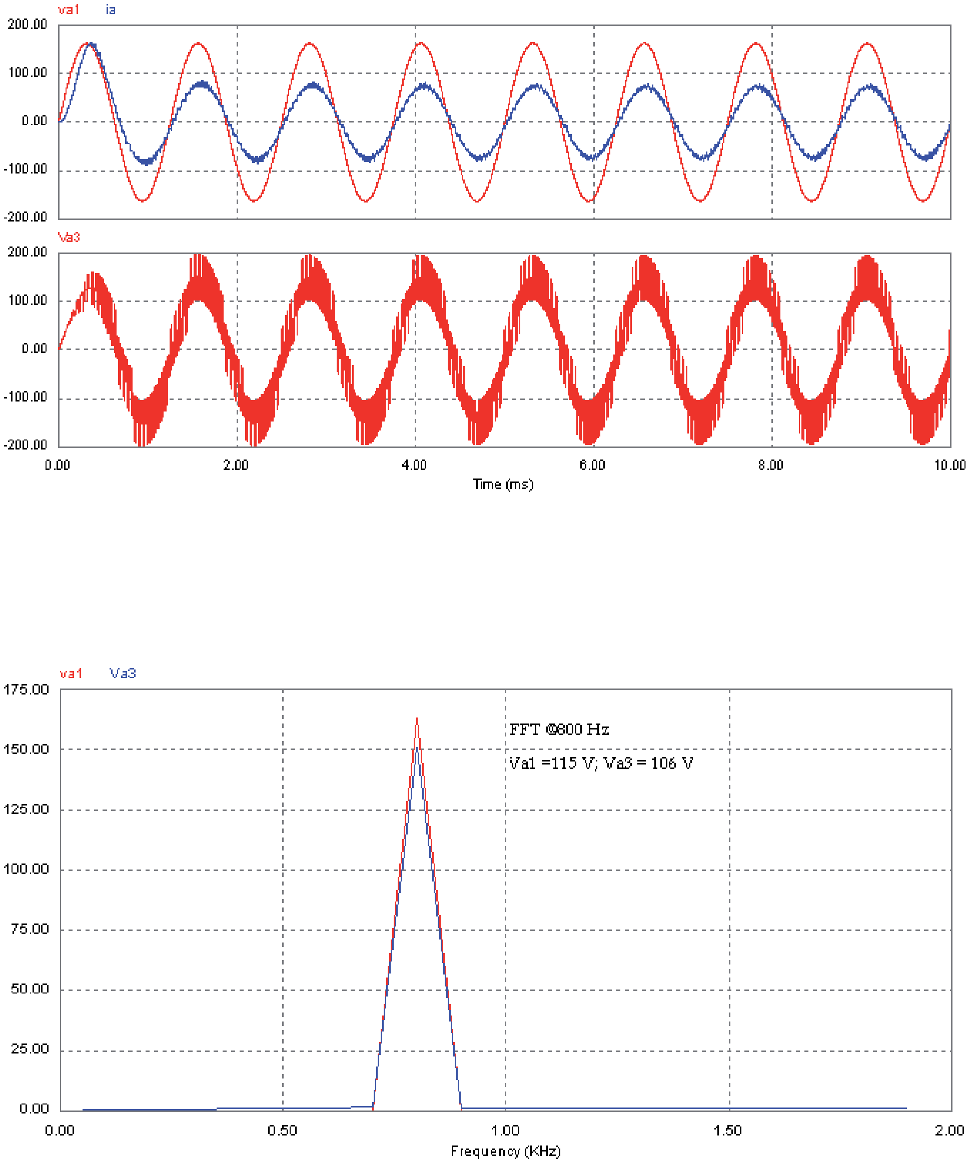


Fig. 22. Results at 800 Hz input frequency for unity power factor.

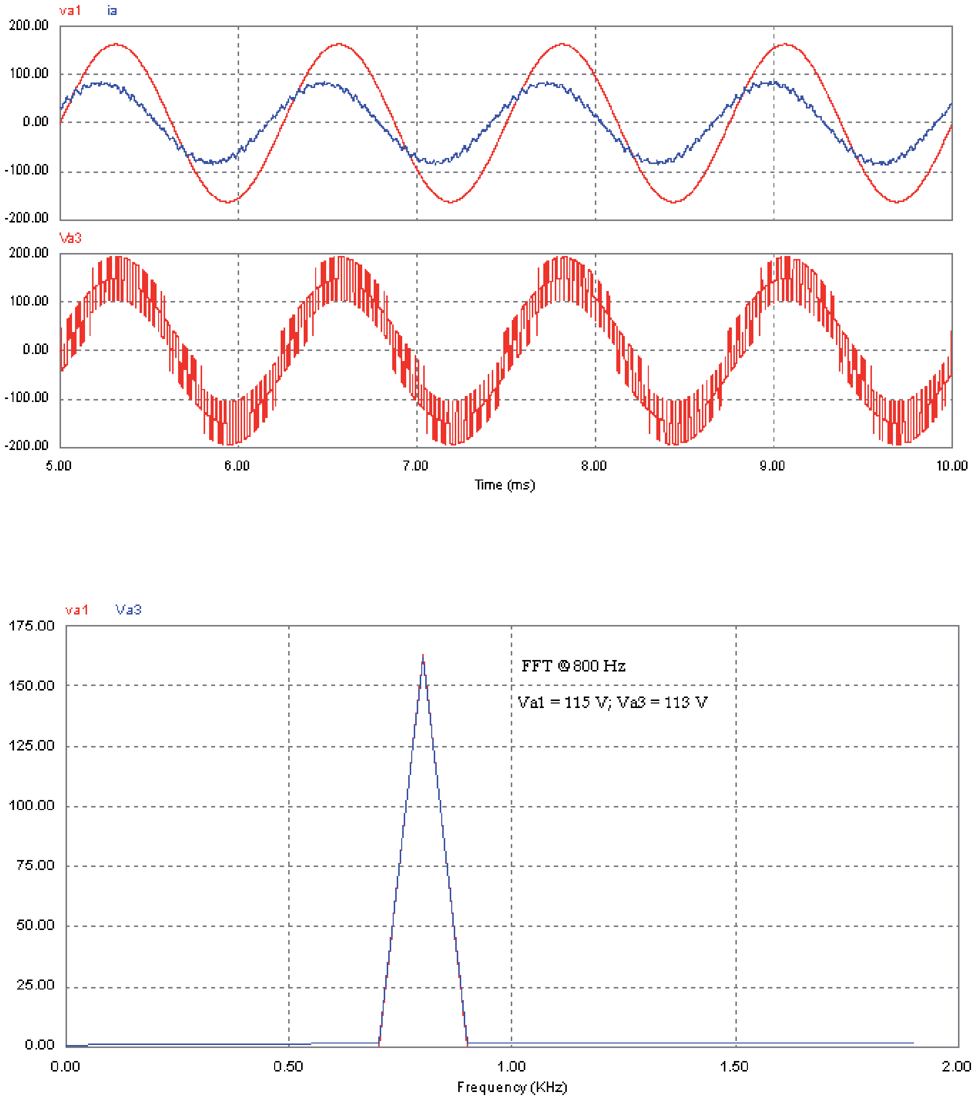


Fig. 23. Results at 800 Hz input frequency for 0.9 power factor.

9. Conclusion

On the basis of the space vector concept a PWM controller was developed. It has been shown that sinusoidal modulation generated in a space vector representation with PI controllers give an adequate performance in steady state and transient condition fast. It has been shown that with the future use of advanced power electronic converters within aircraft equipment, there is the possibility to operate these at variable input frequency and keeping the input current harmonics low.

An AC/DC buck and boost converters with different input frequency and offers low THD (less than 7%) has been described and simulated.

The operation and performance of the proposed topology was verified by simulating a 16 KW with a pure resistive load of 10 Ω and 400 dc voltage for the boost converter and a 3.5 KW with a pure resistive load of 0.5 Ω and 42 dc voltage for the buck converter. The input current is sinusoidal and power factor is unity. The DC voltage is well smoothed.

With the future use of advanced power electronic converters such as active rectifiers and matrix converters within aircraft equipment, there is the possibility to operate these at variable power factor in order to provide system level benefits. These include control of the voltage at the load and improvements in power factor seen at the POR.

10. Acknowledgment

I would like to express my sincere appreciation and respect to the late Prime Minister Rafik El hariri who is entirely responsible for funding my studies in England.

11. References

Faleiro, L. (2005). Beyond the More Electric Aircraft*, Aerospace America,* September 2005, pp 3540

Green, A.; Boys J. & Gates G. (1988). 3-phase voltage sources reversible rectifier, *IEE Proceeding* ,1988,135, pp 362-370, 2002

Green T.; Taha M.; Rahim N.; &Williams B.W. (1997). Three Phase Step-Down Reversible AC-DC Power converter*, IEEE Trans. Power Electron*, 1997,12, pp 319-324

Habetler T. (1993). A space Vector-Bases Rectifier Regulator for AC/DC/AC Converters. I*EEE Trans. Power Electron*, 1993 Vol 8, pp. 30-36

Kazmierkowski M.; Dzeiniakowski M.& Sulkowski W. (1991). Novel space vector based current control for PWM inverters, *IEEE Trans. Power Electron*, 1991, 6, pp 158-166

Taha M (1999). "Power electronics for aircraft application" *Power electronics for demanding applications colloquium, IEE,April 1999, 069 pp 5 -8*

Taha M.; Skinner D.; Gami S.; Holme M. & Raimondi G (2002); *Variable Frequency to constant frequency converter (VFCF) for aircraft application*, PEMD 2002.

Taha M. (2008). Mitigation of Supply Current Distortion in 3- Phase /DC Boost converters For Aircraft Applications*, PEMD 2008.*

Taha, M (2007). Active rectifier using DQ vector control for aircraft power system, *IEMDC 2007* pp 1306-1310

Taha M, Trainer R D (2004); Adaptive reactive power control for aircraft application Power Electronics, Machines and Drives, 2004. (PEMD 2004). Second International Conference on (Conf. Publ. No. 498) 2:469- 474 Vol.2.

Weimer J. (1995). Powe Managemennt and Distribution for More Electric Aircraft, *Proceeding of the 30th Intersociety Energy Conversion Engineering Conference*,pp.273-277

15

Key Factors in Designing In-Flight Entertainment Systems

Ahmed Akl[1,2,3], Thierry Gayraud[1,2] and Pascal Berthou[1,2]
[1]1CNRS-LAAS, *Université de Toulouse*
[2]*UPS, INSA, INP, ISAE; LAAS, F-31077 Toulouse;*
[3]*College of Engineering, Arab Academy for Science,*
Technology, and Maritime Transport, Cairo
[1,2]*France*
[3]*Egypt*

1. Introduction

Most of researches concerning *In-Flight Entertainment (IFE)* systems are done on case bases without a global view that encompasses all IFE components. Thus, we try to highlight the key factors of designing IFE system, and showing how its various components can integrate together to provide the required services for all parties involved with the system.

1.1 Background and historical issues

Flight entertainment started before the First World War by the Graf Zeppelin (see Figure 1). This aircraft had a long, thin body with a teardrop shape; it was about 776 feet long and 100 feet in diameter, filled with hydrogen, and the cabin was located under the hull; five engines were fixed to the hull to power the aircraft.

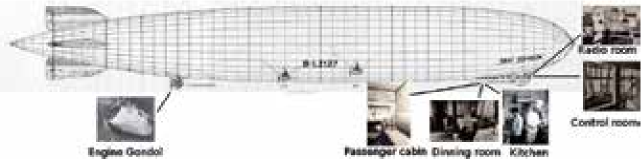


Fig. 1. The Graf Zeppelin aircraft

From the passengers comfort perspective, this model was equipped with a kitchen having electric ovens and a refrigeration unit, a small dinning room, washrooms for men and women, and passenger cabins with a capacity of two passengers each. Unfortunately, the craft was not heated, so passengers were dressing heavy coats and covered with blankets during winter flights. As developments went on, the "*Hindenburg*" aircraft came with heated passenger area, larger dinning room, passengers lounge with a piano as the first audio entertainment, a decorated writing room, a more enhanced passenger cabins, and promenades with seating and windows that can be opened during the flight (Airships.net, Last visit 2011).

In 1949, the "*De Havilland DH 106 Comet*" was the first commercial jet airliner to go into service. It had four jet engines located into the wings. It provided passengers with low-noise pressurized cabin (when compared to propeller-driven airliners), and large windows; hot and cold drinks, and food are serviced through the galley; separate women and men washrooms were available (Davies & Birtles, 1999).

Starting from 1960, *In-Flight Entertainment (IFE)* systems started to attract attention; they were basically a pre-selected audio track that may be accompanied with a film projector. They had shown improvements in both vertical and horizontal dimensions. They expanded horizontally by improving the existing services; audio entertainment moved from using simple audio devices to surround sound and live radio; video display progressed from using a film projector, to CRT displays hanged in the ceiling, to LCD displays dedicated to each passenger. The vertical improvement was noticed through introducing new technologies; cabin telephones allowed passengers to make phone calls during the flight; the system become interactive and allowed passengers to select their own services, while in the past they were forced to follow fixed services; web-based internet services allowed passengers to use some services such as emails and SMS messaging.

The basic idea behind IFE systems was to provide passengers with comfortableness during their long range flights; especially with long transatlantic flights where passengers see nothing but a large blue surface, so that services were initially based on delivering food and drinks to passengers. As passengers demand for more services grows, accompanied with an increase in airlines competition and technology advancement, more services were introduced and modern electronic devices played a remarkable role. This caused a change in the basic concept behind IFE systems; it becomes more than just giving physical comfortableness and providing food. It is extended to provide interactive services that allow passengers to participate as a part of the entertainment process as well as providing business oriented services through connectivity tools. Moreover, it can provide means of health monitoring and physiological comfort.

In recent years, market surveys have revealed a surprising and growing trend in the importance of *IFE* systems with regard to choice of airline. With modern long range aircraft the need for "stop-over" has been reduced, so the duration of flights has also been increased. Air flights, especially long distance, may expose passengers to discomfort and even stress. (Liu, 2007) mentioned that the enclosed environment of the aircraft can cause discomfort or even problems to passengers. This may include psychological and physical discomfort due to cabin pressure, humidity, and continuous engine noise. IFE systems can provide stress reduction entertainment services to the passenger which provides mental distraction to decrease the psychological stress. This can be done by using e-books, video/audio broadcasting, games, internet, and On Demand services. On the other hand, physical problems can range from stiffness and fatigue to the threat of *Deep Vein Thrombosis (DVT)* (Westelaken et al., 2010). IFE systems can provide different solutions such as video guided exercises to decrease fatigue, and seat sensors to monitor the passenger's health status

In fact, passengers from highly heterogeneous pools (i.e., age, gender, ethnicity, etc...) cause an impact on the adaptive interface systems. In non-interactive IFE systems, services (i.e., video and audio contents) are usually implemented based on previous concepts of what passengers may like or require. Using an interactive system based on context-aware services can make passengers more comfortable since they are able to get their own personalized entertainment services. However, such system must be user friendly in terms of easiness to use, and varieties of choice; otherwise, the passenger may get bored and is not able to get the expected satisfaction level.

From the airlines companies' perspective, productivity and profitability are one of the main targets. Achieving these targets is always hindered by the strong competition between companies. Thus, airlines are trying to maximize their attractiveness to get more clients because every empty seat means a revenue loss. IFE systems can play a remarkable role in customer satisfaction and attraction, and it can be used as an efficient portal for in-flight shopping. Moreover, one of the main tasks of aircraft attendants is to keep the passengers calm, unstressed, and to quickly respond to their requests. IFE systems can be a factor of stress elimination, decreasing passenger's movements during the flight, and providing request information quickly to the attendants.

Achieving such level of services requires various technologies and design concepts to be integrated together for implementing such systems. A single networking technology is not capable of providing all types of services. Thus, a good heterogeneous communication network is required to connect different devices and provide multiple services on both system and passenger's levels. For example, a GSM network can provide telephony services; WiFi, Bluetooth, and Infrared to keep passenger's devices connected to the system; LAN and/or *Power Line Communication (PLC)* to form the communication network backbone.

1.2 Chapter structure

Section 2 presents the different types of services provided by IFE systems, and shows the various components which are directly used by passengers as well as the components working at the background, which passengers are not aware of their existence. Section 3 introduces our proposed SysML model that integrates parts of the IFE system to help designers to have a global view of the whole system. Section 4 presents our conclusion. Finally, section 5 discusses future issues of IFE systems.

2. IFE services and components

IFE systems can provide various services for different parties such as airline companies, crew members, and basically passengers. These services are provided through software and hardware components; some components are used directly by passengers, while the others are used indirectly.

2.1 IFE services

IFE services can give solutions for different domains. They can provide health care and monitoring for passengers of health problems, business solutions to advertise products and support business decision making through surveys, and the expected service of entertainment.

2.1.1 Crew services

Although it seems that IFE systems are providing services to passengers only, but it can be extended to provide the cabin attendants with services to facilitate their job. Attendants have to keep a big smile and descent attitude during their work regardless of the current situation, and are burdened with various responsibilities and tasks. We believe that IFE systems can create a dynamic link between passengers and attendants. When an attendant respond to a passenger call, he does not know the reason for the call, so he has to make two moves, one to know the request and the second to fulfill it. An IFE system can allow the passenger to inform the attendant with their request (i.e., drinking water), so that the attendant can finish the service in one move instead of two. Moreover, the IFE system can ask the passenger if he had requested a special meal or not, so the attendant can bring the exact meal to the desired place without moving around with all meals in hand while asking passengers.

The cabin intercommunication service allows the pilot and cabin crew to make announcements to passengers, such as boarding, door closure, take off, turbulence, and landing announcements. These announcements are very important and need to be delivered to all passengers without any interruption; they are usually introduced via a loudspeaker installed in the cabin. If the passenger is wearing his headset, or is not able to understand the announcement language, then few numbers of passengers will comprehend the message. An IFE system can elevate the service through its audio system. When an announcement is introduced while the passenger is running an entertainment service, the entertainment pauses and he hears the announcement through the IFE audio system. Moreover, if it is a standard message such as *"Fasten your seat belt"*, it can be directly translated into the language currently used by the passenger.

Safety demonstrations are used to increase passenger safety awareness. The demonstrations are usually done by crew members. This means that an attendant will stop any current activity and dedicate himself to the demonstration. As an alternative, the IFE system can be used to provide *Aviation safety education for passengers* via multimedia services; insuring accurate instructions, situational awareness, emergency responses, and relevant cabin-safety regulations (Chang & Liao, 2009), so that the attendants can be freed to perform other tasks. Moreover, IFE systems can be used in pre-flight briefing for crew members to improve the quality and availability of information provided to flight crew (Bani-Salameh et al., 2010).

2.1.2 Entertainment services

They are the basic services introduced by IFE systems. They aim at providing multimedia contents for passenger entertainment, audio tracks for different types of music channels, special programs recorded for the airlines, games, and printed media

- ***Video on Demand:*** As mentioned by (Alamdari, 1999), IFE systems usually include screen-based, audio and communication systems. The screen-based products include video systems enabling passengers to watch movies, news and sports. These systems had progressed into *Video on Demand (VoD)*, allowing passengers to have control when they watch movies. The general VoD problem is to provide a library of movies where multiple clients can view movies according to their own needs in terms of when to start and stop a movie. This can be solved by using an *In-flight Management System* to store the pre-recorded contents on a central server, and streams a specific content to passengers privately.

  The service can be enhanced by using subtitles as a textual version of the running dialogue; it is usually displayed at the bottom of the screen with or without added information to help viewers who are deaf or having hearing difficulties, or people who have accent recognition problems to follow the dialogue. In addition, they can be written in a different language to help people who can not understand the spoken dialogue.

- ***Single and multiplayer games:*** Video games are another emerging facet of in-flight entertainment. Gaming systems can be networked to allow interactive playing by multiple passengers. Providing high quality gaming in an aircraft cabin environment presents significant engineering challenges. User expectation of video quality and game performance should be considered because many users had experienced sophisticated computer games with multiplayer capabilities, and high quality three dimensional video rendering. Network traffic characteristics associated with computer games should be studied to help in system design; (Kim et al., 2005) measured the traffic of a *Massively Multi-player On-line Role Playing Game (MMORPG)*, showing the differences in traffic between the server and client side. In a *Massively Multiuser Virtual Environment (MMVE)*, where large number of users can interact in real time, consistency management is required to realize a consistent world view for all users. (Itzel et al., 2010) present an approach that identifies users which actually interact with each other in the virtual world, groups them in consistency sessions and synchronizes them at runtime. On the other hand, there is a trend to use wireless networks in IFE systems; the feasibility of using wireless games is studied in different researches (Khan, 2010; Khan et al., 2010; Qi et al., 2009).
- ***E-documents:*** An in-flight magazine is a free magazine usually placed at the seat back by the airline company. Most airlines are distributing a paper version, and some of them are now distributing their magazines digitally via tablet computer applications. Furthermore, ebooks are widely available electronically with value-added features and search options not available in their print counterparts. Electronic versions are not limited to just text; they may present information in multiple media formats, for example, the text about a type of bird may be accompanied by video depicting the bird in flight and audio featuring its song. Using an electronic version of printed media can change their importance by adding interactive features such as e-commerce services where a passenger can choose his products and buy them instantaneously.

2.1.3 Information services

Air map display provides passengers with up to date information about their travel. They are aware of the plane location and at which part of earth it is passing over. Information telling the outside temperature, speed, altitude, elapsed time, and remaining time gives passengers the sense of movement, because it is difficult at high altitudes, where you can find nothing except blue sky, and sun or moon, to evaluate and sense the aircraft motion. Missing this feeling can be boring for many passengers.

Exterior-view cameras also enable passengers to have the pilot's forward view on take-off and landing on their personal TV screens. The cameras can have different locations. A tail-mounted camera is located in housing atop the vertical stabilizer of the aircraft; it provides a wide-angle view looking forward and typically shows most of the aircraft from above. A belly-mounted camera provides a view looking vertically down, or down at an angle that includes the horizon. A quad-cam belly installation offers a choice of four views covering 360 degrees.

Passengers can pass their time navigating through available entertainment contents to have information about their destination. This can include city maps, sightseeing, languages, and cultural information. Such information will allow passengers to pass a fruitful time and minimize the feeling of being a stranger in a foreign country.

2.1.4 E-business services

Airborne internet communications allows passengers and crew members to use their own WiFi enabled devices, such as laptops, smart phones and PDAs, to surf the Web, send and receive in-flight e-mail with attachments, Instant Message, and access their corporate VPN. Many companies are offering solutions to provide passengers with Internet connectivity. FlyNet (FlyNet, Last visit 2011) is an example for onboard communication service provided by Lufthansa to allow passengers to connect to the Internet during their flight. ROW44 (ROW44, 2011) provides a satellite-based connectivity system that allows airlines to offer uninterrupted broadband service

Mobile phones are one of the most demanded devices by passengers. Many passengers, especially businessmen, are welling to make calls through their personal mobile phone during their flight. However, there are doubts that cell phone signals may endanger aircraft safety by interfering with navigational systems. To overcome this situation, different techniques (i.e., (AeroMobile, Last visit 2011)) were introduced to the market, where an on-board pico cell can connect the mobile phones to the ground stations through the satellite link and managing the signal strength to insure that there is no interference with the navigational systems.

On-board conferencing can turn wasted flight time into productive time for traveling teams of salespersons. Also, it will reduce the effort done by passengers to trade seats after boarding to bring their group together. With the addition of a headset with *Active Noise Cancellation*, the experience can be extended to conversing with someone in the next seat, due to the reduction of ambient noise.

Personal Electronic Devices (PEDs) such as laptop computers (including WiFi and Bluetooth enabled devices), PDAs (without mobile phones), personal music (i.e., iPods), iPads, ebooks and electronic game devices are electronic devices that can be used when the aircraft seat belt sign is extinguished after take-off and turned off during landing. On the other hand, other PEDs using radio transmission such as walkie-talkies, two-way pagers, or global positioning systems are prohibited at all stages of flight, as it may interfere with the aircraft communication and navigation systems.

Power outlets are hardly reached by passenger during traveling to their destination. Spending too much time without a power source can cause PEDs to run out of power, and causing passengers to be frustrated. As a solution for such situation, airlines (AmericanAirlines, 2011; Qantas, 2011) add power outlets to passenger seats. These outlets are usually present in first and business class seats. For safety reasons, some outlets are designed to provide 110 Volt (60 Hz) with 75 watts, however, this may be unsuitable for PCs that consumes more power. Other companies provide 15 volt cigarette lighter outlet, which needs an adapter to connect devices.

2.1.5 E-commerce services

In-flight shopping is dragging more attention from airlines as it is considered as a source of revenue, and a way for passengers to utilize their flight time. (Liou, 2011) presented passenger attitude towards in-flight shopping. He mentioned that customer′s convenience increases when the shopping process takes less time, less effort in planning ahead, and less physical effort to obtain the product or service. Moreover, many factors can affect the decision making process (i.e., to buy a product); this includes pre-purchase information searching, and evaluation of alternatives.

An IFE system can be a remarkable factor for in-flight shopping. It can increase passenger convenience and facilitate decision making. An electronic catalogue viewed through the IFE display unit can provide search options that allow passengers to find other alternatives and make his own comparisons, and it can provide him with exhaustive information about the product. In turn, this will allow passenger to plan ahead without making two much physical effort, and in a relatively shorter time than making the same process in a paper document or through discussion with a crew member. Furthermore, the IFE system can play an extra ordinary role to e-commerce, not only for in-flight shopping, but also for shopping outside the flight. The IFE system can be connected to ground commercial services, so that the passenger can buy products or services (i.e., transport tickets, and duty free products), and receive them directly when he reaches his destination. In addition, multimedia advertising can attract companies to use it as a way to reach passengers.

Surveying is an important part of market evaluation. (Balcombe et al., 2009) held a survey to determine passenger's *Willingness To Pay (WTP)* for in-flight service and comfort level. The survey focused on seat comfort, meal provision, bar service, ticket price, entertainment (i.e., overhead screens for pre-set programs). He reported that older passengers are WTP more for seat comfort, while younger passengers are WTP more for bar and screen services.

However, performing such surveys is very tedious and difficult. (Aksoy et al., 2003) held a survey to evaluate Airline service marketing by domestic and foreign firms from customers viewpoint. The usable responses were 1014 out of 1350 responses, producing a 75.1% response rate. An IFE system can be an effective tool to increase the response rate, where an electronic version of the survey can guarantee that more passengers will participate, erroneous answers can be reduced, and analyzing the results becomes faster and accurate.

2.1.6 Health services

An elevated type of services, which IFE can provide, is health services. Flight conditions may cause the cabin environment to be tough; especially for persons who can face ill conditions. Flight duration, dehydration, pressure, engine noise, and other factors can be reasons of physical and/or psychological problems. A sensory system integrated in IFE system can provide a way to sense bad health conditions of passengers having health problem, and either inform the crew members or perform an action to reduce the effect.

(Schumm et al., 2010) and (Westelaken et al., 2010) suggest solutions based on sensory systems embedded in passenger's seat to sense his current status. (Schumm et al., 2010) introduce the design of smart seat containing sensors to measure *Electrocardiogram (ECG)*, *Electrodermal Activity (EDA)*, respiration, and skin temperature. These measured values can give a good indication about physical and psychological state. ECG is measured in two ways; without skin contact through sensors embedded in the backrest, and with a sensor fixed on the index finger. The second type is more obtrusive, but is more reliable. The same fixation system to the finger includes the EDA, and temperature sensors. The passenger movements can affect the reading quality, so a 3-axis accelerometer is added to compensate the errors. Respiration level is detected through sensors fixed in the seatbelt. The combined reading of these sensors can give a good indication about the passenger's health status.

Physical exercises can reduce physical stress and fatigue. However, the challenge is how to stimulate passengers to do them. (Westelaken et al., 2010) introduces a solution to reduce physical and psychological stress by detecting body movements and gestures to be used as an input for interactive applications in the IFE system. The basic idea for implementing these applications is to allow the passenger to participate in a gaming activity. His movements are captured as inputs for the chosen game. Three techniques were introduced to capture movements; sensors integrated in the floor, sensors integrated in the seat, and video-based gesture recognition. However, each of these techniques has its own pros and cons which need more investigation.

For passengers of special health needs, IFE system can be an effective tool to relief their pain. Passengers of *Spinal Cord Injury (SCI)* are not able to sense pressure acting on certain parts of their body that are cut off nervous system. This may increase the risk of decubitus ulcer, especially for long flights, where passengers may sit for several hours. (Tan, Chen, Verbunt, Bartneck & Rauterberg, 2009) proposed an *Adaptive Posture Advisory System (APAS)* for people of SCI. The passenger's seat plays a great role by having various sensors and actuators. Sensors are used as input source for a central processor connected to a database which is used to record passenger's sitting behavior and conditions. The suitable decision is taken and sent to the actuator to change the seat shape, and softness. This system helps SCI passengers to reposition their sitting posture to shift the points under pressure so that decubitus ulcer risk is minimized.

2.2 IFE components

The IFE components can be categorized into passenger and system components. In (Akl et al., 2011), we identified passenger components as the devices that the passenger uses directly to achieve a service, and system components as the components which are provided by the system and used indirectly by the passenger.

2.2.1 Passenger components

Passenger components are usually designed to be very simple and familiar in appearance and functionality in order to allow passengers of different background to use them; such as display units, remote controls, seat control buttons, headphones, etc...

2.2.1.1 Passenger seat

From the first sight, the passenger's seat may seem to be out of the scope of IFE systems, which are basically designed for entertainment. However, a deep look shows the contrary since passenger's seat is one of the main comfortableness components; especially when we consider that it is the place where the passenger spends most of his travel time. From one side, a poorly designed seat can causes discomfort, which can be extended to a musculoskeletal disorders regardless of the presence of any entertainment or stress reduction techniques; imagine the stay on such a seat for three or two hours, you will think in nothing except the time when the flight ends. Furthermore, when the passenger sits upright and inactive for a long period of time, he may be exposed to several health hazards. The central blood vessels in his legs can be compressed, making it harder for the blood to get back to his heart. Muscles can become tense, resulting in backaches and a feeling of excessive fatigue during, and even after the flight. The normal body mechanism for returning fluid to the heart can be inhibited and gravity can cause the fluid to collect in the feet, resulting in swollen feet after a long flight.

From another side, modern technologies can be used to elevate seat entertainment and comfortable role. Thus, we propose two terms, *Passive Seat (PS)*, and *Active Seat (AS)*. The *PS* is providing the service through its own structural design without any interaction with the passenger. The *AS* is providing the service in response to an intentional or unintentional input captured from the passenger.

- ***Passive Seat*****:** (Nadadur & Parkinson, 2009) discussed different seat design problems. Airlines are aiming at increasing the seats density inside the cabin to increase their revenue. However, such approach diminishes the comfortableness factors in seat design. Increasing the seats density negatively affects the seat pitch, causing a decrease in the passenger's leg room (see Figure 2), which is considered as an important factor especially for tall passengers. He also mentioned that passengers should minimize the pressure between their lower thighs and the surface of the seat to prevent the occurrence of *Deep Vein Thrombosis (DVT)*. This can be achieved by keeping the knees height greater than the seat's height. A design contradiction here arises because lower seat height requires more leg room causing seat pitch to increase, and consequently seats density will decrease. On the other hand, increasing the seat height increases discomfort and the probability to have DVT problems. To find a compromise between these contradictions he proposed a mathematical solution to embed the passenger comfort as a design parameter and link it with the passenger's willingness to pay higher prices. (Vink, 2011) introduced other factors to be considered during seat design such as wider seats, adjustable headrests, space under the armrest, backrest angle, and ideal distribution of pressure over body parts. A better pressure distribution can be achieved by using support under the front part of the legs to spread the load, and ergonomic design of seat back and seat pan. Also, a well designed headrest and neck rest can increase the comfort feeling.

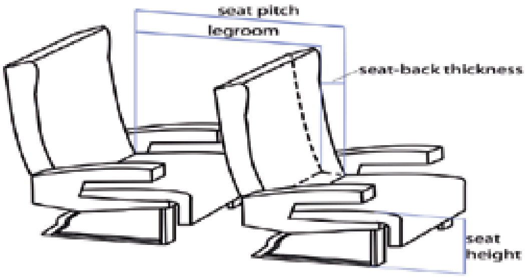


Fig. 2. Some aircraft seat design parameters from (Nadadur & Parkinson, 2009)

Sleeping and sitting posture is an important factor for passenger's comfort especially in long haul flights and it can affect the pressure distribution over the body. (Tan, Iaeng, Chen, Kimman & Rauterberg, 2009) held an analysis of passengers' postures in the economy class to help in seat design. In their study, they identified seven different sleeping positions for passengers. When considering the anthropometry differences between humans of different origins, we can say that it is difficult for a passive seat to achieve all comfort positions of different postures for all passengers, so an active seat with adjustable moving parts is usually required.

- ***Active Seat*****:** A *Passive Seat* provides services of static features. On the contrary, an *Active Seat* is able to get an input from the passenger to change the service it provides. The input can be an activity to change the angle or position of adjustable parts of the seat; for example, the passenger can freely set the backrest angle or adjust the height of headrest to match his posture. In business class, the seat can accommodate a variety of postures for different activities such as watching TV, reading, sleeping, etc... Figure 3 shows a simple mechanical button (in economic class) for changing backrest angle Vs an electronic buttons (in business class) that can easily change the orientation of different parts of the seat using embedded motors.

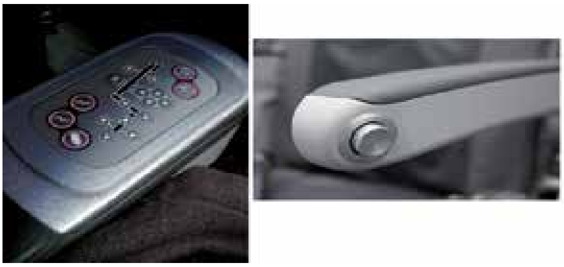


Fig. 3. Electronic Vs Mechanical seat adjustment

In economy class, the degrees of freedom of an *Active Seat* are very limited where minimal parts are allowed to change their orientation due to limited space. For example, the armrest can be moved from the horizontal position to the vertical position to give more space, and the backrest angle can be changed to increase the body inclination and reduce the pressure exerted on the back. However, the inclination angle is usually very small in order not to reduce leg space of the behind seat. On the contrary, the business class seat is featured by large spaces; thus, different parts can be reoriented easily. A premium seat may be in a pod and capable of opening out into a flat sleeping configuration or folding up into a seat for take-off and landing. Moreover, it includes more amenities such as power, task lighting, and has also a design trend towards a higher level of privacy.

2.2.1.2 Visual display units

A *Visual Display Unit (VDU)* is the principal component in the entertainment process. It is the main interface between passengers and the IFE system, as well as their ability to provide interactive services. There are different types of VDUs. At the very beginning, *Cathode Ray Tube (CRT)* displays were used. Although they were able to provide the required service at that time, but were suffering of many drawbacks. They were relatively large in size and heavy in weight, so they were used as a shared display between a set of seats. Furthermore, the ambient lighting may affect the clearness of images. As technology advances, *Liquid Crystal Display (LCD)* units were introduced. They are small in size and light in weight. These characteristics helped greatly in introducing *Video on Demand (VoD)* service, where each passenger has his own display unit to watch his selected items. At the same time LCDs can still be used as shared displays. Nowadays, displays are equipped with an extra feature that allowed them to be used as input devices. Touch screens allow users to choose their own selections by touching the screen in the appropriate location.

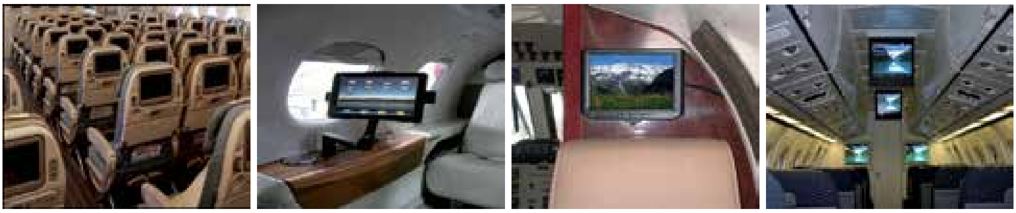
Although a normal VDU is usually sufficient to display the required contents, certain services may have special needs. Table 1 shows the characteristics required to display different media services. With respect to the display quality, Video games do not need high resolution for their images since small moving objects are the main constitute of Video games. On the contrary, movies and virtual reality applications need high resolution to present their high quality images. The interactive feature of Video games and Virtual Reality applications require special input devices, since touch screens are usually suitable for simple selections and not for quick repetitive pressing.

Service	Realistic	Interactive	Immersive	Detailed Character
Video Games	No	Yes	No	Yes
Movies	Yes	No	No	Yes
Virtual reality	Yes	Yes	Yes	Yes

Table 1. Various Display requirements

The VDU location depends on the philosophy of the installed IFE system. If the system is going to present the pre-selected media without any intervention from the user, then a global VDU is installed in the cabin ceiling (see Figure 4(d)). If VoD service are presented with user interaction to select his own media contents, then each passenger seat is provided with a private VDU fixed in the back of the front seat (see Figure 4(a)). Furthermore, seats of special locations such as seats of first row or in the business class may have special VDU placement (see Figure 4(b) & 4(c)).

The VDU viewing angle is an important satisfaction factor. The viewing angle of VDUs fixed at the back of the front seat may change when the front passenger changes the position of his seat back, so that VDUs are usually fixed on a pivot to allow the user to change their inclination; otherwise, the user has to move his head to a fixed position to be able to view the VDU. Another solution is to fix the VDU on a movable axis to give the VDU different degrees of freedom (see Figure 4(b))



(a) Private VDU (b) Movable VDU (c) First seat VDU (d) Ceiling VDU

Fig. 4. Different VDU placements

2.2.1.3 Remote control

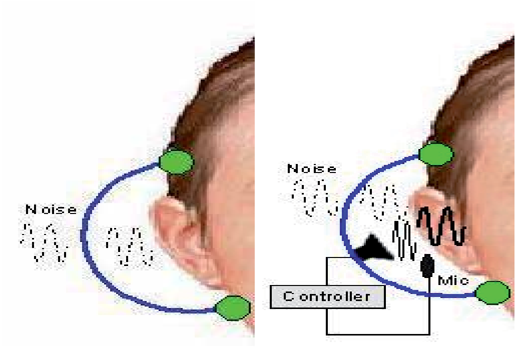
As IFE systems are becoming more and more interactive, a *Remote Control Device (RCD)* is needed to control the surrounding devices. It should be compact and easily held. Moreover, the pocket holding the RCD has to be placed in a way that makes it easily reached and not to affect passenger comfort. At the beginning, RCD used to be fixed aside to the VDU at the back of the front seat. This orientation introduced a problem when the passenger setting beside the window wants to move to the corridor; where all his neighbors have to replace their RCDs to allow him to pass. To overcome this problem, RCDs are now connected to their VDUs through wires passing via their seat. Using wireless technology can minimize such physical complexity (Akl et al., 2011).

Furthermore, passengers of no knowledge about using modern technology must be able to use RDCs easily. Usual control buttons (i.e., Volume, Rewind, Forward, etc...) are known for almost everyone; especial purpose controls such as *Settings*, and *Mode* can be carefully manipulated and, if used, to be provided by explanatory information when possible.

2.2.1.4 Noise canceling headphones

Headphones are used to privatize audio contents, so that each passenger can listen to his own selection without annoying his neighbors or being affected by the surrounding noise. Ordinary headphones are usually enough to do the job. However, modern technology can elevate the service level, by introducing active headphones capable of reducing the effect of surrounding noise (see Figure 5).

Generally, headphone ear cups have passive absorption capability which allows them to block some high frequency noise. However, they are not efficient for attenuating low frequency noise. A *Noise Canceling Headphones (NCH)* can reduce the noise through active noise cancellation techniques (Chang & Li, 2011)



(a) Passiveheadphones (b) Activeheadphones

Fig. 5. Headphones

2.2.1.5 Personal Electronic Device (PED)

Nowadays, people are getting more sticky to their *Personal Electronic Devices (PEDs)* such as laptop, mobile phone, and PDA, so most passengers are traveling with their PEDs. Connecting PEDs do not require special interfaces since modern IFE systems are moving towards wireless communication such as WiFi, Bluetooth, and IrDA, which are already used in most PEDs.

Using PEDs can have several advantages for both Airlines and passengers. Passengers will be able to use their devices to interact with the IFE system. They do not need to use or investigate unknown devices. Also, they can utilize their own data if the system permits them. Furthermore, if the IFE contents can be copied, the passenger can continue it at his hotel.

From the airlines perspective, PEDs can be used to save some dedicated devices of IFE systems. It is cheaper for airlines to remove expensive seatback monitors and let passengers to use their own devices; this is a good option for airlines offering cheap flights. Many companies (Lufthansa, 2011; Thales, 2011) are now offering broadband communication for PEDs.

2.2.2 System components

System components are usually complex to be able to handle the services while keeping simplicity of passenger components. Furthermore, the cabin environment is strict in terms of safety and imposed constrains. These characteristics encouraged the solution of using multiple technologies to form a heterogeneous system where each technology provides a solution for a part of the problem.

A context-aware IFE system can increase passenger satisfaction level. If there are many choices and the interaction design is poor, the passenger tends to get disoriented and is not able to achieve the most appealing contents. This is because most IFE systems are user adaptive systems where the user initiates system adaptation to get his personalized contents. (Liu & Rauterberg, 2007) showed the main architectural components to make a context-aware IFE system which can provide the passenger with entertainment contents based on his personal demographic information, activity, physical and psychological states if the passenger was in stress. Furthermore, the passenger is able to decline the proposed contents, and create his personalized contents.

For IFE networking, wireless technology can introduce different solutions to solve many existing problems as well as providing new services. Nowadays, wired networks are the principal technology of implementing IFE systems. Ethernet is currently the standard for wired communication in different fields. (Thompson, 2004) showed that it is characterized by interesting features such as good communication performance, scalability, high availability, and resistivity to external noise. Using off shelf technologies such as routers can reduce the costs of networking inside the cabin. In spite of all these advantages, IFE system designers are welling to exchange it -or part of it- by wireless technology to achieve more targets. Ethernet cabling is considered a burden for aircraft design because lighter aircrafts consume less fuel, and it imposes difficulties on easiness of reconfiguration and maintenance of the cabin (Akl et al., 2011). Accordingly, using different technologies within the same communication network can introduce a solution to the limitations of using each of them individually.

2.2.2.1 WiFi and Bluetooth

WiFi is a well known technology used in different commercial, industrial, and home devices. It can easily coexist with other technologies to form a heterogeneous network (Niebla, 2003). Moreover, (Lansford et al., 2001) stated that WiFi and Bluetooth technologies are two complementary not a competing technologies. They can cooperate together to provide users with different connecting services.

However, using large number of wireless devices in a very narrow metallic tunnel like the cabin has a dramatic effect on network performance. Furthermore, a major concern for using wireless devices in aircraft cabin is their interference with the aircraft communication and navigation system, especially unintended interference from passenger's *Personal Electronic Devices(PED)*. (Holzbock et al., 2004) said that the installed navigation and communication systems on the aircraft are designed to be sensitive to electromagnetic signals, so they can be protected against passenger's emitters by means of frequency separation. In addition, (Jahn & Holzbock, 2003) mentioned that there are two types of PEDs interference, intentional and spurious. The former is the emissions used to transmit data over the PED allocated frequency band. The latter is the emissions due to the RF noise level. However, indoor channel models mainly investigate office or home environments, thus these models may not be appropriate for modeling an aircraft cabin channel. Attenuation of walls and multi path effects in a normal indoor environment are effects, which are not expected to be comparable to the effect of the higher obstacle density in a metallic tunnel. The elongated structure of a cabin causes smaller losses, than that expected in other type of room shapes. However, the power addition of local signal paths can lead to fading of the signal in particular points. In addition, small movements of the receiver can have a substantial effect on reception. The same opinion was emphasized by (Diaz & Esquitino, 2004).

Different efforts were held to overcome this problem, (Youssef et al., 2004) used the commercial software package *Wireless Insite* to model the electromagnetic propagation of different wireless access points inside different types of aircrafts. (Moraitis et al., 2009) held a measurement campaign inside a Boeing 737-400 aircraft to obtain a propagation development model for three different frequencies, 1.8, 2.1, and 2.45GHz which represent the GSM, UMTS, and WLAN and Bluetooth technologies, respectively. Nowadays, many airline companies allows WiFi devices on their aircrafts such as Lufthanza (FlyNet, Last visit 2011), and Delta Airlines (DeltaAirline, Last visit 2011).

2.2.2.2 Wireless Universal Serial Bus (WUSB)

Universal Serial Bus (USB) technology allows different peripherals to be connected to the same PC more easily and efficiently than other technologies such as serial and parallel ports. However, cables are still needed to connect the devices. This raised the issue of *Wireless USB (WUSB)* where devices can have the same connectivity through a wireless technology. (Leavitt, 2007) stated that although it is difficult to achieve a wireless performance similar to wired USB, but the rapid improvements in radio communication can make WUSB a competent rival. It is based on the *Ultra Wide Band (UWB)* technology. In Europe, it supports a frequency range from 3.1 to 4.8 GHz. Moreover, (Udar et al., 2007) mentioned that UWB communication is suitable for short range communications, which can be extended by the use of mesh networks. Although WUSB was designed to satisfy client needs, but it can also be used in a data centre environment. They discussed how WUSB characteristics can match such environment. This application can be of a great help in IFE systems, which strive to massive data communication to support multimedia services and minimizing connection cables. Moreover, (Sohn et al., 2008) discussed the design issues related to WUSB. He stated that WUSB can support up to 480 Mbps, but in real world it does not give the promised values; and they showed the effect of design parameters on device performance.

2.2.2.3 PowerLine Communication (PLC)

A PLC network can be used to convey data signals over cables dedicated to carry electrical power; where PLC modems are used to convert data from the digital signal level to the high power level; and vice versa. Using an existing wiring infrastructure can dramatically reduce costs and effort for setting up a communication network. Moreover, it can decrease the time needed for reconfiguring cabin layout since less cables are going to be relocated. However, such technology suffers from different problems. A power line cable works as an antenna that can produce *Electromagnetic Emissions (EME)*. Thus, a PLC device must be *Electromagnetic Compatible (EMC)* to the surrounding environment. This means that it must not produce intolerable EME, and not to be susceptible to them. To overcome this problem, the transmission power should not be high in order not to disturb other communicating devices (Hrasnica et al., 2004). However, working on a limited power signal makes the system sensitive for external noise. In spite of this, the PLC devices can work without concerns of external interference due to two reasons. Firstly, the PLC network is divided into segments; this minimizes signal attenuation. Secondly, all cabin devices are designed according to strict rules that prevent EME high enough to interfere with the surrounding devices. (Akl et al., 2010) presented a PLC network dedicated for IFE systems to replace part of the wired communication network, where two PLC devices were used; *Power Line Head Box (PLHB) and Power Line Box (PLB)*. PLHB connects the two terminals of the power line to connect data servers with seats. Each PLHB service a group of seats, which are equipped with PLB per seat (see Figure 6).

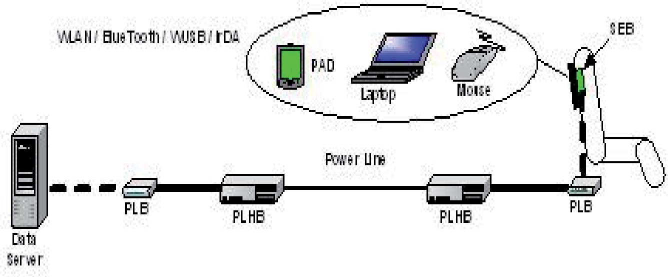


Fig. 6. Heterogeneous network architecture

2.2.2.4 GSM

For several years the aircraft industry has been looking for a technology to provide, at a reasonable cost, an onboard phone service (see Figure 7). Nevertheless, some technical hitches make successful calls via the terrestrial *Global System for Mobile Communications (GSM)* network impossible. The mobiles are unable to make reliable contact with ground-based base stations, so they would transmit with maximum RF power and these RF fields could potentially cause interference with the aircraft communications systems. On the other hand, the high speed of the aircraft causes frequent handover from cell to cell, and in extreme cases could even cause degradation of terrestrial services due to the large amount of control signaling required in managing these handovers. In order to avoid these problems and allow airline passengers to use their own mobile terminals during certain stages of flight, a novel approach called *GSM On-Board (GSMOB)* is used. The GSMOB system consists of a low power base station carried on board the aircraft itself, and an associated unit emitting radio noise in the GSM band, raising the noise floor above the signal level originated by ground base stations. Thus mobiles activated at cruising altitude do not see any terrestrial network signal, but only the aircraft-originated cell. This way, the power level needed is low, which reduces the interference with aircraft systems.

The AeroMobile (AeroMobile, Last visit 2011) is a GSM service provider for the aviation industry that allows passengers to use their mobile phones and devices safely during the flight. Passengers can connect to an AeroMobile pico cell located inside the craft which relays text messages and calls to a satellite link which sends them to the ground network. The AeroMobile system manages all the cellular devices onboard. This system is adopted by Panasonic to be part of its in-flight cellular phone component.

2.2.2.5 Satellite communication

In-cabin communication can be extended by being connected to terrestrial networks through satellite links (see Figure 7). Using satellite channels allow passengers to use their mobile phones, send emails, access internet, and achieve online entertainment services. However, the satellite link is considered as the connection bottleneck, so traffic flow in and out of the cabin must be analyzed (Niebla, 2003). (Radzik et al., 2008) performed a satellite system performance assessment for IFE system and *Air Traffic Control (ATC)*, where the satellite link can be shared between IFE and ATC streams. (Holzbock et al., 2004) presented, in details, two systems that allow in-cabin communication to be connected to assessment networks; the ABATE system (1996-1998), and the WirelessCabin system (2002-2004). Another recent project is the E-CAB project (ECAB, Last visit 2011) which was held by Airbus.

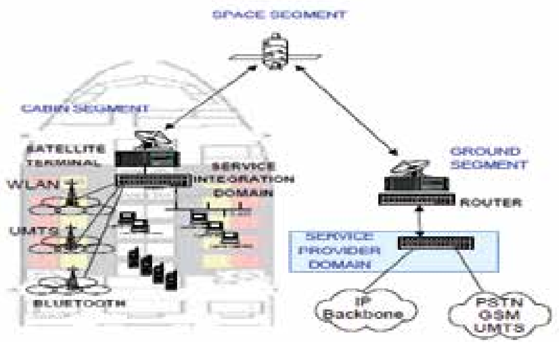


Fig. 7. Satellite link from (Niebla, 2003)

3. Design and evaluation of modern IFE systems

To design an IFE system, different types of requirements need to be defined and constrains must be considered. It is not just adding some entertainment devices, but it is a system which will be located in a very strict environment. This system will have an impact on passengers, airlines, and aircraft design. Therefore, a formal modeling of IFE systems is a paramount need which can be achieved through *System Modeling Language (SysML)*. SysML is a modeling language for representing systems and product architectures, as well as their behavior and functionalities. It is an important tool to have an understanding of a system to prevent complex failure modes leading to costly product recalls. Furthermore, it uses generic language, which is not specific to any engineering discipline, able to present the incremental details of system modeling. Modeling starts by gathering the required functionality until reaching the complex system model. This is achieved by presenting its sub-system structures, and showing their behavior of interacting together as well as with external system. However, we have to stress on the fact that there is no optimum model for any system, but we can have a good model. A good model is the one that fulfills all of system functional and non-functional requirements.

3.1 Proposed IFE model

In this section, we propose a SysML model that takes us through a step by step design process as a systematic design approach to help designers to handle such complex system. The model will show system components, the involved actors, and their interactions with the system. We believe that the model can give the designer an idea on how to adapt his own IFE system to achieve the expected services.

A real IFE system is a large system, where its model can not be fully presented in a book chapter, so we will consider a small case study, and stress on the basic steps and techniques that should be considered during the design process.

Our case study is based on the work done by (Loureiro & Anzaloni, 2011), and our previous work in (Akl et al., 2010) to model the part related to the VoD service and the PLC network. (Loureiro & Anzaloni, 2011) introduced a peer-to-peer networking approach for using VoD for IFE systems and propose two solutions for the problem of content searching in such network. We chose their work because it is a recent research that presents two different techniques to distribute video content over a peer-to-peer network rather than using traditional client-server architecture. The peer-to-peer approach allows passenger IFE units to monitor, store, and serve media contents to each other. This can be achieved by having a *Distribution Table (DT)* containing the video file information (i.e., file ID and IP of storing peer). The work is based on how to build and update the DT. In (Akl et al., 2010), we proposed using a PLC network to replace traditional LAN (see Figure 6). The PLC system consists of a *Power Line Head Box (PLHB)* and a *Power Line Box (PLB)*, where the PLHB connects the two terminals of the power line. Each PLHB service a group of seats which are equipped with PLB per seat. The PLB is responsible for distributing the signal received by the PLHB to the seat attached devices. Each PLHB can support up to 20 PLBs at a rate of 3480 bit/sec each. We will use the model to verify if the technique proposed by (Loureiro & Anzaloni, 2011) can be supported by the PLC network or not.

3.1.1 Use Case diagram

A *Use Case* describes system functionality in terms of how its users (i.e., passengers, crew) use the system to achieve the needed targets. It represents a high level of abstraction to model IFE requirements and interaction with users. Consequently, it typically covers scenarios through which stake holders (i.e., actors) can use the IFE system. Hull et al. (2011) stated that *"A stake holder is an individual, group of people, organization or other entity that has a direct or indirect interest (or stake) in a system"*.

In a Use Case, the system boundaries are identified by a square box to decide what belongs to the system and what does not. For example, a GPS device that provides the IFE system with data used in a map display is considered as a part of an external system (i.e., navigational system).

Figure 8 presents a *Use Case diagram* for our proposed IFE model. There are seven actors; passengers, crew members, a navigational system, a cabin environment, maintenance personnel, airline company, and avionic regulations. The IFE system is enclosed inside the box representing the system boundary. The oval shapes show the interactions of each actor with the system. These interactions are related together through different relations (i.e.,

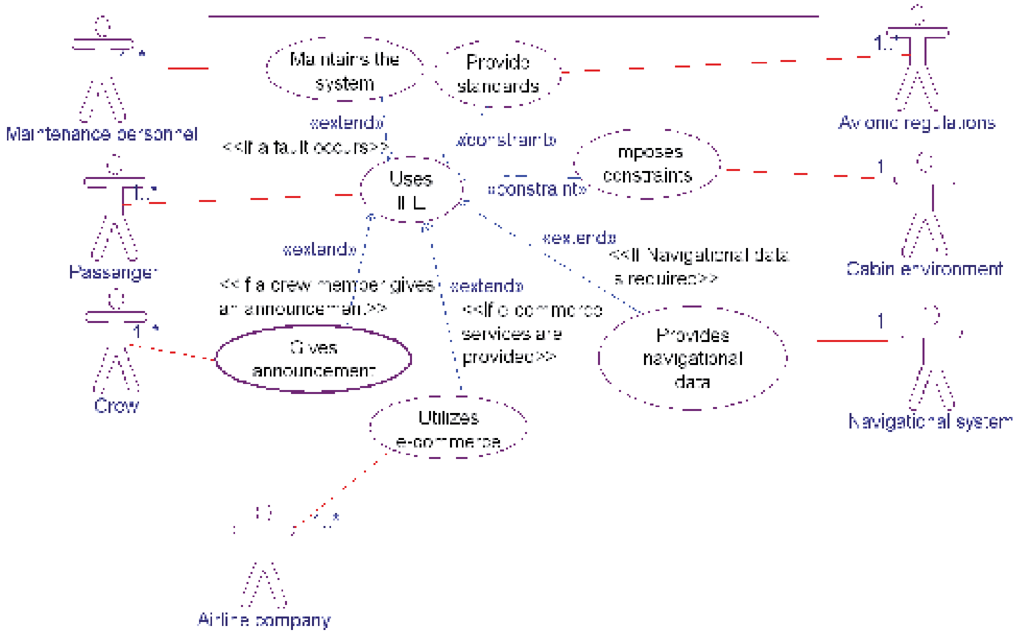


Fig. 8. IFE system Use Case diagram

extended, constrain). *Extend* relationship identify an *extending use case* which is a fragment of functionality that is not considered part of the normal base use case functionality. *Constraint* relationship shows constraints imposed on the system.

The base Use Case is *"Uses IFE system"* which is directly utilized by passengers. It represents the utilization of IFE components (see section 2.2). Its functionality can be extended when a crew member gives an announcement (see section 2.1.1), or the navigational system provides data, or a maintenance personnel performs a maintenance action. Constraints comprise the difficulties imposed by cabin environment, and the standards provided by avionic regulations (i.e., ARINC standard 808, RTCA DO-160E). The next step is to model the requirements needed by stakeholders.

3.1.2 Requirements model

We present a part of the basic requirements related to the entertainment service that can exist in any IFE systems. These requirements are categorized as functional and non-functional requirements. This step helps designers to highlight the basic features of their system.

Defining system requirements seems easy, but in fact, it is not. The defining requirements process is divided into several steps. Firstly, to define stake holders. Second, to start a requirement gathering process, where requirements are collected from stakeholders. Finally, requirements are organized according to well defined rules that guarantee certain requirement characteristics which are essential for requirement analysis. For more information about requirement engineering, we refer readers to (Hull et al., 2011; Young, 2004).

We will assume that the first and second steps are already done, so that our IFE system requirements are already gathered from stakeholders, and we will classify them as functional and non-functional requirements.

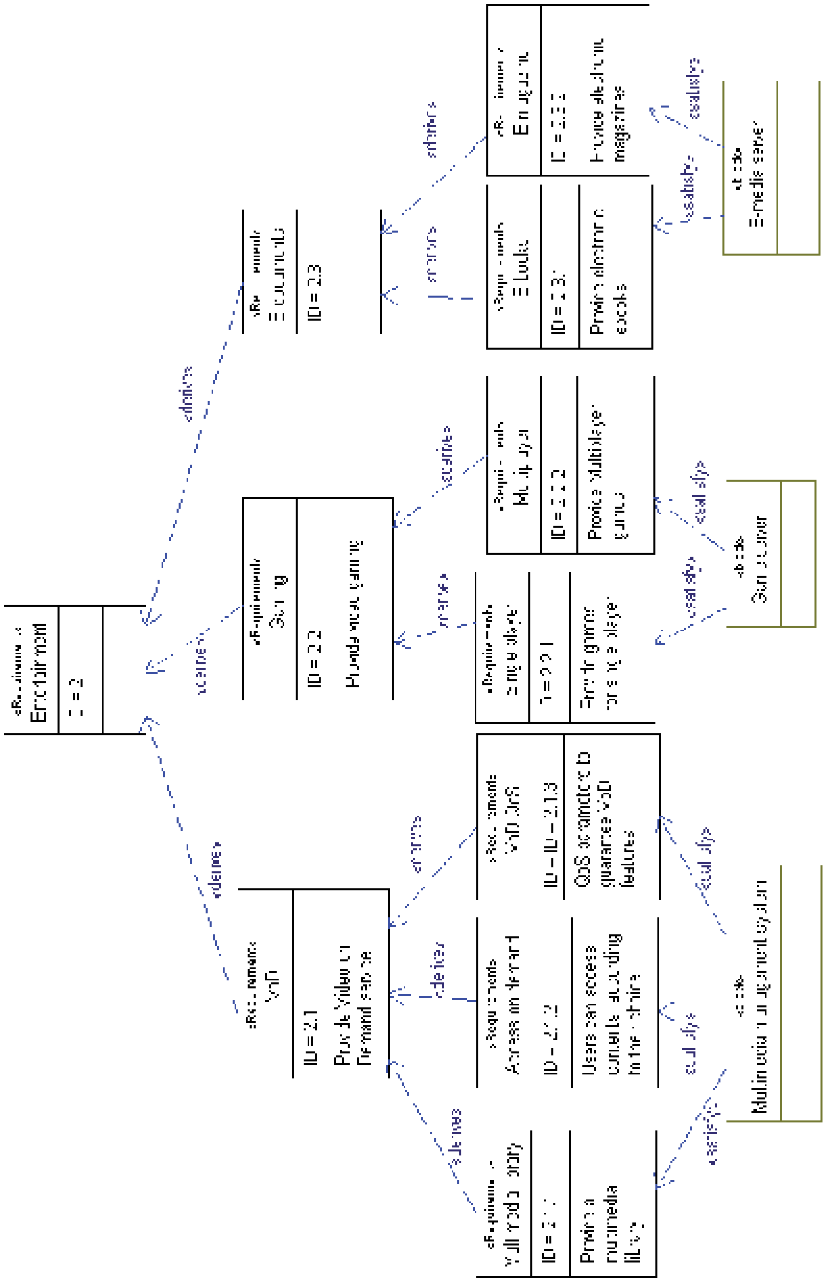


Fig. 9. Requirement diagram of entertainment specifications

3.1.2.1 Functional requirements

Functional requirements describe what the system is supposed to do by defining its behavior (i.e., functions and services). For an IFE system, this includes the different services provided to passengers, and airlines companies (see section 2.1). For each service, there is a dedicated requirement diagram. A group of related requirements are called a specification.

Figure 9 presents the specifications of entertainment service. Each block represents a requirement; showing its name, ID number, and text explaining the purpose of the requirement. The *Derive* relationship shows sub-requirements needed to fulfill the parent requirement. For example, our entertainment service will include VoD, Gaming, and E-documents services. The VoD service will be fulfilled through a *Multimedia Library* to store the VoD contents, and an *Access on Demand* capability. A system component is responsible for satisfying (i.e., represented by the *Satisfy* relationship) these requirements; it is named the *Multimedia Management System*. If necessary, the last level of requirements can decompose into finer levels of derived requirements to show more details of the system. The *Distribution Table* technique (Loureiro & Anzaloni, 2011) will be used to satisfy part of the requirements of VoD service

3.1.2.2 Non-Functional requirements

Non-functional requirements describe constraints and qualities. *Qualities* are properties or characteristics of the system that will affect user's degree of satisfaction. This includes maintainability, reliability, security, and safety issues. Designers usually focus on system functionality and may lately consider the non-functional requirements during the design process. Failing to achieve non-functional requirements may lead to a functional system with undesirable level of satisfaction.

Figure 9 shows QoS parameters as the non-functional requirements needed for the VoD service. (Loureiro & Anzaloni, 2011) identified two main parameters ρ and θ to define the required transmission. ρ and θ represent the amount of information (bytes) that needs to be transmitted across the application layer of the network during system startup, and system normal operation, respectively. They are presented in our model as constraints (explained further in section 3.1.3). ρ and θ are defined as:

$$\rho = nF(c_6 + c_7L) \tag{1}$$

$$\theta = c_5n \tag{2}$$

where n is the total number of peers, F is the number of messages sent between two nodes. L is the number of video files stored in the node's local storage, and c_5, c_6, and c_7 are constants. The next step is to model the system components that satisfy these requirements.

3.1.3 Structural model

Block Definition Diagram realizes the structural aspects of the model. It shows which components exist in the system, and the relation between them. It is formalized and reconciled with both behavior model and requirements. Blocks are used to present components; they are connected through relations, and ports to describe the points at which a block interacts with another block.

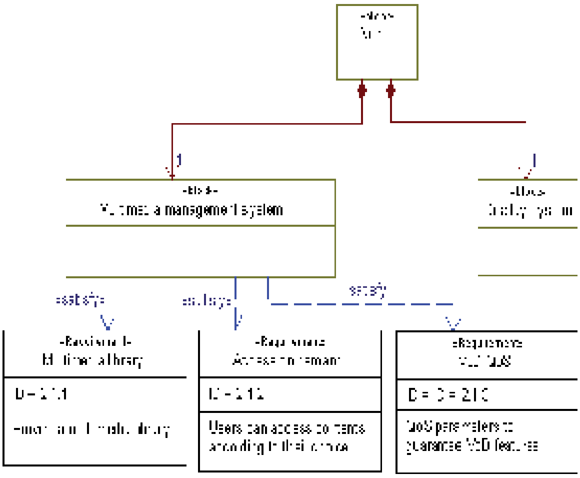

1
Multimedia management system
Display system
«satisfy»
«satisfy»
satisfy
Multimedia library
Access on demand
VoD QoS
Users can access to items according to their choice
QoS parameters to guarantee VoD features

Fig. 10. Block diagram of node structure and satisfied requirements

There are two types of ports: *Flow ports* specify what can flow in and out of blocks (i.e., data or physical items), and *Standard ports* that specify the types of services that a block either require or provide.

Figure 10 shows the main blocks of each node and its relation with the requirements depicted in figure 9. It consists of two main blocks; *Multimedia Management System* and *Display System*. The former manages the multimedia contents, while the later is responsible for displaying multimedia contents and receiving passenger selections. The figure does not show the requirements satisfied by the *Display system* block because we are only interested in the requirements of entertainment service shown in figure 9.

Figure 11 shows the node composition, and its relation with other components (i.e., *Networking System* block). Operations are listed in the *Operations* compartment of the block. However, for readability reasons, we only show the operations of *Multimedia management system* block. *Networking system* is responsible for handling communication between nodes. This is done through the PLHB component that connects different groups of PLBs. The *Multimedia Management System* block consists of three managers; *Content Search Manager*, *Local Storage Manager*, and *Content Selection Manager*. The *Local Storage Manager* handles the local multimedia contents, and defines its location inside the storage device. The *Content Selection Manager* receives the selection request from the *Display System* block, and send back the media content after being received from the *Content Search Manager*. The *Content Search Manager* searches for the requested item in the way mentioned by (Loureiro & Anzaloni, 2011) (the behavior of this technique is modeled in the next section). If the content is not stored locally, a search will be retrieved from neighboring nodes by communicating through the PLB component.

Parametric diagram uses constraint blocks that allow to define and use various system constraints. These constraints represent rules that can constraint system properties, or define rules that the system must conform to. A constraint block consists of constraint name and constraint formula. All variables or constants defined in the formula are linked to the block through an *Attribute* box or through an input from another block.

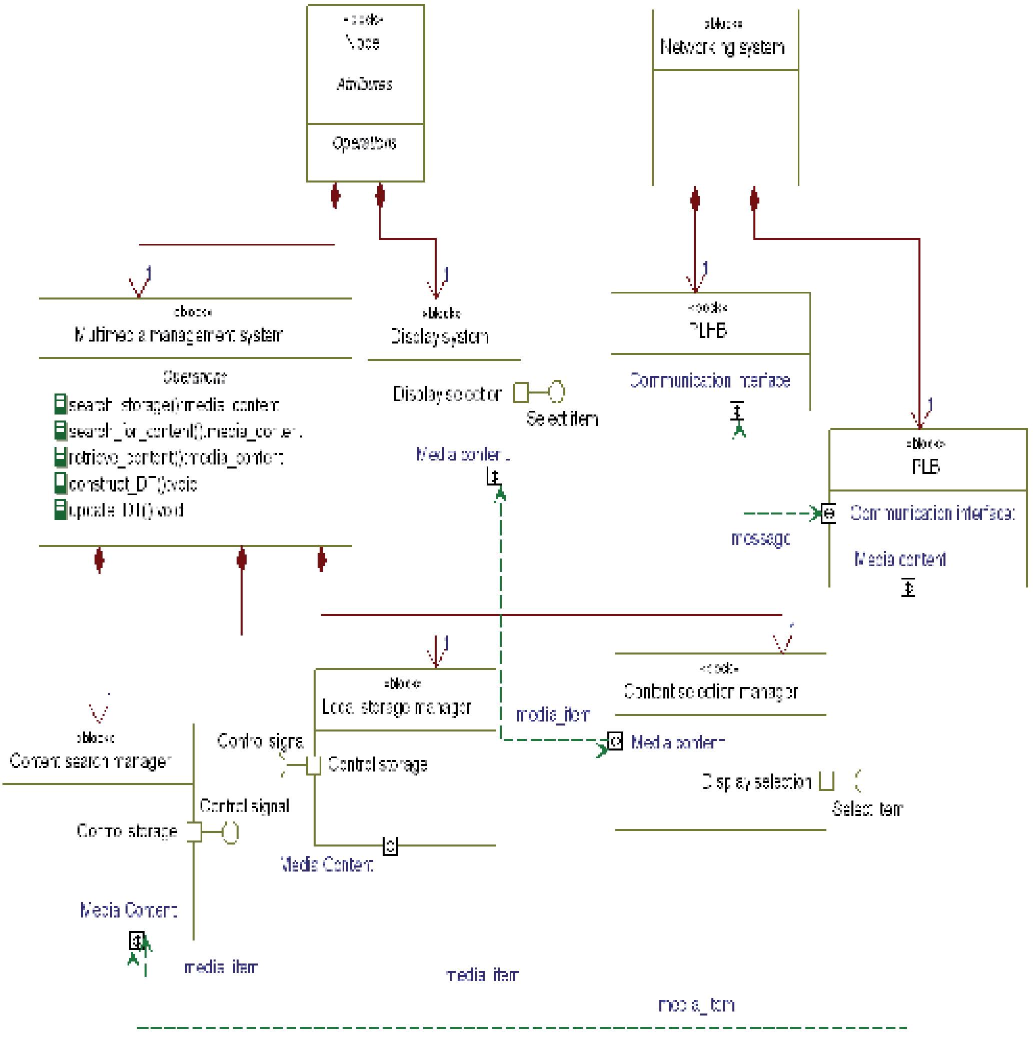


Fig. 11. Block diagram of control signals and flow items

Figure 12 shows three main constraint blocks; *Peer-to-Peer Transmission Rate*, *PLC Parameters*, and *Acceptance Criteria*. The *Peer-to-Peer Transmission rate* defines three formulas (as mentioned in (Loureiro & Anzaloni, 2011)). The PLC parameters define the PLHB maximum bandwidth, as mentioned in (Akl et al., 2010). The output of the two constraints are used to determine the validity of *Criteria 1*. *Criteria 1* is valid when the PLHB maximum bandwidth is greater than the rate of data transmission *B*. This means that the PLC network is able to handle the traffic generated to update the distribution table. *Criteria 2* defines the time taken to transfer data during startup (i.e., constructing the distribution table); this time should be less than a certain threshold defined as $T_{acceptance}$. Table 2 clarifies the meaning of symbols used in figure 12. The next step is to model the behavior of system components to acquire the expected services.

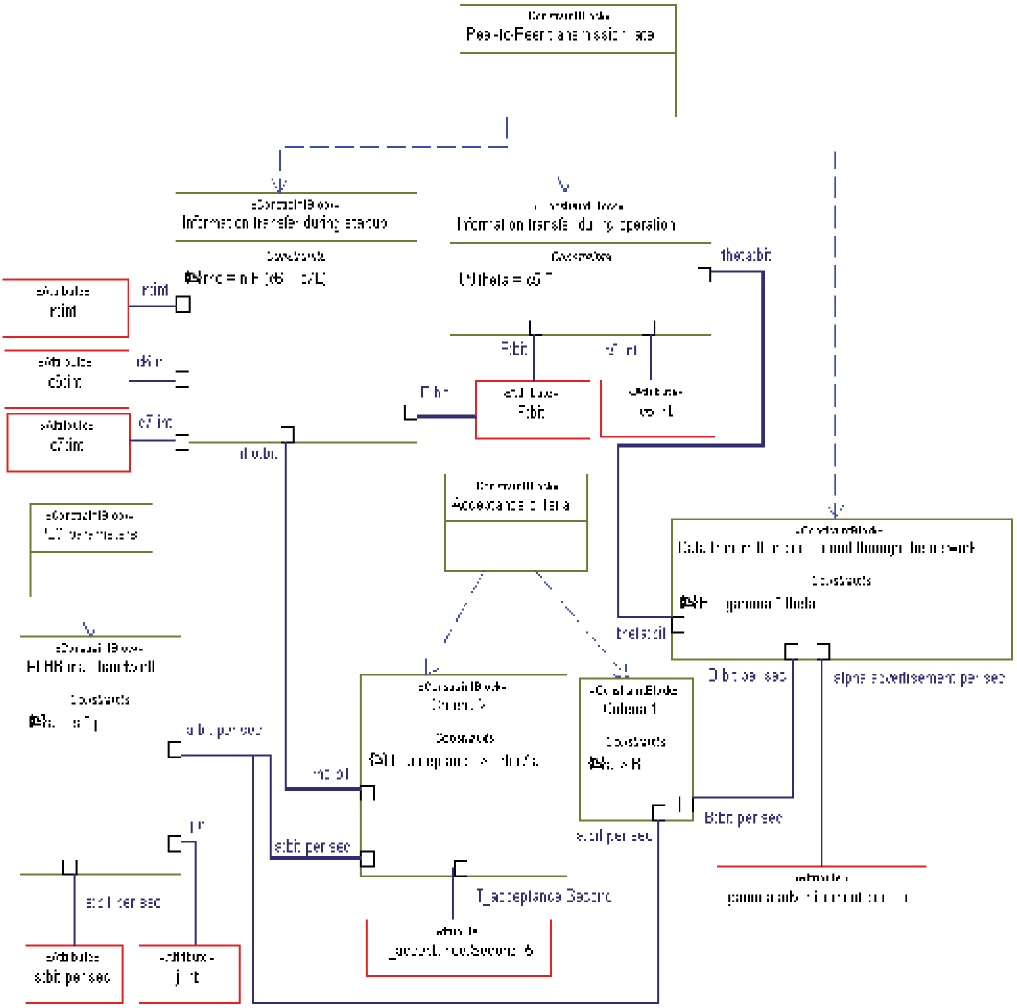

Peer-to-Peer transmission rate
Information transfer during startup
Information transfer during operation
Constraints
theta:bit
Acceptance criteria
Criteria 2
Criteria 1
gamma * theta
B:bit per sec
alpha advertisement per sec
T_acceptance:Second
jitter

Fig. 12. Parametric diagram for system constraints

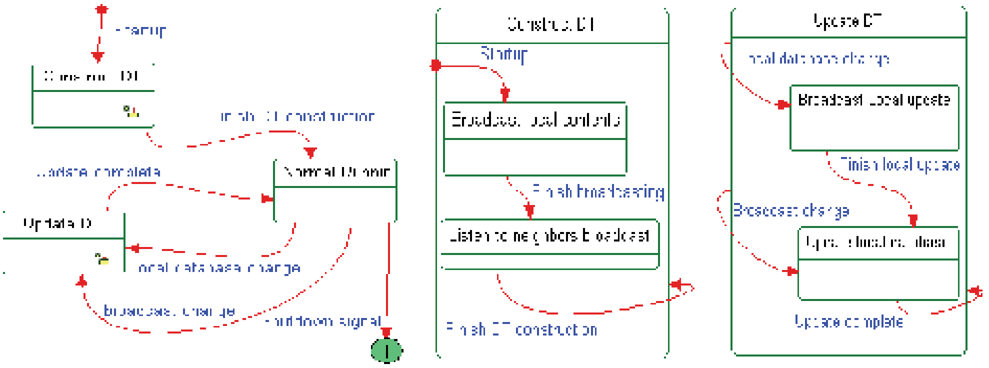
3.1.4 Behavior model

The behavior model is aiming at formalizing system behavior, and reconciling it with other requirements. In SysML, behaviors can be represented in different ways; they can be represented by *Activity diagrams*, *Sequence diagrams*, and *State machine diagrams*. We will show how they can give different views for different parts of the system.

Figure 13(a) shows the state machine representing the states of the decentralized technique. When the system startup, the *Construct DT* state is initiated, and each node starts to broadcast the information of its local video contents. Neighboring nodes receive this information and construct their *Distribution Table (DT)*. The DT contains tuples that consist of a unique video file identifier accompanied with the IP address of the node storing this file. When the construction process completes, the *Normal running* state begins, and nodes start to run normally and exchange video contents. The *Update DT* state is fired in two cases. First, when a node

Symbol	Meaning
rho (ρ)	Total amount of information (bytes) transmitted during construction of DT
n	Total number of peers
F	Messages sent between two nodes
$c_5..c_7$	Constants
theta (θ)	Total amount of information (bytes) transmitted during normal operation when one peer advertise one local database change
B	The amount of data per second transmitted through the network
gamma (γ)	Advertisement per second
L	Number of video files stored in a local storage
a	PLHB maximum bandwidth
s	Maximum bandwidth of a single PLB
j	Number of PLBs
$T_{acceptance}$	Maximum delay needed to complete the transmission of ρ or θ

Table 2. Constraints symbols



(a) First level state machine

(b) Second level of "Construct DT" state

(c) Second level of "Update DT" state

Fig. 13. State Machine Diagram

has a change in its local video contents, it updates its local DT and broadcasts the change to allow other nodes to update their local DT. Second, when it receives a *broadcast change* from neighboring nodes. When the update process finishes, the *Normal running* state is fired by an *Update complete* signal. The system closes when a shutdown signal is detected. As any SysML diagram, a state can decompose into more detailed levels. This is indicated by a small icon at the right bottom corner of the state. The sub-levels are shown in figuers 13(b) and 13(c) to present a deeper presentation of *Construct DT* and *Update DT* states, respectively.

Each system state has its own *Activity diagram*. It includes the actions needed to fulfill the state, the signals required to initiate the state, and signals to fire a transition to another state.

Figure 14 shows the behavior of state *Update DT*. It is initiated by receiving a signal indicating a change in a local file; then an update of a local database is performed, followed by broadcasting an update signal to neighboring nodes. If a *broadcast change* signal is received, the node checks if it is a new message from a neighbor or it was its own broadcast message. It

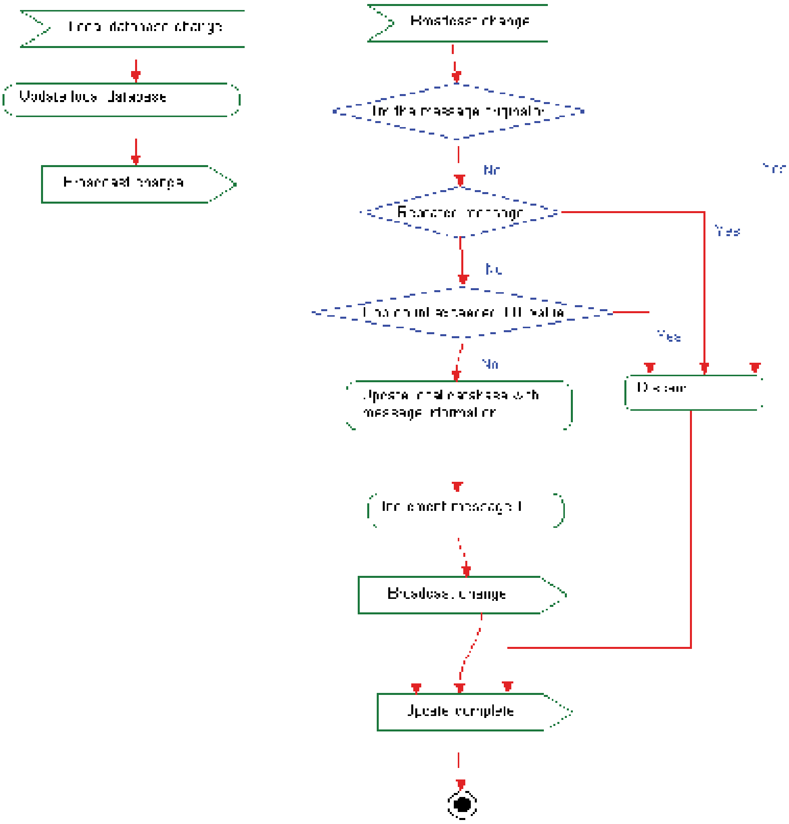


Fig. 14. Activity Diagram

updates its local database, rebroadcast the message, and send an *Update complete* signal to fire a transition to the normal running state.

3.2 System evaluation

We are interested in checking the proposed configuration (i.e., distributed table and PLC network) with the criteria shown in Figure 12 to see if it is feasible or not. According to the values indicated in (Akl et al., 2010) we can calculate the maximum bandwidth supported by each PLHB

$$a = s * j = 3480 * 20 = 69600 bit/sec = 0.06638 Mb/sec = 8.496 KB/sec \tag{3}$$

As shown in (Loureiro & Anzaloni, 2011), when $\gamma = 20 adv/sec$ and $n = 200 peers$, then

$$B = 0.1 Mb/sec \tag{4}$$

This can be interpreted as having 200 passengers, where 20 of them are performing an update to their DT. Since we will compare the value of B with the maximum bandwidth of PLHB, then we are assuming the worst case where all advertisements are initiated at the same PLHB segment.

Furthermore, $\rho = 3371.3KB$ at $n = 200$, so we can deduce its value at $n = 20$, where

$$\rho = (3371.3 * 200)/20 = 337.13KB \tag{5}$$

From 3 and 4, we find that $a < B$, so it does not fulfill the first acceptance criteria in Figure 12.

From 3 and 5, we calculate the time (T) required by the PLHB to transfer the data needed to construct DT is

$$T = \rho/a = 337.13/8.4969 = 39.681sec \tag{6}$$

This is not an accepted value because it must be less than $T_{acceptance}$ (i.e., 5 seconds).

Since both criteria are not fulfilled, then we can say that under this configuration, it is not feasible to use the decentralized technique with this PLC network. The available solutions for this problem are:

- To enhance the performance of the decentralized technique to have a less value for B and T
- To enhance the performance of the PLC network to handle more traffic
- To change the value of $T_{acceptance}$ to allow the system to accept more delay.

To achieve these changes, the designer has to change the behaviour diagrams. He also may change or add or remove some components in the block diagrams. Obviously, $T_{acceptance}$ in parametric diagram needs to be changed if the third solution is considered. If possible, some requirements may be altered to minimize the constraints imposed on the design.

3.3 Discussion

IFE is a large system with various components and parameters, especially in an aircraft environment with strict regulations. SysML provides a solution to model and verify such system. The modeling process starts by defining all parties involved with the system and gathering their requirements; this step helps to have a design that complies with their needs. These requirements are presented in a requirement diagram to show consistency, and relations between requirements and constraints. Moreover, it shows system components that are responsible for satisfying the requirements. System components are modeled using block diagrams. The block diagram shows the relations and connections between different components, define the items flowing between them, and the services they provide or need. The behavior of these components is modeled through different diagrams, where each of them represents a different view of the desired behavior. The behavior diagrams show how components can satisfy needed requirements. During the design of these models, parametric diagrams are considered to model system constraints.

The design process life cycle is not a sequential one; this means that at any step, changes can be done to a previous step. For example, during the behaviour diagram design, changes can be done to block or requirement diagrams. However, changes to requirements must be done after the approval of stakeholders. At the end of the design process, all components and behaviours must fulfill all requirements and constraints.

4. Conclusion

Since the very beginning, IFE systems were targeting passenger comfortableness. This target was the main intention to develop services dedicated to passengers. As time goes, business requirements changes, so IFE systems start to reveal another dimension of services to support crew members and airline companies in order to facilitate crew tasks and increase airline revenue. Recent technological advancements helped designers to offer various designs and services. However, this variations increased system complexity and former design techniques become less efficient.

SysML is offered as indispensable tool for modeling complex systems. It can formalize all parts of the system, so that bug tracking, and future enhancements become more manageable. In this work, we showed the design steps for a part of an IFE system and how it can be modeled. Through SysML capabilities, we were able to integrate two different techniques; the Distribution Table for a peer-to-peer network, and the PLC network. These proposals were done by two independent research teams. However, SysML modeling allowed us to verify if these proposals can be used together in the same system or not, and if not, what are the possible available solutions.

5. Future focus areas for IFE systems

IFE systems are still in their development phase and different topics are still under research. In this section, we propose some ideas to be integrated in future designs. Although IFE system development made a great leap in the past years, but there are still various issues that need further research. These developments range from enhancing current systems to adding new components and services. As technology improves, more advanced devices can be used to enhance current components such as increasing network bandwidth, using more accurate contactless sensors, wireless devices, and lighter components. There is no limit for new services that can be added to IFE systems. Nowadays, a passenger who takes different connections to his destination may not be able to continue his selected IFE content even if he is using the same airline. An attractive service is to allow him to continue unfinished IFE content when changing to the next connection, so he can enjoy the selected service for the whole trip regardless of any flight change. Another service is to create a personal profile through which he can customize his favorite contents before taking the flight, so he does not waste time for selecting items during the flight, and his profile can be used for future travels. For health services, automatic pop-up reminders can be used to stop passengers from being stick to the entertainment content. Using 3D displaying devices can introduce a new sensation to IFE entertainment. Furthermore, hologram images can be used to present safety instructions instead of crew members.

6. References

AeroMobile (Last visit 2011). http://www.aeromobile.net/.

Airships.net (Last visit 2011). http://www.airships.net.

Akl, A., Gayraud, T. & Berthou, P. (2010). Investigating Several Wireless Technologies to Build a Heteregeneous Network for the In-Flight Entertainment System Inside an Aircraft Cabin, *The Sixth International Conference on Wireless and Mobile Communications (ICWMC)* pp. 532–537.

Akl, A., Gayraud, T. & Berthou, P. (2011). A New Wireless Architecture for In-Flight Entertainment Systems Inside Aircraft Cabin, *International Journal on Advances in Networks and Services* 4, no. 1 & 2(ISSN 1942-2644): 159–175.

Aksoy, S., Atilgan, E. & Akinci, S. (2003). Airline services marketing by domestic and foreign firms: differences from the customers' viewpoint, *Journal of Air Transport Management* 9(6): 343–351.

Alamdari, F. (1999). Airline in-flight entertainment: the passengers' perspective, *Journal of Air Transport Management* 5(4): 203–209.

AmericanAirlines (2011). http://www.aa.com/i18n/travelInformation/duringFlight/onboardTechnology.jsp.

Balcombe, K., Fraser, I. & Harris, L. (2009). Consumer willingness to pay for in-flight service and comfort levels: A choice experiment, *Journal of Air Transport Management* 15(5): 221–226.

Bani-Salameh, Z., Abbas, M., Kabilan, M. K. & Bani-Salameh, L. (2010). Design and Development of Systematic Interactive Multimedia Instruction on Safety Topics for Flight Attendants, *Proceeding of the 5th International Conference on e-Learning* pp. 327–342.

Chang, C.-Y. & Li, S.-T. (2011). Active Noise Control in Headsets by Using a Low-Cost Microcontroller, *IEEE Transactions On Industrial Electronics* 58(5): 1936–1942.

Chang, Y.-H. & Liao, M.-Y. (2009). The effect of aviation safety education on passenger cabin safety awareness, *Safety Science* 47(10): 1337–1345.

Davies, R. & Birtles, P. J. (1999). *Comet - The World's First Jet Airliner*, 1st edn, The Crowood Press Ltd.

DeltaAirline (Last visit 2011). http://www.delta.com/traveling_checkin/inflight_services/products/wi-fi.jsp.

Diaz, N. R. & Esquitino, J. E. J. (2004). Wideband Channel Characterization for Wireless Communications inside a short haul aircraft, *Vehicular Technology Conference*, pp. 223–228.

ECAB (Last visit 2011). http://ec.europa.eu/research/transport/projects/items/e_cab_en.htm.

FlyNet (Last visit 2011). http://konzern.lufthansa.com/en/themen/net.html.

Holzbock, M., Hu, Y.-F., Jahn, A. & Werner, M. (2004). Advances of aeronautical communications in the EU framework, *International Journal of Satellite Communications and Networking* 22(1): 113–137.

Hrasnica, H., Haidine, A. & Lehnert, R. (2004). *Broadband Powerline Communications Networks*, John Wiley & Sons, Ltd.

Hull, E., Jackson, K. & Dick, J. (2011). *Requirements Engineering*, 3rd edn, Springer-Verlag London.

Itzel, L., Tuttlies, V., Schiele, G. & Becker, C. (2010). Consistency Management for Interactive Peer-to-Peer-based Systems, *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 1–8.

Jahn, A. & Holzbock, M. (2003). Evolution of aeronautical communications for personal and multimedia services, *IEEE Communications Magazine* 41: 36–43.

Khan, A. M. (2010). *Communication Abstraction for Data Synchronization in Distributed Virtual Environments Application to Multiplayer Games on Mobile Phones*, PhD thesis, Université d'Evry-Val d'Essonne.

Khan, A. M., Arsov, I., Preda, M., Chabridon, S. & Beugnard, A. (2010). Adaptable Client-Server Architecture for Mobile Multiplayer Games, *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques* 11: 1–7.

Kim, J., Choi, J., Chang, D., Kwon, T., Choi, Y. & Yuk, E. (2005). Traffic Characteristics of a Massively Multi-player Online Role Playing Game, *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*, ACM, pp. 1–8.

Lansford, J., Stephens, A. & Nevo, R. (2001). Wi-Fi (802.11b) and Bluetooth: Enabling Coexistence, *IEEE Communications Magazine* pp. 20–27.

Leavitt, N. (2007). For Wireless USB, the Future Starts Now, *IEEE Computer Society* 40(7): 14–16.

Liou, J. J. (2011). Consumer attitudes toward in-flight shopping, *Journal of Air Transport Management* 17(4): 221–223.

Liu, H. (2007). In-Flight Entertainment System: State of the Art and Research Directions, *Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), pp. 241–244.

Liu, H. & Rauterberg, M. (2007). Context-aware In-flight Entertainment System, *Proceedings of Posters at HCI International* pp. 1249–1254.

Loureiro, R. Z. & Anzaloni, A. (2011). Searching Content on Peer-to-Peer Networks for In-Flight Entertainment, *IEEE Aerospace conference* pp. 1–4.

Lufthansa (2011). http://www.lhsystems.com/solutions/infrastructure-services/wireless-in-flight-entertainment.htm.

Moraitis, N., Constantinou, P., Fontan, F. P. & Valtr, P. (2009). Propagation Measurements and Comparison with EM Techniques for In-Cabin Wireless Networks, *Journal EURASIP Journal on Wireless Communications and Networking - Special issue on advances in propagation modelling for wireless systems* 5: 1–13.

Nadadur, G. & Parkinson, M. B. (2009). Using designing for human variability to optimize aircraft seat layout, *SAE International Journal of Passenger Cars-Mechanical Systems* 2: 1641–1648.

Niebla, C. (2003). Coverage and capacity planning for aircraft in-cabin wireless heterogeneous networks, *IEEE Vehicular Technology Conference* pp. 1658–1662.

Qantas (2011). http://www.qantas.com.au/travel/airlines/inflight-communications/global/en.

Qi, H., Malone, D. & Botvich, D. (2009). 802 . 11 Wireless LAN Multiplayer Game Capacity and Optimization, *8th Annual Workshop on Network and Systems Support for Games (NetGames)* pp. 1–6.

Radzik, J., Pirovano, A., Tao, N. & Bousquet, M. (2008). Satellite system performance assessment for In-Flight Entertainment and Air Traffic Control, *Journal of Space Communication* 21: 69–82.

ROW44 (2011). http://row44.com/products-services/broadband/.

Schumm, J., Setz, C., Bächlin, M., Bächler, M., Arnrich, B. & Tröster, G. (2010). Unobtrusive physiological monitoring in an airplane seat, *Personal and Ubiquitous Computing* 14(6): 541–550.

Sohn, J. M., Baek, S. H. & Huh, J. D. (2008). Design issues towards a high performance wireless USB device, *IEEE International Conference on Ultra-Wideband* 3: 109–112.

Tan, C., Chen, W., Verbunt, M., Bartneck, C. & Rauterberg, M. (2009). Adaptive Posture Advisory System for Spinal Cord Injury Patient, *Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE* pp. 1–7.

Tan, C. F., Iaeng, M., Chen, W., Kimman, F. & Rauterberg, G. W. M. (2009). Sleeping Posture Analysis of Economy Class Aircraft Seat, *Proceedings of the World Congress on Engineering (WCE)* 1: 532–535.

Thales (2011). http://www.thalesgroup.com/Case_Studies/Markets/Aerospace/Inno-vating _for_inflight_entertainment_(IFE)/?pid=10295.

Thompson, H. (2004). Wireless and Internet communications technologies for monitoring and control, *Control Engineering Practice* 12(6): 781–791.

Udar, N., Kant, K., Viswanathan, R. & Cheung, D. (2007). Characterization of Ultra Wide Band Communications in Data Center Environments, *Procceedings of ICUWB* pp. 322–328.

Vink, P. (2011). *Aircraft Interior Comfort and Design*, 1st edn, CRC Press (Taylor and Francis Group).

Westelaken, R., Hu, J., Liu, H. & Rauterberg, M. (2010). Embedding gesture recognition into airplane seats for in-flight entertainment, *Journal of Ambient Intelligence and Humanized Computing* 2(2): 103–112.

Young, R. R. (2004). *The requirements engineering handbook*, 1 edn, Artech House Inc.

Youssef, M., Vahala, L. & Beggs, J. (2004). Wireless network simulation in aircraft cabins, *IEEE Antennas and Propagation Society Symposium* 3: 2223–2226.

16

Methods for Analyzing the Reliability of Electrical Systems Used Inside Aircrafts

Nicolae Jula[1] and Cepisca Costin[2]
[1]*Military Technical Academy of Bucharest*
[2]*University Politehnica of Bucharest*
Romania

1. Introduction

This chapter presents two solutions to perform reliability analysis of electrical systems installed on aircrafts. The first method for determining the reliability of electrical networks is based on an analogy between electrical impedance and reliability. The second method is based on application of Boolean algebra to the study of reliability in electrical circuits. By using these research methods we obtain information on operational safety of the electrical systems on board of an airplane, either for the entire system or for each of its components (Jula, 1986). The results allow further optimization of the construction of electrical system used on aircrafts (Aron et al., 1980), (Jula et al., 2008).

2. Calculating electrical impedance and reliability – an analogy

Establishing the reliability of structures resulting from the analysis of electrical systems installed on board of aircrafts can be achieved by direct calculations, but involves a long working time as a result of taking into account all possible situations that can occur during system operation (Reus, 1971), (Hoang Pham ,2003), (Levitin, G. et al., 1997).

A more efficient calculation method for complex structures can be achieved by applying equivalent transformation methods in terms of reliability, similar to the transformation theorems for electrical circuits applied to determine the equivalent impedance between two nodes (Moisil, 1979), (Drujinin,1977), (Billinton, 1996).

2.1 Short presentation of the analogy method

To highlight the approximations introduced by this method of calculation consider a group of elements connected in series, with the likelihood of downtime q_1, q_2, ..., q_n. Using transformation theorem for elements in series, these elements can be replaced with a resultant, a single item that has a probability of downtime q, (Drujinin,1977), given by:

- The exact formula

$$q = 1 - \prod_{i=1}^{n}(1 - q_i) \tag{1}$$

- The approximation of order 1

$$q = \sum_{i=1}^{n} q_i \tag{2}$$

- The approximation of order 2

$$q = \sum_{i=1}^{n} q_i - \frac{1}{2}\sum_{i=1}^{n}\ \sum_{j=1}^{n} q_i q_j \tag{3}$$

For the approximation of order 1, the error made is of the order of magnitude q_i^2, while for 2nd order the approximation error is q_i^3, etc.

Therefore for order 1 approximation, the probabilities of downtimes $q_1, q_2, ..., q_n$ of elements connected in series are added together as if determining the equivalent impedance of a circuit with electrical components connected in series.

A group of elements connected in parallel with the probability of downtimes $q_1, q_2, ..., q_n$ can be replaced by one single element that has a probability of downtime:

$$q = \prod_{i=1}^{n} q_i \tag{4}$$

In this case, the equivalent probability of downtime is achieved as a product of individual probabilities; therefore the result in this case is different from the equivalent impedance of an electrical circuit made of components in parallel.

A group of elements with delta connection, with the likelihood of downtime q_{12}, q_{23}, q_{31} may be replaced by another group of elements connected in star with the probability of downtime q_1, q_2, q_3. The relations for transformation are:

$$\begin{aligned} q_1 &= q_{12}q_{31} \\ q_2 &= q_{23}q_{12} \\ q_3 &= q_{31}q_{23} \end{aligned} \tag{5}$$

with an approximation error proportional with $q_{12} \cdot q_{23} \cdot q_{31}$.

Relation (5) was deducted under the assumption that the reliability of the circuit between two points, for example between point 1 and point 2 - Figure 1 - is the same for both connections in two borderline cases, namely:

- The third point is offline,
- The third point is connected to one of the first two.

Under these conditions the following relationships are obtained:

$$\begin{aligned} q_1 + q_2 - q_1q_2 &= q_{12}(q_{23} + q_{31} - q_{23}q_{31}) \\ q_2 + q_3 - q_2q_3 &= q_{23}(q_{31} + q_{12} - q_{31}q_{12}) \\ q_3 + q_1 - q_3q_1 &= q_{31}(q_{12} + q_{23} - q_{12}q_{23}) \end{aligned} \tag{6}$$

$$\begin{aligned} q_1 + q_{23} - q_1 q_2 q_3 &= q_{12} q_{31} \\ q_2 + q_{31} - q_2 q_3 q_1 &= q_{23} q_{12} \\ q_3 + q_{12} - q_3 q_1 q_2 &= q_{31} q_{23} \end{aligned} \tag{7}$$

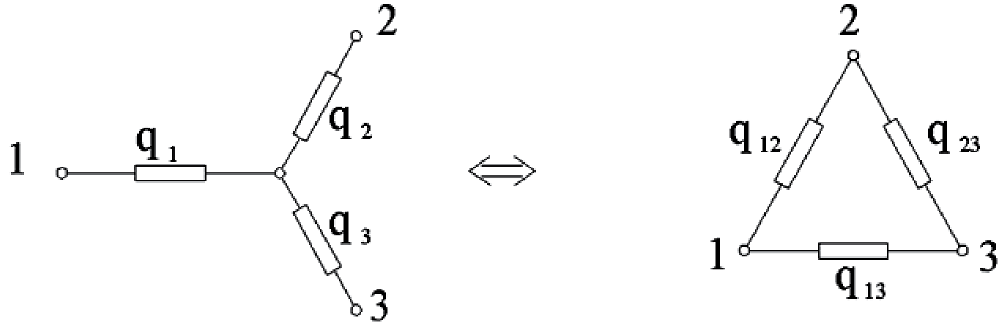


Fig. 1. Star-Delta and Delta - Star transformation for reliability.

It can be seen that the two systems described in (6) and (7) are incompatible. But if you take into account that the components used in electrical circuits on board of an aircraft are characterized by $q << 1$, approximate solutions can be utilized (Aron & Paun, 1980).

Neglecting the smaller higher-order terms of the transformation delta-star, in this case the third order component, equations in (6) become:

$$\begin{aligned} q_1 + q_2 &= q_{12} q_{23} + q_{12} q_{31} \\ q_2 + q_3 &= q_{23} q_{31} + q_{23} q_{12} \\ q_3 + q_1 &= q_{31} q_{12} + q_{31} q_{23} \end{aligned} \tag{8}$$

If the second equation is multiplied by (-1) and all the system equations are added, equation (9) is obtained:

$$q_1 = q_{12} q_{31} \tag{9}$$

Applying the same methodology for the other two remaining equations in (8) results the below equivalence for delta-star transformation:

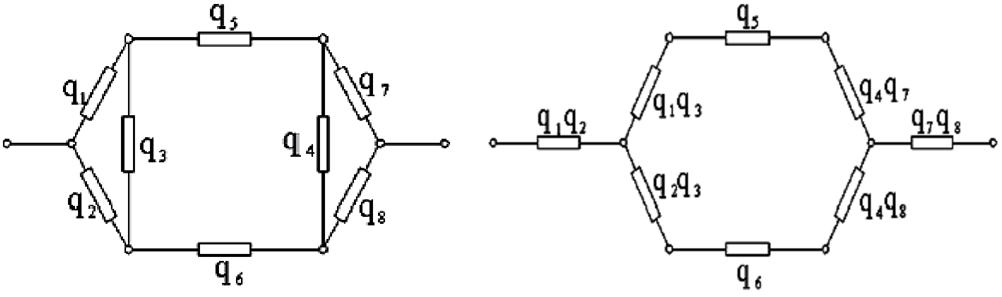
$$\begin{aligned} q_1 &= q_{12} q_{31} \\ q_2 &= q_{23} q_{12} \\ q_3 &= q_{31} q_{23} \end{aligned} \tag{10}$$

From (7) and using the same methodology, relationships for star-delta transformation are obtained (Hohan, 1982):

$$q_{12} = \sqrt{\frac{q_1 q_2}{q_3}} \quad q_{23} = \sqrt{\frac{q_2 q_3}{q_1}} \quad q_{31} = \sqrt{\frac{q_3 q_1}{q_1}} \tag{11}$$

2.2 The analogy method applied for electrical circuits used in aircrafts

Example 1. The diagram presented in Figure 2.a corresponds to a three-phase electrical generator, part of the airplane power system, powered by a three-phase electric motor, both having their stators with delta connection. The transformed version of the diagram according to the analogy method is shown in Figure 2.b.



a) b)

Fig. 2. Delta-star transformation – example 1.

The transformation delta – star applied to q_1, q_2, q_3 and q_4, q_7, q_8 becomes a simple network configuration for which downtime can be established with the specific probability when applying the previously derived relations:

$$Q = q_1q_2 + q_7q_8 + (q_1q_3 + q_5 + q_4q_7)(q_2q_3 + q_6 + q_4q_8)$$

$$Q = q_1q_2 + q_5q_6 + q_7q_8 + q_1q_3q_6 + q_1q_3q_5 + q_4q_6q_7 + q_4q_5q_8$$

If the components have the same probability q, then the probability of downtime Q is:

$$Q = 3q^2 + 4q^3$$

Example 2. Figure 3 shows the diagram of a measurement instrument based on logometric principle, used to measure engine temperature or quantity of existing fuel in the plane tanks (Jula, 1986).

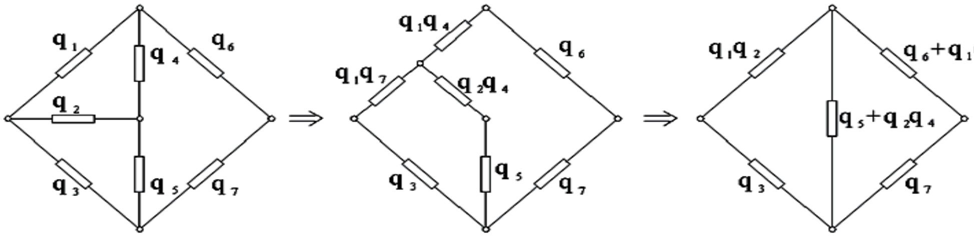


Fig. 3. Transformations for the measurement instrument – example 2.

The relations obtained for the probability of downtime Q after two transformations are:

$$Q = q_1q_2q_3 + q_7(q_6 + q_1q_4) + q_1q_2q_7(q_5 + q_2q_4) + q_3(q_6 + q_1q_4)(q_5 + q_2q_4)$$

$$Q \cong q_6q_7 + q_1q_2q_3 + q_1q_4q_7 + q_3q_5q_6$$

If the components have the same probability of downtime q, it results:

$$Q \cong q^2 + 3q^3$$

Example 3. The diagram in Figure 4 corresponds to an aircraft specific electromagnetic system powered by multiple nodes.

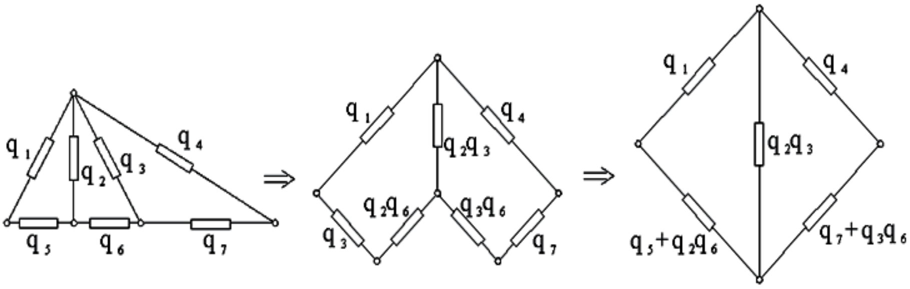


Fig. 4. Successive transformation of the electromagnetic system – example 3.

The downtime probability Q, resulting from the transformations illustrated above is:

$$Q = q_1(q_5 + q_2q_6) + q_4(q_7 + q_3q_6) + q_1q_2q_3(q_7 + q_3q_6) + q_7 + q_2q_3q_4(q_5 + q_2q_6)$$

$$Q \cong q_1q_5 + q_4q_7 + q_1q_2q_6 + q_3q_4q_6$$

If the components have the same probability q of downtime, it results:

$$Q \cong 2q^2 + 3q^3$$

Alternatively, a more efficient transformation is presented in Figure 5.

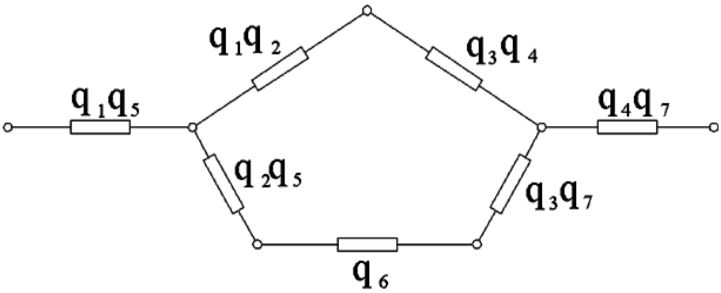


Fig. 5. A version of the final state after the transformation.

A relation for this state is:

$$Q = q_1q_5 + q_4q_7 + (q_6 + q_2q_5 + q_3q_7)(q_1q_2 + q_3q_4)$$

$$Q \cong q_1q_5 + q_4q_7 + q_1q_2q_6 + q_1q_4q_6$$

Whereby the result is identical to the one previously obtained, the calculation time is significantly reduced.

2.3 Conclusions regarding the analogy method

The method draws on the similarity between the calculus for the electrical impedance and the reliability one, allowing the use of simple relationships and reducing the number of equations to be solved. In case of complex networks other methods would lead to difficulties in obtaining results in short time, while the analogy method, with its rather low number of calculations ensures a time efficient way of finding the downtime probability of any electrical circuit.

If one or more circuit elements are less reliable than other parts of the circuit, and therefore its downtime probability is high, the transformation can get more accurate approximations of the real state of the system than other methods, mainly due to the multiplier effect contained.

3. The method based on Boolean logical structures

Large-scale systems reliability analysis is based on the quantification of the failure process at the structural level. Thus, any system downtime is a result of a quantified sequence of states in the failure process. The quantification level can be chosen in accordance with the desired goal and probability, down even to individual components of the system. The more detailed the quantification, the more accurate would be the resulting probability (Reus, 1971) (Muzi, 2008).

The conceptual representation of an emergent downtime is formed by a series of primary events, interconnected through different Boolean logical structures, which indicate the possible combinations of those elements having as result a system failure (Denis-Papin& Malgrange, 1970), (Chern & Jan, 1986). Thus determining the reliability of an aircraft electrical system using Boolean algebra actually means calculating the probability of a “failure” event.

3.1 Principles of the Boolean method

From the structural point of view, for the reliability analysis, we will use the terms:

- Primary elements – components or blocks at the base level of the quantification,
- Primary failures – primary elements failures,
- Unwanted event – system failure state,
- Failure mode – the set of primary elements that when simultaneously in failure mode, drives to a system failure
- Minimal failure mode – the smallest set of primary components that when simultaneously in failure mode, drive to a system failure
- Hierarchic level – all elements that are structurally equivalent and having equivalent positions in the system failure representation.

The method is based on binary logic. Thus, a system function is equivalent to a binary function, which variables are the events (the failures).

This binary function:

$$Y = f\left(X_1, X_2, ..., X_n\right) \tag{12}$$

is synthesized with logical elements AND/OR, using the following symbols and states:

- $\bigcup$ (Reunion) for the function OR
- $\bigcap$ (Intersection) for the function AND

X_i is 1 if the primary element is good and 0 otherwise, and Y is 1 if the system is good and 0 otherwise.

The method representation is depicted in Figure 6. For the reliability function indicators calculus, in the hypothesis of the failure intensity having an exponential distribution, we use the relations:

$$R(t) = \exp\left(-\sum_{i=1}^{n} \lambda_i t\right) = \exp\left(- \wedge t\right) \tag{13}$$

$$R(t) = 1 - \prod_{i=1}^{n} \left[1 - \exp\left(\lambda_i t\right)\right] \tag{14}$$

where: $\wedge = \sum_{i=1}^{n} \lambda_i$.

Relation (13) is used for the serial connection and relation (14) is used for the parallel connection of the elements.

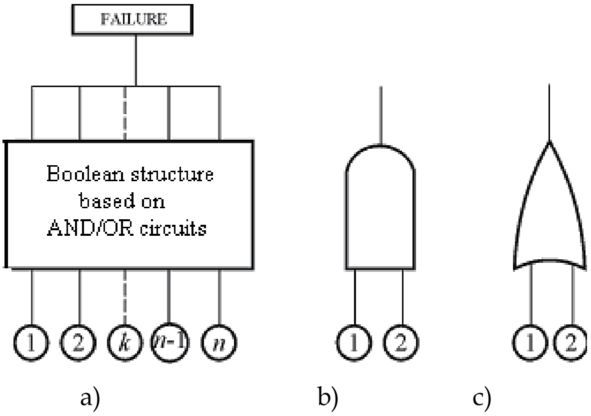


Fig. 6. a) The general concept of the method based on Boolean algebra (1, 2,..., n are independent primary events); b) the schematics of the logic function AND; c) the schematics of the logic function OR.

3.2 Method application for determining the reliability of the aircrafts electric circuits

In order to exemplify the method for the reliability indicators determination, we will focus on the DC electrical power supply system of an aircraft. Figure 7 depicts the electric power supply system of an aircraft.

In principle, this electric power supply system is present (as the main electric power supply system) in a large number of military aircrafts ranging from the MiG family (21, 23, 27, 29,31,35), Su (30,33,34,35,37) to Chengdu (J-10), Shenyang (J-11) and ORAO. The example refers only to a DC electric power supply system nevertheless the method can be used in alternative current and mixed systems set-ups. In Figure 7:

- 1E – starter-generator – startup time of several seconds (as a starter), after a successful start (three attempts permitted) it goes to a generator regime, supplying a 28V DC voltage
- 4E – accumulator switch
- 5E – inverse polarity protection diode
- 13E – accumulator
- 14E – accumulator to DC bar switch
- 24E – generator to DC bar coupler / de-coupler
- 47E – fuse
- 27E – voltage regulator.

The emerging failure state diagram using AND/OR elements is depicted in Figure 8. The failure event is the loss of voltage at the 28V bar.

For the failure intensity λ_i of the components we use the relation:

$$\lambda_i = k\lambda_0 \tag{15}$$

where: k – maintenance and way-of-use coefficient (for aircraft components the coefficient varies between 120 and 160); λ_0 – failure intensity – manufacturer specific data.

The data relative to the electric power supply system are presented in Table 1.

Symbol	Description	$\lambda_0\left[\mathrm{h}^{-1}\right]$	No.	k	$\lambda_i = nk\lambda_0\left[\mathrm{h}^{-1}\right]$	$F_i = 1 - e^{-\lambda_i t}$
4E	Switch	$0.12 \cdot 10^{-6}$	1	160	$\lambda_1 = 1.92 \cdot 10^{-5}$	$F_1 = 1 - e^{-1.92 \cdot 10^{-5} t}$
5E	Diode	$0.6 \cdot 10^{-6}$	1	160	$\lambda_1 = 9.6 \cdot 10^{-5}$	$F_2 = 1 - e^{-9.6 \cdot 10^{-5} t}$
13E	Accumulator	$1.4 \cdot 10^{-6}$	1	160	$\lambda_1 = 22.4 \cdot 10^{-5}$	$F_3 = 1 - e^{-22.4 \cdot 10^{-5} t}$
14E	Coupler	$0.4 \cdot 10^{-6}$	1	160	$\lambda_1 = 6.4 \cdot 10^{-5}$	$F_4 = 1 - e^{-6.4 \cdot 10^{-5} t}$
47E	Fuse	$2.75 \cdot 10^{-6}$	1	160	$\lambda_1 = 44 \cdot 10^{-5}$	$F_5 = 1 - e^{-44 \cdot 10^{-5} t}$
-	Contacts 1	$0.1 \cdot 10^{-6}$	1	160	$\lambda_1 = 16 \cdot 10^{-5}$	$F_6 = 1 - e^{-16 \cdot 10^{-5} t}$

Table 1. Part I

Symbol	Description	$\lambda_0\left[\mathrm{h}^{-1}\right]$	No.	k	$\lambda_i = nk\lambda_0\left[\mathrm{h}^{-1}\right]$	$F_i = 1 - e^{-\lambda_i t}$
1E	Starter-generator	$6 \cdot 10^{-6}$	1	160	$\lambda_1 = 96 \cdot 10^{-5}$	$F_8 = 1 - e^{-96 \cdot 10^{-5} t}$
24E	Coupler / Decoupler	$0.25 \cdot 10^{-6}$	1	160	$\lambda_1 = 4 \cdot 10^{-5}$	$F_9 = 1 - e^{-4 \cdot 10^{-5} t}$
27E	Voltage regulator	$13 \cdot 10^{-6}$	1	160	$\lambda_1 = 208 \cdot 10^{-5}$	$F_{10} = 1 - e^{-208 \cdot 10^{-5} t}$
-	Contacts 1	$0.1 \cdot 10^{-6}$	1	160	$\lambda_1 = 16 \cdot 10^{-5}$	$F_{11} = 1 - e^{-16 \cdot 10^{-5} t}$

Table 1. Part II

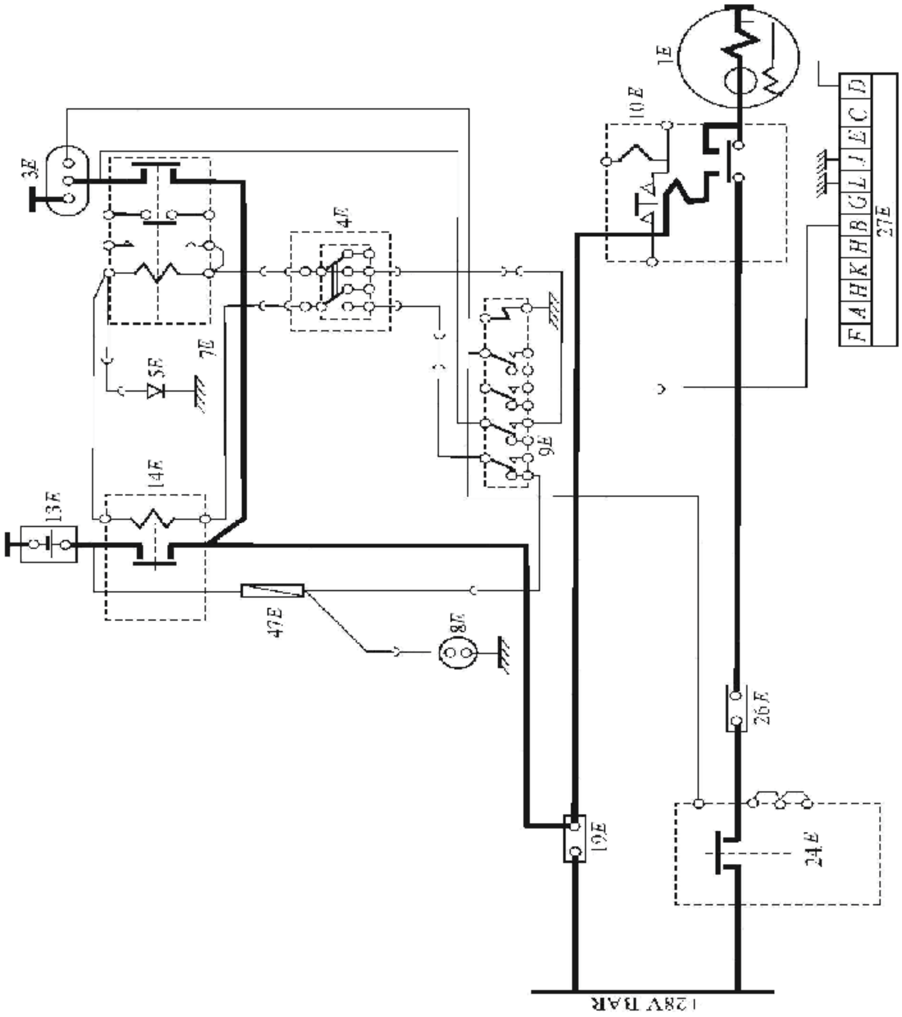


Fig. 7. The electric power supply diagram for a DC main electric supply system aircraft (fragment).

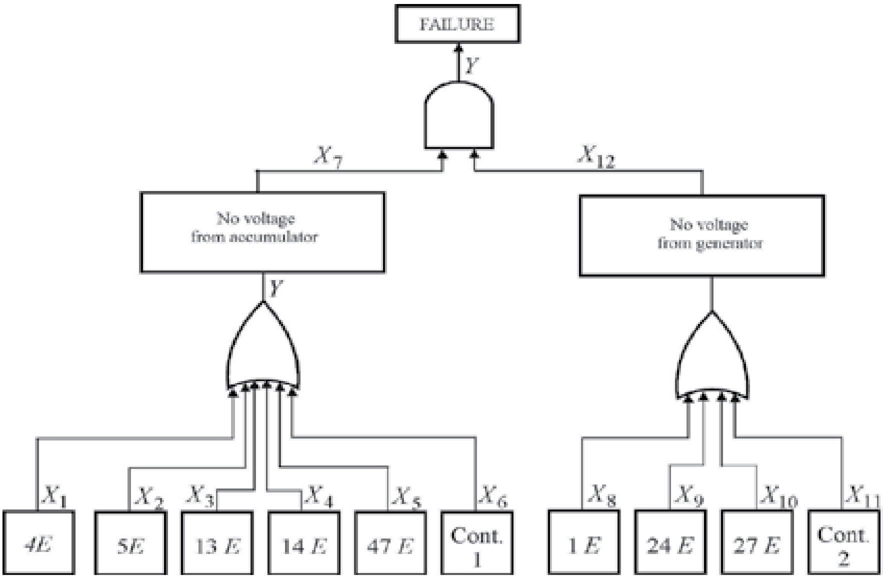


Fig. 8. The logic structure that drives to the system failure status.

In these conditions, the Boolean function associated to the logic structure depicted in Figure 8 has the following form:

$$Y = X_7 \cap X_{12} = \left(X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5 \cup X_6\right) \cap \left(X_8 \cup X_9 \cup X_{10} \cup X_{11}\right) \tag{16}$$

To transform the logic equation into algebraic form we use the following relations

$$X_1 \cap X_2 = X_1 \cdot X_2 \;\; ; X_1 \cup X_2 = X_1 + X_2 - X_1 X_2 \;\; ; \bigcup_{i=1}^{n} X_i = 1 - \prod_{i=1}^{n} \left(1 - X_i\right) \tag{17}$$

Thus, we have

$$Y = \left[1 - (1 - X_1)(1 - X_2)(1 - X_3)(1 - X_4)(1 - X_5)(1 - X_6)\right] \cdot \left[1 - (1 - X_8)(1 - X_9)(1 - X_{10})(1 - X_{11})\right] \tag{18}$$

which is similar to

$$Y = X_7 \cdot X_{12} = \left[1 - \prod_{i=1}^{6} \left(1 - X_i\right)\right] \cdot \left[1 - \prod_{k=8}^{11} \left(1 - X_k\right)\right] \tag{19}$$

Considering the failure intensity as exponential distribution, the system failure probability is given by the following relations:

$$\begin{aligned} F(t) &= \left\{1 - \exp\left[-\left(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6\right) t\right]\right\} \cdot \left[1 - \exp\left(-\lambda_8 - \lambda_9 - \lambda_{10} - \lambda_{11}\right) t\right] = \\ &= 1 - \exp\left[-\sum_{i=8}^{11} \lambda_i t\right] - \exp\left[-\sum_{k=1}^{6} \lambda_k t\right] + \exp\left[-\sum_{\substack{p=1 \\ p \neq 7}}^{11} \lambda_p t\right] \end{aligned} \tag{20}$$

$$R(t) = 1 - F(t) = \exp\left[-\sum_{i=8}^{11} \lambda_i t\right] + \exp\left[-\sum_{k=1}^{6} \lambda_k t\right] - \exp\left[-\sum_{\substack{p=1\\p\neq 7}}^{11} \lambda_p t\right] \tag{21}$$

$$MTBF = \int_0^\infty R(t)\mathrm{d}t = \frac{1}{\sum_{i=8}^{11} \lambda_i t} + \frac{1}{\sum_{k=1}^{6} \lambda_k t} - \frac{1}{\sum_{\substack{p=1\\p\neq 7}}^{11} \lambda_p t} =$$

$$= \frac{1}{(96+4+208+16)\cdot 10^{-5}} + \frac{1}{(1.92+9.6+22.4+6.4+44+16)\cdot 10^{-5}} +$$

$$+ \frac{1}{(1.92+9.6+22.4+6.4+44+16+96+4+208+16)\cdot 10^{-5}}$$

On results *MTBF* = 1069.79 hours.

Thus, mean time between failures in the non improved system may be approximated as follows $MTBF \cong 1070$ hours.

3.3 Reliability optimization of electric power supply in the aircraft industry

We can improve the electric power supply system reliability using a redundant (reserve) subsystem. The proposed improved electric power supply, including the back-up subsystem (dotted lines) is depicted in Figure 9.

Further on we will analyze the improved electric power supply system reliability, using the Boolean method presented in chapter 3.2. This analysis also allows a determination of a relation between the system reliability and the system weight. Such a relation is useful when emphasizing the variation of the system reliability with the total weight of system components.

Through a compared analysis of different reliability improving variants, imposing as minimum condition the component weight, we can obtain an optimal solution. The logic structure that drives to the system failure status (for the improved system schematics) is depicted in Figure 10.

Table 2 presents the values of the failure intensity for the supplementary components from the back-up system, in the exponential distribution hypothesis.

Symbol	Description	$\lambda_0 \left[\mathrm{h}^{-1}\right]$	No.	k	$\lambda_i = nk\lambda_0 \left[\mathrm{h}^{-1}\right]$	$F_i = 1 - e^{-\lambda_i t}$
60E	Coupler	$0.4 \cdot 10^{-6}$	1	160	$\lambda_1 = 6.4 \cdot 10^{-5}$	$F_1 = 1 - e^{-6.4 \cdot 10^{-5} t}$
61E	Switch	$0.12 \cdot 10^{-6}$	1	160	$\lambda_1 = 1.92 \cdot 10^{-5}$	$F_2 = 1 - e^{-1.92 \cdot 10^{-5} t}$
-	Contacts 3	$0.1 \cdot 10^{-6}$	4	160	$\lambda_1 = 6.4 \cdot 10^{-5}$	$F_3 = 1 - e^{-6.4 \cdot 10^{-5} t}$

Table 2.

The Boolean function in this case is:

$$
\begin{aligned}
Y = (X_{16} \cap X_7) \cap X_{12} = (X_{13} \cup X_{14} \cup X_{15}) \cap \\
\cap (X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5 \cup X_6) \cap \\
\cap (X_8 \cup X_9 \cup X_{10} \cup X_{11}).
\end{aligned}
\tag{22}
$$

Transforming in algebraic form, we have:

$$
\begin{aligned}
Y = [1-(1-X_{13})(1-X_{14})(1-X_{15})] \cdot [1-(1-X_1)(1-X_2)(1-X_3)(1-X_4)(1-X_5)(1-X_6)] \cdot \\
\cdot [1-(1-X_8)(1-X_9)(1-X_{10})(1-X_{11})]
\end{aligned}
\tag{23}
$$

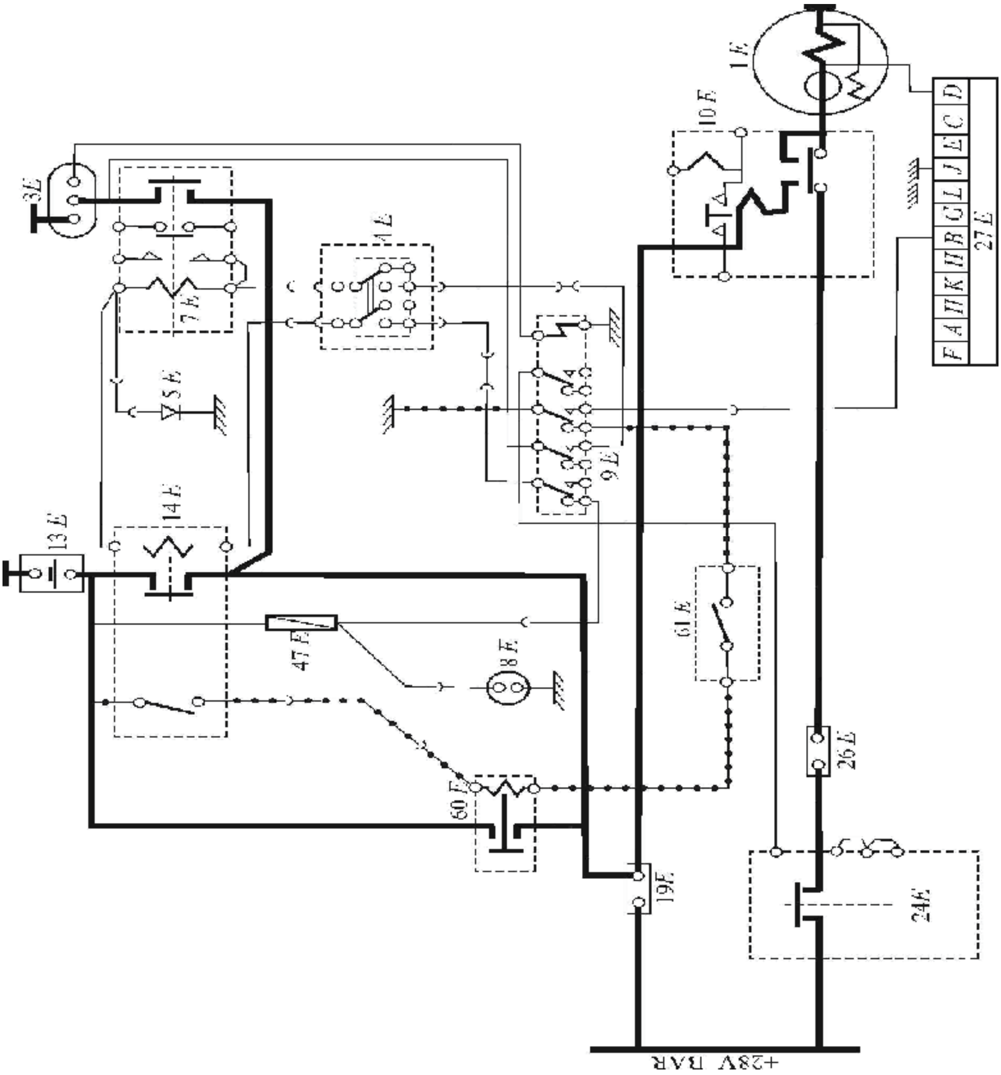


Fig. 9. Electric power supply system of an aircraft including the back-up subsystem (fragment).

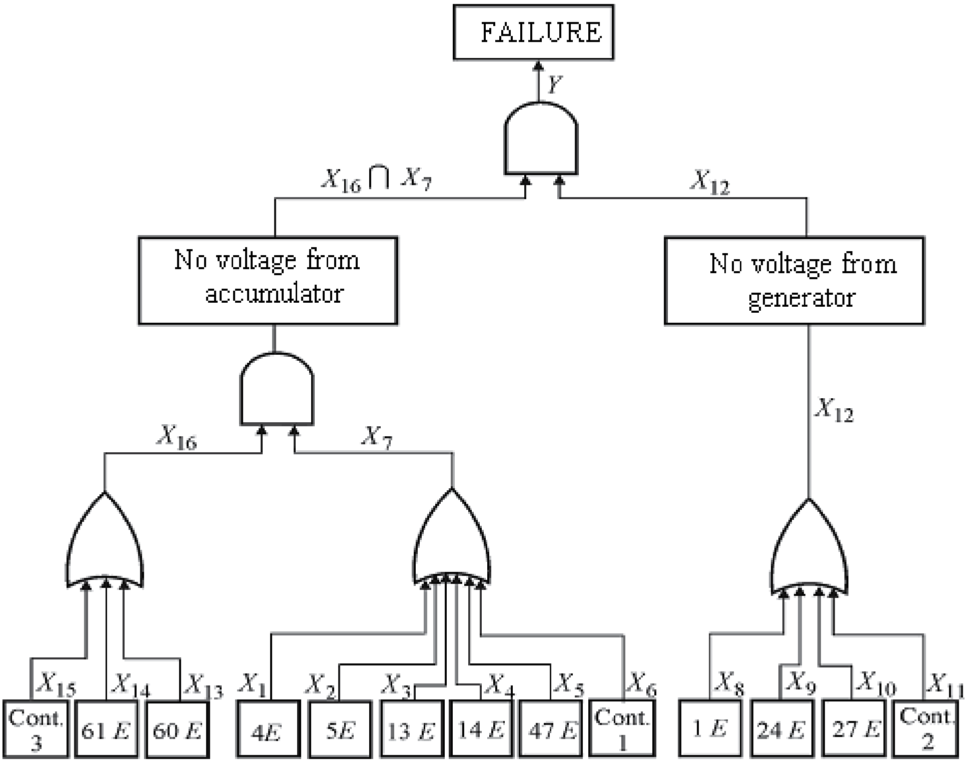


Fig. 10. The logic structure of the electric system presented in fig. 9.

$$Y = \left[1 - \prod_{i=13}^{15}\left(1 - X_i\right)\right] \cdot \left[1 - \prod_{k=1}^{6}\left(1 - X_k\right)\right] \cdot \left[1 - \prod_{p=8}^{11}\left(1 - X_p\right)\right] \qquad (24)$$

From (24) we can determine the system failure probability $F(t)$:

$$\begin{aligned} F(t) = {} & \left[1 - \exp\left(-\sum_{i=13}^{15}\lambda_i t\right)\right] \cdot \left[1 - \exp\left(-\sum_{k=1}^{6}\lambda_k t\right)\right] \cdot \left[1 - \exp\left(-\sum_{p=8}^{11}\lambda_p t\right)\right] = 1 - \exp\left(-\sum_{i=13}^{15}\lambda_i t\right) - \\ & - \exp\left(-\sum_{k=1}^{6}\lambda_k t\right) - \exp\left(-\sum_{p=8}^{11}\lambda_p t\right) + \exp\left(-\sum_{\substack{i=1\\ i\neq 7}}^{11}\lambda_i t\right) + \qquad (25) \\ & + \exp\left(-\sum_{\substack{i=1\\ i\neq 7,8,9,10,11,12}}^{15}\lambda_i t\right) - \exp\left(-\sum_{\substack{i=1\\ i\neq 7\\ i\neq 12}}^{15}\lambda_i t\right) + \exp\left(-\sum_{\substack{i=8\\ i\neq 12}}^{15}\lambda_i t\right) \end{aligned}$$

$F(t)$ and $R(t)$ are complementary functions, thus, for the electric power supply system reliability $R(t)$ we will have the following relation:

$$\begin{aligned}R(t) = \exp\left(-\sum_{i=13}^{15}\lambda_i t\right) + \exp\left(-\sum_{k=1}^{6}\lambda_k t\right) + \\ +\exp\left(-\sum_{p=8}^{11}\lambda_p t\right) - \exp\left(-\sum_{\substack{i=1\\ i\neq 7}}^{11}\lambda_i t\right) - \\ -\exp\left(-\sum_{\substack{i=1\\ i\neq 7,8,9,10,11,12}}^{15}\lambda_i t\right) + \exp\left(-\sum_{\substack{i=1\\ i\neq 7\\ i\neq 12}}^{15}\lambda_i t\right) - -\exp\left(-\sum_{\substack{i=8\\ i\neq 12}}^{15}\lambda_i t\right)\end{aligned} \tag{26}$$

$$\begin{aligned}MTBF = \int_0^\infty R(t)\mathrm{d}t = \frac{1}{\sum\limits_{i=13}^{15}\lambda_i} + \frac{1}{\sum\limits_{k=1}^{6}\lambda_k} + \frac{1}{\sum\limits_{p=8}^{11}\lambda_p} - \\ -\frac{1}{\sum\limits_{\substack{i=1\\ i\neq 7}}^{11}\lambda_i} - \frac{1}{\sum\limits_{\substack{i=1\\ i\neq 7,8,9,10,11,12}}^{15}\lambda_i} + \frac{1}{\sum\limits_{\substack{i=1\\ i\neq 7\\ i\neq 12}}^{15}\lambda_i} - \frac{1}{\sum\limits_{\substack{i=8\\ i\neq 12}}^{15}\lambda_i} \cong 6926\,\text{hours}\end{aligned} \tag{27}$$

3.4 Influence of the maintenance and way-of-use coefficient k on $MTBF$

Taking into account the characteristics of the system failure probability - $F(t)$ and reliability $R(t)$ as in Figure 7 and 9, a simulation was made using a Matlab program (Jula et. Al., 2008), which presents the time evolutions of the variables.

Coefficient k from the equation (15) has the starting value k =160. For this value MTBF was calculated both for the initial and the improved systems. The Matlab program helps conduct a complex analysis of the influence of coefficient k on system failure's probability, its reliability and *MTBF.*

Time characteristics $F(t)$ and $R(t)$, for different values of coefficient k are presented below (k = 120 (blue), k = 130 (red), k = 140 (black), k = 150 (magenta) and k = 130 (green)).

Figures 11 to 13 present the results for the initial system. As it can be seen, the increase of k is directly proportional with function $F(t)$ and inversely proportional with the reliability function $R(t)$. Mean time between failure (*MTBF*) is bigger for small values of the coefficient k.

The same analysis will be conducted for the improved system, in order to compare results. The graphic characteristics are the presented in Figures 14 to 16, while the obtained values both for initial system and improved system are presented in Table 3.

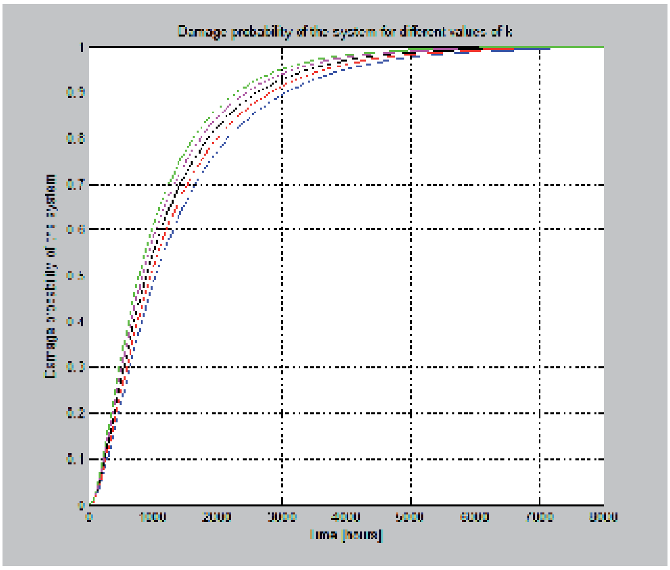


Fig. 11. System failure probability $F(t)$ for different values of k (initial system).

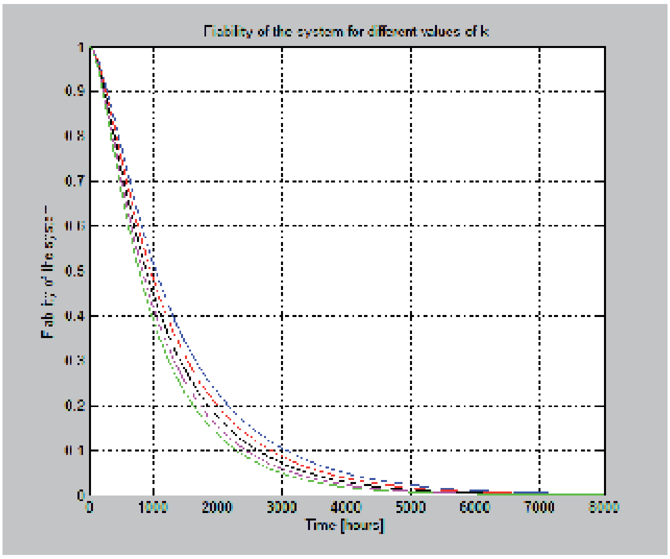


Fig. 12. System's reliability $R(t)$ for different values of k (initial system).

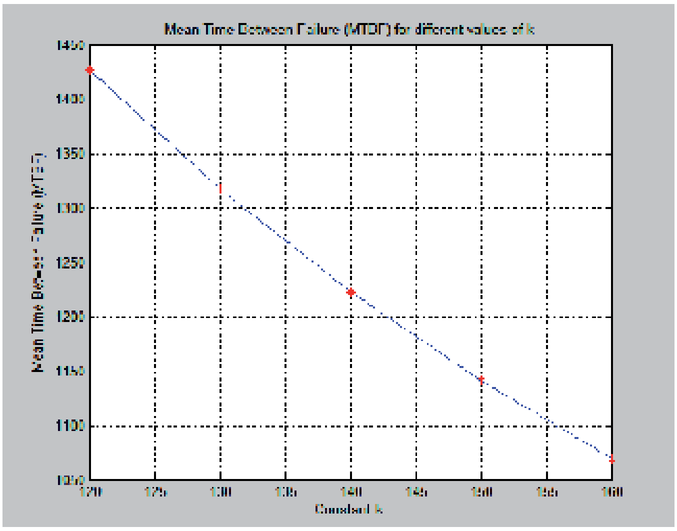


Fig. 13. *MTBF* for different values of *k* (initial system).

MTBF for different *k*	$k = 120$	$k = 130$	$k = 140$	$k = 150$	$k = 160$
Initial system (fig.3)	1426.4 hours	1316.7 hours	1222.6 hours	1141.1 hours	1069.8 hours
Improved system (fig.4)	9.2354 hours	8.5250 hours	7.9160 hours	7.3883 hours	6.9265 hours
$\gamma = \dfrac{(MTBF)_r}{(MTBF)_0}$	6.4746	6.4745	6.4747	6.4747	6.4746

Table 3.

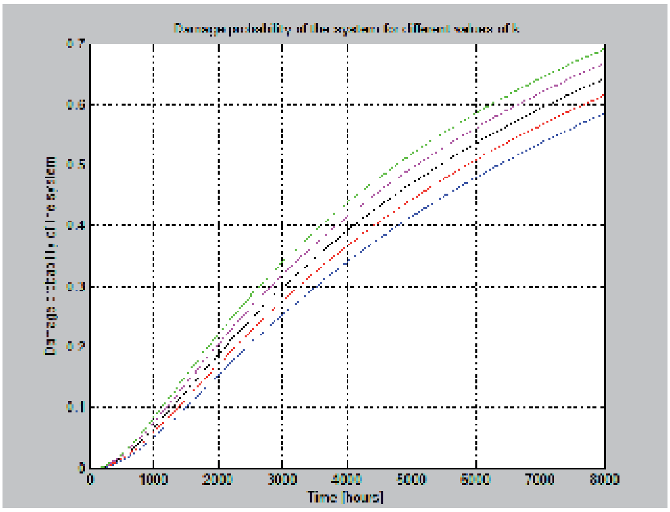


Fig. 14. System failure probability for different values of *k* (improved system).

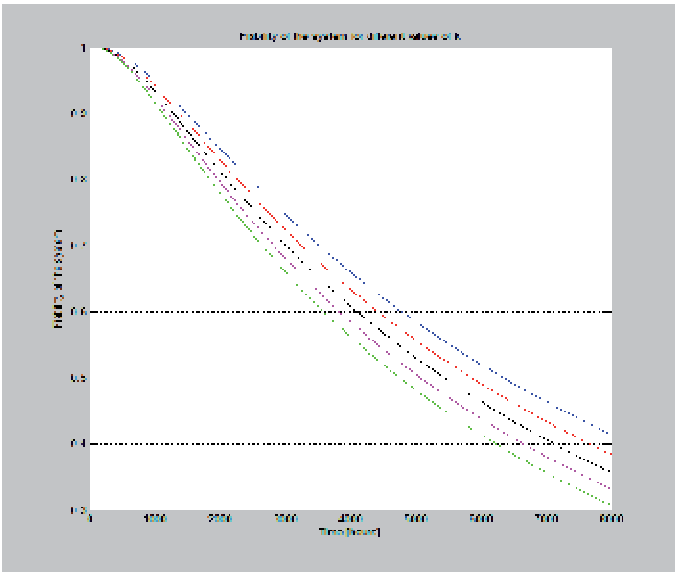


Fig. 15. System's reliability for different values of *k* (improved system).

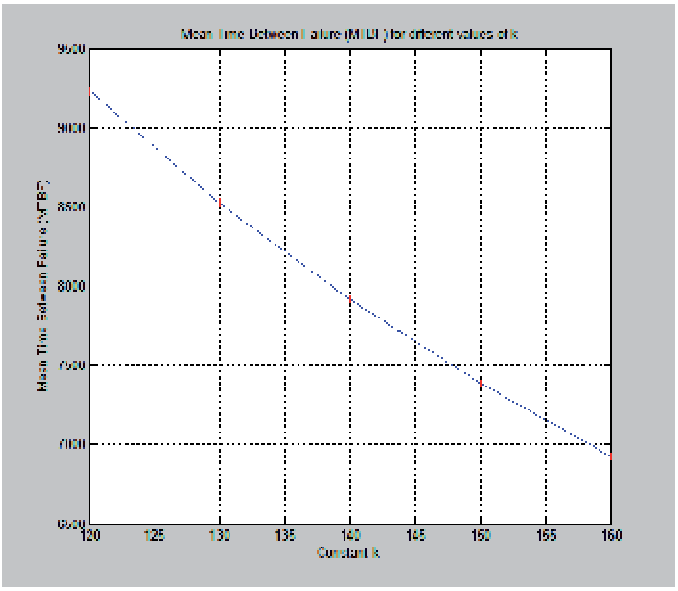


Fig. 16. *MTBF* for different values of *k* (improved system).

A comparative presentation of the two systems' reliability for different values of *k* is depicted in Figure 17 (for initial system with blue lines and red for the improved system).

For the five analyzed values of coefficient *k,* the improved electric supply with a redundant (reserve) subsystem is characterized by superior values of *MTBF* compared to the initial system (fig.18).

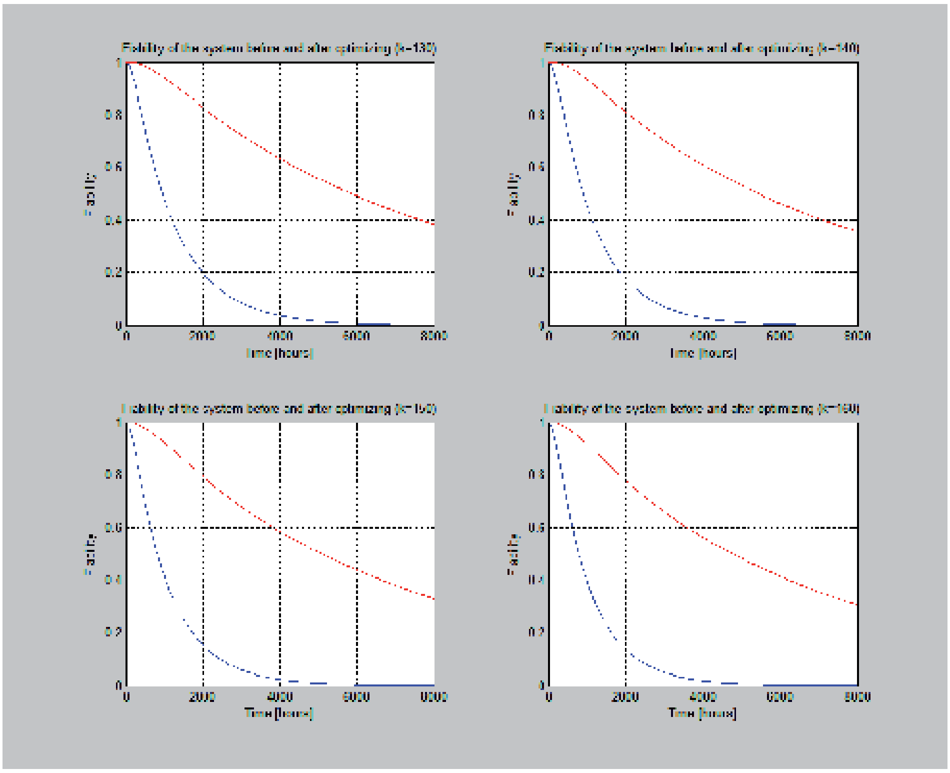


Fig. 17. Comparative analysis of the two systems' reliability for different values of *k*.

In Figure 18 the evolution of *MTBF* for the initial system is represented by a dashed line, while the evolution of *MTBF* for the improved system is represented by a continuous line.

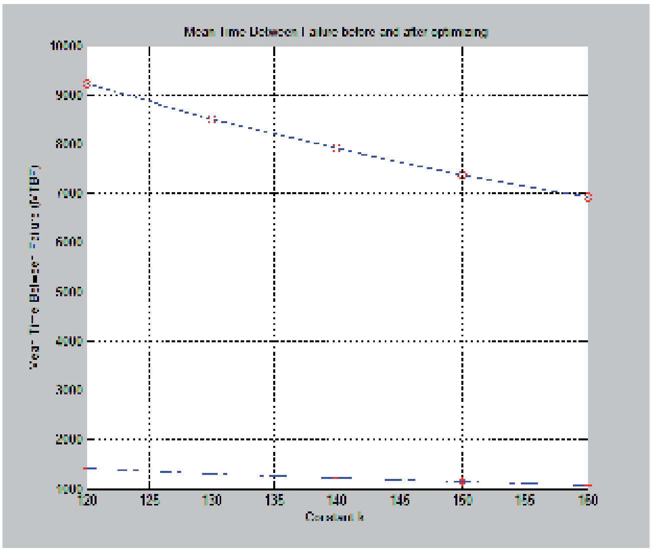


Fig. 18. Evolutions of *MTBF* for the two systems.

3.5 Conclusions regarding the Boolean method

From the analyzed examples and then results obtained for MTBF, we can conclude that the method can be successfully used in the aircraft industry for determining the reliability of the electrical systems. The *MTBF* influencing parameters in the main system nodes (power supply bars and distribution panels) can be calculated and compared.

Through the failure related logic function analysis we can determine the circuits that can improve the system reliability. In the case presented, through the introduction of the components 60E, 61E and corresponding contacts, substantial increase of the reliability (approximately 6 times higher) was obtained for the 28V DC power supply bar.

We have conducted a complex analysis of the influence of the maintenance and way-of-use coefficient *k* on system failure probability, system's reliability and *MTBF.*

4. References

Jula, N. (1986). Contribuții la optimizarea circuitelor electrice de la bordul avioanelor militare. PhD Thesis, Bucureşti, Romania

Moisil, G. (1979). *Teoria algebrică a mecanismelor automate*. Ed. Tehnică, Bucharest

Drujinin, C.V. (1977). *Nadejnot aftometizirovannijh - Sistem*, Energhia, Moskva

Aron, I.; Păun, V. (1980). *Echipamentul electric al aeronavelor*, Editura Didactică şi Pedagogică, Bucureşti, Romania

Hoang Pham (2003). *Handbook of Reliability Engineering*. Springer Verlag

Mathur ,F.P.; De Sousa, P.T.Reliability modeling and analysis of general modular redundant systems, *IEEE Trans.Reliab*. 1975, 24, 296-9

Hohan, I. (1982). *Fiabilitatea sistemelor mari*, E.D.P., Bucharest, Romania

Gnedenko, B.; (1995). *Probabilistic reliability engineering*. New York, John Wiley & Sons

Reus, I. (1971). *Tratarea simbolică a schemelor de comutație*. Ed. Academiei, Bucharest

Muzi, F. Real-time Voltage Control to Improve Automation and Quality in Power Distribution. *WSEAS Transactions on Circuits and Systems,* Issue 6, Vol. 7, 2008

Levitin, G.;Lisnianski, A.; Ben Haim, H.; Elmakis, D. Redundancy optimization for series-paralell multi-state systems. *IEEE Trans. Reliab*. 1998, 47(2), 165-72

Levitin, G.;Lisnianski, A.; Elmakis, D. Structure optimization of power system with different redundant elements. *Electr. Power Syst. Res*. 1997, 43, 19-27

Denis-Papin, M.; Malgrange, Y. (1970), *Exerciții de calcul boolean cu soluțiile lor,* Ed. Tehnică, Bucharest, Romania

Jula, N.; Cepisca ,C.; Lungu, M.; Racuciu, C.; Ursu, T.; Raducanu, D. Theoretical and practical aspects for study and optimization of the aircrafts' electro energetic systems, *WSEAS Transactions on Circuits and Systems,* 12, Vol. 7, 2008, pp.999-1008

Chern, C.S., Jan, R.H. Reliability optimization problems with multiple constraints. *IEEE Trans. Reliab.*,1986,R-35, 431-6

Lyn, M.R. (1996). *Handbook of software reliability engineering*, New York, McGraw-Hill Billinton, R; Allan, R.N. (1996). *Reliability evaluation of power systems*, 2nd ed., New York, Plenum Press

Hecht, H. (2004). *System Reliability and Failure Prevention*, Artech House, London

Part 4

Aircraft Inspection and Maintenance

17

Automatic Inspection of Aircraft Components Using Thermographic and Ultrasonic Techniques

Marco Leo
Consiglio Nazionale delle Ricerche- Istituto di Studi sui Sistemi Intelligenti per l'Automazione Italy

1. Introduction

Safety in aeronautics could be improved if continuous checks were guaranteed during the in-service inspection of aircraft. However, until now, the maintenance costs of doing so have proved prohibitive. In particular, the analysis of the internal defects (not detectable by a visual inspection) of the aircraft's composite materials is a challenging task: invasive techniques are counterproductive and, for this reason, there is a great interest in the development of non-destructive inspection techniques that can be applied during normal routine tests.

Non Destructive Testing & Evaluation (NDT & E) techniques consist of a data acquisition phase (based on any scanning method that does not permanently alter the article being inspected) followed by a data analysis phase carried out by qualified personnel. In particular, transient thermography and ultrasound analysis are two of the most promising techniques for the analysis of aircraft composite materials (Hellier, 2001).

Non-destructive evaluation requires an excessive amount of money and time and its reliability depends on a multitude of different factors. These range from physical aspects of the technology used (e.g., wavelength of ultrasound) to application issues (e.g. probe coupling or scanning coverage) and human factors (e.g. inspector training and stress or time pressure during inspection) (Kemppainen. & Virkkunen, 2011).

Most of the work in the literature concentrates on the study of data acquisition and manipulation processes in order to prove the relationship between data and structural defects or composition of the material (Chatterjee et al., 2011). Unfortunately only some of the work from the literature concentrates on the posterior analysis of the acquired data in order to (fully or partially) delegate, to some computational algorithm, the automatic recognition of material composition, operative conditions, presence of defects, and so on. This is undoubtedly a very attractive research field since it can reduce operational costs, save time and make the process independent from human factors. However, the development of proper algorithms and methodologies is in its infancy and their level of inspection reliability is still inadequate for those sectors (namely, transportation) where an error can have serious health and safety consequences.

The pioneering work on the a posteriori analysis of data dates back to the early 1990s: it suggested that solutions to the problem of automatic ultrasonic NDT data interpretation could be found by expert systems which embody the knowledge of human interpreters (McNab & Dunlop, 1995) (Hopgood et al., 1993) (Avdelidis et al., 2003) (Meola et al., 2006) (Silva et al., 2003). More effective approaches, based on advanced signal processing and artificial intelligence paradigms, have been proposed in the last decade (Benitez et al., 2009) (Wang et al., 2008).

In this chapter, we address the problem of developing an automatic system for the analysis of sequences of thermographic images and ultrasonic signals to help safety inspectors in the diagnosis of problems in aircraft components in all those cases where the defects or the internal damage are not detectable with a visual inspection. In particular thermographic analysis is proposed to automatically discover water insertions whereas ultrasonic inspection aims at revealing solid insertions of brass foil.

The proposed approach considers two main steps for interpreting thermographic and ultrasonic data: in the first step a pre-processing technique is introduced to clean data from noise and to emphasise embedded patterns and the classification techniques used to compare ultrasonic signals and to detect classes of similar points. In the second step two neural networks are trained to extract the information that characterises a range of internal defects starting from ultrasonic and thermographic signals extracted in correspondence to the defective areas. After that the same neural networks are applied to automatically inspect real aircraft components.

Section 2 gives an overview of the proposed approach whereas section 3 and 4 concentrate on the data pre-processing and classification respectively. Finally, section 5 presents the experimental results on real aircraft material and conclusions are derived in section 6.

2. Overview of the system

The proposed system for automatic inspection of aircraft components is schematized in figure 1. The system takes the data extracted by non destructive processes reported in the literature as transient thermography and ultrasound scanning as input.

Transient thermography is a non-contact technique, which uses the thermal gradient variation to inspect the internal properties of the investigated area. The materials are heated by an external source (lamps) and the resulting thermal transient is recorded using an infrared camera. Of course, this kind of analysis is only applicable to materials that have a good thermal conductivity such as metals and carbon composites. Different types of thermal excitation can be used according to the materials and the defects under investigation: for instance uniform heating, spot heating, and line heating.

Ultrasonic inspection uses instead sound signals at frequencies beyond human hearing (more than 20 kHz) to estimate some properties of the irradiated material by analyzing either the reflected (reflection working modality) or transmitted (transmission working modality) signals. A typical ultrasonic inspection system consists of several functional units: pulser, receiver, transducer, and display devices. A pulser is an electronic device that can produce a high-voltage electrical pulse. Driven by the pulser, the transducer generates a high-frequency ultrasonic wave which propagates through the material. In the transmission

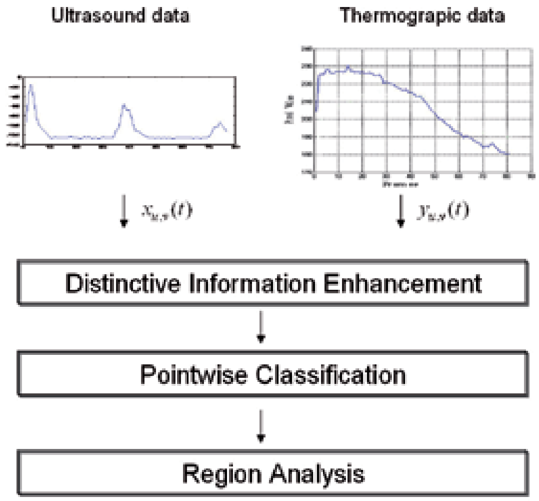


Fig. 1. Scheme of the proposed framework.

modality, the receiver is placed on the opposite side of the material from the pulser, whereas, in the reflection modality, the pulser and the receiver are placed on the same side of the material.

Ultrasonic data can be collected and displayed in a number of different formats. The three most common formats are known in the NDT community as A-scan, B-scan, and C-scan presentations. Each presentation mode provides a different way of looking at and evaluating the region of material being inspected. On the one hand, thermographic analysis is carried out to automatically discover water insertions whereas ultrasonic inspection aims at revealing solid insertions of brass foil.

For thermographic inspection we analyze mono-dimensional signals obtained by considering the time variation of each pixel in the sequence of thermographic images. For each point (i,j) of the material the mono-dimensional signal is generated from the gray levels of the same point in the sequence of images: this signal represents the temperature variation of the material during and after the heating process. This way it is possible to generate spatial-time variant images, the analysis of which allows for the evaluation of the thermal gradient during the heating process.

In Figure 2, the one-dimensional signals extracted from the thermographic sequence of aircraft fuselage are shown: one point belongs to an area affected by the presence of water (red line) whereas the other signal corresponds to non-defective areas (gray lines). From the graph it is clearly evident that a functional description of the intensity variations cannot be easily generalized and the behaviours of points corresponding to defective and non-defective areas are very similar.

For the analysis of ultrasonic data we analyze one-dimensional signals acquired from the reflection working modality and A-scan representation. This means that, for each point of the inspected material, we have a continuous signal that represents the amount of received ultrasonic energy as a function of time.

In figure 3 two ultrasound signals are shown. The signal on top is relative to a non-defective point. Observe that there are large extrema at the beginning and at the end. These changes in ultrasound energy are caused by the transmitted signals being reflected by the boundaries of the material. These boundary extrema are referred to as tool side and bag side peaks, respectively. The ultrasonic signal for an area of material that contains defects is given on the bottom of figure 3. In addition to the boundary extrema, the signals contain extrema at other time locations caused by defective components. The time localization of the additional extrema depends on the defect location in the inspected material.

The temporal evolution of the thermographic and ultrasound signals x(t) is the input to the core of the proposed approach that consists of two main steps: the pre-processing of the data, in order to emphasize the characteristics of the signals belonging to the same class, and the following neural classification.

Pre-processing step allows to discard noise and to enhance the most relevant information for flawed area detection purposes. Two Multi Layer Perceptron (MLP) neural architectures characterized by the presence of an input layer of source nodes, a hidden layer and an output layer, are then used to build an inspection framework that automatically label each signal as belonging to a flawed area or not.

A final connectivity analysis of all the points labelled as belonging to flawed areas is done in order to both discard isolated false positives and to deduce size and shape of the flawed area as a whole.

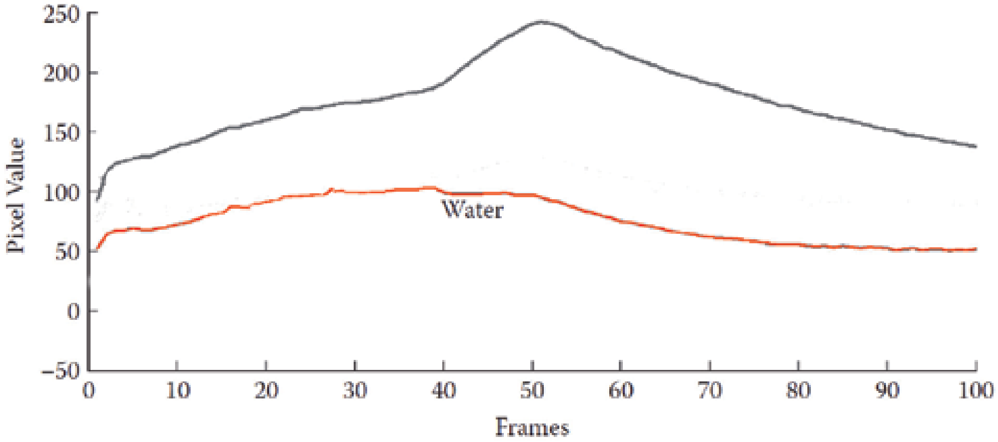


Fig. 2. the one-dimensional signals extracted from the thermographic sequence of aircraft fuselage. The black line corresponds to unflawed areas whereas the red line corresponds to a pixel belonging to water infiltration.

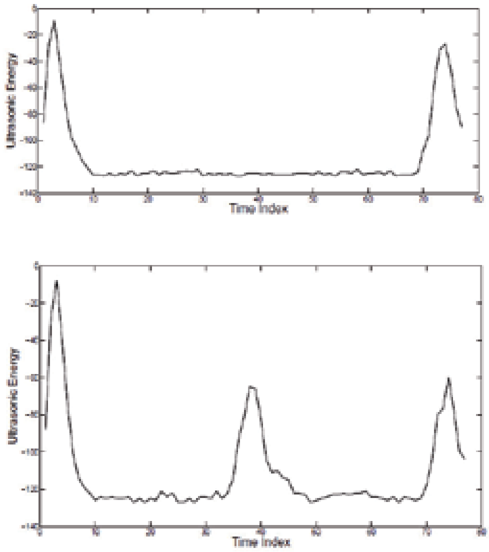


Fig. 3. Two ultrasound signals: the signal on top is relative to a non-defective point. The signal on the bottom is relative to a flawed area.

3. Data pre-processing

The automatic classification of acquired signals as flawed or unflawed is not trivial due to the huge number of intra-class variance: on the one hand ultrasonic and thermal signals relative to unflawed areas can shows different temporal behaviours depending on manufacturing variations in the underlying composite layers or specimen thickness variations. This is evident in figure 4 where different thermographic signals relative to unflawed areas are reported. On the other hand, signals relative to flawed areas can differ since insertions and infiltrations can occur at different locations.

In order to make the classification easier, a pre-processing technique step is then required: on the one hand, it has to increase signal to noise ratio and, on the other hand, to detect and enhance the information that could increase the probability of separating signals belonging to different classes.

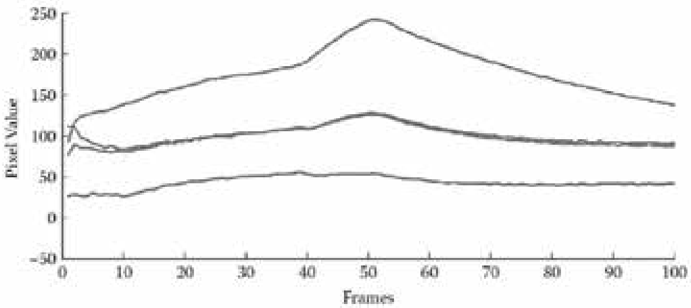
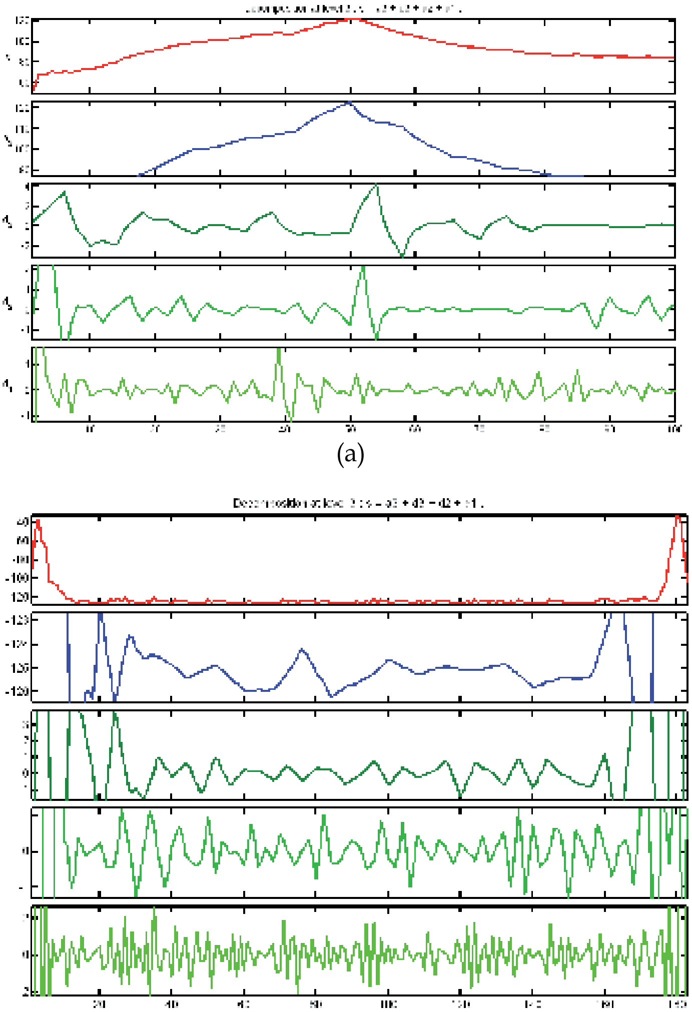


Fig. 4. thermographic signals relative to unflawed areas: their temporal behaviours can strongly differ depending on many factors.

There are many effective signal pre-processing techniques in the literature. Most of them work in a specific domain (time or frequency) whereas a few of them affect both domains simultaneously. In the latter category lies the so-called Wavelet Transform, an extension of Fourier Transform generalized to any wideband transient. For its capability to give a multi-domain representation of the data, the wavelet transform has been used in this work to analyse collected thermographic and ultrasonic data.

In figure 5 the wavelet decomposition (by using Daubechies 3 kernels) at level 3 using a thermographic (top) and ultrasound (bottom) signal are reported.

The next subsection gives some additional theoretical information about the considered pre-processing technique based on Wavelet Transform.



(a)

(b)

Fig. 5. the wavelet decomposition of a thermographic (on top) and ultrasound (at the bottom) signal.

3.1 Wavelet transform

Let us think about our input as a time-varying signal. To analyze signal structures of very different sizes, it is necessary to use time-frequency atoms with different time supports. The wavelet transform decomposes signals over dilated and translated wavelets Mallat (1999). The signal may be sampled at discrete wavelength values yielding a spectrum. In continuous wavelet transform the input signal is correlated with an analyzing continuous wavelet. The latter is a function of two parameters such as scale and position. The widely used Fourier transform (FT) maps the input data into a new space, the basis functions of which are sins and cosines. Such basis functions are defined in an infinite space and are periodic, this means that FT is best suited to signal with these same features. The Wavelet transform maps the input signal into a new space which basis functions are usually of compact support. The term wavelet comes from well- localized wave-like functions.

In fact, they are well-localized in space and frequency i.e. their rate of variations is restricted.

Fourier transform is only local in frequency not space. Furthermore, Fourier analysis is unique, but wavelet not, since there are many possible sets of wavelets which one can choose.

Our trade-off between different wavelet sets is compactness versus smoothness. Working with fixed windows as in the Short Term Fourier Transform (STFT) may bring about problems. If the signal details are much smaller than the width of the window they can be detected but the transform will not localize them. If the signal details are larger than the window size, then they will not be detected properly. The scale is defined by the width of a modulation function. To solve this problem we must define a transform independent from the scale. This means that the function should not have a fixed scale but should vary. To achieve this, we start from a function $\psi(t)$ as a candidate of a modulation function and we can obtain a family starting from it by varying the scale s as follows:

$$\psi_{s,t}(u) = \psi_s(u\ t) = |s|^p \psi(\frac{u\ t}{s}) = \frac{1}{|s|^p}\psi(\frac{u\ t}{s})$$

If ψ has width T then the width of ψ_s is sT . In terms of frequencies, the smaller the s the higher the frequencies ψ_s and vice versa.

The continuous wavelet transform $\tilde{X}$ is the result of the scalar product of the original signal $x(t)$ with the shifted and scaled version of a prototype analysing function $\psi(t)$ called mother wavelet which has the characteristic of a band pass filter impulse response.

The coefficients of the transformed signal represent how closely correlated the mother wavelet is with the section of the signal being analyzed. The higher the coefficient, the more the similarity.

Calculating wavelet coefficients at every possible scale is a fair amount of work, and it generates a great amount of data. If we choose scales and positions based on the power of two (called dyadic scales and positions) then our analysis will be much more efficient. This analysis is called the *discrete wavelet transform*.

In the discrete case, WT is sampled at discrete mesh points and using smoother basis functions. This way a multiresolution representation of the signal $x(t)$ can be achieved.

Notice that the wavelet transform can be written as a convolution product (it is a linear space-invariant filter):

$$\tilde{X}(s,t) = x(u)\psi_{s,t}(u)du = \left\langle \psi_{s,t}, x \right\rangle$$

This leads to a fast and efficient implementation of the wavelet transform for a discrete signal obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut off frequencies at different scales. The wavelet transform for a discrete signal is computed by successive low-pass and high-pass filtering of the discrete time-domain signal. Many filter kernels can be used for this scope and the best choice depends on the features of the input signal that have to be exploited.

At each decomposition level, the half-band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. At the same time, the decimation by 2 doubles the scale. With this approach, the time resolution becomes arbitrarily good at high frequencies, whereas the frequency resolution becomes arbitrarily good at low frequencies.

4. Automatic learning and classification of defective and non-defective patterns

After the pre-processing step the new wavelet based data representations is given as input to an automatic classifier that, after a proper learning phase, is able to label each input stream as belonging to a flawed or unflawed area on the basis of the learned input/output mapping model. One of the most powerful data modelling tools that is able to capture and represent complex input/output relationships is neural network (NN).

4.1 Neural network paradigm

The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. Neural networks resemble the human brain in the following two ways:

1. A neural network acquires knowledge through learning.
2. A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights.

The true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modelled. Traditional linear models are simply inadequate when it comes to modelling data that contains non-linear characteristics.

The most common neural network model is the multilayer perceptron (MLP), having an architecture as reported in figure 6. This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown.

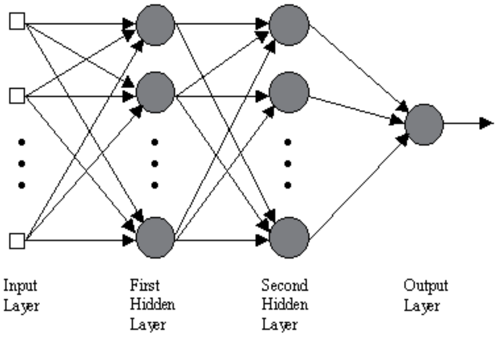


Fig. 6. A feed forward neural network scheme.

The MLP and many other neural networks learn using an algorithm called back propagation. With back propagation, the input data is repeatedly presented to the neural network. With each presentation the output of the neural network is compared to the desired output and an error is computed. This error is then fed back (back propagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output. This process is known as "training".

The hidden layers enable the network to extract higher-order statistics especially when the size of the input layer is large. There is no theoretical limit to the number of hidden layers but, typically, architectures with just one hidden layer are adequate to face the complexity of most of the practical problems . Most used neural architecture have only one hidden layer. Supervised learning involves applying a set of training examples to modify the synaptic weights connecting the neurons of the network. Each example consists of a unique input signal and the corresponding desired response. The network is presented with many examples many times and the synaptic weights are tuned so as to minimize the difference between the desired response and the actual response of the network. The network training is repeated until a steady state is reached, where there are no further significant changes in the synaptic weights.

The input layer has a number of neurons equal to the number of image features. In this work, the features are those extracted after the pre-processing phase. The number of nodes in the output layer depends on the number of classes that the network has to recognize. In our context the network has to recognize the sound point and the defect points (2 output nodes). The number of nodes in the hidden layer is determined by experiment.

There is no quantifiable best answer to the layout of the network for any particular application. There are only general rules picked up over time and followed by most researchers and engineers applying this architecture to their problems.

Rule One: As the complexity in the relationship between the input data and the desired output increases, the number of the processing elements in the hidden layer should also increase.

Rule Two: If the process being modelled is separable into multiple stages, then additional hidden layer(s) may be required. If the process is not separable into stages, then additional layers may simply enable memorization of the training set, and not a true general solution effective with other data.

Rule Three: The amount of training data available sets an upper bound for the number of processing elements in the hidden layer(s). To calculate this upper bound, use the number of cases in the training data set and divide that number by the sum of the number of nodes in the input and output layers in the network. Then divide that result again by a scaling factor between five and ten. Larger scaling factors are used for relatively less noisy data. If you use too many artificial neurons the training set will be memorized. If that happens, generalization of the data will not occur.

5. Experimental setup and results

The composite material used in the experimental tests has an alloy core with a periodic honeycomb internal structure of 128-ply thicknesses (each ply has a thickness of 0.19 mm). The experiments were carried out on two specimens: the first one presents two water infiltrations whereas the second one presents three solid insertions of brass foil (0.02±0.01 mm thickness). One solid insertion was placed two plies from the tool side surface (TOP INSERTION), one at mid part thickness (MIDDLE INSERTION) and the remaining one two plies from the bag side surface (BOTTOM INSERTION). Brass inserts were introduced to represent voids and delamination. In all the cases the defects or the internal damage were not detectable with a visual inspection.

Figure 7 shows the specimens of sandwich material used in the experiments with the graphical information superimposed indicating the exact location of water infiltrations (in blue) and brass foil insertions i.e. top insertion (T) on the left, middle insertion (M) in the centre and bottom insertion (B) on the right.

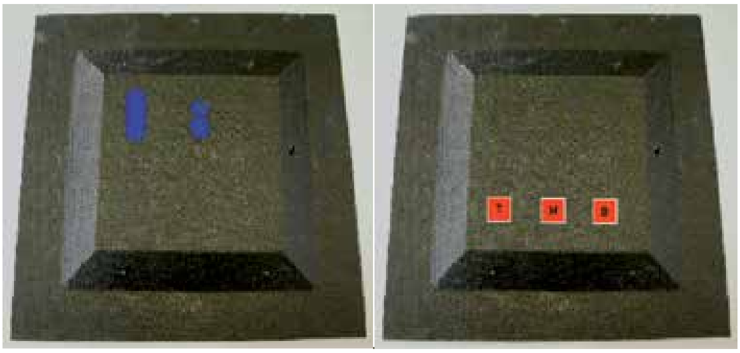


Fig. 7. the sandwich materials used in the experiments with the superimposed graphical information indicating the exact location of water infiltrations and brass foil insertions

The thermographic image sequence was obtained by using a thermo camera sensitive to the infrared emissions. A quasi-uniform heating was used to guarantee a temperature variation of the composite materials around 20C/sec. In figure 8 one of the thermographic images is reported. Only liquid infiltrations become visible due to the larger thermal variation of the water with respect to solid insertions.

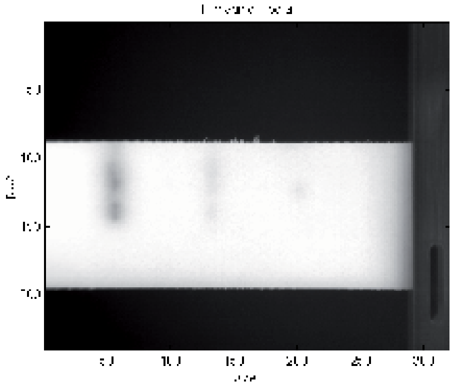


Fig. 8. One of the thermographic images where the liquid infiltrations are visible.

Ultrasonic data were obtained by an ultrasonic reflection technique that uses a single transducer serving as transmitter and receiver (5MHz).

In figure 9, the signal on the left is relative to a non-defective area whereas the signal on the right is relative to a brass insertion placed in the middle of the material thickness and for this reason the corresponding extrema is far from the boundary ones. The signal in the centre of figure 3 is relative to a brass insertion placed very close to the inspected material surface and then the corresponding extrema is mixed with the tool side one. This shows that defective and non-defective areas can have very similar temporal behaviours under ultrasound scanning and this causes traditional NDT techniques to fail.

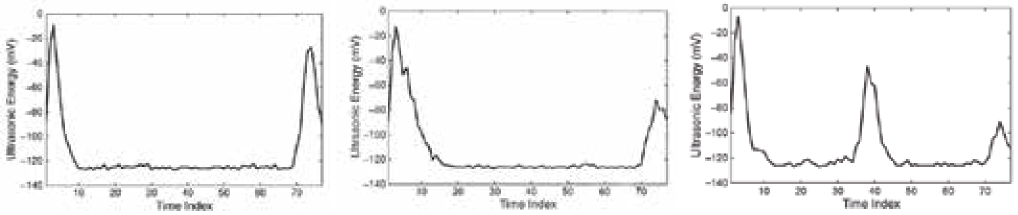


Fig. 9. three ultrasound signals relative to the non-defective area (on the left), a brass insertion placed in the middle of the material thickness (on the right) and very close to the inspected material surface (in the centre).

Acquired experimental data were then represented in the wavelet domain by using Daubechies 3 family of filters and the derived coefficients were given as input to two different neural networks in order to specialize each of them to recognize water infiltration and solid insertion respectively. The defect segmentation step is performed by using neural networks with two output neurons. Each available signal is fed into the net, which classifies it as either relative to defective areas or an unflawed area.

Preliminary experiments aimed at defining the best data model through the selected neural paradigm. In particular they allow the definition of the best number of neurons in the hidden layer and the most suited number of training points. To accomplish this fundamental task different set training examples were built. In particular, for each neural network 3 different training sets consisting of 40, 60 and 80 examples (50% corresponding to unflawed and 50% to flawed areas) were used. At the same time different test sets of points were built for each specimen. In particular, for the specimen with water infiltration two data were built: the first set contained 250 signals relative to unflawed points, the second set contained 250 signals relative to defective areas damaged by the water.

Similarly for the specimen with brass foil insertions 4 data sets were built: the first set contained 250 signals relative to unflawed points, the second set contained 250 signals relative to the defective area corresponding to the brass foil positioned two plies from the tool side surface (Top Insertion), the third set contained 250 signals relative to the defective area corresponding to the brass foil positioned at mid part thickness (Middle Insertion) and finally the fourth set contained 250 signals relative to the defective area corresponding to the brass foil positioned two plies from the bag side surface (Bottom Insertion).

In each experiment a training set was selected and the learned network was then used to classify the data in the corresponding test set. The set of training examples consisted of input–output couples (input signal, corresponding desired response). During the training phase the points of known examples were extracted from the considered materials and continuously fed into the net so that the synaptic weights were tuned to ensure the minimum distance between the actual and the desired output of the net.

Training continues until a steady state is reached, i.e., no further significant change in the synaptic weights could be made to improve net performance. This is repeated also using different configurations of the hidden layer. In particular a number of hidden neurons ranging from 20 to 100 were considered.

The results of this demanding experimental phase are summed up in figure 10 and figure 11.

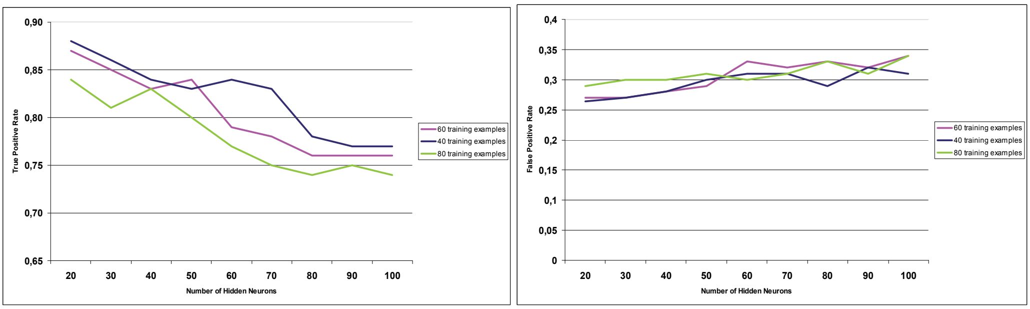


Fig. 10. Experiment results for water infiltration detection using thermal signals when a different number of training examples and hidden neurons were considered

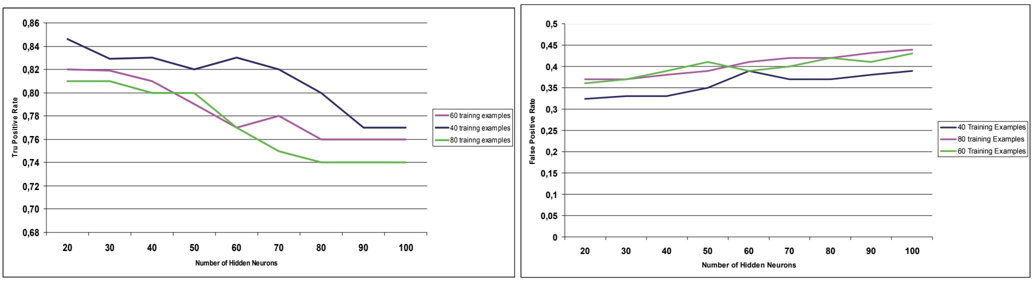


Fig. 11. Experiment results for solid insertion detection using ultrasound signals when a different number of training examples and hidden neurons were considered

Experiments demonstrated that a lower number of hidden layer nodes (i.e 20-30) is a good choice since a larger number of nodes in the inner layer for such situations can drive the classification model to over-fit the training data and to produce a very high failure score. At the same time, experimental proofs pointed out that a limited number of training points (i.e. 40) is the best choice in term of correct classification rate. In other words this is the minimum number of training examples to allow a proper learning of the data distribution and, at the same time, it is the maximum number to avoid data over-fitting case, i.e. to preserve the fundamental capability to classify unknown data (generalization capacity) .

For a better comprehension of experimental results tables I and II report the scatter matrices relative to the experiments performed by using the best network and training set configuration.

	Unflawed	Water insertion
Unflawed	**184/250 (73,6%)**	66/250 (26,4%)
Water insertion	28/250 (11,2%)	**222/250 (88,8%)**

Table I: Scatter matrix derived in the experiments for water infiltration detection.

	Unflawed	Solid Insertion
Unflawed	**169/250 (67,6%)**	81/250 (32,4%)
Brass Foil (top)	36/250 (14,4%)	**214/250 (85,6%)**
Brass Foil (middle)	15/250 (6,0%)	**235/250 (94,0%)**
Brass Foil (bottom)	64/250 (25,6%)	**186/250 (74,4%)**

Table II. Scatter matrix derived in the experiment 1 for brass fail insertion detection.

Table I and II give a quantitative evaluation of the possibility to automatically detect both liquid and solid insertions in composite materials by using thermal and ultrasonic techniques in combination with neural approaches.

In particular, Table II illustrates that brass foil insertions at the mid-thickness level were always better classified than those located either at the top or at the bottom. The defect location is one of the most important factors in ultrasound inspection. The defects placed either at the top or at the bottom of the inspecting structure are in general the most difficult to detect since their echo is mixed with the tool face or the bag side echo. On the contrary, defective areas in the mid-part of the material thickness produce a distinct peak in the signal trend that is straightforward to identify.

In the second part of the experimental phase all the signals extracted by the thermographic and ultrasonic analysis were classified by using the neural networks previously learned.

According to the neural network outputs, a binary image is produced containing black points for defective areas and white points for sound areas.

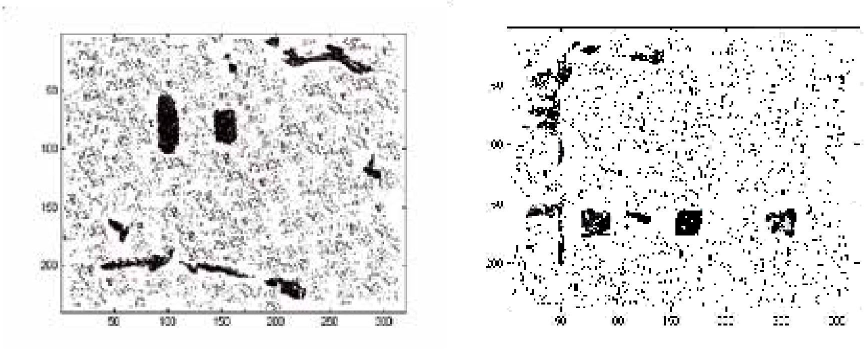


Fig. 12. graphical representation of the raw classification of all the signals extracted from specimens with water infiltration (on the left) and brass foil insertions (on the right).

In figure 12 the graphical representation of the raw classification of all the signals extracted from specimens with water infiltration (on the left) and brass foil insertions (on the right) is reported. Defective areas are correctly detected but there are also a lot of points in the unflawed areas erroneously classified as flawed. For this reason an additional processing step was introduced in order to analyse the output images considering the vicinity of flawed pixels (region analysis). In other words, considering that these false detections were isolated and did not form connected regions having a considerable area value, the elimination of these points was made more straightforward if some a priori knowledge about the minimum expected size of the defective areas is available.

In figure 13 the final outcome is reported after a filtering process based on the connectivity analysis of the detected defective regions and a selection criterion based on removing the regions having an area less than 20 pixels, are shown.

Most of the false flawed points were removed even if some areas in addition to the real defects were still considered flawed. They mainly occurred in correspondence with a variation of the inclination of the surface (see fig. 7): unfortunately, in this unflawed area both thermographic and ultrasound signals changed their slope more evidently with respect to the corresponding signals used to train the net. This problem could be faced by learning the net also on the points belonging to this particular areas. However, this way of proceeding was not considered in this work since it could be counterproductive: the net could miss some real defective areas (or parts of it) and, in our opinion, considering the applicative context, it is critically important to detect all defective points, even at the expense of generating extra false positives

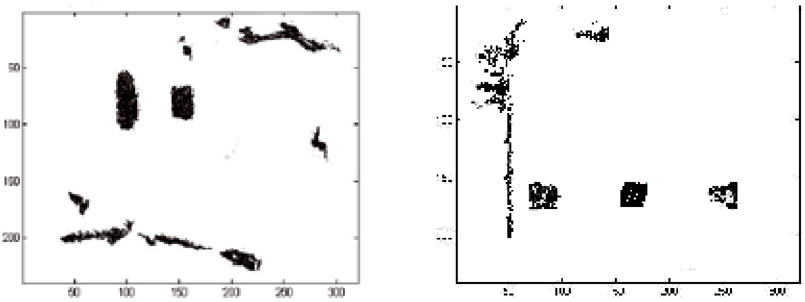


Fig. 13. the result of the cleaning based on the point connectivity analysis on the images reported in figure 8.

6. Conclusion

In this chapter, we address the problem of developing an automatic system for the analysis of sequences of thermographic images and ultrasonic signals to help safety inspectors in the diagnosis of problems in aircraft components.

In particular, thermographic analysis was carried out to automatically discover water insertions whereas ultrasonic inspection aimed at revealing solid insertions of brass foil. Experiments were carried out on real aircraft specimens and demonstrated the capability of the proposed framework to discover flawed areas. A tolerable number of false positive occurrences were also found in correspondence to the part of the specimens having a sloping surface since their points were not included in the learning phase in order to get the best true positive detection rate considering the critical operative context.

Future work will focus on investigating the defect identification capability of the proposed approach. This will be achieved by extending the analysis to material with different thicknesses and different defective insertions. In the future, we will also investigate the possibility of using an unsupervised-learning approach in order to reduce human intervention.

7. References

Kemppainen, M. & Virkkunen, I. (2011). Crack Characteristics and Their Importance to NDE, *Journal of Nondestructive Evaluation,* 2.06.2011 Issn: 0195-9298 Available from http://dx.doi.org/10.1007/s10921-011-0102-z

Chatterjee, K. ; Tuli, S. ; Pickering, S. G. & Almond, D. P. (2011). A comparison of the pulsed, lock-in and frequency modulated thermography nondestructive evaluation techniques, *NDT & E International,* 29.06.2011 Issn 0963-8695 Available from http://www.sciencedirect.com/science/article/pii/S0963869511000892

McNab, A. & Dunlop, I. (1995). A review of artificial intelligence applied to ultrasonic defect evaluation, *Insight*, vol. 37, no. 1, pp. 11–16.

Hopgood, A. A. ; Woodcock, N. ; Hallani, N. J. & Picton, P. (1993). Interpreting ultrasonic images using rules, algorithms and neural networks, *Eur. J. Nondestruct. Test.*, vol. 2, no. 4, pp. 135–149.

Benitez, H. D. ; Loaiza, H. ; Caicedo, E. ; Ibarra-Castanedo, C. ; Bendada, A. & Maldague, X. (2009). Defect characterization in infrared non-destructive testing with learning machines, *NDT & E International*, Volume 42, Issue 7, Pages 630-643, ISSN 0963-8695.

Wang, Y. ; Sun, Y. ; Lv, P. & Wang, H. (2008). Detection of line weld defects based on multiple thresholds and support vector machine, *NDT & E International*, Volume 41, Issue 7, October 2008, pp. 517-524, ISSN 0963-8695.

Hellier, C. (2001). *Handbook of Nondestructive Evaluation*. McGraw-Hill Professional ISBN: 0070281211

Avdelidis, N. P. ; Hawtin, B. C. & Almond, D. P. (2003). Transient thermography in the assessment of defects of aircraft composites, *NDT & E International*, Volume 36, Issue 6, pp. 433-439, ISSN 0963-8695.

Meola, C. ; Carlomagno, G.M., Squillace A. & Vitiello, A. (2006). Non-destructive evaluation of aerospace materials with lock-in thermography, *Engineering Failure Analysis*, Volume 13, Issue 3, pp. 380-388, ISSN 1350-6307

Silva, M. Z. ; Gouyon, R. & Lepoutre, F. (2003) Hidden corrosion detection in aircraft aluminum structures using laser ultrasonics and wavelet transform signal analysis, *Ultrasonics*, Volume 41, Issue 4, pp. 301-305, ISSN 0041-624X.

18

The Analysis of the Maintenance Process of the Military Aircraft

Mariusz Wazny
Military University of Technology
Poland

1. Introduction

This chapter presents the analysis of the maintenance process of a military aircraft with a detailed description of two areas, i.e. the process of maintaining and the process of operating. Each of these processes is briefly characterized. The section also involves methods enabling the determination of: residual durability of specified devices/systems of a military aircraft on the basis of the diagnostic parameters of these devices/systems, and the effectiveness of a combat task execution on the basis of information registered in the process of aiming. Each presented method is illustrated by a computational example.

2. Tasks executed by the military aircraft

A modern military aircraft (MMA) is a hybrid of the most up-to-date achievements in the field of materials engineering (the use of light metal alloys and composite structures), electronic engineering (fast microprocessor systems, modern systems in the field of power electronics), and specialized software supporting the maintenance process (automatic flight control system, integrated diagnostic systems). Due to such combination, tasks executed by MMA comprise a wide range that can be divided into two groups: with the use of aerial combat means and without the use of aerial combat means.

Depending on the nature of a mission, tasks including the use of aerial combat means can be generally classified as:

1. The gaining and maintenance of domination of airspace. This type of task is executed by fast and manoeuvrable aircrafts that are equipped with the most modern armament for aerial combat, i.e. air-to-air missiles and aircraft guns.
2. The support for the operations of ground forces and the navy. As regards this task, aircrafts equipped with air-to-ground weaponry, including rockets, bombs, and aircraft guns, play an important role.
3. The combating of a selected target of an air attack using precision-guided munitions launched from manned and unmanned aircrafts.

When analyzing the use of MMA in respect of the combat task realization without the use of aerial combat means, we can distinguish the following main tasks:

1. Air reconnaissance performed using both aircrafts equipped with specialized apparatus and unmanned flying objects configured for the performance of this type of a mission.
2. Air transport ensuring fast and efficient transfer of both infrastructure elements and soldiers into the area of a new localization for troops.

The support for the operations of different types of forces by means of, among other things, managing a mission on the basis of spatial information obtained via reconnaissance systems installed, for example, on an AWACS-type platform, or enabling the in-flight refuelling.

The analysis of the operations of the armed forces in recent armed conflicts indicates that MMAs are the basic element of the system of military operations. MMAs are used in the first instance to execute all of the above-mentioned tasks.

3. The organization of the maintenance process of the military aircraft

The technical objects maintenance is defined as a set of intentional organizational and economical operations of the people on the technical objects and the relationships between them from the beginning of the object lifecycle up to the end of lifecycle and object disposal. Relationships recognition and identification of the operations which appear between subjects based on the knowledge and experience of the technical objects designers, developers and engineers. The maintenance compliance and utility of product mainly depends on the engineers and designers crew professional competence. However the design presumptions can be altered many times during object lifecycle. These operations are performed to decrease maintenance "waste effect" and maximize "utility effect".

The modern military aircraft, which is the basic technical object in Polish Air Force organization structure, is the complex product including various constructional, technological, engineering and organizational concepts. Design of so sophisticated product based on tactical and technical military requirements which was created after modern battlefield analysis.

The aircraft construction is based on the module structure (Fig. 1) which allows dividing the specified tasks between separate functional blocks. This solution improves the maintenance process and facilitates service and operational use of the aircraft.

The conditions in which the aircrafts are operated are so specific that involves the specified requirements regarding high level of reliability, durability, effectiveness and safety parameters as far as airborne technology is concerned. Required levels of parameters are provided by determining specified functional structure of devices and specified level of redundancy.

Due to specific character of aircraft operations the aircraft maintenance can be performed only within specified system which provides the conditions indispensable for correct aircraft operation. This specified system is called Air System (AR) and contains the aircraft frame, the people who participate in the maintenance process and the devices building the system which ensure process permanence (in functional way) - Fig. 1.

The primary target in military aircraft maintenance process during peace is maintaining both the technical equipment and the personnel on the specified reliability and training level. It is required to provide high level of efficacy and effectiveness during wartime.

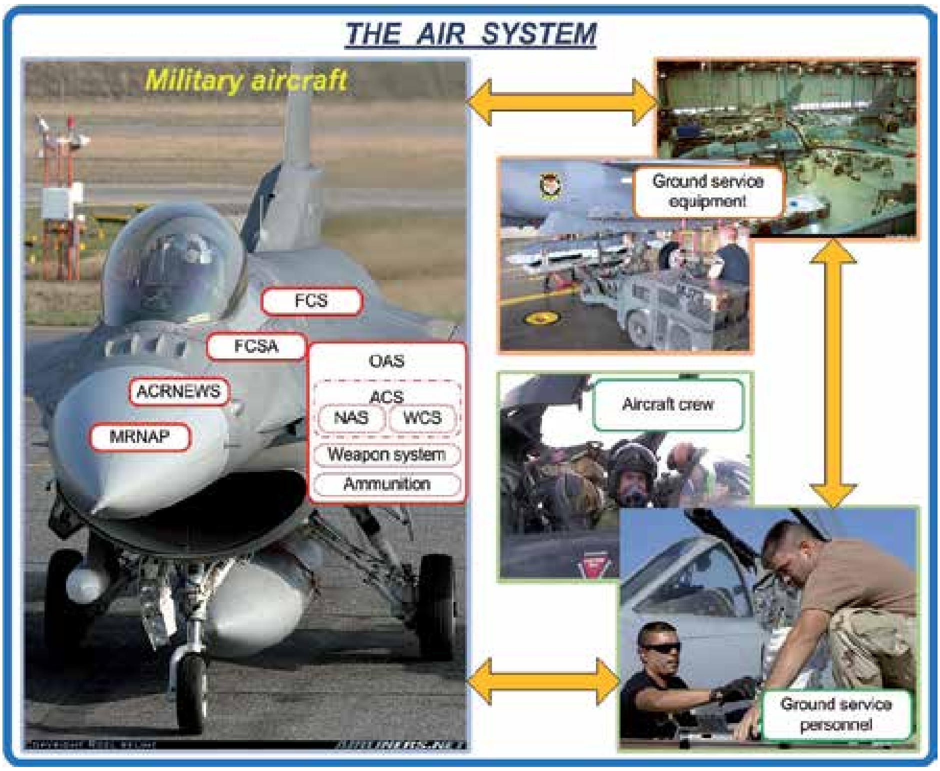


Fig. 1. Structural diagram of the military aircraft and the air system: FCSA – Flight Control System Actuators (frame construction with plating); FCS – Flight Control System; ACRNEWS – Airborne Communication, Radio Navigation and Electronic Warfare Systems.; MRNAP – Multifunctional Radar and Navigation and Aiming Pod; OAS – On-board Armament System; ACS – Armament Control System; WCS – Weapon Control System; NAS – Navigation and Aiming System.

Due to many various external factors, which influence negatively on the specified technical elements of the Air System, it can be claimed, that during operating process the elements are getting "used up". Therefore, due to maintain Air System in the appropriate reliability condition there is required to perform technical service. This action contains adjustment, tuning and replacement of particular devices or whole aggregates, in order to slow down the "using up" process.

In practice there are three aircraft maintenance strategies (Fig. 2.):

1. maintenance system containing prevention services schedule (recurring maintenance).
2. operational maintenance system.
3. preventive/predictive maintenance system.

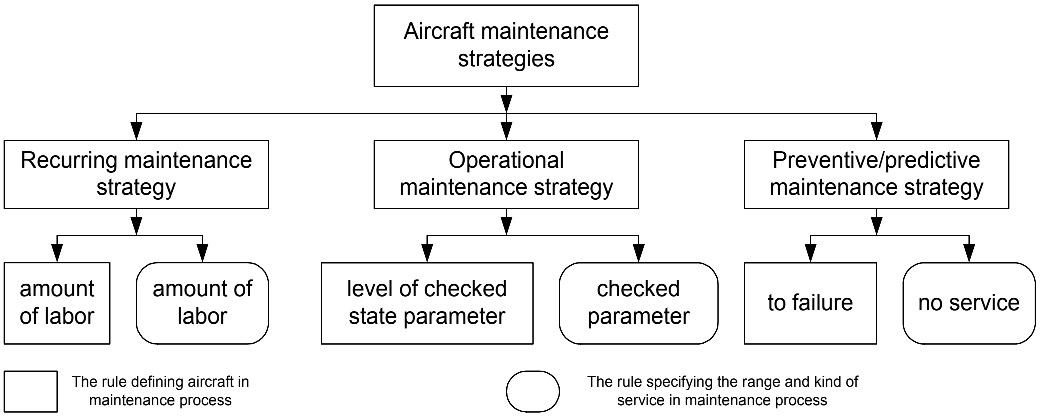

Aircraft maintenance strategies
Recurring maintenance strategy
Operational maintenance strategy
Preventive/predictive maintenance strategy
amount of labor
amount of labor
level of checked state parameter
checked parameter
to failure
no service
The rule defining aircraft in maintenance process
The rule specifying the range and kind of service in maintenance process

Fig. 2. Military aircrafts maintenance strategies.

Organization and scheme of military aircrafts recurring maintenance strategy is presented on Fig. 3. The basis of this maintenance strategy is the measurement of the amount of labor executed by the plant. As far as aircraft is concerned the amount of labor is defined as a number of hours in the sky.

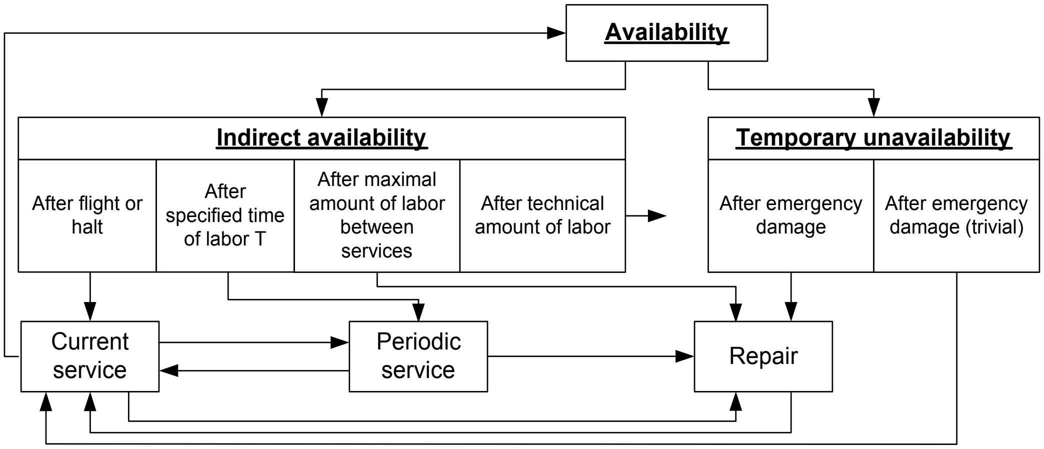

Availability
Indirect availability
After flight or halt
After specified time of labor T
After maximal amount of labor between services
After technical amount of labor
Temporary unavailability
After emergency damage
After emergency damage (trivial)
Current service
Periodic service
Repair

Fig. 3. Recurring maintenance strategy scheme.

One of the maintenance states in the recurring maintenance process is the indirect airworthiness state. The aircraft in this state is mostly working correctly but it lost the flying ability in order to circumstances determined on figure 3. After execution the specified amount of labor (hours of fly) the aircraft lifecycle should be either terminated or directed to the professional service to determine the new amount of labor possible to execute.

As far as operational maintenance strategy is concerned there is the rule the aircraft is in operation as long as the levels of specified parameters do not exceed the specified limits of error. The knowledge about the maintenance state of the device is determining by the external and internal diagnostic equipment. The service operations during this maintenance strategy are executed according to levels of measured diagnostic parameters. The proper control of operational maintenance strategy even for the considerable fleet of aircrafts requires control of every aircraft separately.

The preventive/predictive maintenance strategy defines the reliability as a designed characteristic. The level (value) of the reliability must be provided in the device design and manufacturing process and is maintaining during the device lifecycle. The maintenance schedule which is based on preventive/predictive maintenance strategy provides the desirable or defined levels of both reliability and flight safety. The all of described aircrafts maintenance strategies are followed during the real conditions fleet maintenance process.

Due to the development of diagnostic systems, military aircraft on-board systems include diagnostic procedures enabling the assessment of a current technical state of a given system. The procedure of assessing a given system is performed before an air operation. The procedure results provide information on a technical state of a military aircraft. Based on this information, a pilot decides either to perform a task or to withdraw from performing the task.

Apart from integrated diagnostic systems installed on board, there is a number of devices whose technical state is examined via monitoring and measuring equipment after its disassembly from the board of MMA. During maintenance works, diagnostic parameters of the examined devices are recorded and compared with the range of permissible changes. Any deviation beyond the assumed tolerance limits leads to the implementation of either appropriate maintenance procedures aiming at reducing the resultant deviation or appropriate corrections eliminating the deviation. The ability to predict the service life of MMA when diagnostic parameter tolerance might be exceeded would enable the appropriate management of the maintenance system of MMA. Thus, it is possible to optimize the time when MMA is under certain maintenance works and is not combat ready.

4. The process of maintaining the military aircraft

4.1 The influence of destructive factors on the technical state of devices used on the military aircraft

During the operation process of a military aircraft we can observe the change of technical parameters of selected devices along with the time of their operation. This change causes the deterioration of working conditions of a system and the loss of rated values of technical parameters. Factors influencing the above-mentioned changes include:

- changes of temperature and air-pressure,
- g-forces,
- vibrations,
- ageing process, etc.

The construction of technical systems is based on the assumption that a device fulfils its role when its operational/diagnostic parameters are within acceptable error limits. This assumption depends on the accuracy of work of particular system elements. Thus, in order to assure a faultless functioning of a military aircraft, we cannot allow operational parameters to exceed the acceptable error limits, which can be done in two ways: by frequent checks of operational parameter values of a device/system and its switch off when parameters are close to the fixed limit, or by determining the time after which operational parameters exceed values of the acceptable error.

The first way is onerous with regard to its organization and it is also time consuming and money consuming. Besides, the time spent on checking excludes a military aircraft from its use in a combat task, which consequently leads to a temporal decrease of the fighting efficiency of the air forces.

The second way is based on the use of a particular mathematical method enabling the description of value changes of operational parameters of a device/system and the evaluation of time in which a device/system is in operational state.

It is stated above that military aircrafts undergo changes during the exploitation of operational parameter values of particular devices in avionics system. The changes cause that operational parameter values approximate to the fixed acceptable limit. When parameter values equate with the limit value or exceed it, an adjustment must be done in order to restore nominal conditions of a device/system operation or the operation must be stopped. Figure 4 presents a theoretical course of changes of diagnostic parameter values.

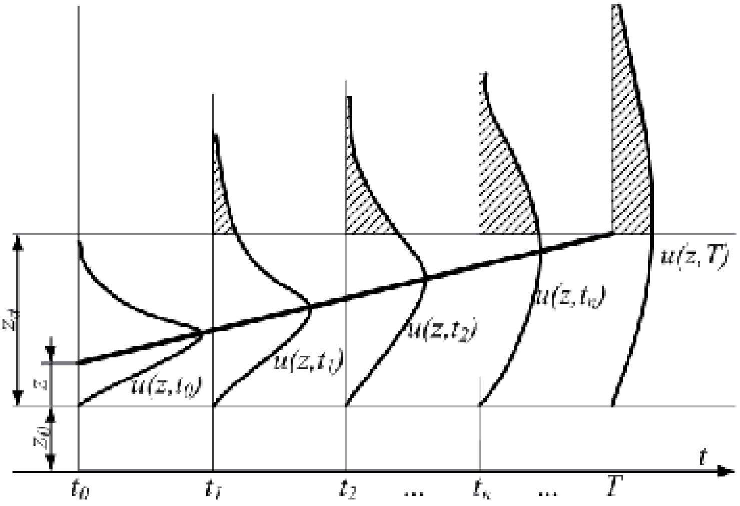


Fig. 4. Diagram of changes of diagnostic parameter values: z_0 – nominal value of a parameter, z – current value of a parameter, z_d – the limit of acceptable changes of parameter values

The second way is based on the use of a particular mathematical method enabling the description of value changes of operational parameters of a device/system and the evaluation of time in which a device/system is in operational state.

4.2 The model of diagnostic parameter changes in the aspect of the occurrence of destructive factors

In the figure, current value of a parameter is marked as "z". If $z < z_d$ then an element is fit for use, but if $z \geq z_d$ the elements losses its operational state. The change of diagnostic parameter values will be of a random character because of a specific character of MA operation process and the influence of destructive processes. So, let's consider "the wear of a device" of avionics system as a random process occurring during the operation of an aircraft.

Getting down to the analytical description of the diagram in Figure 4 and the determination of the density function of the changes of a diagnostic parameter values, the following assumptions were accepted:

1. The technical condition of an element is described by one diagnostic parameter which is marked as „z".
2. The change of the value of the parameter „z" happens only during the operation of a device, i.e. during the flight of an aircraft.
3. The parameter „z" is non-decreasing.
4. The change of the diagnostic parameter „z" is described by the following equation (1).

$$\frac{dz}{dN} = c \tag{1}$$

where:

c - random variable which depends on operational conditions of an element;
N - the number of flights of an aircraft.

1. If $z \in [0, z_d]$ then an element is fit for use, in other case the element is considered as unfit for use.
2. The intensity of flights of an aircraft is described by the following dependence (2).

$$\lambda = \frac{P}{\Delta t} \tag{2}$$

where:

Δt - the range of time in which the flight of an aircraft can be performed with the probability P,
P - the probability of the flight performance within the time interval with length Δt.

The time interval with length Δt shall be selected in such a way as to fulfil the following inequality (3).

$$\lambda \Delta t \leq 1 \tag{3}$$

The intensity of flights λ enables the determination of the number of flights of an aircraft up to the moment t form the following formula:

$$N = \lambda t \tag{4}$$

Using the formula (4), the equation (1) can be written in the following form:

$$\frac{dz}{dt} = \lambda c \tag{5}$$

The dynamics of the changes of a diagnostic parameter can be described by the following difference equation (6).

$$U_{z,t+\Delta t} = (1-\lambda\Delta t)U_{z,t} + \lambda\Delta t\, U_{z-\Delta z,t} \tag{6}$$

where:

$U_{z,t}$ - the probability that in the moment *t* the value of a diagnostic parameter will be *z;*

Δz - the increment of the diagnostic parameter *z* during one flight of an aircraft.

The functional notation of the equation (6) has the following form:

$$u(z,\, t+\Delta t) = (1-\lambda\Delta t)\, u(z,t) + \lambda\Delta t\, u(z-\Delta z, t) \tag{7}$$

where:

$u(z,t)$ - the density function of the probability of the diagnostic parameter value *z* in the moment *t;*

$(1-\lambda\Delta t)$ - the probability that in the time interval with length Δt the flight will not be performed;

$\lambda\Delta t$ - the probability of the flight performance in the time interval with length Δt.

The equation (7) was transformed by substituting the following differential equation (8).

$$\frac{\partial u(z,\, t)}{\partial z} = -\lambda\,\Delta z\, \frac{\partial u(z,\, t)}{\partial z} + \frac{1}{2}\lambda(\Delta z)^2\, \frac{\partial^2 u(z,\, t)}{\partial z^2} \tag{8}$$

where: $\Delta z = c$.

Due to the fact that *c* is a random variable, the following mean value was introduced:

$$E[c] = \int_{c_d}^{c_g} c\, f(c)\, dc \tag{9}$$

where: *f*(*c*) - the density function of the random variable *c;*

c_g, c_u - the limits of variation of *c*.

Taking into consideration the dependence (9), the differential equation (8) can be written in the following form:

$$\frac{\partial u(z,\, t)}{\partial t}\Delta t = -\lambda\, E[c]\, \frac{\partial u(z,\, t)}{\partial z} + \frac{1}{2}\lambda(E[c])^2\, \frac{\partial^2 u(z,\, t)}{\partial z^2} \tag{10}$$

where: $\lambda\, E[c]$ - the mean increment of the parameter value per time unit;

$\lambda(E[c])^2$ - the mean square increment of the value of the diagnostic parameter per time unit.

The solution of the equation (10) is the unknown density function of the probability of the random variable *z* in the following form:

$$u(z,t)=\frac{1}{\sqrt{2\pi\,A(t)}}\,e^{-\frac{(z-B(t))^2}{2A(t)}} \tag{11}$$

where:

$$B(t)=\int_0^t \lambda\,E[c]\,dt=\lambda\,E[c]\,t\,,\quad A(t)=\int_0^t \lambda\big(E[c]\big)^2 dt=\lambda\,E[c]^2\,t \tag{12}$$

Assuming that:

$$b=\lambda\,E[c]\,,\quad a=\lambda\,E[c]^{\,2} \tag{13}$$

the density function (11) has the following form:

$$u(z,t)=\frac{1}{\sqrt{2\pi\,a\,t}}\,e^{-\frac{(z-bt)^2}{2at}} \tag{14}$$

The dependence (14) is the probabilistic characterisation of the increase of the wear in the function of the flying time. However, it is important to know the distribution of the time (the flying time) of the exceedance of the acceptable error value of the parameter *z*.

The probability of the exceedance of the acceptable value by the current value of the diagnostic parameter „*z*" can be written in the following form:

$$Q(t;z_d)=\int_{z_d}^{\infty}\frac{1}{\sqrt{2\pi\,at}}\,e^{-\frac{(z-bt)^2}{2at}}\,dz \tag{15}$$

The density function of the time distribution of the exceedance of the acceptable state z_d has the following form:

$$f(t)=\frac{\partial}{\partial t}\,Q(t;z_d) \tag{16}$$

Thus

$$f(t)=\frac{\partial}{\partial t}\int_{z_d}^{\infty}\frac{1}{\sqrt{2\pi\,at}}\,e^{-\frac{(z-bt)^2}{2at}}\,dz \tag{17}$$

$$f(t)=\int_{z_d}^{\infty}\left\{\frac{\partial}{\partial t}\left[\frac{1}{\sqrt{2\pi\,at}}\,e^{-\frac{(z-bt)^2}{2at}}\right]\right\}dz \tag{18}$$

After calculating the derivative, we obtain:

$$f(t)_{z_d} = \int_{z_d}^{\infty} \left[u(z,t)\left(\frac{z^2 - b^2 t^2 - at}{2\,at^2} \right) \right] dz \tag{19}$$

The original function with regard to the integrand of the dependence (19) has the following form (20).

$$w(z,t) = u(z,t)\left(-\frac{z+bt}{2t} \right) \tag{20}$$

We calculate the integral (19).

$$f(t)_{z_d} = u(z,t)\left(-\frac{z+bt}{2t} \right)\Bigg|_{z_d}^{\infty} = \frac{z_d + bt}{2t} \frac{1}{\sqrt{2\pi\, at}}\, e^{-\frac{(z_d - bt)^2}{2at}} \tag{21}$$

Thus, the dependence (21) determines the density function of the time of the first transition of the current value of the parameter „*z*" through the acceptable state.

Having the above-mentioned data, we can determine the durability of a device with respect to the change of the value of the parameter *z*. For this purpose, we can write down that the formula for the reliability of a device has the following form:

$$R(t) = 1 - \int_0^t f(t)_{z_d}\, dt \tag{22}$$

where the density function $f(t)_{z_d}$ is determined by the formula (21).

The unreliability of a device can be determined from the dependence (23).

$$Q(t) = \int_0^t \frac{z_d + bt}{2t} \cdot \frac{1}{\sqrt{2\pi\, at}}\, e^{-\frac{(z_d - bt)^2}{2at}}\, dt \tag{23}$$

The integral (23) has to be simplified. It can be observed that the integrand can be written in the following form:

$$\frac{z_d + bt}{2t} \cdot \frac{1}{\sqrt{2\pi\, at}}\, e^{-\frac{(z_d - bt)^2}{2at}} = \frac{z_d + bt}{2t} \cdot \frac{1}{\sqrt{2\pi\, at}}\, e^{-\frac{(bt - z_d)^2}{2at}} \tag{24}$$

and now we have to solve the indefinite integral.

$$\int \frac{(z_d + bt)}{2t} \cdot \frac{1}{\sqrt{2\pi\, at}}\, e^{-\frac{(bt - z_d)^2}{2at}}\, dt \tag{25}$$

We make the substitution in the above-mentioned integral.

$$\frac{(bt - z_d)^2}{2at} = u \tag{26}$$

Thus

$$\frac{du}{dt} = \frac{bt + z_d}{2at^2}(bt - z_d) \tag{27}$$

$$dt = \frac{2at^2}{(bt + z_d)(bt - z_d)}\, du \tag{28}$$

After the substitution, the integral (25) has the following form (29).

$$\int \frac{z_d + bt}{2t} \cdot \frac{1}{\sqrt{2\pi at}} e^{-u} \cdot \frac{2at^2}{(bt + z_d)(bt - z_d)}\, du = \frac{1}{2\sqrt{\pi}} \int \frac{1}{\sqrt{u}} e^{-u} du \tag{29}$$

Then, we make the second substitution.

$$\sqrt{u} = w, \rightarrow \frac{dw}{du} = \frac{1}{2\sqrt{u}}, \rightarrow \frac{du}{dw} = 2w, \rightarrow du = 2w\, dw \tag{30}$$

Taking into consideration the above-mentioned dependencies, the integral (29) can be written in the following form:

$$\frac{1}{2\sqrt{\pi}} \int \frac{1}{w} e^{-w^2}\, 2w\, dw = \frac{1}{\sqrt{\pi}} \int e^{-w^2} dw \tag{31}$$

We make one more substitution.

$$w^2 = \frac{y^2}{2}, \rightarrow 2w\, dw = y\, dy, \rightarrow dw = \frac{y}{2w} dy, \rightarrow dw = \frac{y}{\sqrt{2}} \tag{32}$$

Thus, we obtain the integral in the following form:

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{y^2}{2}} dy \tag{33}$$

where:

$$y = \frac{bt - z_d}{\sqrt{at}} \tag{34}$$

Substituting the results into the formula (22) and remembering the appropriate notation of the integration limits, we obtain the formula for the reliability:

$$R(t) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{bt - z_d}{\sqrt{at}}} e^{-\frac{y^2}{2}} dy \tag{35}$$

The distribution function for the standard normal distribution has the following form (36).

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \tag{36}$$

Finally, the formula for the reliability of a system has the form of the following dependence:

$$R^*(t) = 1 - \Phi\left(\frac{b^* t - z_d}{\sqrt{a^* t}}\right) \tag{37}$$

where b^* and a^* are coefficients after the estimation on the basis of data obtained from the exploitation of military aircrafts.

Thus, the risk of a device damage can be determined from the following dependence (38).

$$Q^* = 1 - R^*(t) = \Phi(\gamma) \tag{38}$$

where:

$$\gamma = \frac{b^* t - z_d}{\sqrt{a^* t}} \tag{39}$$

Assuming a specified level of damage risk, we can find γ (by reading values on the tables of the normal distribution). Knowing the value of γ, we can determine the durability (i.e. *t*) from the dependence (39). For this purpose, the dependence (39) was transformed into the following square equation (40).

$$b^{*2} t^2 - \left(\gamma^2 a^* + 2b^* z_d\right) t + z_d^2 = 0 \tag{40}$$

Thus, the durability:

$$T = \frac{\left(\gamma^2 a^* + 2b^* z_d\right) - \sqrt{\left(2b^* z_d + \gamma^* a^*\right)^2 - 4b^{*2} z_d^2}}{2b^{*2}} \tag{41}$$

4.3 A computational example

The efficiency of the chosen system is determined with the help of diagnostic parameters describing the technical condition of particular devices of the system. An aiming head (a navigation and aiming device) is an important device of avionics system. Its technical condition is described by two diagnostic parameters: ε and β which describe the coordinates of position of sight marker.

On the basis of analyzing results of checks of a particular population of aiming heads it was established that as the time of operation goes by and as a result of the influence of destructive factors, the values of these parameters undergo changes. Table 1 presents an exemplary course of changes of values of the diagnostic parameters ε and β during an operation process.

T [months]	0	27	40	57	83	94	102	110	116
ε	0	0,01	0,01	0,01	0,07	0,48	0,48	0,54	0,73
β	0	0,23	0,26	0,26	0,39	0,50	0,53	0,56	0,59

Table 1. Changes of diagnostic parameter values in an aiming head during an operation process

Having data describing the values of deviation of a diagnostic parameter in the following form $[(z_0, t_0), (z_1, t_1), (z_2, t_2), ..., (z_n, t_n)]$, and basing on the following formulas,

$$b^* = \frac{z_n}{t_n}, \quad a^* = \frac{1}{n}\sum_{k=0}^{n-1}\frac{\left[(z_{k+1} - z_k) - b^*(t_{k+1} - t_k)\right]^2}{(t_{k+1} - t_k)} \tag{42}$$

the values of the density function coefficients for both diagnostic parameters were determined:

$$a_\varepsilon^* = 0{,}002; \quad b_\varepsilon^* = 0{,}0063; \quad a_\beta^* = 0{,}0003; \quad b_\beta^* = 0{,}0051 \tag{43}$$

Assuming the following level of reliability $R^*(t) = 0{,}99$, the value of the parameter $\gamma = 2{,}32$ was read on the tables of normal distribution. The parameter z_d was determined on the basis of a technical documentation which is used for service works and includes information on the acceptable values of deviations of the diagnostic parameters.

The values of the parameters a, b, γ, z_d were substituted into the equation (41), and the time after which the values of the diagnostic parameter deviations exceed the limit state was calculated. In this case, the time comes to:

$$T_\varepsilon\text{=5[months]}, \qquad T_\beta\text{=33[months]} \tag{44}$$

since the last check of the diagnostic parameters. The values (44) can be used in technical service depending on the adopted service strategy.

Summing up, we can state that the above-presented method seems to be correct and enables the analysis of a device/system technical condition with respect to the character of changes of values of the diagnostic parameters. The above-presented calculation example enabled the verification of the developed model and showed application qualities of the method. This method can be useful in future work on the improvement of both the operation process and the way of use of aircrafts with avionics system because it enables the determination of time during which a device is fit for use.

Moreover, due to its universal character, the method can be used to determine the residual life of any technical object whose technical condition is determined by analyzing values of the diagnostic parameters.

5. The process of operating the military aircraft

5.1 The influence of destructive factors on the course of the process of operating the military aircraft

The use of military aircrafts concerns mainly the performance of a particular combat task, which often involves the use of aerial combat means. As far as an airborne function of a military aircraft is concerned, the main stages of its operation comprise the take-off, the staying in the air, and the landing. On the other hand, when analyzing the process of the operation of the on-board armament system, we can assume that the operational effect is the sum of the partial effects gained during the flight phase in relation to:

- target detection;
- the execution of the aiming process;
- the execution of the process of attacking.

The level of effect of munitions on a target is the most commonly assumed rate that characterizes the operational effect obtained during the execution of a combat task involving the use of aerial combat means. As regards the on-board armament system, the obtained effect comes down to the determination of the difference between the value of target coordinates and the coordinate values of a drop point of combat armament.

Based on the structural diagram (Fig. 1) and the functions of the on-board armament system, we can assume that the Armament Control System (ACS) is the basic element that affects the value of the operational effect. Both at the stage of maintenance and operation, ACS provides information that is essential for the accurate functioning of the on-board armament system (OAS). In turn, as regards the ACS, its most crucial element involves the navigation and aiming system (NAS). Its basic task comprises the realization of a set of algorithms. Their solution enables – in the maintenance system - the reconstruction of the nominal values of particular initial parameters; - in the operation system – the proper usage of combat means (the intended use). The latter system is the subject of further discussion.

The analysis of the operational effect can be performed on the basis of the assessment of conditions in which NAS is used and the determination of causes that have a negative impact on the final value of the obtained effect. As regards NAS, during the execution of a combat task, the operational effect is the total angular correction represented as an aiming indicator in a pilot's field of view. The process of aiming and attacking is executed on the basis of the total angular correction. Thus, we can assume that the assessment of the operational effect involves the determination of accuracy in defining and reproducing the position of a moving aiming indicator.

The next aspect concerns the use of the aiming correction by a pilot. When the correction is defined and illustrated, the task comes down to the determination of the flight conditions in which an aiming indicator coincides with a target at the moment of using combat means. Based on the conducted analysis, we can assume that the execution of a combat task under real conditions is not an easy process. The causes of errors affecting the value of the operational effect connected with the aiming process execution can be represented as the equation for the pooled error of the aiming process execution Δ_{Σ}:

$$\Delta_{\Sigma} = (\Delta_M + \Delta_K + \Delta_I + \Delta_A) + (\Delta_C + \Delta_W + \Delta_R + \Delta_O) + \Delta_N \tag{45}$$

The error of the method for solving the aiming-related equations Δ_M characterizes two groups of causes:

1. connected with the relative uncertainty resulting from the processing of initial data concerning the aiming process by NAS functional elements, and
2. concerning the error function of equations for aiming.

The system configuration error Δ_K connects with entering invalid control signals (that characterize the combat task being performed) into NAS.

The instrumental error Δ_I connects with the accuracy of determining the operational parameters of NAS by particular information transmitters. This error concerns mainly the measurement error.

The reconstruction error Δ_A characterizes the adequacy of a physical combat situation taking place during the execution of the aiming process to the assumed attack diagram which was used to determine the aiming equations.

The causes of variance between the aiming indicator position and the target Δ_C result from an incorrect approach of an aircraft to an attack path.

The causes of the failure to maintain the required conditions for aiming and attacking Δ_W connect with the failure to keep the required angle of diving, flight speed, bank angle, etc., i.e. the exceeding of the nominal values of particular parameters describing a combat task.

The effect of the weapon position Δ_R on the pooled error value Δ_{Σ}, concerns mainly the process of aiming during the execution of the process of attacking with the use of aerial combat means (that are applied in a time series of particular length).

Environmental conditions determining the value of the error Δ_O significantly influence the execution of the aiming process. Due to the fact that an aircraft moves at high speed in a heterogeneous space, it may encounter various conditions prevailing in space layers or areas, which directly translates into the perturbation of flight-related parameter values.

The general error Δ_N concerns causes which are not included in the presented classification and are the resultant of the lack of possibility to learn or describe them in an analytical way at the present state of knowledge.

All the above-mentioned errors can be of two kinds: determined errors (systematic errors) and probabilistic errors (random errors). So, their accumulated form Δ_{Σ} will be burdened with both types of errors. The phenomenon of the random error occurrence is not precisely determined, that is why an attempt to evaluate its value is fully justified. A random character of compound errors causes that the operational effect of MMA application is burdened with the random error, too.

5.2 The model of the assessment of the execution of a combat mission by the military aircraft

The execution of the aiming process generally comes down to the process of making an aiming indicator coincide with a target. Significant elements of this process include parameters that determine the aiming indicator position and a set of actions aiming at pointing the indicator at a target. Based on these elements, we can consider the process of aiming as the execution of the process of building the aiming triangle using: a pilot – the system operator, an aiming indicator – the quantity describing the appropriate spatial orientation of an aircraft, and a target – the basic point in the execution of the aiming process. The aim of the process is to align these three elements.

The aiming correction is obtained by recording particular parameters (necessary to solve aiming equations) and processing them in NAS. The aiming correction value is represented as the central point of a moving aiming indicator which is displayed on the reflector of the sight head. Due to the effect of various constraints, the aiming indicator can adopt different positions in the assumed flat coordinate system (Fig. 6) placed on the plane of the sight head reflector. The indicator can either move in one out of four directions or move back to the previously occupied position.

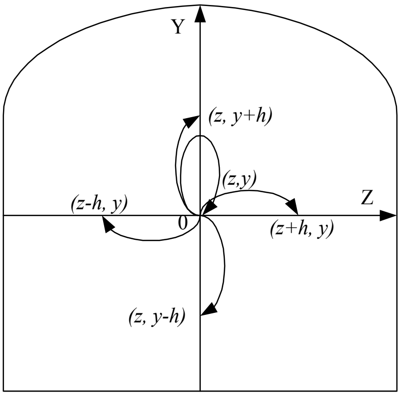


Fig. 6. A graphical representation of the occurrence of possible deviations of the central point of the moving indicator during the execution of the aiming process

$U_{z,y,t}$ denotes the probability that at the moment t the position deviations of the central point of the moving indicator are z and y, where t is the current time of the process of aiming. This probability is characterized by the density function denoted as $U(z,y,t)$. Therefore, using the density function $U(z,y,t)$ we can describe the dynamics of changes in the position deviations of the central point of the moving indicator by a difference equation.

Regarding the issue being discussed above, the difference equation is as follows:

$$\begin{aligned} U(z,y,t+\Delta t) = P_{00}U(z,y,t) + P_{10}U(z-h,y,t) + P_{20}U(z+h,y,t) + \\ + P_{01}U(z,y-h,t) + P_{02}U(z,y+h,t) \end{aligned} \tag{46}$$

where:

U(z,y,t) - the probability density function of deviation values at the moment *t*;
Δt - the time value between the specified deviations;
h - the deviation value along the specified axes;
P_{00} - the probability that the deviation value will not change;
P_{10} - the probability that the deviation value along the OZ axis will change by *-h* at the time *Δt*;
P_{20} - the probability that the deviation value along the OZ axis will change by *h* at the time *Δt*;
P_{01} - the probability that the deviation value along the OY axis will change by *-h* at the time *Δt*;
P_{02} - the probability that the deviation value along the OY axis will change by *h* at the time *Δt*;

When we use expressions obtained from the expansion of the function *U(z,y,t)* in the Taylor series in the surrounding of the point *(z,y)* and the time *t* in accordance with the relationships of the following set of equations:

$$\left.\begin{aligned}
&U(z,y,t+\Delta t)=U+\frac{\partial U}{\partial t}\Delta t\\
&U(z-h,y,t)=U-\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\\
&U(z+h,y,t)=U+\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\\
&U(z,y-h,t)=U-\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\\
&U(z,y+h,t)=U+\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}
\end{aligned}\right\} \tag{47}$$

where *U=U(z,y,t)*, and the fact that $P_{00}+P_{10}+P_{20}+P_{01}+P_{02}=1$, the equation (47) takes the following form:

$$\begin{aligned}
U+\frac{\partial U}{\partial t}\Delta t=P_{00}U+P_{10}\left(U-\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\right)+P_{20}\left(U+\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\right)+\\
+P_{01}\left(U-\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\right)+P_{02}\left(U+\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\right)
\end{aligned} \tag{48}$$

When adding and subtracting *U* in the equation (48) and multiplying appropriate expressions in the brackets and taking the parameter *U* outside the brackets, the following result was obtained:

$$\begin{aligned}
\frac{\partial U}{\partial t}\Delta t=-U+\left(P_{00}+P_{10}+P_{20}+P_{01}+P_{02}\right)U+P_{10}\left(-\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\right)+\\
+P_{20}\left(\frac{\partial U}{\partial z}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2}\right)+P_{01}\left(-\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\right)+\\
+P_{02}\left(\frac{\partial U}{\partial y}h+\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\right)
\end{aligned} \tag{49}$$

Using the assumption that the sum of all probabilities describing the weapon angular position equals one, the equation (49) takes the following form:

$$\begin{aligned}\frac{\partial U}{\partial t}\Delta t = -P_{10}\frac{\partial U}{\partial z}h + P_{10}\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2} + P_{20}\frac{\partial U}{\partial z}h + P_{20}\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2} - P_{01}\frac{\partial U}{\partial y}h + \\ + P_{01}\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2} + P_{02}\frac{\partial U}{\partial y}h + P_{02}\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\end{aligned} \qquad (50)$$

After grouping the quantities from the above equation, the following equation was obtained:

$$\begin{aligned}\frac{\partial U}{\partial t}\Delta t = -P_{10}\frac{\partial U}{\partial z}h + P_{20}\frac{\partial U}{\partial z}h + P_{10}\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2} + P_{20}\frac{1}{2}h^2\frac{\partial^2 U}{\partial z^2} - P_{01}\frac{\partial U}{\partial y}h + \\ + P_{02}\frac{\partial U}{\partial y}h + P_{01}\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2} + P_{02}\frac{1}{2}h^2\frac{\partial^2 U}{\partial y^2}\end{aligned} \qquad (51)$$

After dividing both sides of the equation (51) by *Δt*, the following result was obtained:

$$\begin{aligned}\frac{\partial U}{\partial t} = -\frac{(P_{10} - P_{20})h}{\Delta t}\frac{\partial U}{\partial z} + \frac{(P_{10} + P_{20})\frac{1}{2}h^2}{\Delta t}\frac{\partial^2 U}{\partial z^2} + \\ \text{-}\frac{(P_{01} - P_{02})h}{\Delta t}\frac{\partial U}{\partial y} + \frac{(P_{01} + P_{02})\frac{1}{2}h^2}{\Delta t}\frac{\partial^2 U}{\partial y^2}\end{aligned} \qquad (52)$$

By introducing the following denotations:

$$b_1 = \frac{(P_{10} - P_{20})h}{\Delta t}, \quad \mathrm{b}_2 = \frac{(P_{01} - P_{02})h}{\Delta t} \qquad (53)$$

$$a_1 = \frac{(P_{10} + P_{20})h^2}{\Delta t}, \quad a_2 = \frac{(P_{01} + P_{02})h^2}{\Delta t} \qquad (54)$$

and substituting them into the equation (52), the following differential equation was obtained:

$$\frac{\partial U}{\partial t} = -b_1\frac{\partial U}{\partial z} - b_2\frac{\partial U}{\partial y} + \frac{1}{2}a_1\frac{\partial^2 U}{\partial z^2} + \frac{1}{2}a_2\frac{\partial^2 U}{\partial y^2} \qquad (55)$$

The following function is the solution of the above equation:

$$U(z, y, t) = \frac{1}{\sqrt{2\pi a_1 t}\sqrt{2\pi a_2 t}}e^{-\frac{1}{2}\left(\frac{(z - b_1 t)^2}{a_1 t} + \frac{(y - b_2 t)^2}{a_2 t}\right)} \qquad (56)$$

Assuming that the probabilities P_{10} and P_{20} are of the same order, i.e. $P_{10}=P_{20}$, we can write that the coefficient $b_1 \approx 0$. Similarly, we can assume that the probabilities P_{01} and P_{02} are also of the same order, so the coefficient $b_2 \approx 0$. Given these assumptions, the equation (55) takes the following form:

$$\frac{\partial U}{\partial t} = \frac{1}{2} a_1 \frac{\partial^2 U}{\partial z^2} + \frac{1}{2} a_2 \frac{\partial^2 U}{\partial y^2} \tag{57}$$

The following form of the density function is the solution of the equation (57):

$$U(z,y,t) = \frac{1}{\sqrt{2\pi a_1 t}\sqrt{2\pi a_2 t}} e^{-\frac{1}{2}\left(\frac{z^2}{a_1 t} + \frac{y^2}{a_2 t}\right)} \tag{58}$$

The explicit form of the density function (58) requires determining the equation coefficients (57) and connects with:

- obtaining input data;
- determining the density function (58);
- determining the likelihood function L enabling the determination of the parameter estimates a_1 and a_2:

$$L = \frac{1}{(2\pi)^n (a_1 a_2)^{\frac{n}{2}}} \prod_{k=1}^{n-1} \frac{1}{(t_{k+1}-t_k)} \exp\left\{-\frac{1}{2}\left[\frac{(z_{k+1}-z_k)^2}{a_1(t_{k+1}-t_k)} + \frac{(y_{k+1}-y_k)^2}{a_2(t_{k+1}-t_k)}\right]\right\} \tag{59}$$

To determine the parameters a_1 and a_2 we can use the method of the maximum likelihood. The method consists in finding the parameter values a_1 and a_2 that maximize the likelihood function. So, we seek the solution of the set of equations

$$\begin{cases} \dfrac{\partial \ln L}{\partial a_1} = 0 \\ \dfrac{\partial \ln L}{\partial a_2} = 0 \end{cases} \tag{60}$$

Therefore, the logarithm of the likelihood function L takes the following form:

$$\begin{aligned} \ln L = -n \ln 2\pi - \frac{n}{2} \ln a_1 - \frac{n}{2} \ln a_2 + \\ + \sum_{k=1}^{n-1}\left[\ln(t_{k+1}-t_k) + \left[-\frac{1}{2}\left(\frac{(z_{k+1}-z_k)^2}{a_1(t_{k+1}-t_k)} + \frac{(y_{k+1}-y_{*k})^2}{a_2(t_{k+1}-t_k)}\right)\right]\right] \end{aligned} \tag{61}$$

By determining the derivatives of the function L relative to specified parameters, the following set of equations was obtained:

$$\begin{cases} -\dfrac{n}{2a_1}+\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(z_{k+1}-z_k\right)^2}{2a_1^2\left(t_{k+1}-t_k\right)}=0 \\ -\dfrac{n}{2a_2}+\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(y_{k+1}-y_k\right)^2}{2a_2^2\left(t_{k+1}-t_k\right)}=0 \end{cases} \tag{62}$$

which after transformation provides the following equations (63):

$$\begin{cases} a_1=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(z_{k+1}-z_k\right)^2}{\left(t_{k+1}-t_k\right)} \\ a_2=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(y_{k+1}-y_k\right)^2}{\left(t_{k+1}-t_k\right)} \end{cases} \tag{63}$$

Therefore, the parameters a_1 and a_2 can be defined on the basis of the above set of equations. When analyzing the function notation (58), it can be assumed that in order to determine the variance characterizing the distribution of the indicator central point, the parameters a_1 and a_2 must be multiplied by time, which leads to the following result:

$$\begin{cases} \sigma_z^2\left(t_n\right)=a_1t_n=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(z_{k+1}-z_k\right)^2}{\left(t_{k+1}-t_k\right)}\sum_{k=1}^{n-1}\left(t_{k+1}-t_k\right) \\ \sigma_y^2\left(t_n\right)=a_2t_n=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n-1}\dfrac{\left(y_{k+1}-y_k\right)^2}{\left(t_{k+1}-t_k\right)}\sum_{k=1}^{n-1}\left(t_{k+1}-t_k\right) \end{cases} \tag{64}$$

The determination of the function parameters (58) will allow defining the probability density function of the correct position of the indicator central point.

As regards the case described, it is assumed that the probability of the occurrence of deviations in any direction of the assumed coordinate axes is the same. Such situation takes place when the process of aiming is performed correctly, i.e. when at the beginning of the aiming process, an aiming indicator coincides with a target and any dislocation of the indicator is compensated with its resetting on the target. A real process of aiming often involves the indicator dislocation relative to a target. The occurrence of such dislocation causes that the probability of the indicator dislocation in a specified direction is higher than the indicator dislocation in an opposite direction. Thus, the values of the parameters b_1 and b_2 are not 0. Therefore, the differential equation describing the aiming process takes the form of the equation (55). Its solution is the density function (56). The parameters b_1, b_2, a_1 and a_2 need to be determined for the function. Using the above-described technique, the likelihood function (65) was determined. It was used to estimate the sought parameters:

$$L=\frac{1}{(2\pi)^n\left(a_1a_2\right)^{\frac{n}{2}}}\prod_{k=1}^{n-1}\frac{1}{\left(t_{k+1}-t_k\right)}\exp\left\{-\frac{1}{2}\left[\frac{\left(\left(z_{k+1}-z_k\right)-b_1\left(t_{k+1}-t_k\right)\right)^2}{a_1\left(t_{k+1}-t_k\right)}+\frac{\left(\left(y_{k+1}-y_k\right)-b_2\left(t_{k+1}-t_k\right)\right)^2}{a_2\left(t_{k+1}-t_k\right)}\right]\right\} \tag{65}$$

The process of determining the function parameter (65) is analogous to the way of determining the equation coefficients (59). By determining the derivatives of the function logarithms (65) relative to specified coefficients and comparing them to 0, the following relationships were obtained:

$$\begin{aligned} b_1 &= \frac{z_n}{t_n}, \qquad\qquad b_2 = \frac{y_n}{t_n} \\ a_1 &= \frac{1}{n}\sum_{k=1}^{n-1}\frac{\left[\left(z_{k+1}-z_k\right)-b_1\left(t_{k+1}-t_k\right)\right]^2}{\left(t_{k+1}-t_k\right)} \\ a_2 &= \frac{1}{n}\sum_{k=1}^{n-1}\frac{\left[\left(y_{k+1}-y_k\right)-b_2\left(t_{k+1}-t_k\right)\right]^2}{\left(t_{k+1}-t_k\right)} \end{aligned} \tag{66}$$

By determining the values of the above coefficients and substituting them into the equation (56), we can determine the density function of the indicator position during the aiming process involving the indicator dislocation relative to a target.

The indicator path relative to a target (described for subsequent moments *t₀, t₁, t₂, ..., tₙ,*) can be characterized by horizontal coordinates *z₀, z₁, z₂, ..., zₙ* and vertical coordinates *y₀, y₁, y₂, ..., yₙ* of the assumed coordinate system. When converting these quantities to current data, the time of recording the position of the aiming indicator can be replaced by the number of the registered positions (next coordinate values will constitute the sum of previous coordinates). Thus, the indicator position will be characterized by:

1. the number of registered positions: *0, 1, 2, ..., n;*
2. the deviation toward the 0Z axis: $0, z_1, (z_1+z_2), (z_1+z_2+z_3), ..., \sum_{i=1}^{n} z_i$;
3. the deviation toward the 0Y axis: $0, y_1, (y_1+y_2), (y_1+y_2+y_3), ..., \sum_{i=1}^{n} y_i$.

Based on the above, we can determine the following parameters:

$$b_1^* = \frac{\sum_{i=1}^{n} z_i}{n}, \qquad\qquad b_2^* = \frac{\sum_{i=1}^{n} y_i}{n} \tag{67}$$

$$\begin{aligned} \sigma_1^2 &= a_1^* = \frac{1}{n}\sum_{k=1}^{n-1}\left[\left(\hat{z}_{k+1}-\hat{z}_k\right)-\left(\frac{1}{n}\sum_{i=1}^{n} z_i\right)\right]^2 \\ \sigma_2^2 &= a_2^* = \frac{1}{n}\sum_{k=1}^{n-1}\left[\left(\hat{y}_{k+1}-\hat{y}_k\right)-\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\right]^2 \end{aligned} \tag{68}$$

where:

$$\begin{aligned} \widehat{z}_{k+1} &= \sum_{i=1}^{k+1} z_i, \quad \widehat{z}_k = \sum_{i=1}^{k} z_i \\ \widehat{y}_{k+1} &= \sum_{i=1}^{k+1} y_i, \quad \widehat{y}_k = \sum_{i=1}^{k} y_i \end{aligned} \tag{69}$$

Because

$$\widehat{z}_{k+1} - \widehat{z}_k = z_{k+1} \quad \text{and} \quad \widehat{y}_{k+1} - \widehat{y}_k = y_{k+1} \tag{70}$$

therefore:

$$\begin{aligned} \sigma_1^2 &= \frac{1}{n}\sum_{k=1}^{n}\left[z_k - \frac{1}{n}\sum_{i=1}^{n} z_i \right]^2 \\ \sigma_2^2 &= \frac{1}{n}\sum_{k=1}^{n}\left[y_k - \frac{1}{n}\sum_{i=1}^{n} y_i \right]^2 \end{aligned} \tag{71}$$

The above relationships can be used to describe the process of aiming under real-life conditions.

5.3 A computational example

The execution of a combat task with the use of aerial combat means is characterized by the fact that the possibility of their use is determined by conditions that constitute a set of various factors enabling the performance of a combat task at the required level and with the consideration of a current tactical, navigational, meteorological, and radio-technical situation. The basic determinants of these conditions involve combat capabilities of an aircraft and the level of competence among aircrew members. The essence of the aiming process comes down to the controlling of an aircraft in such a way that it reaches the point in space where the applied weapon will hit a target. This procedure is performed in the NAS environment on the basis of the following data:

- motion parameters of an aircraft executing an attack, a target, and parameters of the centre where an aircraft motion is executed;
- the required coordinates of a target;
- the actual coordinates of a target;
- the comparison between actual and required coordinates of a target.

A common method for analyzing the aiming process during an attack is the recorded material analysis (using either the film placed in a photo-control apparatus located in front of the sight head or a camera recording a tactical situation in front of MMA.) Based on the recorded material, it is possible to determine a mutual position of an aiming indicator and a target at the moment of a weapon use.

Having the material registered by photo-control devices (Fig. 6) and using the above-mentioned method, it is possible to define coordinates of the mutual position of a target and indicator in successive moments of the attacking process.

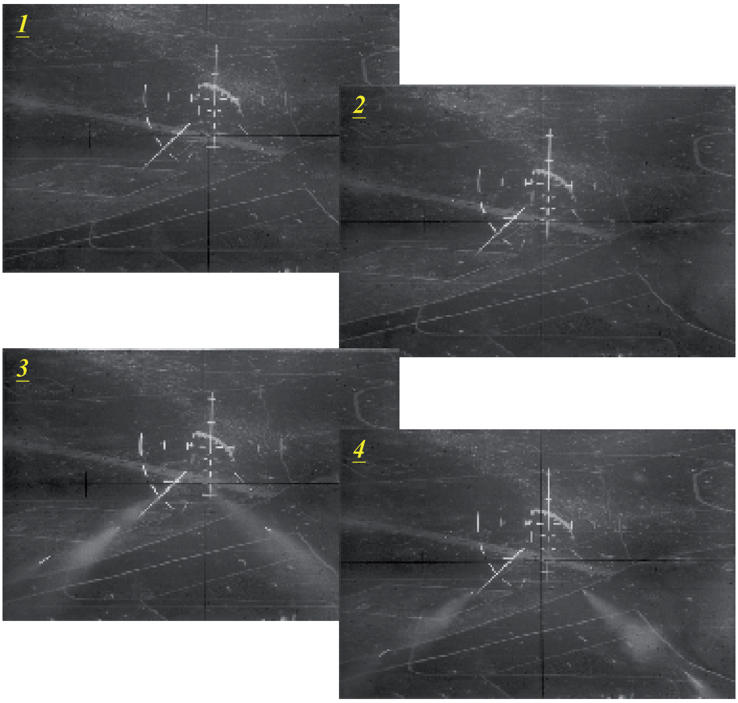


Fig. 6. Photos taken with a photo-control apparatus during the realization of the attacking process with the use of non-guided missiles

Based on the obtained data, it was possible to determine the aiming indicator path relative to a target. Figure 7 depicts the path. When analyzing the position of the central point of the aiming indicator, we can assume that the position adopting the chaotic motion of the indicator was the proper position that completely reflects the nature of the real process.

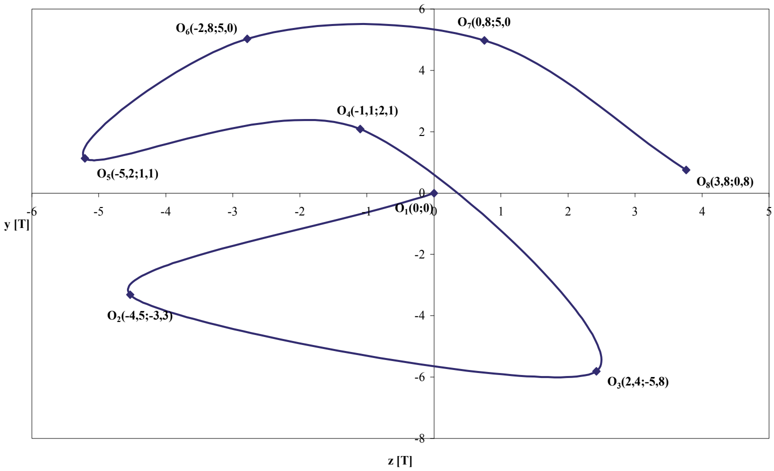


Fig. 7. The course of changes in the position of the aiming indicator relative to a target during the realization of the aiming process with the use of non-guided missiles.

The variance values were determined for the data presented in Fig. 7. The values are as follows:

$$\sigma_z^2 = 14{,}24\left[T^2\right], \qquad \sigma_y^2 = 22{,}80\left[T^2\right] \tag{72}$$

By substituting the above equation values (58) and on the basis of the recorded data, it was possible to determine a graphical form of the probability density function (Fig. 8) that characterizes the concurrence of the aiming indicator with a target during the execution of the aiming process.

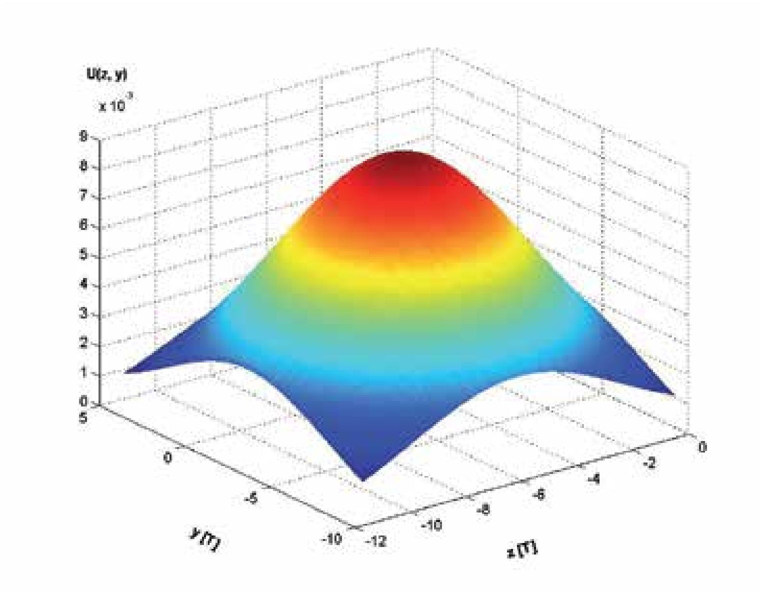


Fig. 8. A graph of the probability density function of indicator deviations during the execution of the aiming process with the use of non-guided missiles

6. Summary

Works carried out during the process of maintaining aim to ensure the required level of safety concerning aircraft engineering and to maintain it in good working condition. This is achieved by carrying out planned works and systematic checks of diagnostic parameter values. Apart from identification, diagnostic testing includes two more aspects concerning the technical state genesis and prediction. That is why, for safety and reliability reasons, it is important to develop methods enabling prediction of the technical state of devices on the basis of information obtained during the maintenance process. The 4rd chapter comprises the presentation of the probabilistic method for the determination of residual durability of devices on the basis of their diagnostic parameter changes registered during the process of maintaining. The application of the above-mentioned method may facilitate the military aircraft maintenance process by limiting the number of stoppages through the indication of a time of next maintenance works for a specified device/system. It shall be emphasized that the presented method is universal as it can be applied to the maintenance process modernization not only in respect of aircraft engineering but also in respect of any field where device/system diagnostic parameters are registered.

The process of operating is inevitably connected with "an operational effect" which results from the completion of a particular combat mission. Depending on a combat mission, this effect will concern, for example hitting the target, intercepting an enemy, identifying the target to attack, etc. The operational effect is always obtained during flight. Due to flying conditions of the military aircraft, we can list a number of destructive factors reducing the value of the obtained operational effect. Analyzing the process of operating, we can state that one of the most significant "cells" in this process is the flying military personnel – a pilot. His task involves the appropriate configuration of the military aircraft systems and the performance of the aiming process that generally comes down to the process of making an aiming indicator coincide with a target. The method presented in the 5th chapter enables the quantitative assessment of the aiming process quality. The results obtained in this way and supported by parameters describing conditions in which a combat task was conducted may constitute the basis for the evaluation of the realization of both a current combat task and the progress in training (considering the series of tasks of a given type in a specified time interval).

7. References

Fisz M. (1958). Probability Calculus and Mathematical Statistics. PWN, Warsaw, Poland

Kaczmarski W. (1990). Aircraft Weapons. Part II. Aircraft Sights Handbook. DWL, Poznan, Poland

Moir I.; Seabridge A. (2006). Military Avionics System. Chichester, England: Wiley

Olearczuk E.; Sikorski M.; Tomaszek H. (1978). Aircrafts maintenance. MON, Warsaw, Poland

Skomra A., Tomaszek H.; Wroblewski M. (1999). Tactical and Technical Characteristics and the Effectiveness of Combat Air Munitions. Military Academy of Technology-Textbook, Warsaw, Poland

Su-22M4 Handbook 7. Weapons. Part VII. Technology of Periodic Service Works. DWLiOP, 1986, Poznan, Poland

Tomaszek H.; Wazny M. (2008). The outline of the assessment of durability against surface wear of a construction element with the use of the distribution of time of the exceedence of limit state (admissible state). ZEM, Vol. 3(155) 2008. pp. 47-59, ISSN: 0137-5474, Radom, Poland

Tomaszek H.; Zurek J.; Loroch L. (2004). The outline of a method of estimation reliability and durability of aircraft's structure elements on the basis of destruction process description. ZEM, Vol. 3(139) 2004. pp. 73-85, ISSN 0137-5474, Radom, Poland

Wazny M. (2003). The analysis of operating causes of the dispersion of selected munition and their influence on the air weapons effectiveness. Military Academy of Technology 2003, Warsaw, Poland

Wazny M. (2008). The method of determining the time concerning the operation of a chosen navigation and aiming device in the operation system. Maintenance and Reliability Nr2/2008, 2(38), pp. 4-11. ISSN: 1507-2711, Lublin, Poland

Wazny M.; Wojtowicz K. (2008). The analysis of the military aircraft maintains system and the modernization proposal.: Maintenance and Reliability Nr3/2008, 3(39), pp. 4-11, ISSN: 1507-2711, Lublin, Poland

www.airliners.net

Part 5

Miscellaneous Topics

Review of Technologies to Achieve Sustainable (Green) Aviation

Ramesh K. Agarwal
Department of Mechanical Engineering and Materials Science
Washington University in St. Louis, St. Louis, MO,
USA

1. Introduction

Among all major modes of transportation, people travel by airplanes and automobiles continues to experience the fastest growth. As shown in Figure 1 [1], the travel as measured by Passenger - Kilometers (PKM) is forecasted to more than double from the current 2010 level of ~ 40 trillion PKM to approximately 103 trillion PKM by 2050. Among these two modes of transportation, air travel is experiencing the faster growth. The number of Passenger – Kilometers Travelled (PKT)/ capita by various modes of transportation in different countries is shown in Figures 2(a) - 2(d) [1]. Figures 2(a) and 2(c) also show that the use of personal vehicles compared to public transport (in PKT) is highest in U.S. followed by the wealthier nations. Furthermore, as the per capita income of a nation increases, the travel demand will increase (Figure 3) [1] resulting in greater demand for personal vehicles as well as for air transportation as shown in Figure 1. These projections are based on 3% growth in world Gross Domestic Product (GDP), 5.2% growth in passenger traffic and 6.2% increase in cargo movement. Only major policy changes and intervention by governments through development of infrastructure for public transportation is likely to slow down these trends shown in Figure 1. Most of the energy for transportation is currently provided by the fossil fuels (primarily petroleum). Figure 4 shows the oil consumption for transportation in U.S. and its forecast for the future [2]. Figure 5 shows the relative percentage of fuel consumption by various categories of vehicles in U.S [2]. The consequence of burning fossil fuels is well established in their long term impact on climate and global warming due to Greenhouse Gas (GHG) emissions, primary being the CO_2 and NOx. Table I gives the current level of CO_2 emissions worldwide by ground and air transportation [3] and Figure 6 shows the forecast for the future if the current Business as Usual (BAU) scenario continues [3]. The reduction in GHG emissions due to the burning of fossil fuels is the major goal of “Green Transportation.” The “Sustainability” goal is to explore both the technological solutions to increase the efficiency of transportation as well as the alternative carbon neutral fuels (e.g. biofuels among others).

2. Sustainable (green) air transportation

Most of the material presented in this section has been taken from the author’s William Littlewood Award Lecture [4]. This section provides an overview of issues related

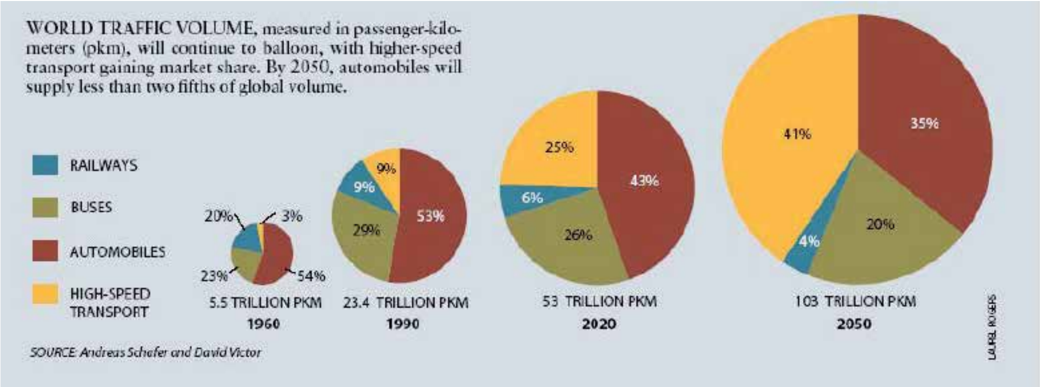
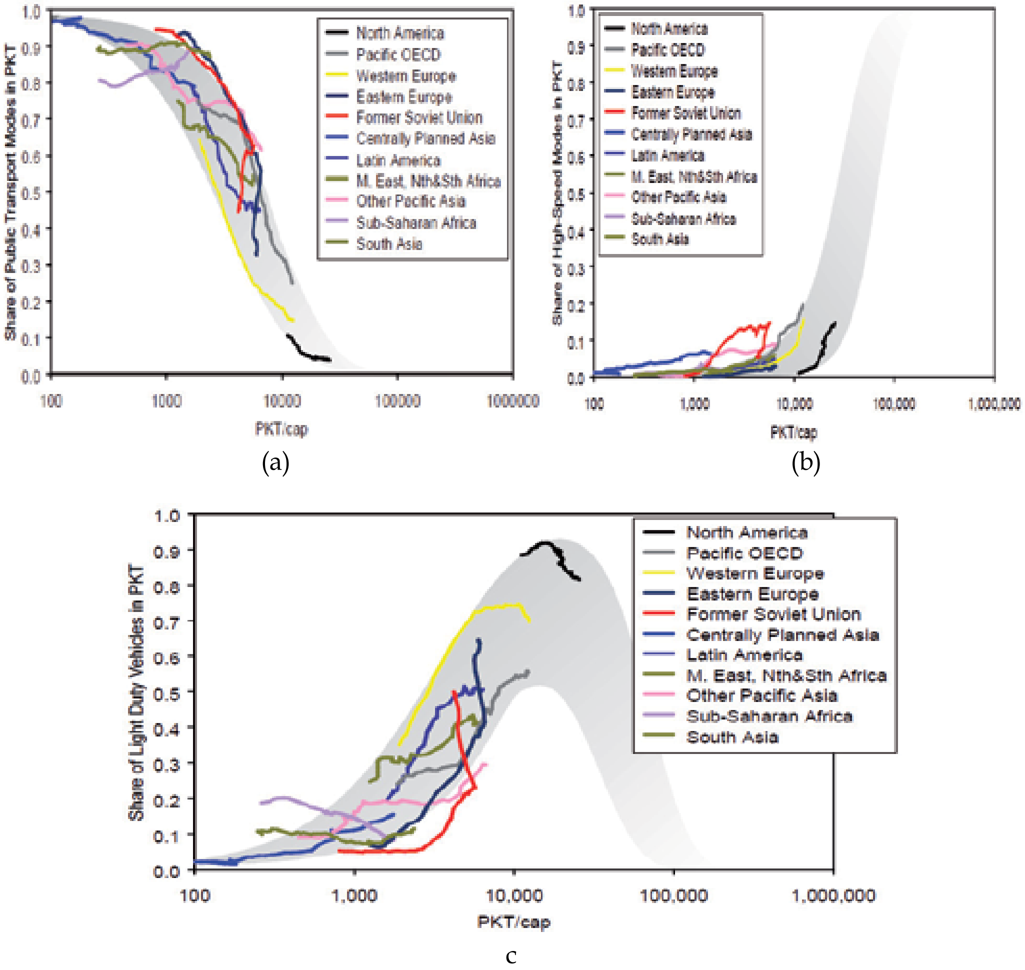


Fig. 1. Global mobility trends from various modes of transportation [1].



(a)

(b)

c

Fig. 2. a: % share of public transport in various countries; b: % share of high speed transport in various countries; c: % share of light-duty vehicle transport in various countries [1].

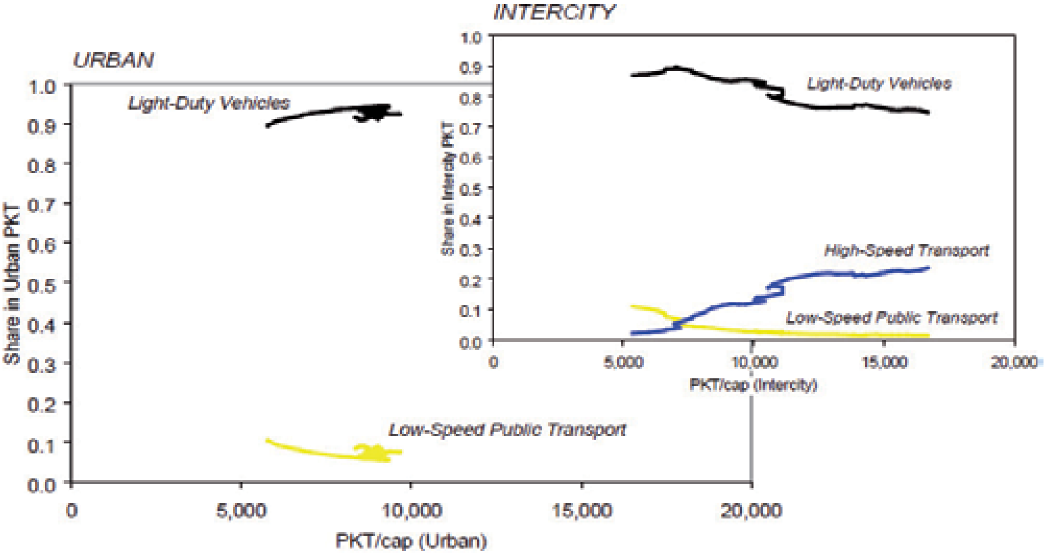
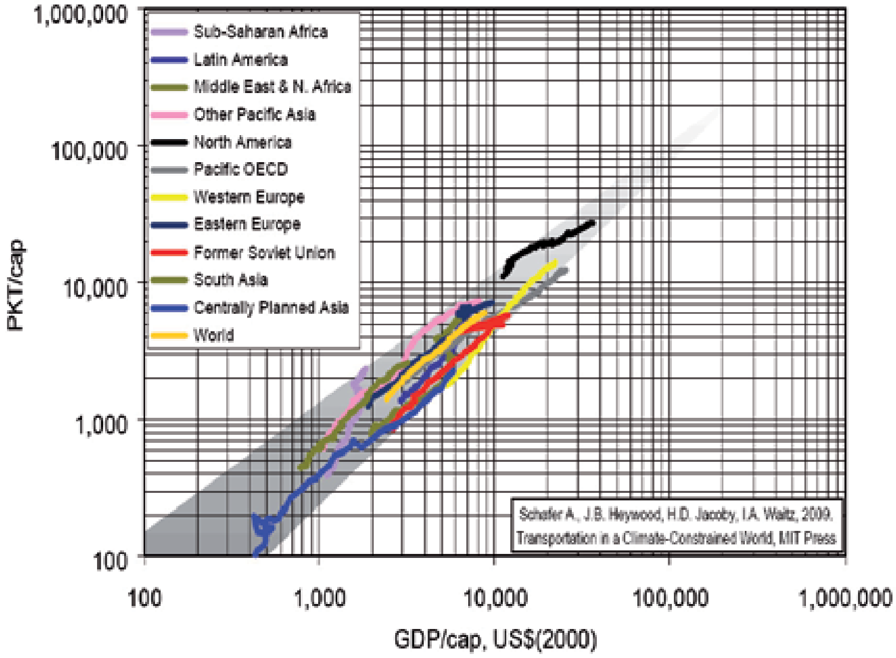
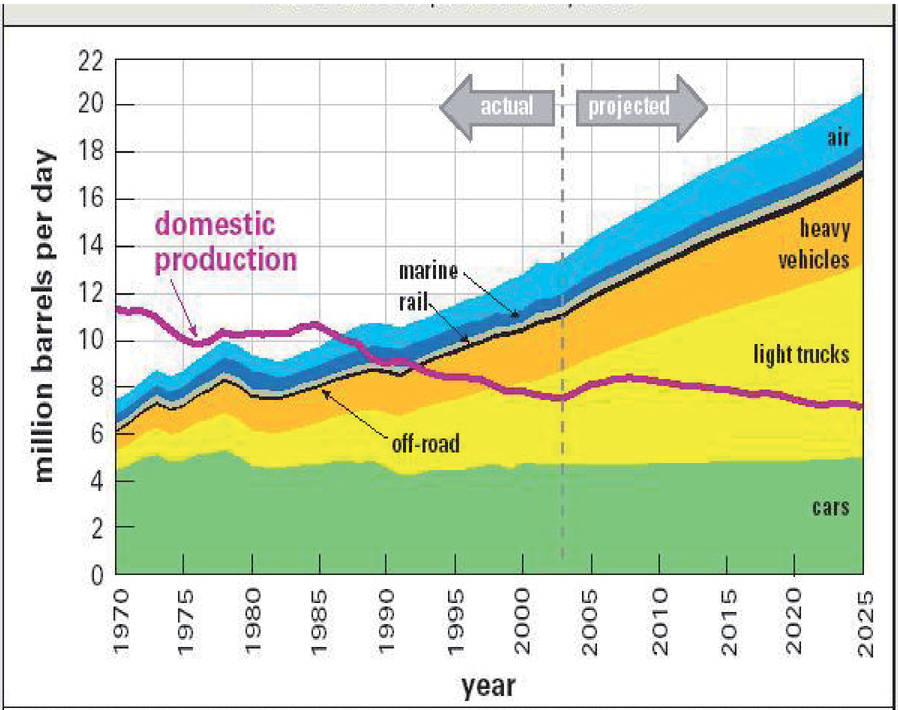


Fig. 2(d). % share of various modes of transportation for inter-city travel in U.S. [1].



Schafer A., J.B. Heywood, H.D. Jacoby, I.A. Waitz, 2009. Transportation in a Climate-Constrained World, MIT Press

Fig. 3. Travel demand/capita with increase in GDP/capita of nations [1].



Source: *Transportation Energy Data Book: Edition 23*, DOE/ORNL-6970, October 2003; and *EIA Annual Energy Outlook 2004*, January 2004.

Fig. 4. Fuel consumption in U.S by transport vehicles [2].

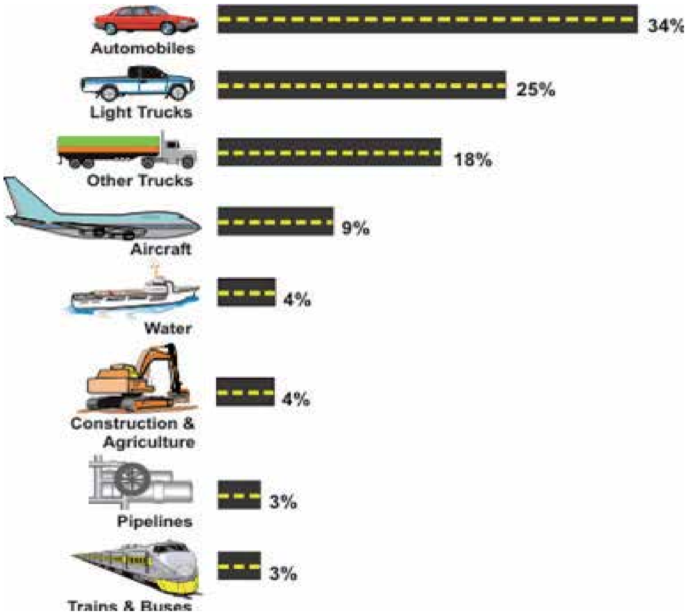
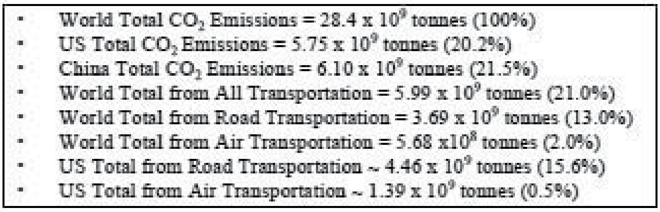


Fig. 5. Relative fuel consumption in U.S by various categories of vehicles [2].



- World Total CO_2 Emissions = 28.4 x 10^9 tonnes (100%)
- US Total CO_2 Emissions = 5.75 x 10^9 tonnes (20.2%)
- China Total CO_2 Emissions = 6.10 x 10^9 tonnes (21.5%)
- World Total from All Transportation = 5.99 x 10^9 tonnes (21.0%)
- World Total from Road Transportation = 3.69 x 10^9 tonnes (13.0%)
- World Total from Air Transportation = 5.68 x10^8 tonnes (2.0%)
- US Total from Road Transportation ~ 4.46 x 10^9 tonnes (15.6%)
- US Total from Air Transportation ~ 1.39 x 10^9 tonnes (0.5%)

Table 1. Current level of CO_2 emissions from air and ground transportation [3].

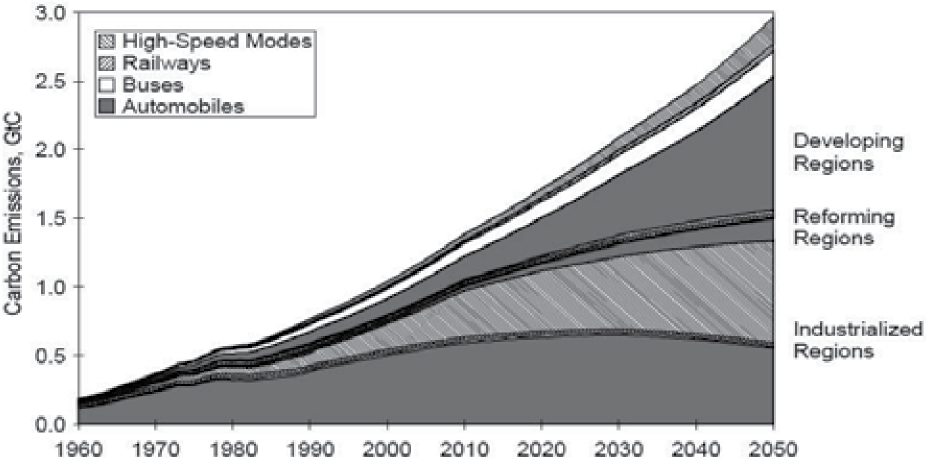


Fig. 6. CO_2 emissions due to world passenger travel in Business as Usual (BAU) scenario [3].

to air transportation and its impact on environment. The environmental issues such as noise, emissions and fuel burn (consumption), for both airplane and airport operations, are discussed in the context of energy and environmental sustainability. They are followed by the topics dealing with noise and emissions mitigation by technological solutions including new aircraft and engine designs/technologies, alternative fuels, and materials as well as examination of aircraft operations logistics including Air-Traffic Management (ATM), Air-to-Air Refueling (AAR), Close Formation Flying (CFF), and tailored arrivals to minimize fuel burn. The ground infrastructure for sustainable aviation, including the concept of 'Sustainable Green Airport Design' is also covered.

As mentioned in the 'Introduction', in the next few decades, air travel is forecast to experience the fastest relative growth among all modes of transportation, especially due to many fold increase in demand in major developing nations of Asia and Africa. Based on these demands for air travel, Boeing has determined the outlook for airplane demand by 2025 as shown in Figure 7 [5]. Figure 8 shows various categories of 27,200 airplanes that would be needed by 2025 [5]. The total value of new airplanes is estimated at $2.6 trillion. As a result of three fold increase in air travel by 2025, it is estimated that the total CO_2 emissions due to commercial aviation may reach between 1.2 billion tonnes to 1.5 billion tonnes annually by 2025 from its current level of 670 million tonnes. The amount of nitrogen oxides around airports, generated by aircraft engines, may rise from 2.5 million tonnes in 2000 to 6.1 million tonnes by 2025. The number of people who may be seriously affected by aircraft noise may rise from 24 million in 2000 to 30.5 million by 2025. Therefore there is urgency to address the problems of emissions and noise abatement through technological innovations in design and operations of the commercial aircraft.

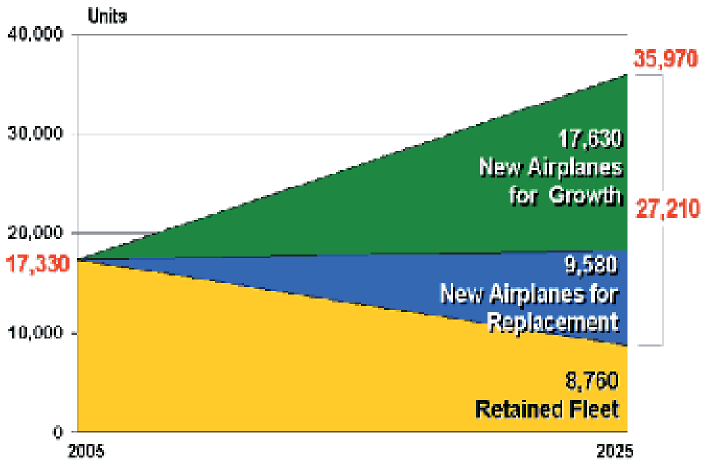


Fig. 7. Boeing market forecast for new airplanes [5].

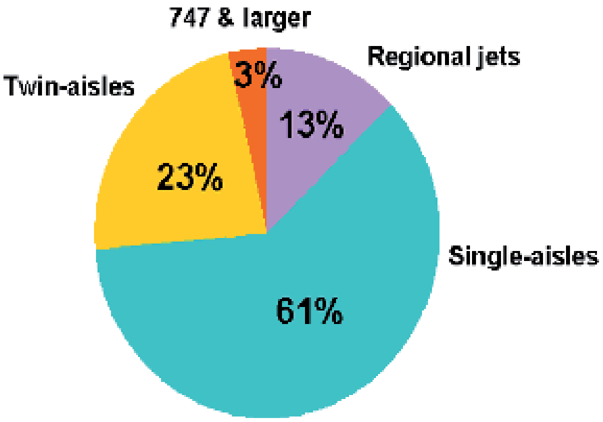
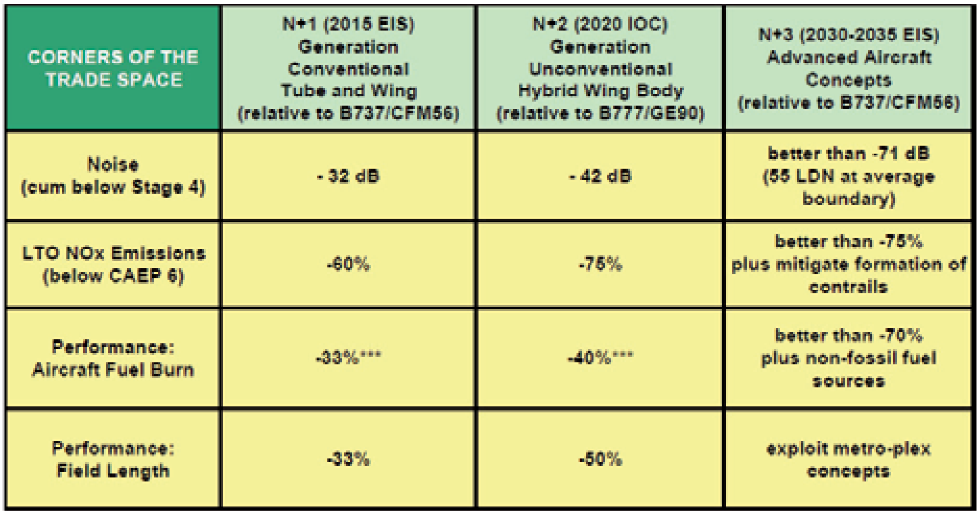


Fig. 8. Boeing demand forecast for various types of Airplanes by 2025 [5].

2.1 Environmental challenges

To meet the environmental challenges of the 21st century, as a result of growth in aviation, the Advisory Committee for Aeronautical Research in Europe (ACARE) has set the following three goals for reducing noise and emissions by 2020; (a) reduce the perceived noise to one half of current average levels, (b) reduce the CO_2 emissions per passenger kilometer (PKM) by 50%, and (c) reduce the NOx emissions by 80% relative to 2000 reference [6]. NASA has similar objectives for 2020 as shown in Figure 9 for N+2 generation aircraft [7]. It is expected that the technology readiness level (TRL) of N+1, N+2 and N+3 generation will be between 4 and 6 in 2015, 2020 and 2030 timeframes respectively. The

NASA definitions of TRL are given in Reference [8]. TRL 4-6 implies that the key technologies readiness will be somewhere between component/subsystem validation in laboratory environment to system/subsystem model or prototyping demonstration in a relevant environment.



CORNERS OF THE TRADE SPACE	N+1 (2015 EIS) Generation Conventional Tube and Wing (relative to B737/CFM56)	N+2 (2020 IOC) Generation Unconventional Hybrid Wing Body (relative to B777/GE90)	N+3 (2030-2035 EIS) Advanced Aircraft Concepts (relative to B737/CFM56)
Noise (cum below Stage 4)	- 32 dB	- 42 dB	better than -71 dB (55 LDN at average boundary)
LTO NOx Emissions (below CAEP 6)	-60%	-75%	better than -75% plus mitigate formation of contrails
Performance: Aircraft Fuel Burn	-33%***	-40%***	better than -70% plus non-fossil fuel sources
Performance: Field Length	-33%	-50%	exploit metro-plex concepts

*** An additional reduction of 10% may be possible through improved operational capability; metro-plex concepts will enable optimal use of runways at multiple airports within the metropolitan area

Fig. 9. NASA subsonic fixed wing system level metric for improving noise, emission and performance using technology & operational improvements [7].

The achievement of these goals will not be easy; it will require the cooperation and involvement of airplane manufactures, airline industry, regulatory agencies such as ICAO and FAA, R & D organizations, as well as political will by many governments and support of public. However, these challenges can be met with concerted efforts as stated beautifully by the Chairman, President and CEO of Boeing Company, W. J. McNerney, “Just as employees mastered "impossible" challenges like supersonic flight, stealth, space exploration and super-efficient composite airplanes, now we must focus our spirit of innovation and our resources on reducing greenhouse- gas emissions in our products and operations.”

2.2 A List of new technologies and operational improvements for green aviation

Recently, Aerospace International, published by the Royal Aeronautical Society of U.K., has identified 25 new technologies, initiatives and operational improvements that may make air travel one of the greenest industries by 2050 [9]. These 25 green technologies/concept areas are listed below from Reference [9].

1. “*Biofuels* – These are already showing promise; the third generation biofuels may exploit fast growing algae to provide a drop-in fuel substitute.
2. *Advanced composites* – The future composites will be lighter and stronger than the present composites which the airplane manufacturers are just learning to work with and use.

3. *Fuel cells* - Hydrogen fuel cells will eventually take over from jet turbine Auxiliary Power Units (APU) and allow electrics such as in-flight entertainment (IFE) systems, galleys etc. to run on green power.
4. *Wireless cabins* – The use of Wi-Fi for IFE systems will save weight by cutting wiring - leading to lighter aircraft.
5. *Recycling* - Initiatives are now underway to recycle up to 85% of an aircraft's components, including composites - rather than the current 60%. By 2050 this could be at 95%.
6. *Geared Turbofans (GTF)* - Already under testing, GTF could prove to be even more efficient than predicted, with an advanced GTF providing 20% improvement in fuel efficiency over today's engines.
7. *Blended wing body aircraft* - These flying wing designs would produce aircraft with increased internal volume and superb flying efficiency, with a 20-30% improvement over current aircraft.
8. *Microwave dissipation of contrails* – Using heating condensation behind the aircraft could prevent or reduce contrails formation which leads to cirrus clouds.
9. *Hydrogen-powered aircraft* - By 2050 early versions of hydrogen powered aircraft may be in service - and if the hydrogen is produced by clean power, it could be the ultimate green fuel.
10. *Laminar flow wings* – It has been the goal of aerodynamicists for many decades to design laminar flow wings; new advances in materials or suction technology will allow new aircraft to exploit this highly efficient concept.
11. *Advanced air navigation* - Future ATC/ATM systems based on Galileo or advanced GPS, along with international co-operation on airspace, will allow more aircraft to share the same sky, reducing delays and saving fuel.
12. *Metal composites* - New metal composites could result in lighter and stronger components for key areas.
13. *Close formation flying* - Using GPS systems to fly close together allows airliners to exploit the same technique as migrating bird flocks, using the slip-stream to save energy.
14. *Quiet aircraft* - Research by Cambridge University and MIT has shown that an airliner with imperceptible noise profile is possible - opening up airport development and growth.
15. *Open-rotor engines* - The development of the open-rotor engines could promise 30%+ breakthrough in fuel efficiency compared to current designs. By 2050, coupled with new airplane configurations, this could result in a total saving of 50%.
16. *Electric-powered aircraft* - Electric battery-powered aircraft such as UAVs are already in service. As battery power improves one can expect to see batteries powered light aircraft and small helicopters as well.
17. *Outboard horizontal stabilizers (OHS) configurations* – OHS designs, by placing the horizontal stabilizers on rear-facing booms from the wingtips, increase lift and reduce drag.
18. *Solar-powered aircraft* - After UAV applications and the Solar Impulse round the world attempt, solar-powered aircraft could be practical for light sport, motor gliders, or day-VFR aircraft. Additionally, solar panels built into the upper surfaces of a Blended-Wing-Body (BWB) could provide additional power for systems.
19. *Air-to-air refueling of airliners* - Using short range airliners on long-haul routes, with automated air-to-air refueling could save up to 45% in fuel efficiency.

20. *Morphing aircraft* - Already being researched for UAVs, morphing aircraft that adapt to every phase of flight could promise greater efficiency.
21. *Electric/hybrid ground vehicles* – Use of electric, hybrid or hydrogen powered ground support vehicles at airports will reduce the carbon footprint and improve local air quality.
22. *Multi-modal airports* - Future airports will connect passengers seamlessly and quickly with other destinations, by rail, Maglev or water, encouraging them to leave cars at home.
23. *Sustainable power for airports* - Green airports of 2050 could draw their energy needs from wave, tidal, thermal, wind or solar power sources.
24. *Greener helicopters* - Research into diesel powered helicopters could cut fuel consumption by 40%, while advances in blade design will cut the noise.
25. *The return of the airship* - Taking the slow route in a solar-powered airship could be an ultra 'green' way of travel and carve out a new travel niche in 'aerial cruises', without harming the planet."

Some of the ideas listed above require technological innovation in aircraft design and engines, use of alternative fuels and materials while others require operational improvement. Some concepts such as electric, solar and hydrogen powered aircraft are currently feasible but are unlikely to become viable for mass air transportation by 2050. In what follows, we describe the current levels of noise, CO_2 and NOx emissions due to air transportation and possible strategies for their mitigation to achieve the ACARE and NASA goals.

2.3 Noise & its abatement

Historically, the reduction in airplane noise has been a major focus of airplane manufacturers because of its health effects and impact on the quality of life of communities, especially in the vicinity of major metropolitan airports. As a result, there has been a significant progress in achieving major reduction in noise levels of airplanes in past five decades as shown in Figure 10 [10]. These gains have been achieved by technological innovations by the manufacturers in reducing the noise from airframe, engines and undercarriage as well as by making changes in the operations. Worldwide, there has been ten fold increases in number of airports since the 1970s that now impose the noise related restrictions as shown in Figure 11 [11]. The airports have imposed operating restrictions and also there has been special attention paid to the planning, development and management of airports for sustainability. Since 1980, FAA has invested over $5billion in airport noise reduction.

In recent years, the joint MIT/Cambridge University project on "Silent Aircraft" has produced an innovative aircraft/engine design, shown in Figure 12 that has imperceptible noise outside an urban airport [12]. In order to meet the ACARE and NASA goals of reducing the perceived noise by 50% of the current level by 2020, several new technology ideas are being investigated by the airplane and engine manufacturers to both reduce and shield the noise sources as shown in Figure 13 in the chart by Reynolds [13]. The most promising for the near future are the chevron nozzles, shielded landing gears and the ultra high bypass engines with improved fan (geared fan and contra fan) and fan exhaust duct-liner technology. In addition, new flight path designs in ascent and descent flight can reduce the perceived noise levels in the vicinity of the airports.

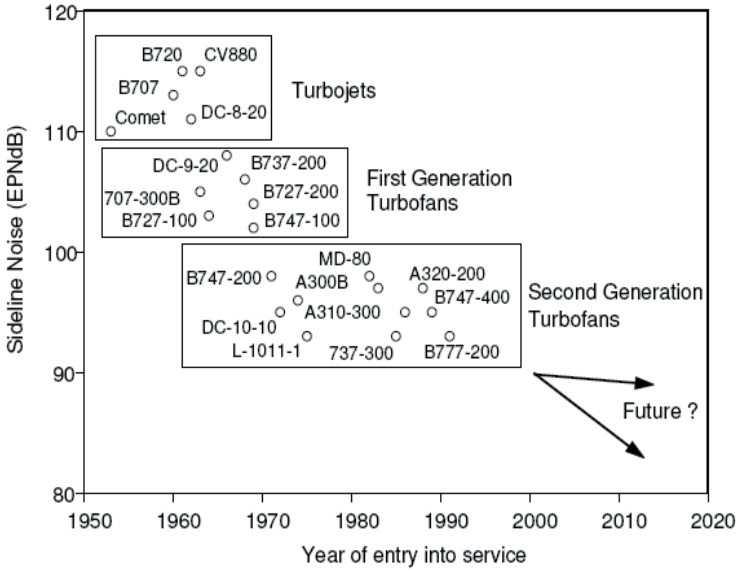


Fig. 10. Reductions in noise levels of aircrafts in past thirty years [11].

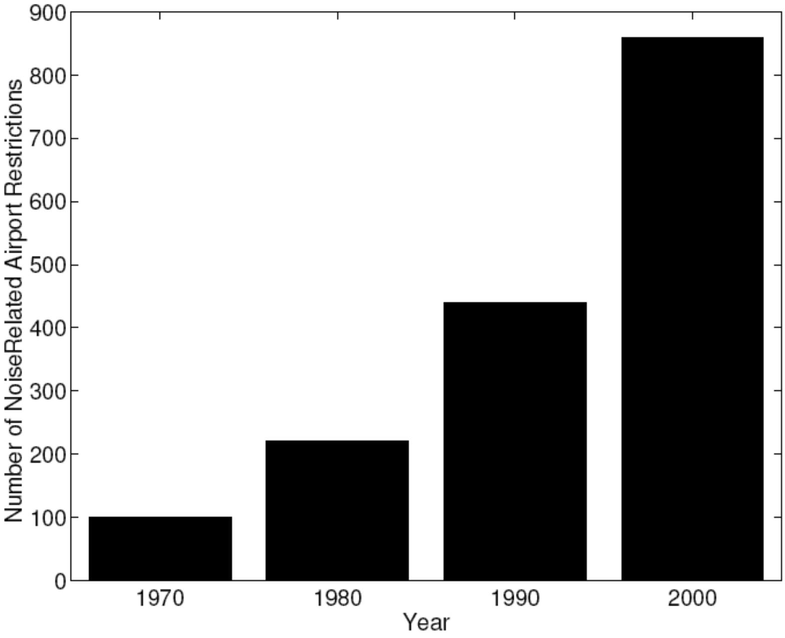


Fig. 11. Number of airports with noise related restrictions in past fifty years [10].

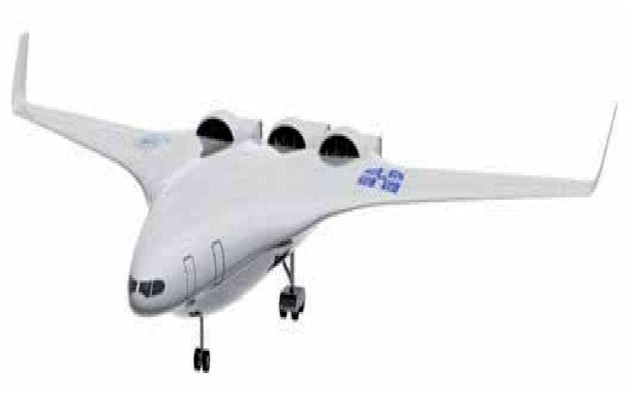


Fig. 12. Silent aircraft SAX – 40: (joint MIT/Cambridge University design) [12].

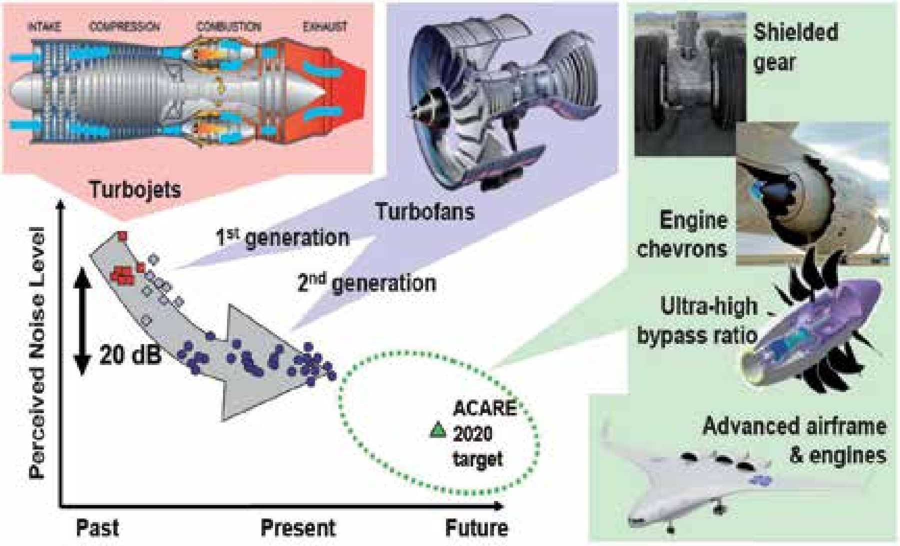


Fig. 13. Evolution of noise reduction technologies [13].

2.4 Emissions and fuel burn

Aviation worldwide consumes today around 238 million tonnes of jet-kerosene per year. Jet-kerosene is only a very small part of the total world consumption of fossil fuel or crude oil. The world consumes 85 million barrels/day in total, aviation only 5 million. At present, aviation contributes only 2-3% to the total CO_2 emissions worldwide [14] as shown in Figure 14. However, it contributes 9% relative to the entire transportation sector. With 2050 forecast of air travel to become 40% of total PKT (Figure 1), it will become a major contributor to GHG emissions if immediate steps towards reducing the fuel burn by innovations in technology and operations, as well as alternatives to Jet-kerosene are not sought and put into effect.

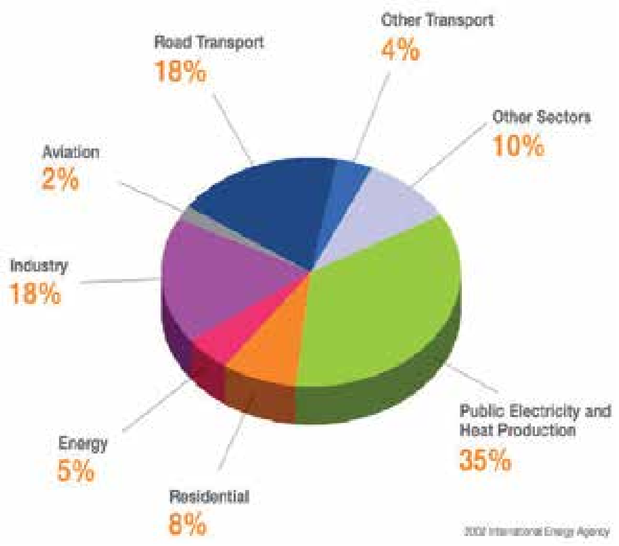


Fig. 14. CO_2 emissions worldwide contributed by various economic sectors [14].

Fig. 15. Contrails & Cirrus Clouds.

Of the exhausts emitted from the engine core, 92% are O_2 and N_2, 7.5% are composed of CO_2 and H_2O with another 0.5% composed of NOx, HC, CO, SOx and other trace chemical species, and carbon based soot particulates. In addition to CO_2 and NOx emissions, formation of contrails and cirrus clouds (Figure 15) contribute significantly to radiative forcing (RF) which impacts the climate change. This last effect is unique to aviation (in contrast to ground vehicles) because the majority of aircraft emissions are injected into the upper troposphere and lower stratosphere (typically 9-13 km in altitude). The impact of burning fossil fuels at 9-13 km altitude is approximately double of that due to burning the same fuels at ground level [15]. The present metric used to quantify the climate impact of aviation is radiative forcing (RF). Radiative forcing is a measure of change in earth's radiative balance associated with atmospheric changes. Positive forcing indicates a net warming tendency relative to pre-industrial times. Figures 16 and 17 show the IPCC (Intergovernmental Panel for Climate Change) estimated increase in total anthropogenic RF due to aviation related emissions (excluding that due to contrails and cirrus clouds) from 1992 to 2050 [16]. It should be noted that in Figures 16 and 17, RF scale is given in W/m2. It is usually given in mW/m2; then the numbers in Figures 16 and 17 should be multiplied by 1000 as shown. The horizontal line in Figures 16 and 17 is indicative of the current level of scientific understanding of the impact of each exhaust species.

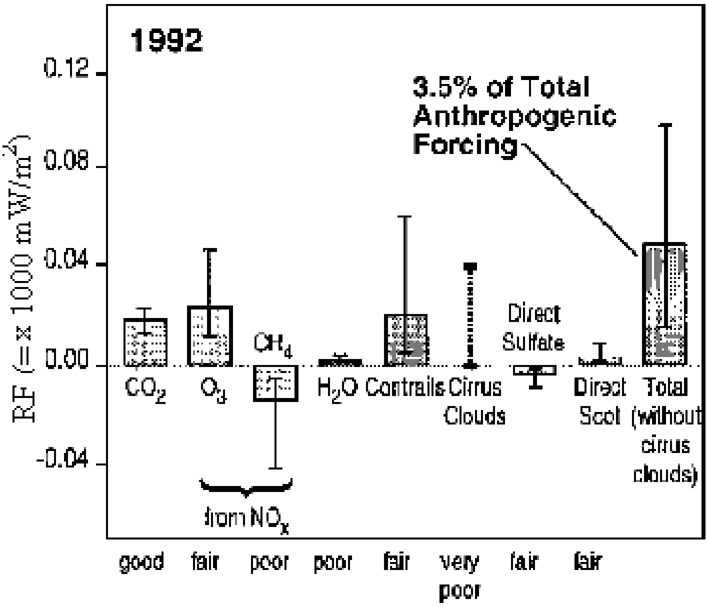


Fig. 16. IPCC estimated Radiative Forcing (RF) due to Emissions – 1992 [16].

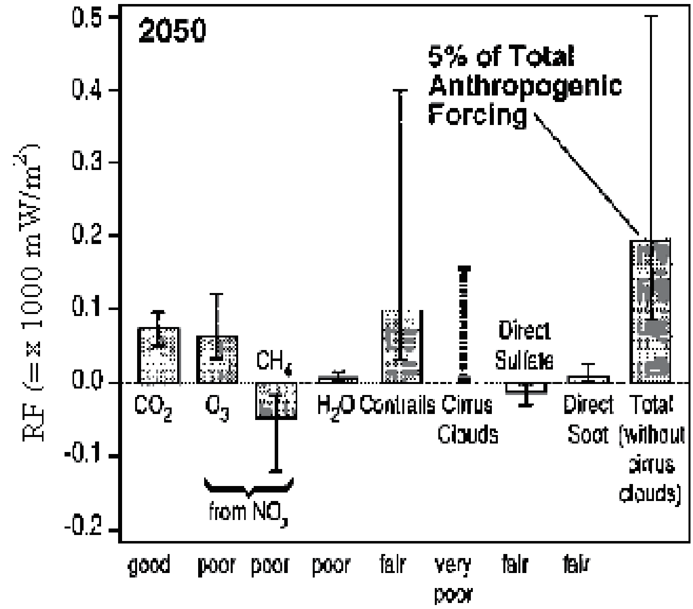


Fig. 17. IPCC estimated Radiative Forcing (RF) due to emissions – 2050 [16].

It should be noted that the RF estimates for 2050 in Figure 17 are based on several assumptions about the growth in aviation, state of technology etc. which are most likely to change. Based on the RF estimates shown in Figures 16 and 17, aviation is expected to account for 0.05K of the 0.9K global mean surface temperature rise expected to occur between 1990 and 2050 [15]. However, RF is not a good metric for weighing the relative importance of short-lived and long-lived emissions. Most importantly, the range of uncertainty about the climate impact of contrails and cirrus cloud remains substantial. According to recent IPCC report, the best estimates for RF in 2005 from linear contrails were 10 (3-30)mW/m^2 and 30(10-80)mW/m^2 from total aviation induced cloudiness, the numbers in bracket give the range of the 2/3 confidence limit [17]. As noted in Reference [17], "the tradeoff estimate of the CO_2 RF in 2000 was 23.5mW/m^2. Despite the growth in CO_2 RF between 2000 and 2005, aviation induced cloudiness remains the greatest contributor to RF according to these estimates. Because of doubts of RF as a metric as well as data spread in cloudiness related RF, the relative contribution of the two (CO_2 and cloudiness) to climate change can not be ascertained with confidence at present time. However, the atmospheric conditions under which an aircraft will generate a persistent contrail – the Schmidt-Appleman criterion [18] – are well understood and can be predicted accurately for a particular aircraft.

Currently there is no technological fix to prevent contrail formation if the atmospheric conditions and engine exhaust characteristics satisfy the Schmidt-Appleman criterion. One assured way of reducing the persistent contrail formation is to reduce aircraft traffic through regions of supersaturated air in which the persistent contrail can form, by flying under, over or around these regions. However, this approach may not be acceptable commercially because of increase in fuel burn, disruption in airline schedule, added ATM workload, and additional operating costs as well as increase in CO_2 and NOx emissions. Because contrail reduction involves an increase in CO_2 and NOx emissions, the best environmental solution is not the complete avoidance of contrails, but a balanced result that minimizes climate impact. This requires a better understanding of the relationship between the properties of the atmosphere (temperature, humidity etc.), the size of the aircraft, the quantity of its emissions (water and particulates), and extent of the persistent contrail and subsequent cirrus formation that results. The adoption of synthetic kerosene produced by Fischer-Tropsch or some similar process offers the prospect of substantial reduction in sulfate and black carbon particulate emissions. This is likely to reduce the extent of contrail and cirrus formation, but the extent of reduction as well as to what extent it would reduce the fuel burn penalty of operational avoidance measures requires further research. Based on the current status, it appears that fuel additives do not offer a significant reduction in contrail formation. The contrail avoidance measures e.g. making modest changes in altitude can reduce contrail formation appreciably with a small penalty in additional fuel burn." Increasing the cruise altitude and higher engine pressure ratio can reduce CO, HC, and CO_2 emissions as well as decrease the fuel burn (improve the fuel efficiency) and facilitate noise reduction. Since higher pressure ratio requires higher flame temperature, the NOx formation rate increases. On the other hand, decreasing the cruise altitude and reducing the engine overall pressure ratio can reduce the NOx but increase the CO_2 emissions. This should be an important consideration in the optimization of future aircraft and engine designs. Research is needed in understanding the impact of cruise altitude on climate. *In addition, there is a need for new optimized aircraft and engine designs that provide a compromise* *between minimizing the fuel burn and reducing the climate impact.* The lower NOx emissions can possibly be achieved by new combustor concepts such as flameless catalytic combustor and technological improvements in fuel/air mixers using alternative fuels (biofuels), aided by active combustion control. These concepts/technologies should make it possible to meet the N+1 and N+2 generation goals (Figure 9) of achieving the LTO NOx reductions by 60% and 75% respectively below the ICAO standard adapted at CAEP 6 (Committee on Aviation Environmental Protection). It should result in reducing the steepness of the trade-off between NOx and CO_2 emissions and should therefore also help in making a significant contribution to the aircraft performance goal by reducing the fuel burn by 33% and 40% for the N+1 and N+2 generation aircraft respectively. Thus, there are three key drivers in emissions reductions as shown in Figure 18 [19]: (a) innovative engine technologies and aircraft designs, (b) the improvement in ATM and operations, and (c) the alternative fuels e.g. biofuels. The three-prong approach can achieve the goals enunciated by ACARE and NASA by 2020 and beyond. These are discussed in next few sections.

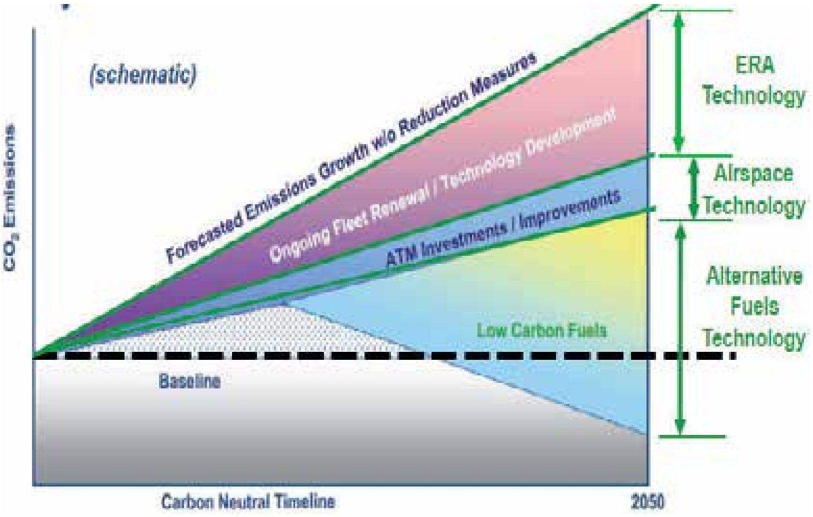


Fig. 18. Key drivers for emissions reductions [19].

2.5 Innovative engine technologies

In cruise condition, the amount of fuel burn varies in inverse proportion to propulsion efficiency and lift-to-drag ratio. Aircraft and engine manufacturers in U.S. and Europe along with several research organizations are developing new engine technologies aimed at improving the propulsion efficiency to reduce the fuel burn and also to simultaneously reduce NOx emissions and noise. The greatest gains in fuel burn reduction in the past sixty years (since the appearance of jet engine) have come from better engines. The earliest engines were turbojets in which all the air sucked in at the front is compressed, mixed with fuel and burned, providing thrust through a jet out the back (see Figure 13). Afterwards, more efficient turbofans were designed when it was realized that greater engine efficiency could be achieved by using some of the power of the jet to drive a fan that pushes some of the intake air through ducts around the core (see Figure 13). Other boosts in efficiency have come from better compressors and materials to let the core burn at higher pressure and temperature. As a result, according to International Airport Transport Association (IATA), new aircraft are 70% more fuel efficient than they were forty years ago. In 1998, passenger aircraft averaged 4.8 liters of fuel/100km/passenger; the newest aircraft – Airbus A380 and Boeing B787 use only three liters. Figure 19 shows the relative improvement in fuel efficiency of various aircraft engines since 1955 [20]. The current focus is on making turbofans even more efficient by leaving the fan in the open. Such a ductless "open rotor" design (essentially a high-tech propeller) would make larger fans possible; however one may need to address the noise problem and how to fit such engines on the airframe. In the short-to-medium-haul market, where most fuel is burned, the open rotor offers an appreciable reduction in fuel burn relative to a turbofan engine of comparable technology, but at the expense of some reduction in cruise Mach number. It is worth noting here that in mid 1980's GE invested significant effort in advanced turbo-prop technology (ATP). The unducted fan (UDF) on a GE36 ultra high bypass (UHB) engine on MD-81 at Farnborough air show in 1988 (Figure 20 [21]) created enormous buzz in the air transportation industry. The author of this paper was at McDonnell Douglas during that period and played a small role in the airframe – engine integration study of MD81 with GE36 ATP. However, in spite of its potential for 30% savings in fuel consumption over existing turbofan engines with comparable performance at speeds up to Mach 0.8 and altitudes up to 30,000 ft, for a variety of technical and business reasons, the advanced turboprop concept never quite got-off the ground [22].

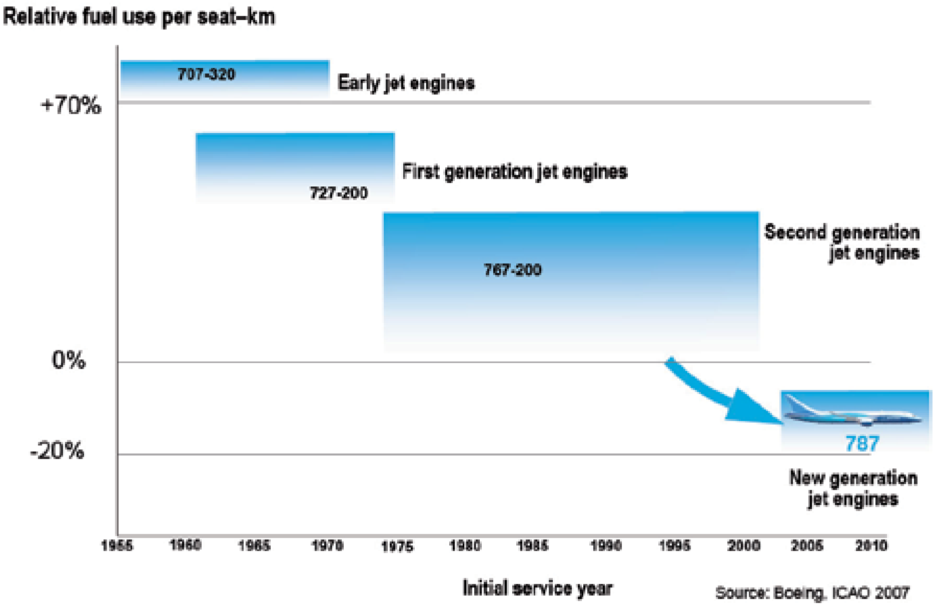


Fig. 19. Relative improvement in fuel efficiency of various aircraft engines from 1955 to 2010 [20].

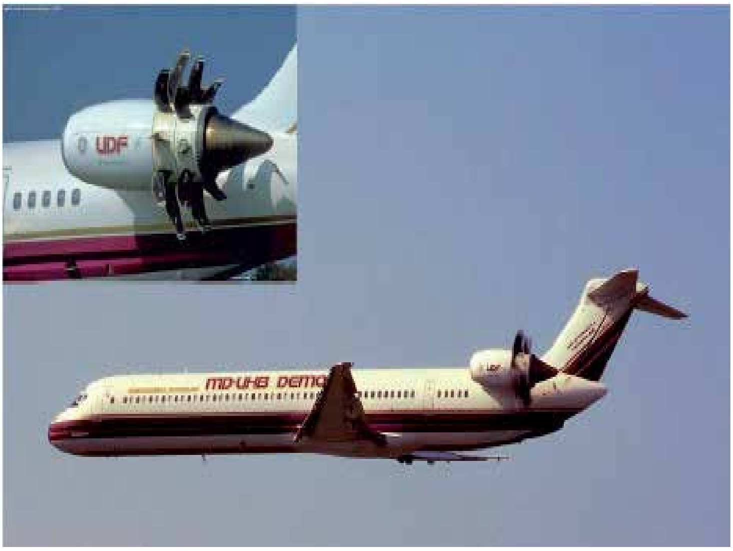


Fig. 20. GE36 Turbo-Prop demonstrator engine on MD-81 aircraft [21].

At present in Europe, under the auspices of NACRE (New Aircraft Concept Research Europe), Rolls-Royce and Airbus are making a joint study of the open rotor configurations (Figure 21), including wind-tunnel investigations of power plant installation effects. A key issue in future engine design is how to balance the conflicting aims of reducing fuel burn and NOx emissions (along with the other conflicting aims of reducing noise, weight, initial investment cost and maintenance cost). The results of these types of current and future projects should provide a sounder basis for making decisions between turbofan and open rotor engines for future aircraft. They should also take engine technology well towards its contribution to the goal of a 20% improvement in the installed engine fuel efficiency by 2020.

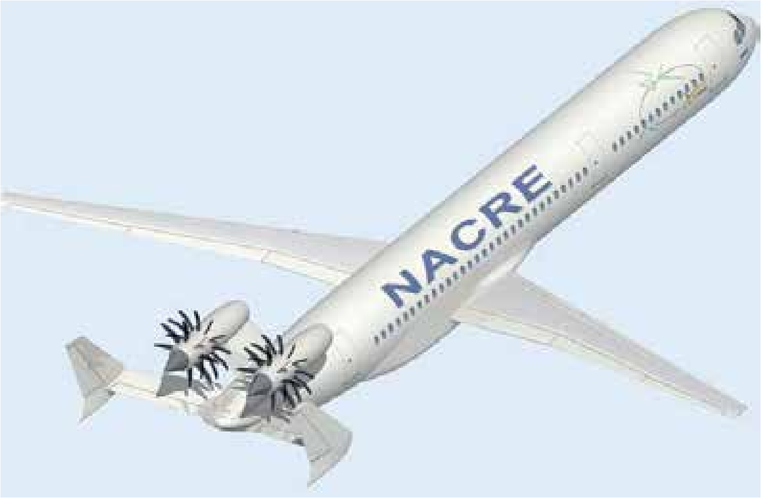


Fig. 21. Open-Rotor version of pro-active Green Aircraft in NACRE study [17].

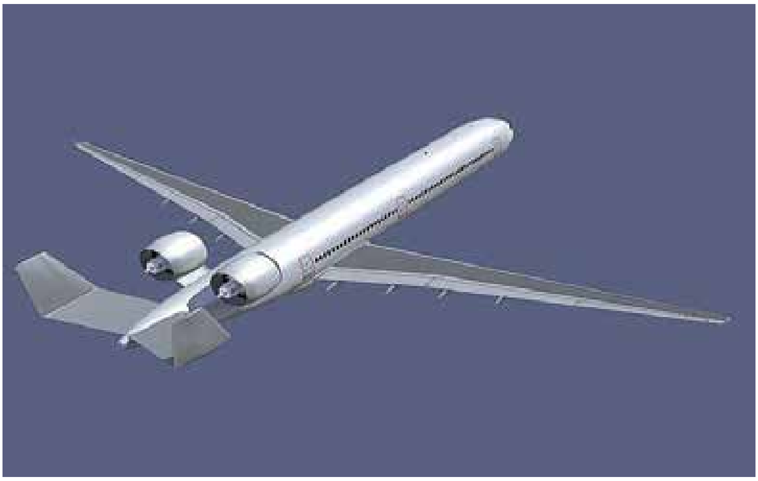


Fig. 22. Turbofan version of pro-active Green Aircraft in NACRE study [17].

2.6 Innovative aircraft designs

As noted in Reference [17], "the classic swept-winged aircraft with a light alloy structure has been evolving for some sixty years and the scope for increasing its lift-to-drag ratio (L/D), if its boundary layers remain fully turbulent, is by now exceedingly limited. Nevertheless, it is well established that increasing L/D is one of the most powerful means of reducing fuel burn. The three ways of increasing L/D are to (a) increase the wing span, (b) reduce the vortex drag factor κ and (c) reduce the profile drag area. The vortex drag factor is a measure of the degree to which the span-wise lift distribution over the wing departs from the theoretical ideal. Current swept-wing aircraft are highly developed and there is little scope for further improvement. A flying wing may enable some additional small reduction in κ, however realistically; there is no real prospect of a significant reduction in fuel burn by altering span-wise loading distributions. Furthermore, increasing the wing span increases wing weight. Current long-range aircraft are optimized to minimize the fuel burn at current cruise Mach numbers. In a successful design the balance between the wing span and wing weight is close to optimum. However, the change to advanced composite materials for the wing structure should result in an optimized wing of greater span; both the B787 and Airbus A350 reflect this. If cruise Mach number is reduced, reducing wing sweep also enables the wing to be optimized at a greater span. The turbofan version of Pro-Active Green Aircraft (Figure 22) included in the NACRE study features a slightly forward swept wing optimized at a significantly higher than usual span. This aircraft is aimed at an appreciable increase in L/D at the expense of some reduction in cruise Mach number. The third option for increasing L/D is to reduce the profile drag of the aircraft. This is seen as the option with the greatest mid-term and long-term potential. For large aircraft, the adoption of a blended wing-body (BWB) layout reduces profile drag by about 30%, providing an increase of around 15% in L/D (estimates of 15% - 20% have been published)." The work on such configurations, both by Boeing (the X-48B, wind tunnel and flight tested at model scale by NASA [Figure 23]) and by Airbus within the NACRE project are proceeding. At present, the first applications of the Boeing BWB are envisaged to be in military roles or as a freighter, with 2030 suggested as the earliest entry to service date for a civil passenger aircraft.

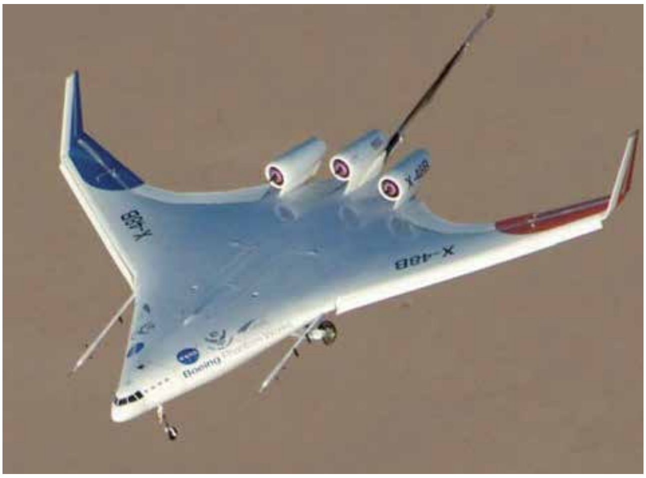


Fig. 23. Boeing/NASA X-48B BWB technology demonstrator aircraft [23].

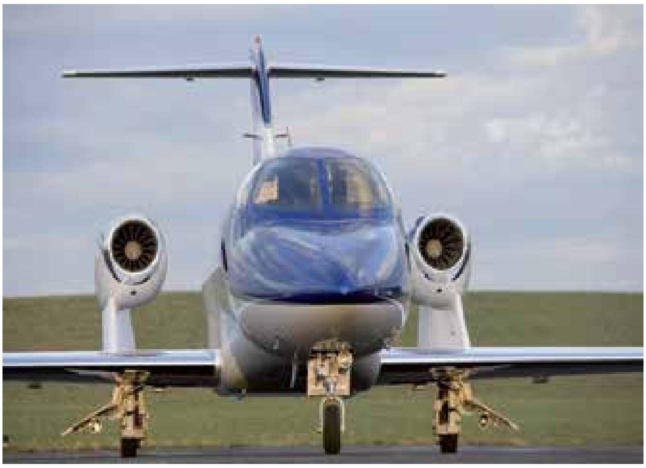
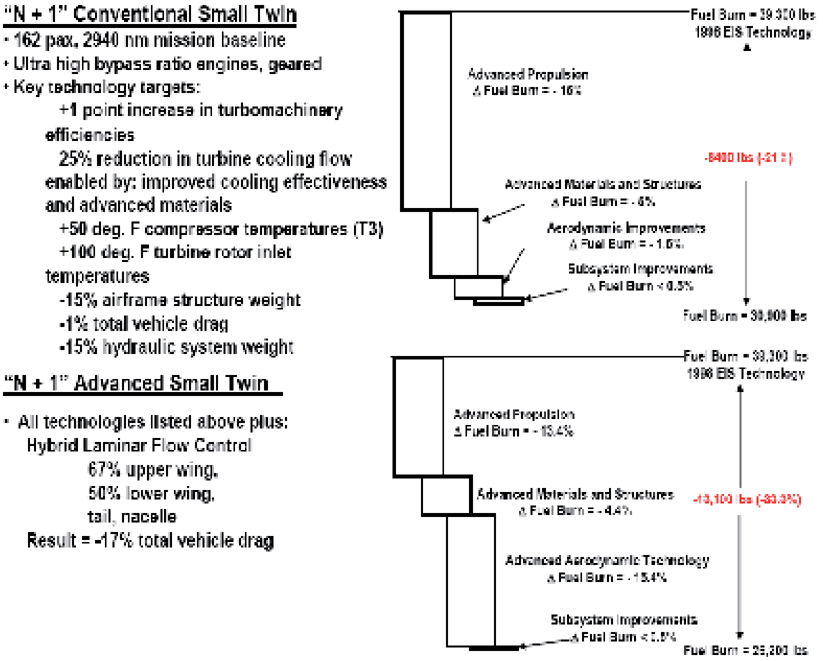


Fig. 24. Honda Jet [24].

The other well known approach of reducing the profile drag is by the use of laminar flow control in one of its three forms - natural, hybrid or full. Natural laminar flow control was applied with great success in World War II on the P-51 Mustang fighter to give it an exceptional range. As a result there was significant effort devoted to the development of laminar flow airfoils after the end of World War II. In these airfoils, the reduction in friction drag was achieved by moving the transition farther back on the airfoil. In addition, the location of the maximum airfoil thickness was at about 60% of the chord which moved the shock system farther back and reduced the effects of boundary layer thickening and separation caused by it. However in spite of a large number of studies, the success in the laboratory in reducing the drag was never realized on medium size aircraft with swept wings. Therefore, its application has been restricted by a combination of size and wing sweep either to small aircraft with swept wings or medium-sized aircraft with zero or very little sweep. The Pro-Active Green Aircraft in the NACRE project (Figures 21 & 22) is designed to exploit natural laminar flow control and has slightly swept forward wings, to avoid contamination of the flow over the wing by the turbulent boundary layer on the fuselage. "Hybrid laminar flow control employs suction over the forward upper surface of the wing to stabilize the boundary layer. This enables the drag reducing principles that underlie natural laminar flow control to be applied to larger, swept-winged aircraft up to typically the size of the A310. The use of suction to maintain laminar flow over the first half of an airfoil surface has been successfully demonstrated in flight on a B757 wing and an A320 fin. The aerodynamic principles are well understood but the engineering of efficient, reliable, lightweight suction systems requires further work. Thereafter, demonstration of the practicality of the system and assessment of the maintenance and other operational problems that it may encounter will require an extended period of operational validation. The application of suction to maintain laminar flow over the entire surface of a flying wing airliner was proposed by Handley Page in the early 1960s. The proposal was based on the substantial body of research into full laminar flow control, including flight demonstrations, over the preceding decade. Full laminar flow control may have potential to double L/D relative to current standards [17]." Recently unveiled "Honda Jet" (Figure 24) has combined several innovative aircraft and engine design features, namely a combination of over the wing (OTW) engine mount design, natural laminar flow wing (NLF), all composite fuselage, HF – 120 turbofan engine, which give it a 30-35% more fuel efficiency and higher cruise speed than conventional light business jets. This is the range of efficiency that can be achieved for the N+1 generation conventional tube and wing aircraft by 2015. Saeed et al. [25] have recently conducted the conceptual design study of a Laminar Flying Wing (LFW) aircraft capable of carrying 120 passengers. They have estimated that, subject to the constraint of a low cruise Mach number of 0.58, LFC has the potential to reduce aircraft fuel-burn by just over 70%, to about 6 gram per passenger-km (PKM), with a trans-Atlantic range of 4125 nautical miles. Studies of this nature do show the promise of innovative aircraft designs to reduce the fuel burn.

Figure 9 shows the NASA goals of achieving a 33% and 40% reduction in fuel burn for N+1 and N+2 generation aircrafts respectively by using the advanced propulsion technologies, advanced materials and structures, and by improvements in aerodynamics and subsystems. Collier [26] from NASA Langley has provided a detailed outline as to how such savings in fuel burn can be achieved. He has estimated that for a N+1 generation conventional small twin aircraft (162 passengers and 2940nm range), 21% reduction in fuel burn can be achieved by using advanced propulsion technologies, advanced materials and structures, and by improvements in aerodynamics and subsystems. For an advanced small twin, additional 12.3% savings in fuel burn can be achieved by using hybrid laminar flow control as shown in Figure 25.

For a N+2 generation aircraft (300 passengers and 7500 nm range) flying at cruise Mach of 0.85, 40% saving in fuel burn relative to baseline B777-200ER/GE90 can be achieved by a combination of hybrid wing-body configuration (with all composite fuselage), advanced engine and airframe technologies, embedded engines with BLI inlets and laminar flow as shown in Figure 22 [24]. For the baseline aircraft, the fuel burn at Mach 0.85 with 300 passengers for a 7500nm mission range is 237,000 lbs. The N+2 generation aircraft should require 141,100lbs of fuel. As discussed in next few sections, additional savings of 10% in fuel burn can be achieved by operational improvements.



"N + 1" Conventional Small Twin

- 162 pax, 2940 nm mission baseline
- Ultra high bypass ratio engines, geared
- Key technology targets:
  - +1 point increase in turbomachinery efficiencies
  - 25% reduction in turbine cooling flow enabled by: improved cooling effectiveness and advanced materials
  - +50 deg. F compressor temperatures (T3)
  - +100 deg. F turbine rotor inlet temperatures
  - -15% airframe structure weight
  - -1% total vehicle drag
  - -15% hydraulic system weight

"N + 1" Advanced Small Twin

- All technologies listed above plus:
  Hybrid Laminar Flow Control
  67% upper wing,
  50% lower wing,
  tail, nacelle
  Result = -17% total vehicle drag

Fig. 25. Reduction in fuel burn for N+1 generation aircraft relative to baseline B737/CFM56 using advanced technologies [26].

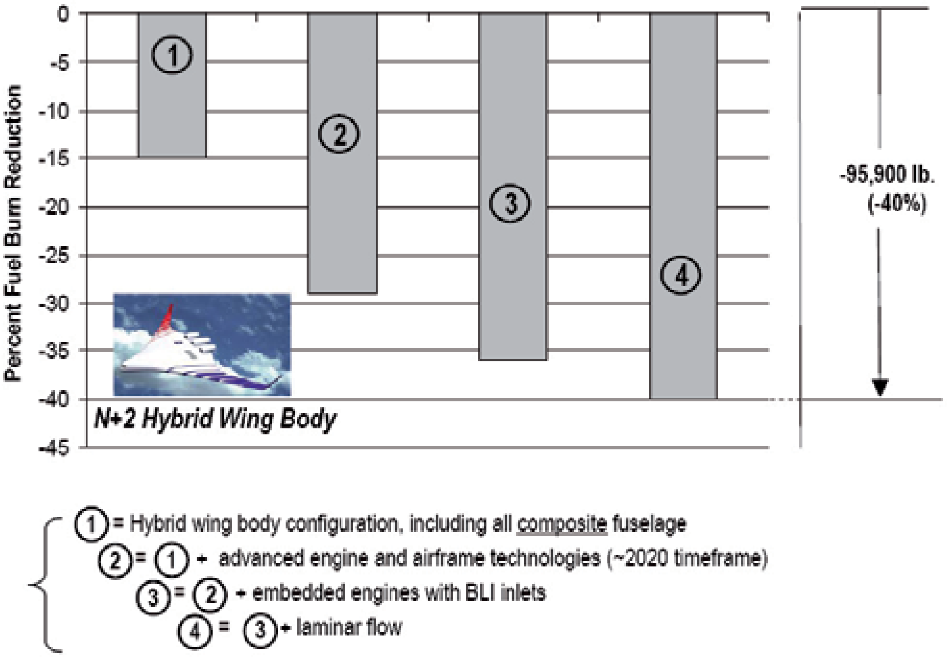


Fig. 26. Reduction in fuel burn for N+2 generation aircraft relative to baseline B777-200ER/GE96 using advanced technologies [26].

2.7 Operational improvements/changes

2.7.1 Improvement in air traffic management (atm) infrastructure

There are many improvements in operations that are being introduced, or will be introduced in the relatively near future that can reduce CO_2 emissions significantly. Foremost among these is the reduction of inefficiencies in ATM, which give rise to routes with dog-legs, stacking at busy airports, queuing for a departure slot with engines running, etc. U.S. Next Generation Air Transportation System (NextGen) architecture and the European air traffic control infrastructure modernization program, SESAR (Single European Sky ATM Research Program), are an ambitious and comprehensive attack on this problem. As described in the U.S. National Academy of Science (NAS) report [27], "NextGen is an example of active networking technology that updates itself with real time-shared information and tailors itself to the individual needs of all U.S. aircraft. NextGen's computerized air transportation network stresses adaptability by enabling aircraft to immediately adjust to ever-changing factors such as weather, traffic congestion, aircraft position via GPS, flight trajectory patterns and security issues. By 2025, all aircraft and airports in U.S. airspace will be connected to the NextGen network and will continually share information in real time to *improve efficiency, safety, and absorb the predicted increase in air transportation.*" Here it is worth noting that operational measures, which can apply to almost the entire world fleet, can have a greater impact, sooner, than the introduction of new aircraft and engine technologies, which can take perhaps 30 years to fully penetrate the world fleet.

2.7.2 Air-to-air refueling (aar) with medium range aircraft for long-haul travel

One particular operational measure that has been advocated is the use of medium-range aircraft, with intermediate stops, for long-haul travel. It has been estimated, using a simple parametric analysis, that undertaking a journey of 15,000km in three hops in an aircraft with design range of 5,000km would use 29% less fuel than doing the trip in a single flight in a 15,000km design. Hahn [28] and Creemers & Slingerland [29] have performed analyses to address this issue using sophisticated aircraft design synthesis methods. Hahn [28], analyzing the assessment for a 15,000km journey in one stage or three, predicted a fuel saving of 29%. Creemers & Slingerland [29], considering a B747-400 (range 13,334km) as the baseline long-range aircraft, designed an aircraft with the same fuselage and passenger capacity (420) but for half the design range (6,672km). This aircraft was predicted to do the long-haul journey in two hops with a 27% fuel saving and at a fuel cost of $70 per barrel, a DOC saving of 9%. Nangia [30] has shown that fuel burn savings of as much as 50% were achievable by using a 5,000km design for a 15,000km journey, since a medium range aircraft can carry a much higher share of their maximum payload as passengers. This difference — which appears essentially to be the difference between medium-range single and long-range twin-aisle aircraft — was not a feature of either the study of Hahn [28] or Creemers & Slingerland [29], which used the same fuselage for both long and medium range designs. This highlights the importance of cabin dimensions and layouts in considering future designs in which, both environmentally and commercially, seat-kilometers per gallon becomes an increasingly important objective. The full system assessment of this proposition, using optimized medium-range aircraft needs further investigation. In order to avoid the intermediate refueling stops, air-to-air refueling (AAR) (Figure 27) has been suggested as a means of enabling medium-range designs to be used on long-haul operations. Nangia has now published a number of papers reporting his work on AAR, which indicate substantial fuel burn savings even after the fuel used by the tanker fleet is taken into account [30, 31].

Fig. 27. Air-to-Air Refueling [30].

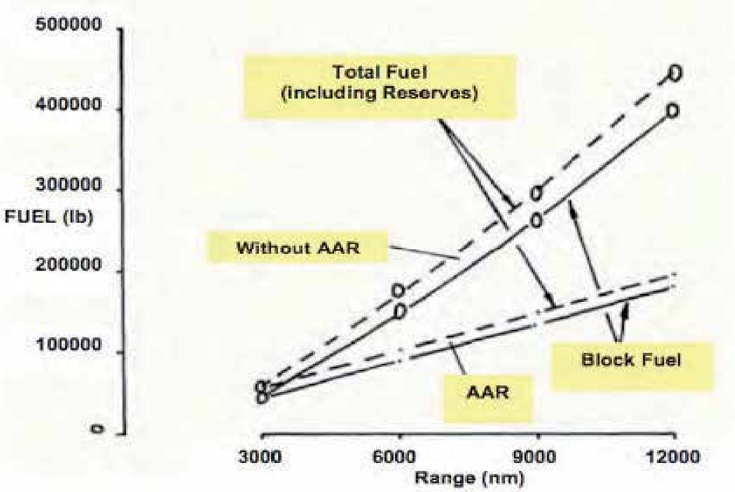


Fig. 28. Savings in fuel burn with Air-to-Air Refuelling (AAR) for long haul flights [31].

Nangia [31] has shown (Figure 28) that an aircraft with *L/D* = 20, would require 46,147 lbs, 161,269 lbs, and 263,073 lbs of fuel to cover a range of 3,000, 6,000 and 9,000 nautical miles (nm) respectively. With AAR, it will require 92,294 lbs and 138, 441 lbs of fuel for a range of 6,000 and 9,000 nm respectively indicating a savings of 43% and 47% in fuel burn relative to that required without AAR. Accounting for the fuel required by the air tanker – 9,000 lbs for one refueling for a range of 6,000nm and 18,000 lbs for two refueling for a range of 9,000nm, the net savings in fuel burn with AAR are 37% and 41% for a range of 6,000nm and 12,000 nm respectively. However it is paramount that with AAR, the absolute safety of the aircraft is assured.

2.7.3 Close Formation Flying (CFF)

The possibility of using CFF to reduce fuel burn or to extend range is well known. As stated by Nangia [31], “aircraft formations (Figure 29) occur for several reasons e.g. during displays or in AAR but they are not maintained for any significant length of time from the fuel efficiency perspective.” The reason is that flying in formation will require extreme safety measures by use of sensors coupled automatically to control systems of individual aircrafts. Furthermore, flying a close formation through clouds or in gusty environment may not be practical. The obvious benefit of flying in formation is a more uniform downwash velocity field, which minimizes the energy transferred into it from propulsive energy consumption. Another benefit is the cancellation of vortices shed from the wing-tips of individual airplanes, except the two outermost ones. How effective this cancellation will be would depend upon the practicality of achievable spacing among the aircrafts. There would also be a substantial benefit in elimination of vortex contrails and cirrus clouds. Recently, NASA conducted tests on two F/A-18 aircraft formations [32]. It was shown that the benefits of CFF occur at certain geometry relationships in the formation, namely the trailing aircraft should overlap the wake of the leading aircraft by 10-15% semi-span in this case. Jenkinson [33] suggested that the CFF of several large aircrafts is more efficient in comparison with flying a very large aircraft. The aircrafts could take-off from different airports and then fly in formation over large distances before peeling off for landing at required destinations. Bower at al. [34] have recently investigated a two aircraft echelon formation and a three aircraft formation of three different aircraft and analyzed the fuel burn. Their study determined the fuel savings and difference in flight times that result from applying CFF to missions of different stage lengths and different spacing between the cities of origin. For a two aircraft formation, the maximum fuel savings were 4% with a tip-to-tip gap between the aircraft equal to 10% of the span and 10% with a tip overlap equal to 10% of the span. For the three aircraft inverted-V formation, the maximum fuel savings were about 7% with tip-to-tip gaps equal to 10% of the span and about 16% with tip overlaps equal to 10% of the span.

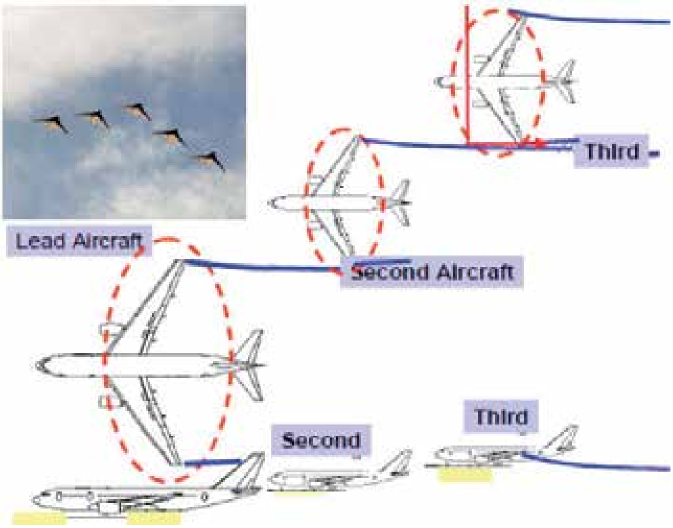


Fig. 29. Three different aircraft type in CFF [31].

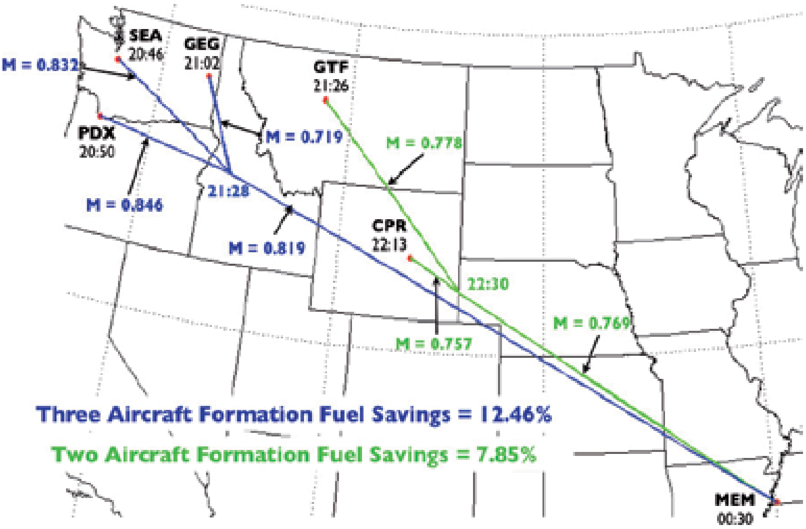


Fig. 30. Five FedEx aircraft in Formation Flight enroute from Pacific Northwest to Memphis [34].

Bower et al. [34] conducted a case study to examine the effect of formation flight on five FedEx flights from the Pacific Northwest to Memphis, TN. The purpose of this study was to quantify the fuel burn reduction achievable in a commercial setting without changing the flight schedule. With tip-to-tip gaps of about 10% of the span it was shown that fuel savings of approximately 4% could be achieved for the set of five flights. With a tip-to-tip overlap of about 10% of the span the overall fuel savings were about 11.5% if the schedule was unchanged. This translated into saving of approximately 700,000 gallons of fuel per year for this set of five flights. Figure 30 shows the three types of aircrafts employed in the study – two Boeing B 727-200, two DC 10-30 and one Airbus A300 – 600F. It should be noted that in CFF, each aircraft will experience off-design forces and moments. It is important that these are adequately modeled and efficiently controlled. Simply using aileron may trim out the induced roll but at the expense of drag. But as Bower et al. [34] have shown, it is possible to realize savings in fuel burn by using the existing aircraft by suitably tailoring the formation.

2.7.4 Tailored arrivals

Boeing [35] is working with several airports, airlines and other partners around the world in developing tools for "tailored arrivals" which can reduce fuel burn, lower the controller workload and allow for better scheduling and passenger connections (Figure 31). To optimize tailored arrivals, additional controller automation tools are needed. Boeing completed the trial of Speed and Route Advisor (SARA) with Dutch air traffic control agency (LVNL) and Eurocontrol in April/May 2009. SARA delivered traffic within 30 seconds of planned time on 80% of approaches at Schiphol airport in Netherlands compared to within 2 minutes on a baseline of 67%. At San Francisco airport, more than 1700 complete and partial tailored arrivals have been completed between December 2007 and June 2009 using the B777 and B747 aircraft. It has been found that tailored arrivals save an average of 950 kg of fuel and approximately $950 per approach. Complete tailored arrivals saved approximately 40% of the fuel used in arrivals. For one year period, four participating airlines saved more than 524,000 kg of fuel and reduced the carbon emissions by 1.6 million kg.

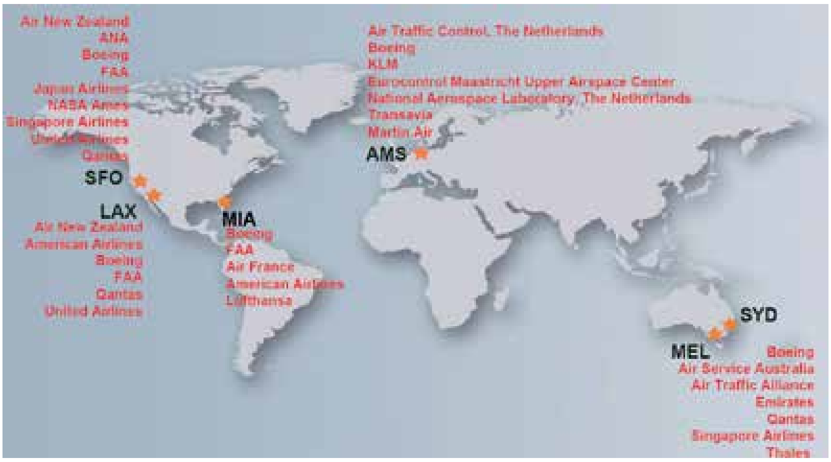


Fig. 31. Airports and Partners participating in the concept of Tailored Arrivals [35].

2.8 Savings in fuel burn by aircraft weight reduction

It is well known that substantial savings in fuel burn can be achieved by reducing the ratio of the empty weight to payload of an aircraft. It can be accomplished by the development and use of lighter and stronger advanced composites, and by reducing the design range and cruise Mach number.

2.8.1 Aircraft weight reduction by use of advanced composites

Reducing the weight of an aircraft is one of the most powerful means of reducing the fuel burn. Boeing and Airbus, as well as other Business and General Aviation aircraft manufacturers are investing in advanced composites which have the prospects of being lighter and stronger than the present carbon fiber composites (CFC). The replacement of structural aluminum alloy with carbon fiber composite is the most powerful weight reducing option currently available to the aircraft designer working towards a given payload-range requirement. The Boeing B787 and Airbus A350 have both taken this step, having wings and fuselage made with CFC. Most new designs are likely to take this path.

2.8.2 Aircraft weight reduction by reducing the design range

Although the historic trend has been in the opposite direction, another powerful means of reducing the weight of an aircraft is to reduce its design range. The study by Hahn [28] has shown that by reducing the design range from 15,000km to 5,000km, with the fuselage and passenger accommodation fixed, it is possible to reduce the operational empty weight (OEW) by 29%. The study by Creemers & Slingerland [29] noted a 17% reduction in OEW by halving the design range from 13,334km to 6,672km. Nangia [30, 31] has also shown that, with the fuselage and number of passengers fixed, wing area increases rapidly to contain the fuel needed and to maintain C_L as the design range increases. Figure 32 shows the aircraft designs and maximum take-off weight MTOW for design range from 3,000 to 12,000 nm. From Nangia's study [31], it is clear that 3,000nm aircraft can provide substantial savings in fuel burn by having less weight and can be used for long range flight by using AAR. In past twenty years, each new aircraft type has achieved 10-15% gain in fuel efficiency. Additional achievements in fuel efficiency by improvements in airframe and engine design will take some time, however, several studies have shown that it is possible to reduce fuel burn significantly by instituting operational measures such as more efficient Air-Traffic Management (ATM), Air-to-Air Refueling (AAR), Close Formation Flying (CFF), Tailored Arrivals, and by reducing the ratio of empty weight to payload.

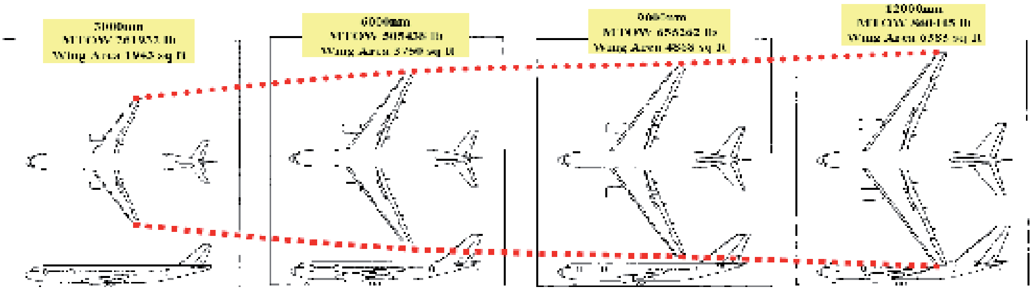


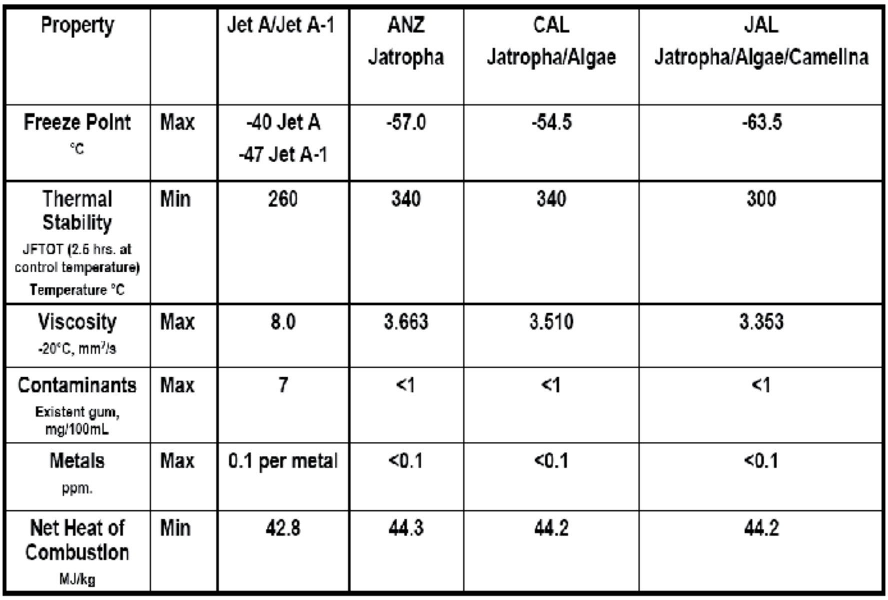
Fig. 32. Aircraft designs, with fixed fuselage, 250 passengers and $C_{L,}$ for different ranges of operation [30, 31].

2.9 Alternative fuels

All forms of powered ground and air transportation are experiencing the pressure of the need to mitigate greenhouse gas (GHG) emissions to arrest their impact on climate change. In addition the high price of fuel (oil reaching $149/barrel during summer of 2008) as well as the need for energy security are driving an urgent search for alternative fuels, in particular the biofuels. There is emphasis on both the improvements in energy efficiency and new alternative fuels. Aviation is particularly sensitive to these pressures since, for many years, no near term alternative to kerosene has been identified. Until recently, biofuels have not been considered cost competitive to kerosene. An important much desired characteristic of an alternative fuel is whether it can be used without any change to the aircraft or engines. The attractions of such a *drop-in fuel* are clear: it does not require the delivery of new aircraft but the environmental impact of all aircraft flying today can be significantly reduced. Non-drop-in fuels, such as hydrogen or methane hydrates, are unlikely to be used before 2050. The key criteria in identifying that a new alternative fuel would be beneficial in reducing CO_2 emissions should be based on the life cycle analysis of CO_2; the life-cycle CO_2 generation must be less than that of kerosene. Many first generation biofuels have performed poorly against this criterion, though second generation biofuels appear to be far more promising. Furthermore, it is important that there are no adverse side-effects arising from production of the feedstock for biofuel generation, such as adverse impact on farming land, fresh-water supply, virgin rain-forests and peat-lands, food prices, etc. Algae and halophytes (salt-tolerant plants irrigated with sea/saline water) are emerging as potential sustainable feedstock solutions. The alternative fuels need to meet specific aviation requirements and essentially should have the key chemical characteristics of kerosene, that is they won't freeze at flying altitude and they would have a high enough energy content to power an aircraft's jet engine. In addition, the alternative fuel should have good high-temperature thermal stability characteristics in the engine and good storage stability over time.

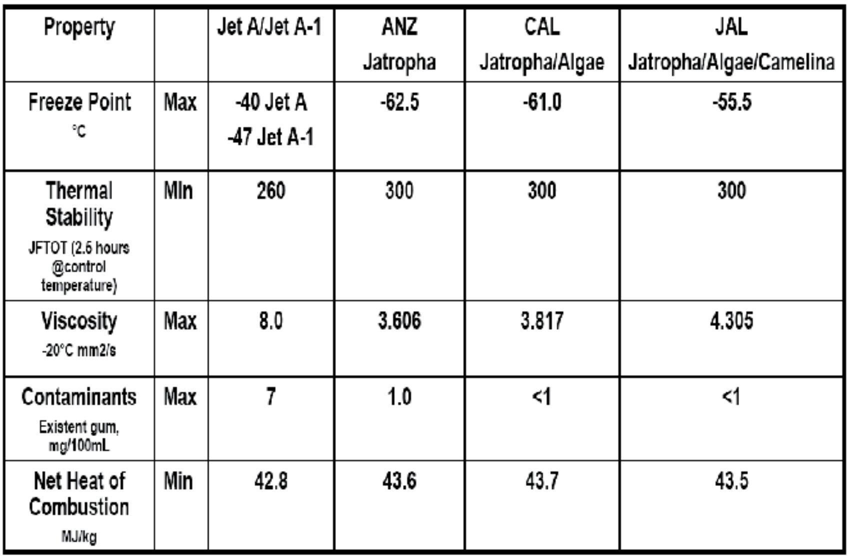
Interest in biofuels for civil aircraft has increased dramatically in recent years and the focus of the aviation industry on what is and what is not credible in this arena has sharpened. It is clear that a *'drop-in'* replacement for kerosene i.e. the synthetic kerosene appears to be the only realistic possibility in the foreseeable future. The potential of such bio-derived synthetic paraffinic kerosene (Bio-SPK) to reduce the net CO_2 emissions from aviation may well match or exceed that of advances in airframe and engine technologies, and perhaps may achieve reductions across the world fleet sooner than new technologies. In addition, since synthetic kerosene produces substantially less black carbon and sulphate aerosols than kerosene from oil wells, there is a possibility that its use will reduce contrail and cirrus formation as well.

Boeing, Airbus and the engine manufacturers believe that the present engine technology can operate on biofuels (tests are very promising) and that within 5 to 15 years, the aviation industry can convert to biofuels. On 19 June 2009, Billy Glover of Boeing made a presentation to the press at the Paris air show [35] describing the Boeing's "Sustainable Biofuels Research and Technology Program." Tables I and II show the comparisons of key fuel properties of currently used Jet A/Jet A-1 fuel with those with Bio-SPK fuel derived from three different feed-stocks (Jatropha, Jatropha/Algae, and Jatropha/Algae/Camelina) for neat fuel and blends respectively. All Bio-SPK blends met or exceeded the aviation jet fuel requirements. In this presentation, Boeing declared that they are preparing a comprehensive report on Bio-SPK fuels for submittal to ASTM International and expect an approval in 2010. Boeing is working across the industry on regional biofuel commercialization projects. There have already been a few experimental flights operated by several airlines using the biofuel blends and many more are planned in the near future.



Property		Jet A/Jet A-1	ANZ Jatropha	CAL Jatropha/Algae	JAL Jatropha/Algae/Camelina
Freeze Point °C	Max	-40 Jet A -47 Jet A-1	-57.0	-54.5	-63.5
Thermal Stability JFTOT (2.5 hrs. at control temperature) Temperature °C	Min	260	340	340	300
Viscosity -20°C, mm^2/s	Max	8.0	3.663	3.510	3.353
Contaminants Existent gum, mg/100mL	Max	7	<1	<1	<1
Metals ppm.	Max	0.1 per metal	<0.1	<0.1	<0.1
Net Heat of Combustion MJ/kg	Min	42.8	44.3	44.2	44.2

Table I. Key Biofuel (Neat) and Jet/Jet A-1 Fuel properties comparison [35].



Property		Jet A/Jet A-1	ANZ Jatropha	CAL Jatropha/Algae	JAL Jatropha/Algae/Camelina
Freeze Point °C	Max	-40 Jet A -47 Jet A-1	-62.5	-61.0	-55.5
Thermal Stability JFTOT (2.5 hours @control temperature)	MIn	260	300	300	300
Viscosity -20°C mm2/s	Max	8.0	3.606	3.817	4.305
Contaminants Existent gum, mg/100mL	Max	7	1.0	<1	<1
Net Heat of Combustion MJ/kg	Min	42.8	43.6	43.7	43.5

ANZ = Air New Zealand, CAL = Continental Airline, JAL = Japan Airline

Table II: Key Biofuel (Blend) and Jet/Jet A-1 fuel properties comparison [35].

On 24 February 2008, Virgin Atlantic operated a B747-400 on a 20% biofuel/80% kerosene blend on a short flight between London-Heathrow and Amsterdam. This was the first time a commercial aircraft had flown on biofuel and it was the result of a joint initiative between Virgin Atlantic, Boeing and GE. On 30 December 2008, Air New Zealand (ANZ) conducted a two hour test flight of a B747-400 from Auckland airport with one-engine powered by 50-50 blend (B50) of biofuel (from Jatropha) and conventional Jet-A1 fuel. B50 fuel was found to be more efficient. ANZ has announced plans to use the B50 for 10% of its needs by 2013. The test flight was carried out in partnership with Boeing, Rolls-Royce and Honeywell's refining technology subsidiary UOP with support from Terasol Energy. On January 7th, Continental Airline (CAL) completed a 90-minute test flight using biofuel derived from algae and Jatropha. B737-800 flew from Houston with one engine operating on a 50-50 blend of biofuel and conventional fuel (B50) and the other using all conventional fuel for the purpose of comparison. The biofuel mix engine used 3,600 lbs of fuel compared to 3,700 lbs used by the conventional engine. On January 30, 2009, Japan Airline (JAL) became the fourth airline to use B50 blend of Jatropha (16%), algae (<1%) and Camelina (84%) on the third engine of a 747-300 in one-hour test flight. It was again reported that biofuel was more fuel efficient than 100% jet-A fuel. It should be noted that in all the above demos, biofuel came from sustainable feedstocks (see Tables I and II), sources that neither compete with staple food crops nor cause deforestation. It is worth mentioning that on 1 February 2008, Airbus A380 flew from Filton, U.K. to Toulouse, France with one of its Rolls-Royce engines powered by an alternative, synthetic gas-to-liquid (GTL) jet fuel. Airbus and Qatar Airways are now partners in a GTL consortium which also includes Shell International Petroleum to investigate the use of GTL neat/blend vis-à-vis conventional jet fuel. From an environmental standpoint, it is encouraging and very hopeful that both major manufacturers – Boeing and Airbus are positioning themselves to be at the forefront of alternative and bio-jet fuels. It is surmised that by 2050, with the use of synthetic kerosene derived from biomass, the world fleet CO_2 emissions per passenger-kilometer (PKM) could be lower at least by a factor of three, NOx emissions lower by a factor of 10 and contrail and contrail-induced cirrus formation lower by a factor of 5 to 15.

2.10 Electric, solar or hydrogen powered green aircraft

For many years, there have been several exploratory studies in academia and industry to build and fly aircraft using sources of energy other than Jet-kerosene or synthetic kerosene (biofuels). There have been several success stories in recent years. In March 2008, Boeing successfully conducted a test flight of a manned aircraft powered by PEM hydrogen fuel cells [36], shown in Figure 33. Since fuel cells convert hydrogen directly into electricity and heat without the products of combustions such as CO_2, they use a clean or green source of energy. Fuel cells propelled aircraft is also often called as “an all electric aircraft.”

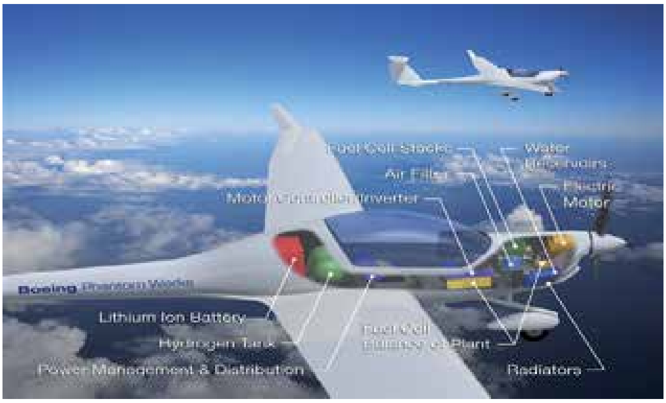


Fig. 33. Boeing PEM Fuel Cell Powered Electric Aircraft [36].

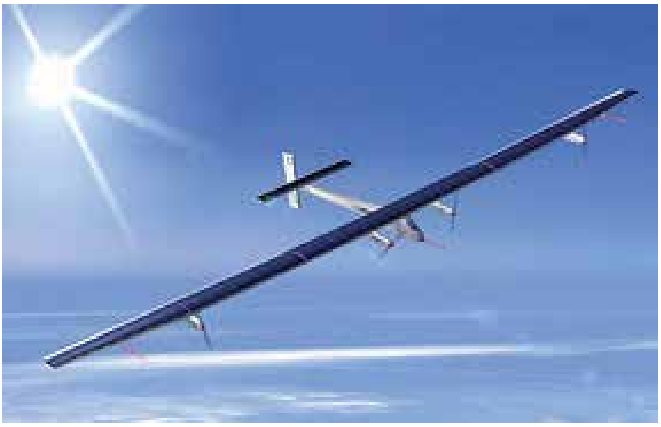


Fig 34. Solar Power Aircraft HB-SIA from SOLAR IMPULSE [37].

Recently in June 2009, the prototype of a new solar-powered manned aircraft was unveiled in Switzerland by the company SOLAR IMPULSE [37]. The airplane is designed to fly both day and night without the need for fuel. The aircraft has a wing span equal to that of a Boeing 747 but weighs only 1.7 tons. It is powered by 12,000 solar cells mounted on the wing to supply renewable solar energy to the four 10HP electric motors. During the day, the solar panels charge the plane's lithium polymer batteries, allowing it to fly at night. To be sure, the fuel-cell propelled electric aircraft and the solar energy driven aircraft are not likely to become feasible for mass air transportation. However, they can become viable for recreation and personal transportation, and possibly as business aircraft in not too distant future. The idea of using liquid hydrogen as a propellant has been around for many decades, but is unlikely to become feasible for commercial aircraft, at least before 2050, because of many challenges that would have to be overcome. Figure 35 shows the artist's rendering of a hydrogen-powered version of A310 Airbus [38]. It is also called a "Cryoplane" because of the very visible cryogenic hydrogen tank located above the passengers. Cryogenic hydrogen is the only possibility for the airplane since the high pressure tanks would be too heavy. The physical properties of the liquid hydrogen determine the appearance of the Cryoplane. Liquid hydrogen occupies 4.2 times the volume of jet fuel for the same energy; therefore the tanks will have to be huge. Jet fuel weighs 2.9 times more than liquid H_2 for the same energy. The reduced weight partly compensates for the increased aerodynamic drag of the tanks. The Cryoplane would have less range and speed than A310. It will have higher empty weight. Furthermore, whatever energy source is used, 30% will be lost in hydrogen liquefaction. In addition, the cost, infrastructure and passenger acceptance issues would have to be addressed. The main advantage of using a hydrogen powered airplane is the reduced emissions as shown in Figure 36 from Penner [39]. Since the use of H_2 does not produce any CO_2, it is dubbed as clean fuel.

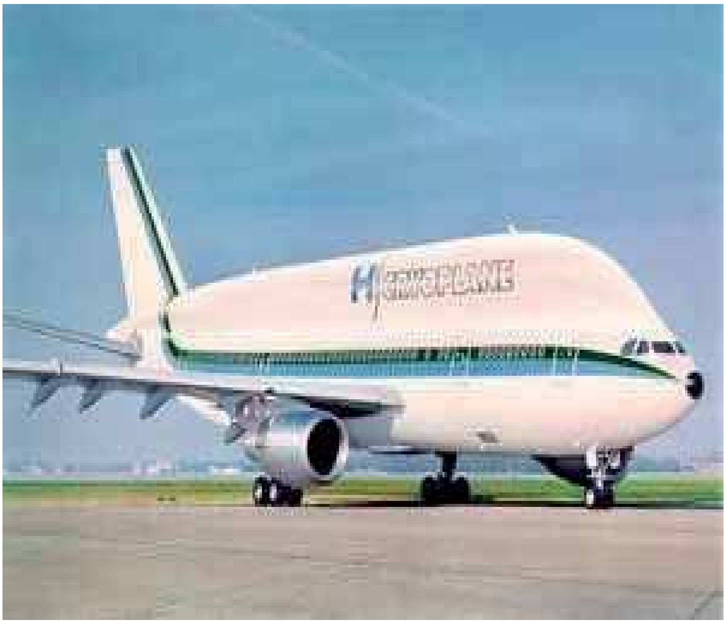


Fig. 35. Artist's rendering of a Hydrogen powered version of A310 Airbus [38].

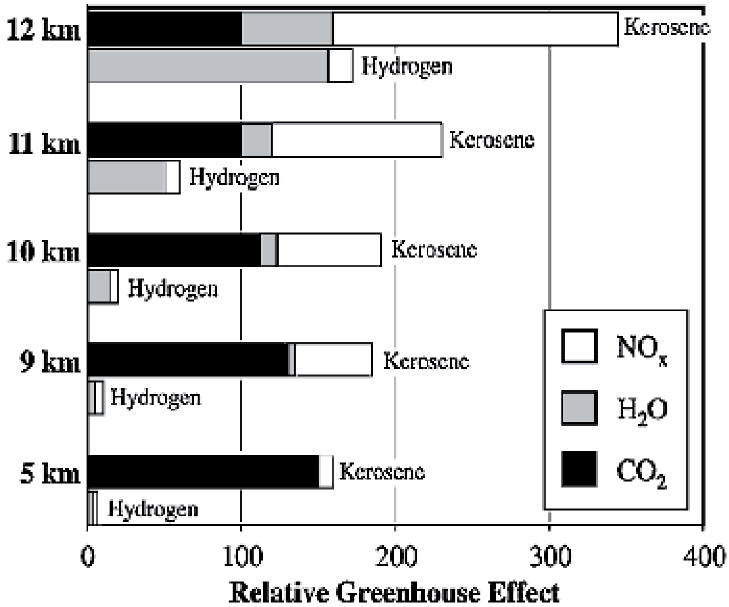


Fig. 36. Relative emissions from Jet-kerosene and Hydrogen at various altitudes [39].

2.11 Modeling environmental & economic impacts of aviation

2.11.1 Cambridge university aviation integrated modeling project (AIM)

Institute for Aviation and the Environment at Cambridge University in U.K. has developed one of the most comprehensive projects – called the Aviation Integrated Modeling (AIM) project to develop a policy assessment capability to enable comprehensive analyses of aviation, environment and economic interactions at local and global levels. It contains a set of inter-linked modules of the key elements which include models of aircraft/engine technologies, air transport demand, airport activity and airspace operations, all coupled to global climate, local environment and economic impact blocks. A major benefit of AIM architecture is the ability to model data flow and feedback between the modules allowing for the policy assessment to be conducted by imposing policy effects on upstream modules and determining the implications through down stream modules to the output metrics, which can then be compared to the baseline case [40].

These modules include: (a) an *Aircraft Technology and Cost Module* to simulate aircraft fuel use, emissions production and ownership/operating costs for various airframe/engine technology evolution scenarios which are likely to have an effect during the period of the forecast; (b) an *Air Transport Demand Module* to predict passenger and freight demand into the future between origin-destination pairs within the global air transportation network; (c) an *Airport Activity Module* to investigate the air traffic growth as a function of passenger and freight growth, to calculate delays and future airline response to them, and to model ground and low altitude operations and congestion to determine LTO emissions as a function of growth in air traffic operations within the vicinity of the airport; (d) an *Aircraft Movement Module* to simulate airborne trajectories between city-pairs, accounting for airspace inefficiencies and delays for given Air Traffic Control (ATC) scenarios and to identify the locations of emissions release from aircraft in flight; (e) a *Global Climate Module* to investigate global environmental impact of aircraft movements in terms of multiple emissions species and contrails; (f) a *Local Air Quality and Noise Module* to investigate local environmental impacts from dispersion of critical air pollutants and noise from landing and take-off (LTO) operations; and (g) a *Regional Economics Module* to investigate positive and negative economic impacts of aviation in various parts of the world, including the increase in direct and indirect employment opportunities in the region. The schematic of the AIM general architecture is shown in Figure 37 [40].

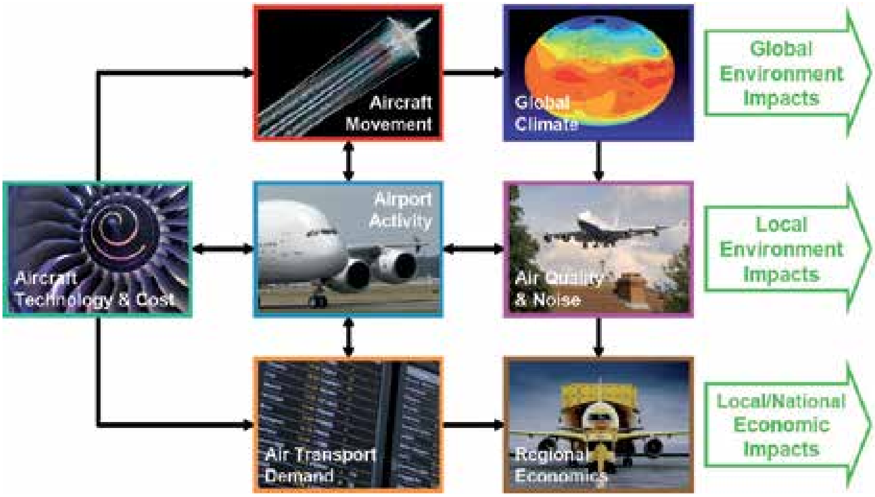

Aircraft Movement
Global Climate
Global Environment Impacts
Aircraft Technology & Cost
Airport Activity
Air Quality & Noise
Local Environment Impacts
Air Transport Demand
Regional Economics
Local/National Economic Impacts

Fig. 37. AIM Architecture [40].

The details of the seven modules and interaction among them are not given here but can be found in many papers listed on the website of the Institute for Aviation and the Environment of Cambridge University in U.K (http://www.iae.damtp.cam.ac.uk/ innovation.html). Here we briefly describe the power of the AIM architecture by reproducing some results from Reynolds et al. [40]. Employing the AIM architecture, Reynolds et al. [40] have performed a case study of the U.S. transportation system, which provides a forecast of air transport passenger demand between 50 major airports in U.S. from 2000 to 2030. The flights between these 50 airports represent over 40% of U.S. scheduled domestic departures in 2000 and nearly 20% the world's scheduled flights. Reynolds et al. [40] conducted simulations under three scenarios: 1. Unconstrained/No Feedback (air transport passenger demands and resulting operations were assumed to grow unconstrained), 2. Feedback of Delay Effects (a simplified airline response to delay is modeled by assuming that the 50% of the cost incurred by the airlines due to delays are passed directly to passengers in the form of higher fares), and 3. Feedback of Delay Effects Plus Per-Km Tax Policy (This is same as scenario 2 , but with a per-Km tax applied to tickets from 2020 onwards with the objective of reducing the Revenue Passenger Km (RPKM) demand in 2020 to 2000 levels, so that the resulting delays and emissions can be directly compared). *Reynolds et al.* [40] *state that these three scenarios, their associated forecasts and environmental impact results are for illustrative purposes only to show the capabilities of AIM; they do not represent realistic evolutions of the U.S. air transportation system.* The main focus of the scenarios is on interactions between the Air Transport Demand and the Airport Activity Modules. However, one can calculate the en route and local emissions utilizing the capabilities of other modules in AIM integrated structure as given in [40]. Details of the data and assumptions used in the simulation are not presented here. The reader is referred to the paper by Reynolds et al. [40].

Forecasts from 2000 to 2030 for annual demand in terms of Revenue Passenger-Km (RPKM) from the Air Transport Demand Module; and total system aircraft operations, system average arrival delay and local NOx emissions at Chicago O'Hare (ORD) from the Airport Activity Module for the above three scenarios are presented in Figures 38 – 41 from Reynolds et al. [40]. The demand forecasts in Figure 38 include those from Airbus (for U.S market), and Boeing, ICAO and AERO-MS for the North American (NA) market for the purpose of comparison. Since they apply to different route groups and time periods, the start year total RPKM value in each case has been normalized to the historical value for the 50 airports extracted from U.S department of transportation T100 data. Figure 38 shows that for scenario 1, the demand growth measured by increase in RKPM will be 3.5 times the 2000 level by 2030. This is higher than the published estimates as expected given the unconstrained nature of the scenario 1. In scenario 2, the relatively modest feedback of 50% of the increased operating cost to the passenger has a significant effect, particularly over longer time frames. Demand forecast shows a 20% reduction (Figure 38), annual systems operations show a 15% reduction (Figure 39) and average arrival delays show a 50% reduction (Figure 40). Under scenario 3, Figures 38-40 show the effects of distance-based tax; in order to reduce the RPKM demand to 2000 levels in 2020, a 7.7 cents/km charge is required, equating to an additional $300 on a ticket from New York to Los Angeles. Figure 41 shows the annual local emissions at Chicago O'Hare (ORD); all scenarios show an initial gradual increase in emissions which can be explained in conjunction with Figures 38-40 accounting for the increase in RPKM, aircraft operations and arrival delays. The sharp decrease in emissions in scenario 3 in 2020 is due to the reduced operations caused by the introduction of distance-tax policy. The Local Air Quality and Noise Module of AIM architecture can provide results for local air quality at ORD e.g. the annual average NOx concentration at ORD as well as en route CO_2 emissions and global radiative forcing. These results demonstrate that significant insights about environmental and economic impact of aviation can be gained by AIM architecture. It should be noted that many improvements and enhancements to AIM architecture are currently under development at Cambridge.

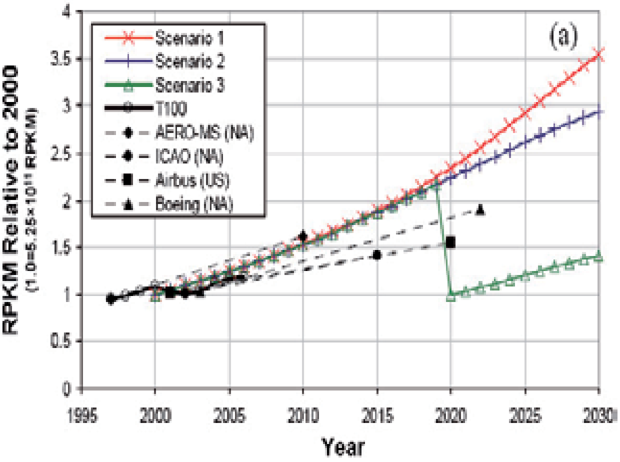


Fig. 38. Forecast of system Revenue Passenger – Km (RPKM) growth at O'Hare [40].

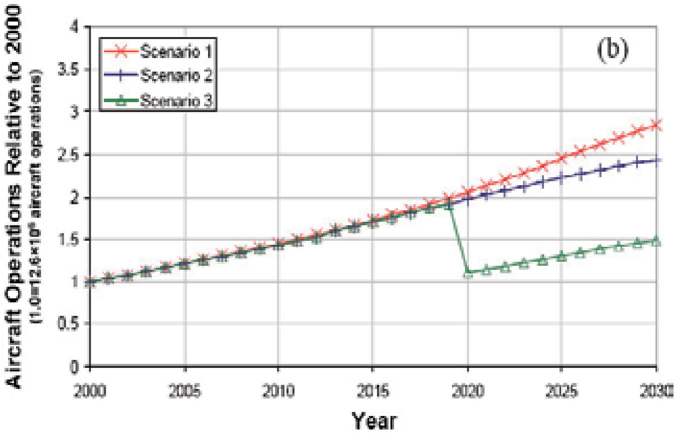


Fig. 39. Forecast of total system aircraft operations at O'Hare [40].

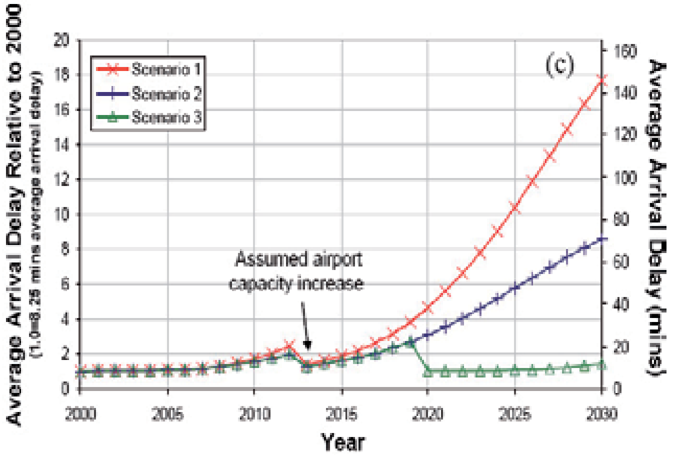


Fig. 40. System average arrival delays at O'Hare [40].

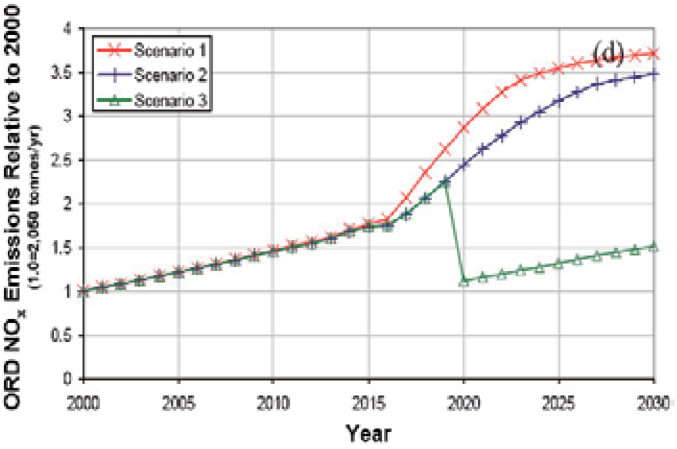


Fig. 41. LTO NOx emissions at O'Hare [40].

2.12 Sustainable airports

The airports and associated ground infrastructure constitute an integral part of Green Aviation. To address the issues of energy and environmental sustainability, the Clean Airport Partnership (CAP) was established in U.S. in 1998 [41] and is the only not-for-profit corporation in the U.S devoted exclusively to improving environmental quality and energy efficiency at airports. CAP believes "that efficient airport operations and sound environmental management must go hand in hand. This approach can reduce costs and uncertainty of environmental compliance; facilitate growth, while setting a visible leadership example for communities and the nation." The airport expansion and the development of new airports should include both the environmental costs and life-cycle costs. Sustainable growth of airports requires that they be developed as inter-modal transport hubs as part of an integrated public transport network. The ground infrastructure development should include low emission service vehicles; LEEDS certified green buildings with low energy requirements, and recyclable water usage. There should be effective land use planning of the area around the airports (including securing land for future development) with active investments into the surrounding communities. Airport expansion must also consider the issue of noise and its impact on the surrounding communities, and should be involved in its mitigation by engaging in the flight path design. The air quality near the airports should be monitored and measures for its continuous improvement should be put in place. In addition, there should be regulatory requirements to set risk limits.

3. Opportunities and future prospects

It is clear that the expected three fold increase in air travel in next twenty years offers enormous challenge to all the stakeholders – airplane manufacturers, airlines, airport ground infrastructure planners and developers, policy makers and consumers to address the urgent issues of energy and environmental sustainability. The emission and noise mitigation goals enunciated by ACARE and NASA can be met by technological innovations in aircraft and engine designs, by use of advanced composites and biofuels, and by improvements in aircraft operations. Some of the changes in operations can be easily and immediately put into effect, such as tailored arrivals and perhaps AAR. Some innovations in aircraft and engine design, use of advanced composites, use of biofuels, and overhauling of the ATM system may take time but are achievable by concerted and coordinated effort of government, industry and academia. They may require significant investment in R&D. It is now recognized by the industry (airlines and manufacturers) as well the relevant government agencies and the policy makers that there is urgent need for action to meet the challenges of climate change; aviation is becoming an important part of it. It is worth noting that in July 2008 in Italy, G8 countries (U.S, Canada, Russia, U.K., France, Italy, Germany and Japan) called for a global emission reduction target of "at least 50%" by 2050, which is in line with goal established by IATA members at their June 2009 Annual General Meeting in Kuala Lumpur, Malaysia. IATA further committed to carbon-neutral traffic growth by 2020. These challenges provide opportunities for breakthrough innovations in all aspects of air transportation.

4. Acknowledgements

The author wants to acknowledge several individuals and sources for their help and permission to use the material from their presentations and papers. The author is grateful to

Dr. Tom Reynolds of the Institute of Aviation and Environment at Cambridge University to allow the use of material used in the section on "Modeling Environmental & Economic Impacts of Aviation." All the material in this section has been taken from Reference [40]. The author is also grateful to Dr. Raj Nangia for helpful discussions and allowing the use of material from several of his papers [30, 31]. Reference [17] has also provided a significant amount of material for several sections. The author is thankful to Dr. Richard Wahls of NASA Langley for his permission to use the material from NASA Langley presentations on 'Environmentally Responsible Aviation.' The author would like to thank Professor Raimo J. Hakkinen for reading the manuscript and for making many helpful suggestions that have improved the paper. Finally, it should be noted that the material used in this review paper has been used from a variety of sources listed in the references; any omission is completely unintentional.

5. References

[1] Schafer, A., Heywood, J.B., Jacoby, H.D., and Waitz, I.A., *Transportation in a Climate Constraint World,* MIT Press, Cambridge, MA, 2009.

[2] Salari, K.,"DOE's Effort to Reduce Truck Aerodynamic Drag Through Joint Experiments and Computations," LLNL-PRES-401649, 28 February 2008.

[3] www.pewclimate.org

[4] Agarwal, R.K., "Sustainable (Green) Aviation: Challenges and Opportunities (2009 William Littlewood Lecture)," SAE Int. J. Aerospace, Vol. 2, pp. 1-20, 2009.

[5] http://www.boeing.com/randy/archives/2006/07/in_the_year_202.html

[6] http://www.acare4europe.com/

[7] NRC Meeting of Experts on NASA's Plans for System-Level Research in Environmental Mitigation, National Harbor, MD, 14 May 2009; Presentation by R.A. Wahls; http://www.aeronautics.nasa.gov/calendar/20090514.htm

[8] Mankins, J.C.,"Technology Readiness Levels," http://www.hq.nasa.gov/office/codeq /trl

[9] Aerospace International, *The Green Issue*, Aerosociety, U.K., March 2009.

[10] Smith, M.J.T., *Aircraft Noise*, Cambridge University Press, Cambridge, U.K., 1989.

[11] Erickson, J.D., "Environmental Acceptability" Office of Environment and Energy, Presented to FAA, 2000.

[12] http://silentaircraft.org/

[13] Reynolds, T.G., "Environmental Challenges for Aviation – An Overview," Presented to Low Cost Air Transport Summit, London, 11-12 June 2008.

[14] www.iea.org

[15] Lee, J.J., Lukachko, S.P., Waitz, I. A., and Schafer, A., "Historical & Future Trends in Aircraft Performance, Cost, and Emissions," Annu. Rev. Energy Environ, Vol. 26, pp. 167-200, 2001.

[16] Penner, J.E., *Aviation and the Global Atmosphere*, Cambridge University Press, Cambridge, U.K., pp. 76-79, 1999.

[17] Royal Aeronautical Society Annual Report, "Air travel - Greener by Design Annual Report 2007-2008," April 2008 (http://www.greenerbydesign.org.uk/).

[18] Schumann, U., "On Conditions for Contrail from Aircraft Exhaust," Meteor. Zeitsch, Vol. 5, pp. 3-22, 1996.

[19] NRC Meeting of Experts on NASA's Plans for System-Level Research in Environmental Mitigation, National Harbor, MD, 14 May 2009; Presentation by A. Strazisar; http://www.aeronautics.nasa.gov/calendar/20090514.htm

[20] www.boeing.com

[21] www.b-domke.de/AviationImages/Propfan/0815

[22] www.flightglobal.com/articles/2007/06/12/214520

[23] http://www.dfrc.nasa.gov/Gallery/Photo/X-48B/HTML/ED08-0092-13.html

[24] http://hondajet.honda.com/

[25] Saeed, T.I, Graham, W.R., Babinsky, H., Eastwood, J.P., Hall, C.A., Jarrett, J.P., Lone, M.M. and Seffen, K.A., "Conceptual Design of a Laminar Flying Wing Aircraft," AIAA 2009-3616, 27th AIAA Applied Aerodynamics Conference, San Antonio, TX, 22-25 June 2009.

[26] Collier, F.S., NASA Langley, "Progress in Environmental Aeronautics," Presentation at Aviation & Environment – A Primer for North American Stakeholders Meeting; http://www.airlines.org/NR/rdonlyres/A78FA93B-986C-4D95-BA87-B4DD961CC369/0/11collier.pdf

[27] National Academy of Science (NAS) Report, "Assessing the Research and Development plan for the Next Generation Air Transportation System: Summary of a Workshop," (http://www.nap.edu/catalog/12447.html), 2008.

[28] Hahn. A.S., "Staging Airliner Service," AIAA 2007-7759, 7th AIAA ATIO Conference, Belfast, 18-20 Sept. 2007.

[29] Creemers, W.L.H. and Slingerland, R., "Impact of Intermediate Stops on Long-Range Jet-Transport Design," AIAA 2007-7849, 7th AIAA ATIO Conference, Belfast, 18-20 Sept. 2007.

[30] Nangia, R.K., "Air to Air Refueling in Civil Aviation," Paper #9, Royal Aeronautical Soc. "Greener by Design" Conference, London, 7 October 2008.

[31] Nangia, R.K., "Way Forward to a Step Jump for Highly Efficient & Greener Civil Aviation – An Opportunity for the Present and a Vision for the Future," Personal Publication RKN-SP-2008-120, September 2008.

[32] Wagner, E., Jacques, D., Blake, W., and Pachter, M., "Flight Test Results for Close Formation Flight for Fuel Savings," AIAA 2002-4490, AIAA Atmospheric Flight Mech. Conf., Monterey, CA, 5-8 August 2002.

[33] Jenkinson, L.R., Caves, R.E, and Rhodes, D.R., "A Preliminary Investigation into the Application of Formation Flying to Civil Operation," AIAA 1995-3898, 1995.

[34] Bower, G.C., Flanzer, T.C. and Kroo, I.M., "Formation Geometries and Route Optimization for Commercial Formation Flight," AIAA 2009-3615, 27th AIAA Applied Aerodynamics Conference, San Antonio, TX, 22-25 June 2009.

[35] Boeing Presentation at Paris Air Show by Billy Glover, June 2009 (http://www.boeing.com/paris2009/media/presentation/june17/glover_enviro_briefing/).

[36] www.boeing.com

[37] www.solarimpulse.com

[38] http://www.planetforlife.com/h2/h2vehicle.html

[39] Penner, J.E., *Aviation and the Global Atmosphere*, Cambridge University Press, Cambridge, U.K., p. 257, 1999.

[40] Reynolds, T.G., Barrett, S., Dray, L.M., Evans, A.D., Kohler, M.O., Morales, M.V., Schafer, A., Wadud, Z., Britter, R., Hallam, H., and Hunsley, R., " Modeling Environmental & Economic Impacts of Aviation: Introducing the Aviation Integrated Modeling Project," AIAA 2007-7751; 7th AIAA Aviation Technology, Integration and Operations Conference, Belfast, 18-20 Sept. 2007.

[41] http://www.cleanairports.com

20

Synthetic Aperture Radar Systems for Small Aircrafts: Data Processing Approaches

Oleksandr O. Bezvesilniy and Dmytro M. Vavriv
Institute of Radio Astronomy of the National Academy of Sciences of Ukraine
Ukraine

1. Introduction

The synthetic aperture radar (SAR) is considered now as the most effective instrument for producing radar images of ground scenes with a high spatial resolution. The usage of small aircrafts as the platform for the deployment of SAR systems is attractive from the point of view of many practical applications. Firstly, this enables for a substantial lowering of the exploitation costs of SAR sensors. Secondly, such solution provides a possibility to perform a rather quick surveillance and imaging of particular ground areas. Finally, the progress in this direction will allow for a much wider application of SAR sensors.

However, the formation of high-quality SAR images with SAR systems deployed on small aircrafts is still a challenging problem. The main difficulties come from significant variations of the aircraft trajectory and the antenna orientation during real flights. These motion errors lead to defocusing, geometric distortions, and radiometric errors in SAR images.

In this chapter, we describe three effective approaches to the SAR data processing, which enable the solution of the above problems:

1. Time-domain SAR processing with clutter-lock and geometric correction by resampling,
2. Time-domain SAR processing with built-in geometric correction and multi-look radiometric correction,
3. Range-Doppler algorithm with the 1-st and 2-nd order motion compensation.

The proposed solutions have been successfully implemented in Ku- and X-band SAR systems developed and produced at the Institute of Radio Astronomy of the National Academy of Sciences of Ukraine. The efficiency of the proposed algorithms is illustrated by SAR images obtained with these SAR systems.

The chapter is organized as follows. In Section 2, basic principles of SAR data processing is described. In Section 3, the problem of motion errors of airborne SAR systems is considered, and the appearance of geometric distortions and radiometric errors in SAR images is discussed. The three data processing approaches are considered in details in Sections 4, 5, and 6. Section 7 describes the RIAN-SAR-Ku and RIAN-SAR-X systems used in our experiments. The conclusion is given in Section 8.

2. Principles of SAR data processing

The synthetic aperture technique is used to obtain high-resolution images of ground surfaces by using a radar with a small antenna installed on an aircraft or a satellite. The radar pulses backscattered from a ground surface and received by the moving antenna can be considered as the pulses received by a set of antennas distributed along the flight trajectory. By coherent processing of these pulses it is possible to build a long virtual antenna – the synthetic aperture that provides a high cross-range resolution. A high range resolution is typically achieved by means of a pulse compression technique that involves transmitting long pulses with a linear frequency modulation or a phase codding.

2.1 Concept of the synthetic aperture

Practical SAR systems are produced to operate in one or several operating modes. Depending on the mode, they are referred as the strip-map SAR, the spot-light SAR, the inverse SAR, the ScanSAR, and the interferometric SAR (Bamler & Hartl, 1998; Carrara at al., 1995; Cumming & Wong, 2005; Franceschetti & Lanari, 1999; Rosen at al., 2000; Wehner, 1995). We shall consider mainly the most popular and practically useful strip-map SAR operating mode. However, the presented further results are applicable to other modes to a large extent.

In the strip-map SAR mode, the radar performs imaging of a strip on the ground aside of the flight trajectory. Geometry of the strip-map mode is shown in Fig. 1. The aircraft flies along the straight line above the x -axis with the velocity V at the altitude H above the ground plane (xy).

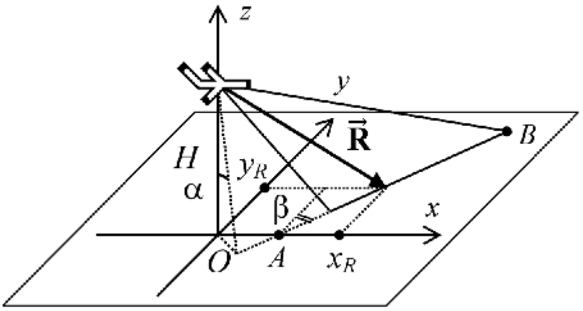


Fig. 1. Geometry of the strip-map SAR mode.

The orientation of the real antenna beam is described by the pitch angle α and the yaw angle β that are measured with respect to the flight direction. The line AB in Fig. 1 is the intersection of the elevation plane of the real antenna pattern and the ground plane. This line is called the Doppler centroid line. The coordinates of the point (x_R, y_R) on this line at the slant range R from the aircraft are given by

$$x_R = H\tan\alpha\cos\beta + \sin\beta\sqrt{R^2 - H^2 - (H\tan\alpha)^2} \ , \qquad (1)$$

$$y_R = -H\tan\alpha\sin\beta + \cos\beta\sqrt{R^2 - H^2 - (H\tan\alpha)^2} \,. \tag{2}$$

In order to form the synthetic aperture and direct the synthetic beam to the point (x_R, y_R), the signal $s_R(\tau + t)$ backscattered from this point should be summed up coherently on the interval of synthesis $-T_S / 2 \le \tau \le T_S / 2$ taking into account the propagation phase $\varphi(\tau)$ (Cumming & Wong, 2005; Franceschetti & Lanari, 1999):

$$I(t, x_R, y_R) = \left| \frac{1}{T_S} \int\limits_{-T_S/2}^{T_S/2} s_R(\tau + t) h_R(\tau) d\tau \right|^2 , \tag{3}$$

$$h_R(\tau) = w_R(\tau)\exp\left[-i\varphi(\tau)\right], \quad \varphi(\tau) = -\frac{4\pi}{\lambda}R(\tau) \,. \tag{4}$$

Here $I(t, x_R, y_R)$ is the SAR image pixel, t is the time when the aircraft is at the centre of the synthetic aperture $(0, 0, H)$, τ is the time within the interval of synthesis, $h_R(\tau)$ is the azimuth reference function in the time domain, $w_R(\tau)$ is the weighting window applied to improve the side-lobe level of the synthetic aperture pattern, λ is the radar wavelength, and $R(\tau)$ is the slant range to the point:

$$R(\tau) = \sqrt{(x_R - V\tau)^2 + y_R^2 + H^2} = \sqrt{R^2 - 2x_R V\tau + (V\tau)^2} \,. \tag{5}$$

If the slant range $R(\tau)$ changes during the time of synthesis T_S more than the size of the range resolution cell, then the target signal "migrates" through several range cells. This effect known as the range migration should be taken into account during the aperture synthesis. The one-dimensional backscattered signal $s_R(\tau + t)$ should be obtained from the two-dimensional "azimuth – slant range" matrix of the range-compressed radar data by the interpolation along the migration curve (5).

The instant Doppler frequency of the received signal is, approximately,

$$f(\tau) = -\frac{2}{\lambda}\frac{dR(\tau)}{dt} \approx F_{DC} + F_{DR}\tau \,, \tag{6}$$

where the Doppler centroid F_{DC} and the Doppler rate F_{DR} are given by

$$F_{DC} = \frac{2}{\lambda}V\frac{x_R}{R} \,, \tag{7}$$

$$F_{DR} = -\frac{2}{\lambda}\frac{V^2}{R}\left[1 - \left(\frac{x_R}{R}\right)^2\right] . \tag{8}$$

It is useful to note that the Doppler centroid determines the synthetic beam direction, whereas the Doppler rate is responsible for the beam focusing.

From the point of view of signal processing, the formation of the synthetic aperture (3) is the matched filtering of linear frequency modulated signals (6). Such filtering can be performed either in the time or in the frequency domain. Accordingly, there are time- and frequency-domain SAR processing algorithms.

It is easy to show that the azimuth resolution ρ_X is given by (Cumming & Wong, 2005; Carrara at al., 1995)

$$\rho_X = K_w \frac{V}{\Delta F_D} . \tag{9}$$

Here $\Delta F_D = |F_{DR}| T_S$ is the Doppler frequency bandwidth that corresponds to the interval of the synthesis. The coefficient K_w describes the broadening of the main lobe of the synthetic aperture pattern caused by windowing.

In order to improve the quality of SAR images, a multi-look processing technique is used in most modern SAR systems (Moreira, 1991; Oliver & Quegan, 1998). According to such technique, a long synthetic aperture is divided on shorter intervals that are processed independently to build several SAR images of the same ground scene, called SAR looks. It can be considered as building the synthetic aperture with multiple synthetic beams. A non-coherent averaging of the SAR looks into one multi-look image is used to reduce speckle noise and to reveal fine details in SAR images. Multi-look processing can be used for other applications, for example, for measuring the Doppler centroid with a high accuracy and high spatial resolution and retrieving 3D topography of ground surfaces (Bezvesilniy et al., 2006; Bezvesilniy et al., 2007; Bezvesilniy et al., 2008; Vavriv & Bezvesilniy, 2011b).

In the next sections we consider peculiarities of the realization of SAR processing algorithms in time and frequency domains.

2.2 SAR processing in time domain

The SAR processing in the time domain is performed according to the relations (3)-(5). The block-scheme of the algorithm is shown in Fig. 2.

The received range-compressed radar data are stored in a memory buffer. The buffer size in the range corresponds to the swath width; the buffer size in the azimuth is determined by the time of synthesis. The basic step of the SAR processing procedure for a given range R includes the following calculations:

1. Calculation of the Doppler centroid (7), the Doppler rate (8), and the required time of synthesis (9),
2. Interpolation along the migration curve (5),
3. Multiplication by the reference function with windowing (4), and
4. Coherent summation (3).

As the result, a single pixel of the SAR image is obtained representing the ground point on the Doppler centroid line at the range R. This basic step is repeated for all ranges within the swath producing a single line of the SAR image in the range direction. In order to form the next line of the SAR image, the data in the buffer is shifted in the azimuth and supplemented with new data, and the computations are repeated.

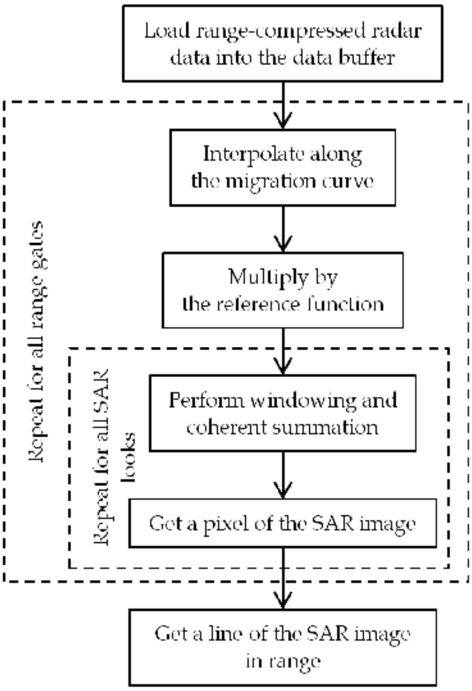
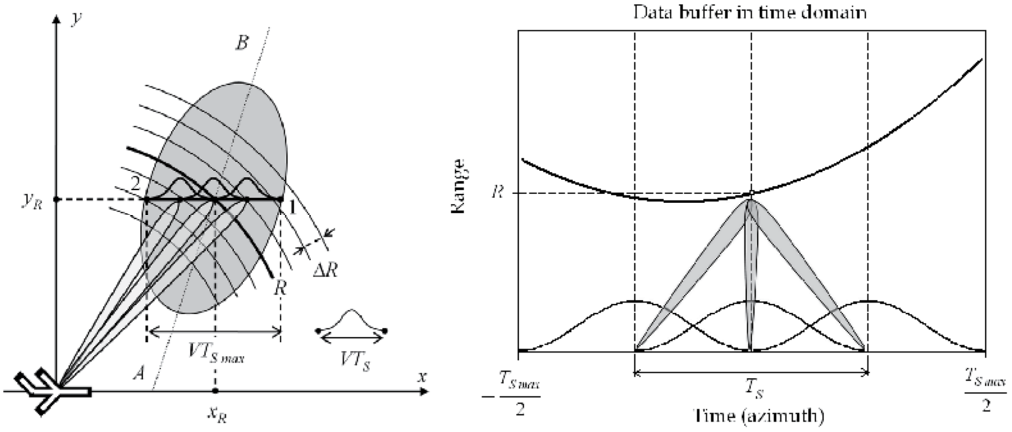

Load range-compressed radar data into the data buffer
Interpolate along the migration curve
Multiply by the reference function
Perform windowing and coherent summation
Get a pixel of the SAR image
Repeat for all SAR looks
Repeat for all range gates
Get a line of the SAR image in range

Fig. 2. SAR processing in time domain.



(a) (b)

Fig. 3. Multi-look processing in the time domain: (a) the antenna footprint consideration, (b) the data buffer consideration.

The multi-look processing in the time domain is usually performed directly following the definition (Moreira, 1991; Oliver & Quegan, 1998). Namely, the reference functions and range migration curves are built for the long interval of synthesis $T_{S\max}$, which is the time required for the ground target to cross the antenna footprint from point 1 to point 2 in Fig.

3a. The multi-look processing is performed by splitting the long interval of the synthesis $T_{S\max}$ on several sub-intervals T_S, forming in this manner multiple synthetic beams pointed to the same point on the ground at the different moments of time, as shown in Fig. 3b. The number of the looks for a scheme with the half-overlapped sub-intervals is given by

$$N_L = \operatorname{int}\left\{\frac{T_{S\max}}{T_S / 2}\right\} - 1 . \tag{10}$$

2.3 SAR processing in frequency domain

The SAR data processing can be also performed effectively in the frequency domain. It is known that the convolution of two signals in the time domain is equivalent to the multiplication of their Fourier pairs in the frequency domain. The corresponding computations are efficient due to the application of the fast Fourier transform (FFT). A number of FFT-based SAR processing algorithms have been so far developed (Cumming & Wong, 2005).

In particular, the range-Doppler algorithm (RDA) (Cumming & Wong, 2005) is a relatively simple and widely-used FFT-based algorithm. The processing steps of this algorithm are shown in Fig. 4 and illustrated also in Fig. 5. The received radar data are stored in a large memory buffer. The buffer size in the range direction corresponds to the swath width, and the buffer size in the azimuth direction is equal to the length of the FFT that covers many intervals of synthesis. First, the range-compressed data are transformed into the range-Doppler domain by applying FFT in the azimuth. The frequency scale is limited by the pulse repetition frequency (PRF). Then, the range migration correction is performed in the frequency domain. By using the relation (6) between the instant frequency f and the time τ within the interval of synthesis (preserving the square-root law for the slant range) one can derive the formula for the migration curve in the frequency domain from the migration curve (5) in the time domain:

$$R(f) = R\sqrt{1 - \frac{\lambda^2 F_{DC}^2}{4V^2}} \Bigg/ \sqrt{1 - \frac{\lambda^2 f^2}{4V^2}} . \tag{11}$$

After that, the phase compensation and windowing are applied for the azimuth compression. By using the principle of stationary phase (Cumming & Wong, 2005) an expression for the reference function in the frequency domain is obtained:

$$h_R(f) = w_R(f)\exp[-i\theta(f)], \; \theta(f) = -\frac{4\pi}{\lambda} R\left[\sqrt{1-\left(\frac{\lambda f}{2V}\right)^2}\sqrt{1-\left(\frac{\lambda F_{DC}}{2V}\right)^2} + \left(\frac{\lambda f}{2V}\right)\left(\frac{\lambda F_{DC}}{2V}\right)\right] . \tag{12}$$

Finally, the SAR image is formed by applying the inverse FFT in the azimuth. Thus, the basic processing step in the frequency domain performed for a given range gives the line of the SAR image in the azimuth. This basic step is repeated for all ranges within the swath producing the complete SAR image of the ground scene presented in the data frame.

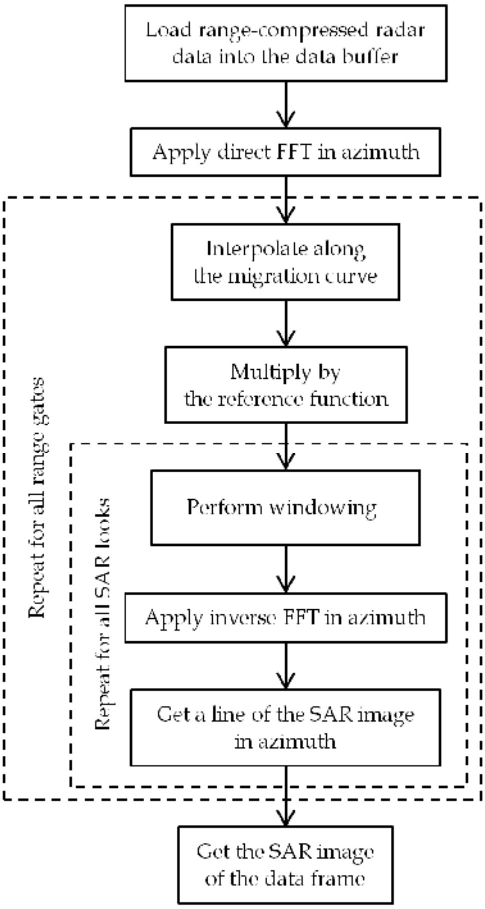

Load range-compressed radar data into the data buffer
Apply direct FFT in azimuth
Interpolate along the migration curve
Multiply by the reference function
Perform windowing
Apply inverse FFT in azimuth
Get a line of the SAR image in azimuth
Get the SAR image of the data frame
Repeat for all range gates
Repeat for all SAR looks

Fig. 4. SAR processing in frequency domain.

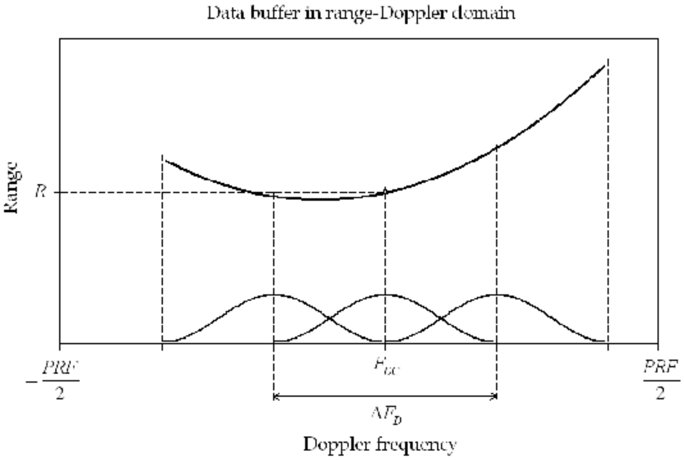


Fig. 5. Multi-look processing in frequency domain.

In the cases of a significantly squinted geometry (a large antenna yaw angle) or a very high resolution, or a large number of looks, an additional processing step called "secondary range compression" is required (Cumming & Wong, 2005).

The multi-look processing is performed in the frequency domain by dividing the whole Doppler band $\Delta F_{D\max} = |F_{DR}| T_{S\max}$ of the backscattered radar signals into the sub-bands ΔF_D for the separate azimuth compression (Cumming & Wong, 2005; Carrara at al., 1995):

$$\Delta F_D = \frac{\Delta F_{D\max}}{(N_L + 1) / 2} . \tag{13}$$

For the multi-look processing scheme with the half-overlapped sub-bands, the central frequencies of the SAR looks with respect to the Doppler centroid are given by

$$\Delta F_C(R, n_L) = F_{DC}(R) - n_L \frac{\Delta F_D}{2} , \tag{14}$$

where $n_L = -N_L / 2, ..., N_L / 2 - 1$ is the SAR look index. Since the Doppler rate (8) is always negative, the first sub-interval in the time domain corresponds to the last sub-band in the frequency domain. Therefore, we write the minus sign in (14).

3. Problem of aircraft motion errors

Deviations of the aircraft flight trajectory and instabilities of the aircraft orientation significantly complicate the formation of SAR images. Such motion errors lead to defocusing, geometric distortions, and radiometric errors in SAR images (Blacknell et al., 1989; Buckreuss, 1991; Franceschetti & Lanari, 1999; Oliver & Quegan, 1998). In this section, we shall discuss these problems and their solutions in details.

3.1 Aircraft flight with motion errors

The trajectory of an aircraft may deviate from a straight line significantly in real flights. The orientation of the aircraft could also be unstable. These motion errors should be measured and compensated in order to produce high-quality SAR images. We assume that the navigation system is capable of measuring the aircraft trajectory and the aircraft velocity vector. We suppose also that the orientation of the real antenna beam with respect to the velocity vector is known.

Usually, the final product of the strip-map SAR system is a sequence of SAR images of a particular dimension, built in a projection to the ground plane, with indication of the north direction and the latitude-longitude position. Later, if necessary, several consequent images can be stitched together to produce a larger map of a particular ground area of interest. Thus, the received radar data is processed by data frames. Each frame gives one SAR image from the image sequence. The data frames are usually overlapped to guarantee successful stitching of the produced SAR images without gaps.

In order to produce the SAR image from the data frame, it is needed to define a reference flight line for this frame, the averaged flight altitude H_{ref} and the averaged velocity V_{ref}. Under unstable flight conditions, the reference flight line should be close to the actual curvilinear flight trajectory of the aircraft. Also, the reference antenna pitch and yaw angles α_{ref} and β_{ref}, which describes the averaged orientation of the real antenna beam during the time of the data frame acquisition, should be introduced.

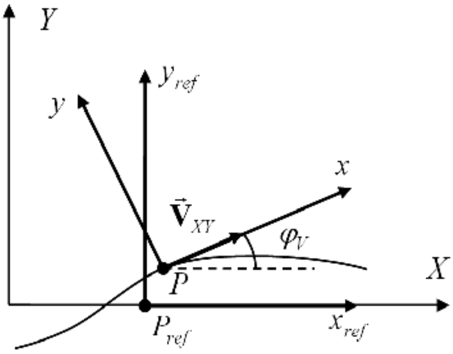


Fig. 6. The scene coordinate system, the reference local coordinate system, and the actual local coordinate system.

Let us define the scene coordinate system (X, Y, Z) so that the reference flight line goes exactly above the X axis. The final-product SAR image is to be sampled on the coordinate grid of the ground plane (X, Y) of this coordinate system. The scene coordinate system is shown in Fig. 6 together with the actual local coordinate system (x, y, z), which slides along the real aircraft flight trajectory, and the reference local coordinate system $(x_{ref}, y_{ref}, z_{ref})$, which slides along the X axis (that is along the reference flight line). The current flight direction is described by the angle φ_V between the horizontal component of the velocity vector $\vec{\mathbf{V}}_{XY}$ and the X axis.

The aircraft trajectory $(X_A(t), Y_A(t), Z_A(t))$ is described in the scene coordinate system. The actual local coordinates (x, y) and the reference local coordinates (x_{ref}, y_{ref}) are related to each other as follows:

$$x = [x_{ref} - X_A(t) + V_{ref}t]\cos\varphi_V(t) + [y_{ref} - Y_A(t)]\sin\varphi_V(t), \tag{15}$$

$$y = -[x_{ref} - X_A(t) + V_{ref}t]\sin\varphi_V(t) + [y_{ref} - Y_A(t)]\cos\varphi_V(t). \tag{16}$$

The pitch $\alpha(t)$ and yaw $\beta(t)$ angles describe the antenna beam orientation with respect to the current aircraft velocity vector or, in other words, with respect to the actual local coordinate system. It means that when the synthetic beam is directed to the point (x_R, y_R) on the Doppler centroid line by using the Doppler centroid (7), the Doppler rate (8), and the migration curve (5) under unstable flight conditions, the coordinates (x_R, y_R) are given in the actual local coordinate system. In order to find the scene coordinates (or the reference local coordinates) of this point, the above relations (15), (16) should be used.

An example of motion errors typical for a light-weight aircraft AN-2 is shown in Fig. 7. In the figure, one can see the coordinate grid of the radar coordinates "slant range – azimuth" projected onto the ground plane (X, Y). The horizontal curves are the curves of the constant slant range from the aircraft. They are curved because of deviations of the trajectory from the straight line. The vertical lines are the central lines of the antenna footprint (the Doppler centroid lines) for the consequent aircraft positions. As it is seen, the central lines are not equidistant and not parallel because of variations of the antenna orientation.

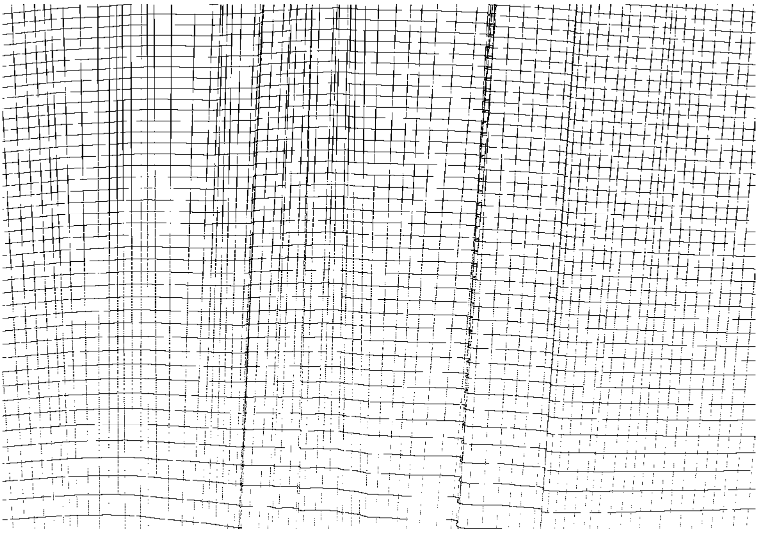


Fig. 7. Trajectory deviations and orientation instabilities illustrated by the coordinate grid in the radar coordinates "slant range – azimuth" on the ground plane.

3.2 Geometric distortions in SAR images

The direction of the synthetic beam is determined by the used Doppler centroid with respect to the current velocity vector. In other words, the Doppler centroid controls the direction of the synthetic beam with respect to the actual local coordinate system. Therefore, if the deflections of the velocity vector from the reference flight direction (described by the angle φ_V in Fig. 6) are not compensated properly, then the synthetic beam is moving forward or backward along the flight path with respect to the scene coordinate system. It means that the scene will be sampled non-uniformly in the azimuth direction resulting in geometric distortions in SAR images. For example, if the synthetic beams are pointed to the centre of the real beam, i.e. to the Doppler centroid line, then the scene will be sampled on a non-uniform grid like that shown in Fig. 7.

If the aircraft trajectory and the orientation of the synthetic aperture beams are known, the geometric distortions can be corrected by resampling of the obtained SAR images to a rectangular grid on the ground plane. This resampling procedure is described in Section 4. However, this approach could be inefficient in the case of significant geometric distortions.

Alternatively, geometric errors can be avoided if the orientation of the synthetic beams is adjusted at the stage of synthesis by using the trajectory information. The purpose of this adjustment is to keep the beam orientation constant with respect to the reference flight direction. This is the idea of the built-in geometric correction discussed in Section 5.

The correction of the phase errors and range migration errors caused by trajectory deviations can be applied to the raw data before the aperture synthesis. After such compensation, the raw data look like be collected from the reference straight line. After such motion compensation, the synthetic beams will be set with respect to the reference local coordinate system. Such approach is widely used with the SAR processing algorithms working in the frequency domain. This motion compensation technique is considered in Section 6 with application to the range-Doppler algorithm.

3.3 Radiometric errors in SAR images

The problem of radiometric errors is illustrated in Fig. 8. If there are no orientation errors, the synthetic beam of the central look is directed to the centre of the real antenna beam, and all SAR look beams are within the main lobe of the real antenna pattern, as shown in Fig. 8a. The antenna orientation errors lead to the situation when the SAR beams are directed outside the real antenna beam to not-illuminated ground areas, as shown in Fig. 8b, resulting in radiometric errors.

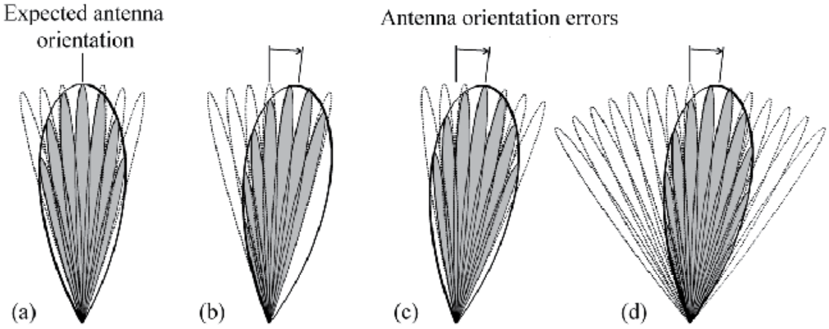


Fig. 8. Multi-look processing without antenna orientation errors (a) and with orientation errors: (b) without clutter-lock, (c) with clutter-lock, (d) with extended number of looks.

Instabilities of the aircraft orientation can be compensated by the antenna stabilization by mounting it on a gimbal. It helps to keep the constant antenna beam orientation. However, this approach is rather complicated and expensive.

The application of a wide-beam antenna firmly mounted on the aircraft is a less expensive way to guarantee a uniform illumination of the ground scene despite of instabilities of the platform orientation. Several shortcomings of this approach should be admitted. The application of a wide antenna beam means some degradation of the radar sensitivity. Also, it calls for a higher PRF to sample the increased Doppler frequency band. Moreover, only the central part of the antenna footprint will be illuminated uniformly limiting the number of looks that can be built without an additional radiometric compensation.

The clutter-lock technique (Li at al., 1985; Madsen, 1989) is usually used to avoid radiometric errors in SAR images. According to the clutter lock technique, the azimuth reference functions are built adaptively so that the synthetic beams track the direction of the real antenna beam staying within the main lobe of the real antenna pattern as shown in Fig. 8c. However, the variations of the synthetic beam orientation due to the clutter-lock naturally lead to geometric distortions in SAR images.

The clutter lock technique is effective if the variations of the antenna beam orientation are slow in time and small as compared to the real antenna beam width in the azimuth. In this case, the geometric distortions can be corrected by re-sampling. Provided that the orientation instabilities are fast and significant, the clutter-lock leads to strong geometric distortions in SAR images which cannot be easily corrected by re-sampling.

We have proposed an alternative radiometric correction approach, which is based on a multi-look SAR processing with an extended number of SAR looks (Bezvesilniy et al., 2010c; Bezvesilniy et al., 2010d; Bezvesilniy et al., 2011b; Bezvesilniy et al., 2011c). This technique can be used instead of the clutter-lock. The idea of the approach consists in the formation of an extended number of looks to cover directions beyond the main lobe of the real antenna pattern as illustrated in Fig. 8d. In such approach, some of the SAR look beams are always presented within the real antenna beam despite of the orientation errors. In Section 5, we describe how to combine these extended SAR looks to produce the multi-look SAR image without radiometric errors. This approach is appropriate for the cases when the clutter-lock cannot be applied because of fast orientation instabilities or for SAR processing algorithms that cannot be used together with the clutter-lock. The proposed method also allows correcting the radiometric errors in SAR images if the antenna orientation is not known accurately.

3.4 Dilemma: geometric distortions vs. radiometric errors

From the above considerations, one can conclude that an attempt to avoid geometric errors by the appropriate pointing of the synthetic beams leads to radiometric errors. And vice versa, the clutter-lock results in geometric errors. So the dilemma of “geometric distortions vs. radiometric errors” should be resolved when developing any SAR data processing approach for SAR systems with motion errors.

We describe three alternative approaches to this problem. In the first approach, described in Section 4, the priority is set to avoiding radiometric errors and the clutter-lock is applied. Geometric errors are corrected by resampling of the obtained SAR images. In the second approach, considered in Section 5, the geometric accuracy of SAR images is the primary goal and we implement a synthetic beam control algorithm called “built-in geometric correction” to point the beams to the nodes of a correct rectangular grid on the ground plane. Radiometric errors are corrected by multi-look processing with extended number of looks. In the third approach, discussed in Section 6, a range-Doppler algorithm with the 1-st and 2-nd order motion compensation is considered, which allows obtaining SAR images without significant geometric errors. The application of a wide-beam real antenna could be a solution of the problem of radiometric errors for this approach.

4. Time-domain SAR processing with clutter-lock and geometric correction by resampling

In this section, we consider a time-domain SAR data processing algorithm assuming that the aircraft flight altitude and velocity, as well as the antenna beam orientation angles are changed slowly in the sense that they can be considered constant during the time of the synthesis. The main steps of the algorithm are the same as in the case of the straight-line motion with a constant orientation. These steps are described in the block-scheme shown in

Fig. 2. At each step of the synthesis, the reference function and migration curves are adjusted according to the estimated orientation angles of the real antenna beam providing the clutter-lock. Due to the clutter-lock, it is possible to avoid radiometric errors. Geometric errors are corrected by resampling of the obtained SAR images on the post-processing stage.

4.1 Estimation of the antenna orientation angles from Doppler centroid measurements

According to the clutter-lock technique, the synthetic beams are built adaptively to track the direction of the real antenna beam. The orientation angles of the aircraft can be measured by a navigation system. The commonly used navigation systems are based on Inertial Measurement Unit (IMU) or on a combination of IMU and attitude GPS. They are typically rather expensive and do not always provide the required accuracy and the needed rate of measurements. We have proposed an effective method for the estimation of the antenna orientation angles – pitch and yaw – from the Doppler measurements. The application of this technique has allowed us to simplify the navigation system by reducing it to a simple GPS receiver to measure the platform velocity and coordinates only.

The mathematical background of this technique is as follows. The dependence of the Doppler centroid on the slant range is given by

$$F_{DC} = \frac{2}{\lambda}\frac{(\vec{\mathbf{R}}\cdot\vec{\mathbf{V}})}{R} = \frac{2}{\lambda}\frac{x_R V_x - HV_z}{R} \,. \tag{17}$$

The slant range vector $\vec{\mathbf{R}} = (x_R, y_R, -H)$ is directed from the antenna phase centre to the point (x_R, y_R) on the Doppler centroid line as shown in Fig. 1, and the aircraft velocity vector is $\vec{\mathbf{V}} = (V_x, 0, V_z)$. In opposite to (7), formula (17) accounts for the possible vertical direction of the aircraft motion. Substituting in (17) the expression (1) for the coordinate of a point on the Doppler centroid line, we rewrite the above dependence as

$$F_{DC}(R,\alpha,\beta) = \frac{2}{\lambda}\frac{V_x}{R}\left[H\tan\alpha\cos\beta + \sin\beta\sqrt{R^2 - H^2 - (H\tan\alpha)^2} - H\frac{V_z}{V_x}\right]. \tag{18}$$

The behaviour of the Doppler centroid on range depends strongly on particular values of the antenna pitch and yaw angles as illustrated in Fig. 9. It means that theoretically the antenna beam orientation angles can be estimated via an analysis of the dependence of the measured Doppler centroid on range. However, for the practical implementation of this idea, it was needed to answer the questions: Would it be a reliable estimate? Is it possible to achieve the required accuracy of the angle measurements? And, is it possible to realize this estimation in real time? Fortunately, we have found solutions which provide positive answers on the above questions.

We have found that the pitch and yaw angles can be estimated by fitting the theoretical dependence of the Doppler centroid on range (18) into a set of Doppler centroid values $F_{DC}^{[n]} = F_{DC}(R_n)$ roughly estimated from the received data at each range gate from the Doppler spectra calculated by using the FFT. Here n is the range gate index.

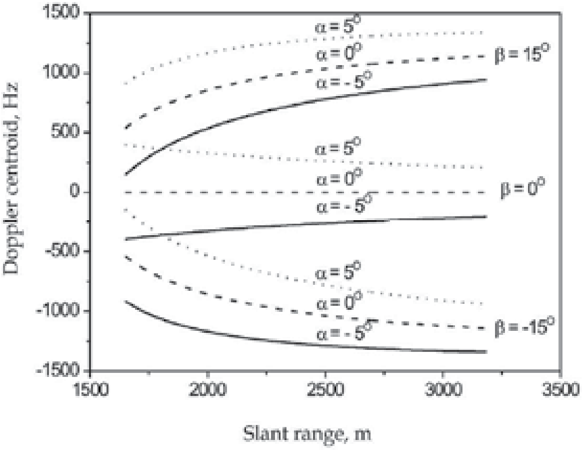


Fig. 9. Dependence of the Doppler centroid on slant range.

We have developed the following fast and effective fitting procedure. By introducing new variables X_n^i, Y_n as

$$X_n^i = \sqrt{R_n^2 - H^2 - (H \tan \alpha_{i-1})^2} \ , \ Y_n = \frac{\lambda F_{DC}^{[n]}}{2V_x} R_n + H \frac{V_z}{V_x} \ , \tag{19}$$

the dependence (18) is transformed into the equation of a straight line:

$$Y_n = (H \tan \alpha_i \cos \beta_i) + X_n^i (\sin \beta_i) \ . \tag{20}$$

Thus, the problem of fitting of the non-linear dependence (18) is turned into the well-known task of fitting of a line into a set of experimental points. The only difficulty is that the unknown pitch angle appears in the transformation of the coordinates (19). We have solved this difficulty by using an iteration procedure. The fitting is performed iteratively with respect to the pitch angle considered as a small parameter. The index $i = 1, 2, 3, ...$ in (19), (20) is the iteration index. At the first iteration, the pitch angle is assumed to be zero: $\alpha_0 = 0$. It has been found that two iteration are typically enough to achieve the required accuracy of about 0.1° in real time. The method has been implemented in SAR systems developed and produced at the Institute of Radio Astronomy (Vavriv at al., 2006; Vavriv & Bezvesilniy, 2011a; Vavriv at al., 2011).

4.2 Correction of geometric distortions in SAR images by resampling

In the considered time-domain SAR processing algorithm with the clutter-lock, each line of a SAR image in the range direction represents the ground scene on the Doppler centroid line determined by the current antenna beam orientation angles $\alpha(t)$ and $\beta(t)$ in the actual local coordinate system (see Fig. 6). Thus, the application of the clutter-lock under unstable flight conditions leads to geometric distortions in SAR images as illustrated in Fig. 7. Such geometric distortions can be corrected by resampling of the images from the radar native coordinates "slant range – azimuth" to a correct rectangular grid on the ground plane (X, Y) by taking into account the measured aircraft trajectory and the orientation of the synthetic aperture beams.

The resampling procedure consists of the following steps.

1. Define the reference flight line and the reference parameters, as well as the corresponding scene coordinate system for a given SAR image frame as it was described in Section 3.1.
2. Perform the resampling (interpolation) of the SAR image $SAR(X_A, R)$ from the slant range to the ground range in four steps:
   2.1. Calculate the coordinates of the image pixels $SAR(X_A, R)$ in the actual local coordinate system: $(x_{SAR}(X_A, R), y_{SAR}(X_A, R))$.
   2.2. Re-calculate the coordinates of the image pixels from the actual local coordinate system to the scene coordinate system according to (15), (16) and obtain the coordinates $(X_{SAR}(X_A, R), Y_{SAR}(X_A, R))$.
   2.3. Perform a one-dimensional interpolation of the SAR image line-by-line in the range direction from the uniform grid in the slant range to the uniform grid in the ground range. As the result, we obtain the image $SAR(X_A, Y)$.
   2.4. Find the coordinates $X_{SAR}(X_A, Y)$ of the image samples $SAR(X_A, Y)$ in the scene coordinate system from the coordinates $X_{SAR}(X_A, R)$ by the same one-dimensional interpolation in the range direction.
3. Perform the interpolation of the SAR image in the azimuth direction in the following two steps:
   3.1 Perform a joint sorting of the pairs of the range-interpolated image samples $SAR(X_A, Y)$ and their azimuth coordinates $X_{SAR}(X_A, Y)$ in the ascending order with respect to the X_A-coordinate. This step is required to correct significant forward-backward sweeps of the synthetic beam caused by motion errors.
   3.2 Perform a one-dimensional interpolation of the SAR image samples $SAR(X_A, Y)$ from the initial non-uniform grid of the along-track azimuth coordinate X_A to the uniform grid X. The result is the desired image $SAR(X, Y)$ in the ground scene coordinates.

The above described resampling algorithm is typically performed as a post-processing procedure.

4.3 Experimental results

The described in Sections 4.1 and 4.2 SAR processing approach has been implemented in the airborne RIAN-SAR-Ku system (Vavriv at al., 2006; Vavriv & Bezvesilniy, 2011a; Vavriv at al., 2011). The light-weight aircrafts Antonov AN-2 and Y-12 were used as the platform.

An example of a single-look SAR image built by the described SAR processing algorithm with the clutter-lock is shown in Fig. 10a. This is the SAR image before the correction of the geometric distortions by resampling. The image resolution is 3 m. The "forward-backward-

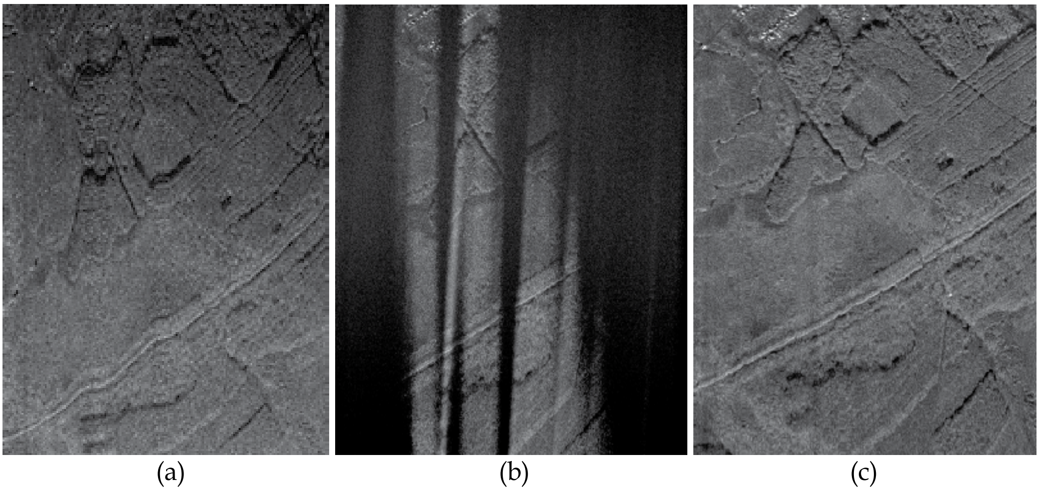


Fig. 10. Geometric distortions in a single-look SAR image built by using the clutter-lock (a), radiometric errors in the multi-look SAR image built without the clutter lock (b), the multi-look SAR image without errors after the resampling procedure (c).

forward" motion of the antenna beam leads to the evident distortions of the road lines and the contours of the forest areas in this image.

A 5-look SAR image formed without the clutter-lock is shown in Fig. 10b. The characteristic amplitude of the antenna beam orientation instabilities was larger than the 1-degree antenna beam width what resulted in significant radiometric errors. It should be noted that the proposed clutter-lock method based on the estimation of the antenna beam from the Doppler centroid measurements is efficient enough to avoid these radiometric errors in Fig. 10a.

The SAR images in Figs. 10a and 10b illustrates the dilemma of "geometric distortions vs. radiometric errors". Radiometric errors are removed due to the clutter-lock in Fig. 10a at the expense of geometric errors. And, vice versa, geometric errors are eliminated in Fig. 10b built without the clutter-lock, but at the cost of significant radiometric errors.

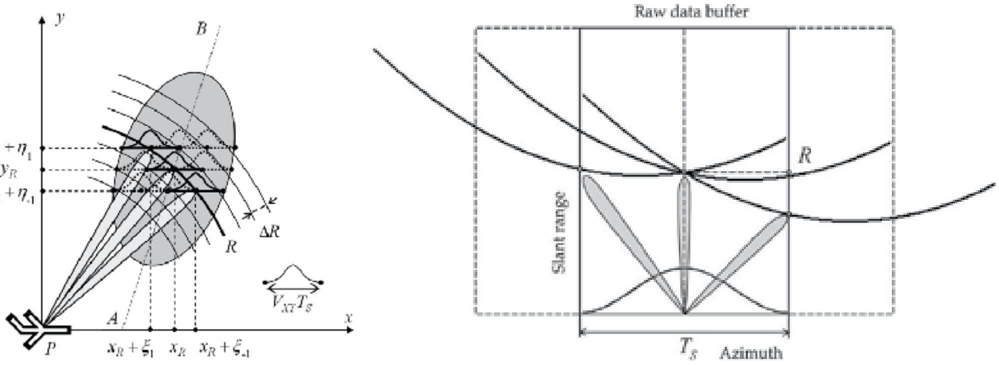
The application of the proposed resampling procedure resolves the dilemma, as it is illustrated in Fig. 10c. In this figure, both geometrical and radiometric errors are corrected.

5. Time-domain SAR processing with built-in geometric correction and multi-look radiometric correction

In this section, we describe a SAR processing approach, in which the correction of geometric distortions in SAR images is considered as the primary goal. We proposed (Bezvesilniy et al., 2010a; Bezvesilniy et al., 2010b; Bezvesilniy et al., 2010d; Bezvesilniy et al., 2011a) an algorithm called "built-in geometric correction" to control the synthetic beam direction so that the beams are pointed to the nodes of a correct rectangular grid on the ground plane. As the result, the SAR images are geometrically correct after the synthesis. The synthetic beams are obviously set to the nodes regardless of the real antenna beam orientation. The radiometric errors that arise in this case are corrected by a multi-look processing with extended number of looks.

5.1 Multi-look SAR processing on a single-look interval of synthesis

The multi-look processing in the time domain is usually performed by the coherent processing on sub-intervals of a long interval of the synthesis as described in Section 2.2. According to such approach, it is assumed that there are no significant uncompensated phase errors during the long time of the synthesis $T_{S\max}$ determined by (10). However, as a matter of fact, in order to achieve the desired azimuth resolution it is sufficient to perform the coherent processing on the short time interval T_S given by (9). This fact gives an alternative realization of the multi-look processing in the time domain, which is more preferable in the case of significant motion errors. The idea of the algorithm is to process the data collected during the short time of synthesis T_S with a set of different reference functions and migration curves to form the SAR look beams. We have called this approach "the multi-look processing on a single-look interval of synthesis". The proposed approach is illustrated in Fig. 11.



(a) (b)

Fig. 11. The multi-look processing on a single-look interval of synthesis: (a) the antenna footprint consideration, (b) the raw data buffer consideration.

The reference functions of the different SAR looks should be built with the central frequencies (14) similar to the multi-look processing scheme in the frequency domain. The SAR look beam formed with the central frequency $\Delta F_C(R, n_L)$ is directed to some point $(x_R + \xi(R, n_L), y_R + \eta(R, n_L))$, which appears at the same slant range R at the centre of the short interval of the synthesis as illustrated in Fig. 11a. Let us derive formulas for these coordinates. The position of the point in the azimuth direction is related to its Doppler centroid (17), so we can write:

$$\Delta F_C(R, n_L) = \frac{2}{\lambda} \frac{(x_R + \xi(R, n_L))V_x - HV_z}{R} . \tag{21}$$

Substituting the expressions (14) and (17) into (21), we obtain:

$$\xi(R, n_L) = -n_L \frac{\lambda R}{2V_x} \frac{\Delta F_D}{2} . \tag{22}$$

Since the points, to which the synthetic beams of the SAR looks are directed, appear at the same slant range at the centre of the short interval of the synthesis, we can write:

$$x_R^2 + y_R^2 = (x_R + \xi(R, n_L))^2 + (y_R + \eta(R, n_L))^2 , \tag{23}$$

and, finally,

$$\eta(R, n_L) = \sqrt{x_R^2 + y_R^2 - (x_R + \xi(R, n_L))^2} - y_R . \tag{24}$$

Thus, in order to form the set of the synthetic beams of the different SAR looks on the short interval of synthesis for the slant range $R(n_L, X, Y)$, we should first calculate the points $(x_R + \xi(R, n_L), y_R + \eta(R, n_L))$ from (22) and (24), which correspond to the required central frequencies (14). Then, we should process the same raw data on the interval of the synthesis T_S with the appropriate range migration curves (5), the Doppler centroids (7) and the Doppler rates (8), by substituting the calculated coordinates $(x_R + \xi(R, n_L), y_R + \eta(R, n_L))$ instead of the coordinates (x_R, y_R) in these formulas.

The described approach to the multi-look processing has the following benefits. First, it is much easier to keep a low level of the phase errors on the short interval of synthesis, as compared to the long coherent processing time for all looks. Second, the orientation of the real antenna beam does not change significantly during the short processing time. This fact simplifies considerably the calculation of the orientation of all SAR look beams with respect to the real antenna beam for the subsequent radiometric correction. Third, a more accurate motion error compensation can be introduced in this processing scheme as compared to FFT-based algorithms. The compensation is performed based on the measured aircraft trajectory individually for each pixel of the SAR image accounting for both the range and the azimuth dependence of the phase and migration errors without any approximations. In other words, the accuracy of the motion error compensation is limited only by the accuracy of the trajectory measurements.

In the described approach, all SAR look beams are aimed at different points on the ground. It means that the obtained SAR look images are sampled on different grids. Therefore, the SAR look images should be first resampled to the same ground grid and only then they can be averaged to produce the multi-look image. Deviations of the aircraft trajectory introduce further complexity into the re-sampling process. We have proposed (Bezvesilniy et al., 2010a; Bezvesilniy et al., 2010b; Bezvesilniy et al., 2010d; Bezvesilniy et al., 2011a) an algorithm named "the built-in correction of geometric distortions" to solve this problem. This algorithm is described in the next section.

5.2 Built-in geometric correction

In order to avoid the interpolation steps in the above-described multi-look processing approach, the reference functions and the migration curves should be specially designed to point the multi-look SAR beams exactly to the nodes of a rectangular grid on the ground plane. The grid nodes to which the multi-look SAR beams should be pointed can be found as follows. The radar data are processed frame-by-frame forming a sequence of overlapped SAR images. For each frame, we define the reference flight line and the reference parameters before the synthesis of the aperture. The reference parameters of the data frame are used to calculate the Doppler centroid values $F_{DC}(R)$, the central Doppler frequencies $\Delta F_C(R, n_L)$ of the SAR looks, and the coordinates $(x_R^{ref} + \xi(R, n_L), y_R^{ref} + \eta(R, n_L))$ of the corresponding points on the ground in the reference local coordinate system. The found points are situated on the central frequency lines, which are similar to the Doppler centroid line AB in Fig. 1. The synthetic beams of the SAR looks should be pointed to the grid nodes, which are closest to the corresponding frequencies lines.

To point the SAR look beam to the found grid node, it is needed to recalculate the coordinates of this node from the reference local coordinate system to the actual local coordinate system by using (15), (16), taking into account the actual aircraft position and the orientation of the aircraft velocity vector. This recalculation is performed at each step of the synthesis. After that, the appropriate range migration curves (7), the Doppler centroids (8), and the Doppler rates (9) can be determined. Finally, the synthetic beam is formed to be directed to this node.

The proposed built-in geometric correction algorithm cannot be combined with the clutter-lock technique since the SAR beams do not follow the orientation of the real antenna beam. Therefore, the algorithm works well without an additional radiometric correction only for a wide-beam antenna and only for the central SAR looks. In order to use all possible SAR looks to form a multi-look SAR image without radiometric errors, we have proposed an effective radiometric correction technique based on multi-look processing with extended number of looks (Bezvesilniy et al., 2010c; Bezvesilniy et al., 2010d; Bezvesilniy et al., 2011b; Bezvesilniy et al., 2011c).

5.3 Radiometric correction by multi-look processing with extended number of looks

Let us denote an error-free SAR image to be obtained as $I(X, Y)$, where (X, Y) are the ground coordinates of the image pixels. This image is not corrupted by speckle noise and not distorted by radiometric errors. Whereas, a real SAR look image $I(n_L, X, Y)$ (n_L is the index of the SAR looks) is corrupted by speckle noise $S(n_L, X, Y)$ and distorted by radiometric errors $0 < R(n_L, X, Y) \le 1$ so that

$$I(n_L, X, Y) = I(X, Y) \cdot S(n_L, X, Y) \cdot R(n_L, X, Y) . \tag{25}$$

The speckle noise in a single-look SAR image (Oliver & Quegan, 1998) is a multiplicative noise with the exponential probability density function with the mean and the variance, correspondingly,

$$\mu\{S(n_L, X, Y)\} = 1 , \quad \sigma\{S(n_L, X, Y)\} = 1 . \tag{26}$$

The speckle noise is different for all SAR looks what is indicated here by the SAR look index n_L. The radiometric errors caused by instabilities of the antenna orientation can be considered as low-frequency multiplicative errors. The highest spatial frequencies of the radiometric error function $R(n_L, X, Y)$ are inversely proportional to the width of the real antenna footprint in the azimuth direction. Similar to the speckle noise, the radiometric errors are different for different SAR looks.

In order to compensate the radiometric errors, they should be estimated. For this purpose, we use a low-pass filtering $\mathbf{F}$ to measure the local brightness of the SAR images. This filter is designed to pass the radiometric errors and, at the same time, to suppress the speckle noise to some extent:

$$\mathbf{F}\{R(n_L, X, Y)\} \approx R(n_L, X, Y)\,, \quad \mathbf{F}\{S(n_L, X, Y)\} \approx 1\,. \tag{27}$$

The application of this filter to the SAR look image (25) gives, approximately:

$$I_{LF}(n_L, X, Y) = \mathbf{F}\{I(n_L, X, Y)\} \approx I_{LF}(X, Y) \cdot R(n_L, X, Y)\,. \tag{28}$$

Here $I_{LF}(X, Y)$ is the low-frequency component of the error-free SAR image to be reconstructed. The corresponding components of the real SAR looks $I_{LF}(n_L, X, Y)$ (28) contain information about the radiometric errors and they are almost not corrupted by speckle noise. These images can be used to compare radiometric errors on different SAR looks and, via such comparison, to estimate the radiometric errors. The idea of this empirical approach to the radiometric correction is based on the fact that one of many looks is pointed very closely to the centre of the real antenna beam. This look demonstrates the maximum power (brightness) among all looks, and this power is not distorted by radiometric errors.

Let us denote the number of looks to be summed up into the multi-look image as N_L^{pro}. This number of looks is slightly less than the number of the looks within the real antenna beam N_L since the orientation instabilities may corrupt the side looks considerably. By using the low-pass filter, it is possible to select the brightest (best-illuminated) parts of the scene among all extended SAR looks with the indexes $n_L = 1, ..., N_L^{ext}$ and compose only N_L^{pro} SAR looks (called the composite looks) for further processing. It is convenient to build the following sequence of the pairs of the composite looks and their low-frequency components:

$$\{I^{pro}(n_L^{pro}, X, Y), I_{LF}^{pro}(n_L^{pro}, X, Y)\}\,, \quad n_L^{pro} = 1, ..., N_L^{pro}\,. \tag{29}$$

This sequence is kept in the ascending order with respect to the brightness:

$$I_{LF}^{pro}(n_L^{pro}, X, Y) \le I_{LF}^{pro}(n_L^{pro} + 1, X, Y)\,. \tag{30}$$

After processing of all the extended SAR looks, the brightest composite look is the look with the index $n_L^{pro} = N_L^{pro}$. These brightest values are obtained with the synthetic beams that are directed very closely to the centre of the real beam. Therefore, these brightness values are not distorted by the radiometric errors and give the estimate of the low-frequency component of the error-free SAR image to be reconstructed:

$$I_{LF}^{pro}(N_L^{pro}, X, Y) \approx I_{LF}(x, y) \cdot \tag{31}$$

This image can be used as the reference to estimate the radiometric error functions for all SAR looks:

$$R(n_L, X, Y) \approx \frac{I_{LF}(n_L, X, Y)}{I_{LF}^{pro}(N_L^{pro}, X, Y)}\,. \tag{32}$$

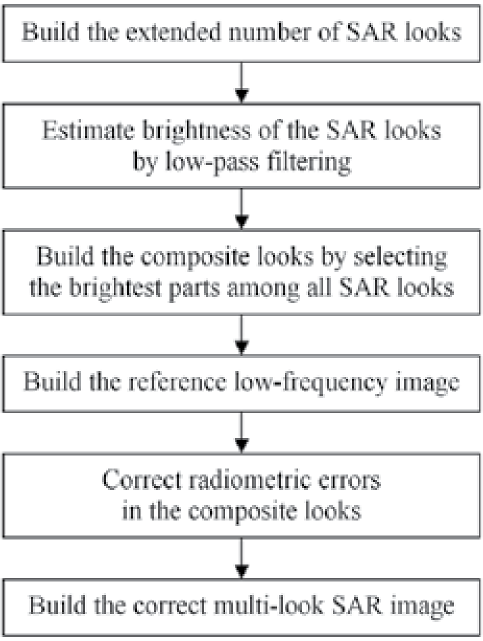

Build the extended number of SAR looks
Estimate brightness of the SAR looks by low-pass filtering
Build the composite looks by selecting the brightest parts among all SAR looks
Build the reference low-frequency image
Correct radiometric errors in the composite looks
Build the correct multi-look SAR image

Fig. 12. The main steps of the multi-look radiometric correction algorithm.

By using the estimated radiometric error functions, radiometric errors for all SAR looks can be corrected before combining them into the multi-look SAR image. The main steps of the described algorithm are shown in Fig. 12.

If the navigation system is capable of measuring accurately the fast variations of the real antenna beam orientation, and if the real antenna pattern is known, the radiometric error functions (32) can be calculated directly from the relative orientation of the synthetic beam and the real antenna beam. This approach is more rigorous and accurate than the above-described empirical approach with the image brightness estimation. Nevertheless, with this approach, it is still necessary to build extended number of SAR looks, select the best parts of SAR images among all looks, and form the composite looks for multi-look processing.

5.4 Experimental results

The proposed approach has been used for post-processing of the radar data obtained with the RIAN-SAR-Ku and RIAN-SAR-X systems described in Section 7.

The performance of the built-in geometric correction is illustrated in Fig. 13. The SAR image shown in Fig. 13a is built by using the clutter-lock technique. One can see geometric distortions caused by instabilities of the antenna orientation. The undistorted SAR image shown in Fig. 13b is formed by using the algorithm with the built-in geometric correction. Both images have 3-m resolution and are built of 3 looks. The accuracy of the geometric correction is illustrated in Fig. 13c, where the SAR image built of 45 looks and formed by using the built-in geometric correction is imposed on the Google Map image of the scene.

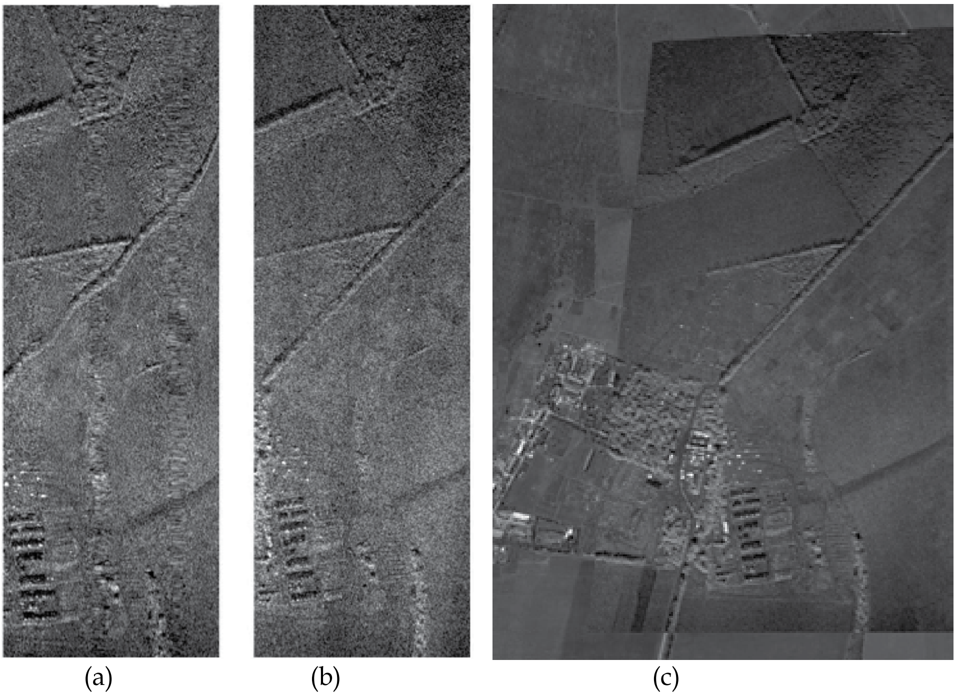


Fig. 13. Illustration of the geometric correction: (a) the 3-look SAR image built by using the clutter-lock technique, (b) the 3-look SAR image formed by using the built-in geometric correction, and (c) the 45-look SAR image formed by using the built-in geometric correction is imposed on the Google Maps image of the scene.

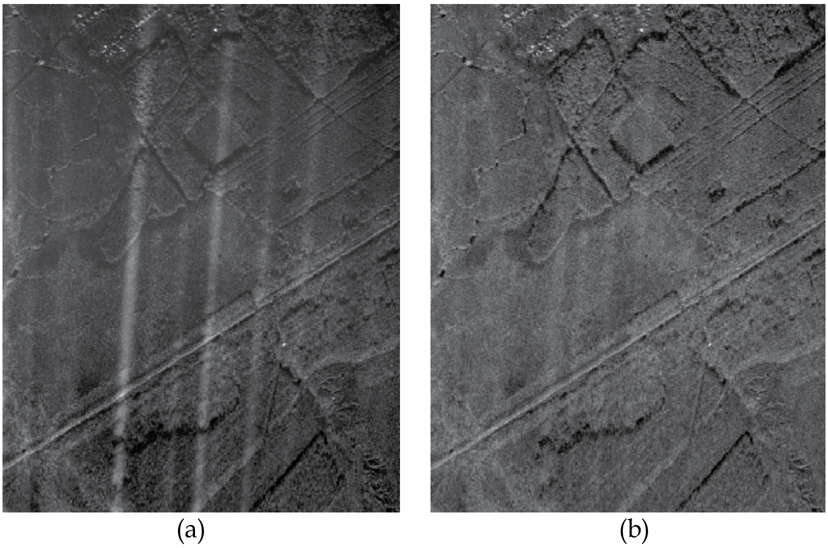


Fig. 14. Radiometric errors in the SAR image built by simple averaging of all extended SAR looks (a). SAR image formed of 5 composite SAR looks by using the proposed radiometric correction with extended number of looks.

The performance of the proposed radiometric correction by the multi-look processing with extended number of looks is illustrated in Fig. 14. The SAR image in Fig. 14a is built by simple averaging of all extended SAR looks. The image demonstrates good geometric accuracy; however the radiometric errors are presented. One can see dark and light strips in the image caused by the non-uniform illumination of the scene. The dark areas are due to the illumination for a short time when the real antenna footprint quickly moves to the neighbour areas of the scene. The light areas are correspondently illuminated for a longer time. The SAR image shown in Fig. 14b is built by using the proposed method of the multi-look radiometric correction with extended number of looks. The image is built of 5 composite SAR looks. One can see that the radiometric errors have been corrected successfully.

The obtained results prove that the described SAR processing approach can be effectively used for SAR systems installed on light-weight aircrafts with a non-stabilized antenna. An important advantage of the algorithm is that the produced SAR images are already geometrically correct at once after the synthesis, and there is no need in any additional interpolation. Another important advantage of the algorithm is the reduced requirements to the SAR navigation system. Although the aircraft velocity vector should be measured quite accurately to point the synthetic beams at the proper points on the ground, the aircraft trajectory should be measured and compensated with the high accuracy of a fraction of the radar wavelength only during the short time of the synthesis of one look. There is no need to keep so high accuracy of the trajectory measurement during the long time of the data acquisition for all looks.

6. Range-Doppler algorithm with the 1-st and 2-nd order motion compensation

The range-Doppler algorithm (RDA) is one of the most popular SAR processing algorithms. A high computational efficiency and a simplicity of the implementation are its main advantages. This algorithm belongs to the frame-based SAR processing algorithms, which use the FFT and work in the frequency domain. The motion compensation within the data frame is required. The SAR images are geometrically correct but they are originally produced in the radar coordinates “slant range – azimuth”. Therefore, the ground mapping by an interpolation is required followed by stitching of the obtained image frames into the SAR image of the ground strip. Possible radiometric errors should be additionally corrected.

6.1 The 1-st and 2-nd order motion compensation

The geometry of the motion compensation problem is illustrated in Fig. 15. The point $A(0,0,H)$ indicates the expected position of the aircraft on the reference straight line trajectory. The point $A_E(\Delta x_E, \Delta y_E, H+\Delta z_E)$ corresponds to the actual position on the real trajectory. The slant range error for the synthetic beam directed to the point $P(x_R, y_R, 0)$ on the Doppler centroid line (1), (2) at the slant range R can be written as

$$\Delta R_E(x_R, y_R) = R_E(x_R, y_R) - R, \tag{33}$$

$$R_E(x_R, y_R) = \sqrt{(\Delta x_E - x_R)^2 + (\Delta y_E - y_R)^2 + (H + \Delta z_E)^2}\ . \tag{34}$$

These relations describe both the range migration errors and the corresponding phase errors

$$\varphi_E(x_R, y_R) = -\frac{4\pi}{\lambda}\Delta R_E(x_R, y_R) \qquad (35)$$

caused by the trajectory deviations $\vec{\mathbf{r}}_E = (\Delta x_E, \Delta y_E, \Delta z_E)$.

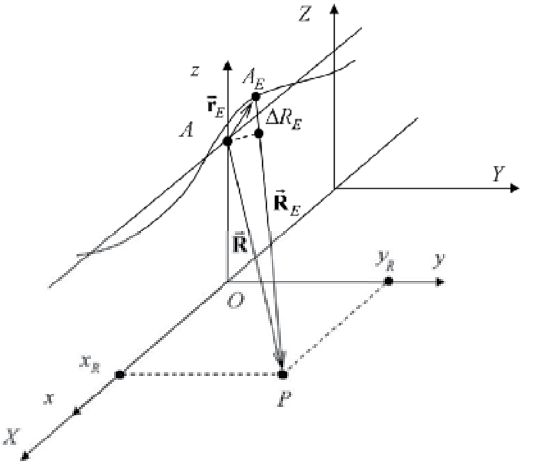


Fig. 15. Geometry of trajectory deviations.

In order to compensate the motion errors, we should correct the range migration errors (33), (34) by introducing an addition interpolation in the range direction and also correct the phase errors (35) in the azimuth direction. The corresponding correction should be performed individually for each pulse on the interval of the synthesis in the accordance with the current aircraft position error $\vec{\mathbf{r}}_E = (\Delta x_E, \Delta y_E, \Delta z_E)$. The problem is that the range error ΔR_E depends not only on the slant range, but also on the direction to the point $P(x_R, y_R, 0)$. It means that the motion errors depend on both range and azimuth and are different for different points on the scene. In other words, the same radar pulses on two overlapped intervals of the synthesis should be compensated individually for the neighbour points in the azimuth direction. Such complete and accurate motion error compensation is possible only in those SAR processing algorithms, which allow the application of an individual reference function and range migration curve for each point of SAR image. It is possible, for example, in the time-domain SAR processing algorithms considered here. However, for the most SAR processing algorithms, including the range-Doppler algorithm, the dependence of the error ΔR_E on the azimuth must be disregarded and the range dependence is taken into account only.

The motion error correction should not interfere with the range and azimuth compression. Any range-dependent motion compensation can not be applied before the range compression of the received radar pulses. Otherwise, the range LFM waveform of the transmitted pulse will be distorted. Also, the range-dependent compensation cannot be applied before the range migration correction step of the SAR processing algorithm.

Otherwise, different corrections applied for the neighbour range bins will introduce phase errors in azimuth direction.

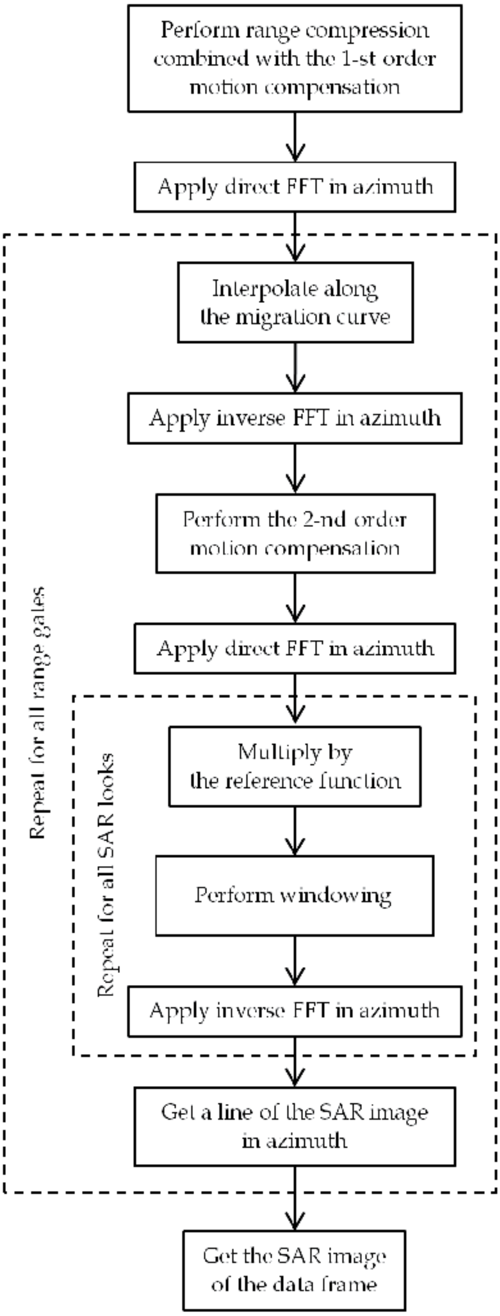

Perform range compression combined with the 1-st order motion compensation
Apply direct FFT in azimuth
Interpolate along the migration curve
Apply inverse FFT in azimuth
Perform the 2-nd order motion compensation
Apply direct FFT in azimuth
Multiply by the reference function
Perform windowing
Apply inverse FFT in azimuth
Get a line of the SAR image in azimuth
Get the SAR image of the data frame
Repeat for all range gates
Repeat for all SAR looks

Fig. 16. Range-Doppler algorithm with the 1-st and 2-nd order motion compensation.

To cope with the above problems, the motion compensation procedure for the range-Doppler algorithm (and similar FFT-based algorithms) is usually divided on two steps (Franceschetti & Lanari, 1999):

1. First-order range-independent motion compensation,
2. Second-order range-dependent motion compensation.

The first-order motion compensation includes the range delay (33), (34) of the received pulses (with interpolation) and the phase compensation (35), which are calculated for some reference range, for example, for the centre range of the swath R_C:

$$\varphi_E^{(I)}(R_C,t)=\exp\left[-i\frac{4\pi}{\lambda}\Delta R_E^{(I)}(R_C,t)\right]. \tag{36}$$

Here t is the flight time. The first-order motion compensation can be incorporated into the range compression step but it should be performed before any processing step in the azimuth, in particular, before the range migration correction in the range-Doppler algorithm, as shown in Fig. 16.

The second-order range-dependent motion compensation is performed after the range compression and the range migration correction steps. It includes the phase compensation and may (or may not) include the following range interpolation step:

$$\Delta R_E^{(II)}(R,t)=\Delta R_E(R,t)-\Delta R_E^{(I)}(R_C,t)\,. \tag{37}$$

$$\varphi_E^{(II)}(R,t)=\exp\left[-i\frac{4\pi}{\lambda}\Delta R_E^{(II)}(R,t)\right]. \tag{38}$$

Since the motion errors depend on time, it is needed to return from the range Doppler domain into the time domain by the inverse FFT, apply the corrections (37), (38), and come back into the range-Doppler domain by applying the direct FFT again, as shown in Fig. 16. After that, we can perform the azimuth compression.

After the compensation, the raw data seem like they are collected from the reference straight line trajectory, and the range-Doppler processing is performed by using the reference parameters of the data frame.

6.2 Problem of radiometric errors caused by motion compensation

The RDA performs processing of data blocks in the azimuth frequency domain assuming that the aircraft goes along a straight trajectory with a constant orientation during the time of the data frame accumulation. Therefore, instabilities of the antenna beam orientation within the data frame lead to radiometric errors in SAR images formed by the RDA.

After applying the above-described 1-st and 2-nd order motion compensation procedures, the corrected raw data demonstrate the range migration and phase behaviour as if the data were collected from the reference straight trajectory. However, the illumination of the scene by the real antenna is not changed, and radiometric errors are still presented.

Moreover, the application of the motion compensation can make the problem of radiometric errors even worse (Bezvesilniy et al., 2011c). After the motion compensation, the location of the antenna footprint on the ground should be described with respect to the position of the aircraft on the reference trajectory by the orientation angles α_{MoCo} and β_{MoCo} which are different from the angles α and β of the actual local coordinate system. The Doppler centroid values of the corrected radar data are apparently different from the Doppler centroid values before the motion compensation. For example, even if the antenna orientation is constant with respect to the actual local coordinate system, the orientation of the antenna beam can demonstrate variations with respect to the reference flight line. It means that the raw data with the constant Doppler centroid could demonstrate Doppler centroid variations after applying the motion compensation. This effect becomes more significant in the case of notably curved trajectories.

The problem of the correction of radiometric errors in SAR images formed by using the range-Doppler algorithm can be solved by the multi-look processing with extended number of looks as it was described in Section 5. It should be pointed out that the clutter-lock based on the estimation of the antenna beam orientation angles from the Doppler centroid measurements can be used together with the range-Doppler algorithm only to estimate the reference orientation angles for each data frame.

6.3 Experimental results

The range-Doppler algorithm with the 1-st and 2-nd order motion compensation procedures was implemented in the airborne RIAN-SAR-X system, what allows us to obtain multi-look SAR images in real time. The application of a wide-beam antenna enables avoiding radiometric errors in real time. An example of a 7-look SAR image with a 2-m resolution is given in Fig. 17. The application of the multi-look radiometric correction with the extended number of looks can be applied as a post-processing task to the recorded raw data.

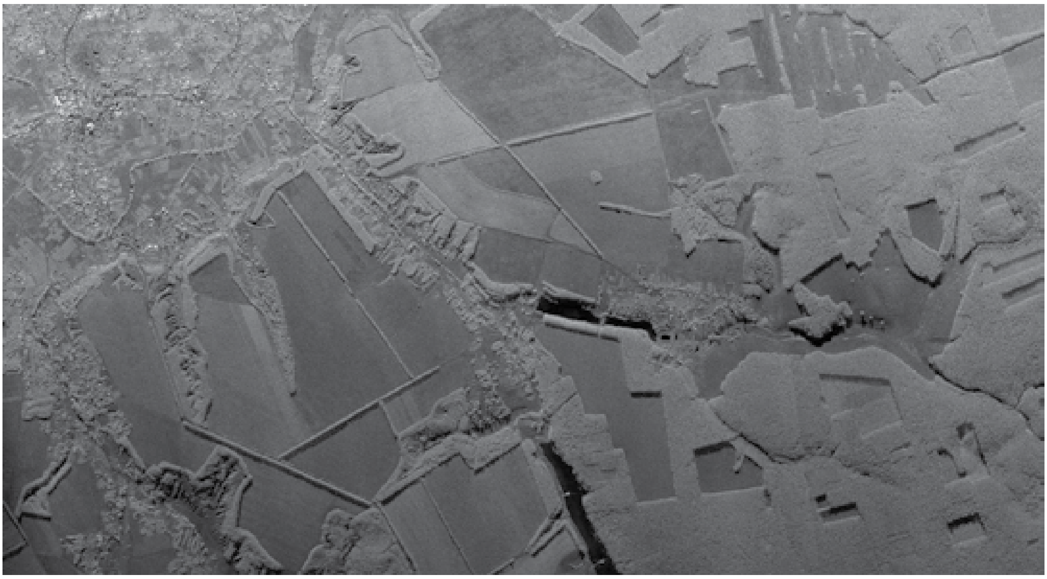


Fig. 17. An example of a 7-look SAR image obtained with the X-band SAR system.

7. Practical Ku- and X-band SAR systems

In this section we describe the design and basic technical characteristics of the mentioned already the Ku- and X-band SAR systems developed and produced at the Institute of Radio Astronomy of the National Academy of Sciences of Ukraine (Vavriv at al., 2006; Vavriv & Bezvesilniy, 2011a; Vavriv at al., 2011). The SAR systems were designed to be deployed on a light-weight aircraft. The systems were successfully operated from AN-2 and Y-12 aircrafts.

7.1 Airborne system RIAN-SAR-Ku

The Ku-band SAR system RIAN-SAR-Ku Ukraine (Vavriv at al., 2006; Vavriv & Bezvesilniy, 2011a; Vavriv at al., 2011) operates in a strip-map mode producing single-look SAR images with a 3-meter resolution in real time. The radar can perform measurements at two linear polarizations. The system has also a Motion Target Indication (MTI) capability. Characteristics of the system hardware are listed in Table 1.

7.1.1 Hardware solutions

The radar transmitter is based on a traveling-wave tube power amplifier (TWT PA). The radar transmits long pulses with the duration of 5 μs. The binary phase codding technique is used for the pulse compression to achieve a 3-meter range resolution. The M-sequences of the length of 255 are used for phase codding. The transmitted pulse bandwidth is 50 MHz.

A high pulse repetition frequency (PRF) of 20 kHz is required in the system for detection of moving targets. The application of binary phase codding allows us to simplify dramatically the hardware realization of the range compression as compared to the well-known pulse compression technique of pulses with linear frequency modulation (LFM). It is critical to manage the range compression in real time at the high PRF of 20 kHz.

The pulse repetition frequency is adjusted continuously to keep the ratio of the aircraft velocity to the PRF constant. It means that the aircraft always flights the same distance during the pulse repetition period. Such approach is used to simplify the further SAR processing.

A sensitive receiver with the noise figure of 2.5 dB is used in the SAR. The system losses are 4.0 dB. The received data are sampled with two 12-bit ADCs at the sampling frequency of 100 MHz.

For the detection of moving targets, we used the following simple principle: All signals, which are detected outside of the Doppler spectrum of the ground echo, are assumed to be signals of moving targets. This approach calls for using of a narrow-beam antenna so that the Doppler spectrum from the ground is narrow. Therefore, a long slotted-waveguide antenna of the length of 1.8 m with a 1-degree beam has been used. The antenna is actually built of two separate antennas so that the SAR system can operate at two orthogonal linear polarizations.

The usage of such narrow-beam antennas is not common for airborne SAR systems. It imposes the following limitations on SAR imaging.

First, the azimuth resolution in the strip-map mode is limited by the half of the antenna length that is about 1 m for this system. If we degrade this resolution to 3 m, which is equal to the range resolution, it is possible to build only 5 half-overlapped SAR looks.

Parameter	**RIAN-SAR-Ku**	**RIAN-SAR-X**
Transmitter		
Transmitter type	TWT PA*	SSPA**
Operating frequency	Ku-band	X-band
Transmitted peak power	100 W	120 W
Pulse repetition frequency	5 – 20 kHz	3 – 5 kHz
Pulse repetition rate	< 200 Hz / (m/s)	Not used
Pulse compression technique	Binary phase codding (M-sequences)	Linear frequency modulation
Pulse bandwidth	50 MHz	100 MHz
Pulse duration	5.12 μs	5 – 16 μs
Receiver		
Receiver type	Analogue	Digital
Receiver bandwidth	100 MHz	100 MHz
Receiver noise figure	2.5 dB	2.0 dB
System losses	4.0 dB	1.5 dB
ADC sampling frequency	100 MHz	200 MHz
ADC capacity	12 bit	14 bit
Antenna		
Antenna type	Slotted-waveguide / Horn	Slotted-waveguide
Antenna beam width in azimuth	1° / 7°	10°
Antenna beam width in elevation	40° / 40°	40°
Antenna gain	30 dB / 21 dB	20 dB
Polarization	HH or VV / VV	VV
SAR Platform		
Aircraft flight velocity	30 – 80 m/s	30 – 80 m/s
Aircraft flight altitude	1000 – 5000 m	1000 – 5000 m
Aircrafts used	AN-2, Y-12	AN-2

* TWT PA is an acronym for a traveling-wave tube power amplifier.
** SSPA is an acronym for a solid-state power amplifier.

Table 1. Characteristics of the SAR hardware systems.

Second, the antenna beam orientation should be measured with a high accuracy of about 0.1° (that is 1/10th of the antenna beam width) to avoid radiometric errors in SAR images. The application of the antenna with a wider beam would simplify this requirement. An alternative horn antenna with a 7-degree beam was used to make the system capable of producing high-quality SAR images with many looks by processing of the recorded data.

7.1.2 Signal processing solutions

Radar data processing is performed with a special PCI-board equipped with a DSP and an FPGA. Characteristics of the SAR data processing system are given in Table 2.

Parameter	RIAN-SAR-Ku	RIAN-SAR-X
Range processing		
Range resolution	3 m	2 m
Range sampling interval	1.5 m	1.5 m
Number of range gates	1024	2048 (processed) / 4096 (raw)
Range swath width	1536 m	3072 m
Azimuth processing		
SAR processing algorithm	Time-domain convolution (stream-based)	Range-Doppler algorithm (frame-based)
Real-time motion error compensation (trajectory)	No	Yes, 1st- and 2nd-order MOCO
Clutter-lock*	Line-by-line	Frame-by-frame
Pre-filtering	Yes	Yes
Azimuth resolution	3.0 m	2.0 m
Number of looks (in real time)	1	1 – 15
Ground mapping of SAR images	Post-processing	In real time
Data recording		
Raw data	Range-compressed, 7-times decimated	Uncompressed, no decimation
Recorded raw data rate	12 MB/s	80 MB/s
Pre-filtered data, navigation data, SAR images, etc.	Yes	Yes
Other capabilities		
Detection and indication of moving targets	Yes	No

* Estimation of the antenna beam orientation angles from the backscattered radar data and updating the SAR reference functions.

Table 2. Characteristics of the SAR data processing systems.

The procedure of pre-filtering was implemented to reduce the high input data rate by a coherent accumulation and down-sampling of the data in azimuth from 20 kHz to about 100 Hz that is determined by the antenna beam width.

The time-domain convolution-based SAR processing algorithm with range migration correction by interpolation is implemented, as described in Section 4. This algorithm forms each pixel of the SAR image with a separate reference function and a migration curve. Therefore, the algorithm works well under unstable flight conditions. The algorithm is fast enough for the operation in real time, if the length of the convolution is not too long. With the narrow-beam antenna and the pre-filtering procedure, this requirement has been satisfied. The SAR processing system is able to build single-look SAR images with 3-meter resolution in real time. The number of range gates is 1024 resulting in 1536-meter range swath width.

In order to measure accurately the antenna orientation, the algorithm described in Section 4 for the estimation of the antenna orientation angles directly from Doppler frequencies of backscattered radar signals was introduced. The accuracy of the estimation is about 0.1°. The angles are updated about 10 times per second what is sufficient to track fast variations of the antenna beam orientation.

The estimated angles are used to realize the clutter-lock. The pre-filter and the SAR reference functions are updated rapidly to track variations of the antenna orientation, and thus to avoid radiometric errors in SAR images.

The radar system is able to record the range-compressed data at the data rate of about 12 MB/s to hard disk drives for post-processing. A 7-times decimation of the input data stream is used to reduce the data rate for recording. The pre-filtered radar data, the navigation data, and SAR images are recorded as well.

7.2 Airborne system RIAN-SAR-X

The X-band SAR system RIAN-SAR-X Ukraine (Vavriv & Bezvesilniy, 2011a; Vavriv at al., 2011) is capable of producing high-quality multi-look SAR images with a 2-meter resolution in real time. The system is designed to operate from light-weight aircraft platforms in side-looking or squinted strip-map modes. Characteristics of the radar hardware and the signal processing systems are listed in Tables 1 and 2.

7.2.1 Hardware solutions

The radar operates in the X-band. The transmitter is based on a modern solid-state power amplifier (SSPA). The peak transmitted power is 120 W. The radar transmits long pulses with a linear frequency modulation. A direct digital synthesizer (DDS) provides frequency sweeping. The pulse duration can be chosen from 5 to 16 µs. The transmitted pulse bandwidth is 100 MHz. It gives the range resolution of 2 m. The pulse repetition frequency is from 3 kHz to 5 kHz, and that guarantees an unambiguous data sampling in the azimuth.

A digital receiver technique has been implemented. The noise figure of the receiver is 2 dB. The system losses are 1.5 dB. We have used one 200-MHz ADC with a 14-bit capacity.

The radar uses a compact slotted-waveguide antenna with a 10-degree beam. The wide beam is used, first, to avoid radiometric errors during the formation of SAR images in real time, and, second, to enable building of high-quality SAR images with a large number of looks at a post-processing stage. The antenna is firmly mounted on the aircraft; however it can be installed either into a side-looking or a 40-degree-squinted position.

The SAR system is designed to be operated from a light-weight aircrafts. During test flights, the SAR system was successfully deployed on an AN-2 aircraft. The aircraft flight altitude could be from 1000 m to 5000 m, and the aircraft flight velocity is expected to be from 30 m/s to 80 m/s. The implemented SAR processing algorithms can operate beyond of these intervals of flight parameters with minor adjustments.

7.2.2 Signal processing solutions

A strip-map SAR processing is performed by using a frame-based range-Doppler algorithm with motion compensation, as described in Section 6. The SAR system is capable of producing SAR images with a 2-meter resolution formed of up to 15 looks in real time. A scheme with half-overlapped frames is implemented to provide continuous surveillance of the strip without gaps despite of possible motion instabilities.

The SAR navigation system is based on a simple GPS-receiver capable of measuring the aircraft position and the aircraft velocity vector. The measured position is used to link the obtained SAR images to ground maps, and also to know the flight altitude above the ground. The aircraft flight trajectory is integrated from the measured aircraft velocity with a sufficient accuracy to perform the motion compensation. The antenna beam orientation is estimated from Doppler frequencies of the backscattered radar signals. The pitch and yaw antenna orientation angles are used both for motion compensation and for the aperture synthesis. Such angle estimation is a kind of clutter-lock processing allowing to track variations of the antenna beam orientation by adjusting the SAR data processing algorithm from one radar data frame to another.

The signal processing system is divided on two main parts. The first part of the system performs: 1) range compression of LFM pulses combined with the 1st-order motion compensation, 2) calculation of Doppler centroid values for each range gate (by FFT in azimuth) and estimation of the antenna orientation angles, and 3) pre-filtering of the range-compressed data. This processing is performed in a special PCI board with a DSP and an FPGA.

The second part of the data processing system forms multi-look SAR images by using a range-Doppler algorithm with the 2nd-order motion compensation. This processing is performed on a PC with an Intel Quad Core CPU (the above-mentioned PCI board is installed on this PC). It gives a flexibility in setting the azimuth processing parameters and allows using the developed SAR system as a suitable test-bed for testing new modifications of various frame-based SAR algorithms.

Stitching of the obtained SAR images into a continuous strip map can be performed on a client PC (or a notebook), while viewing the data in real time or offline.

The SAR system is capable of recording the original uncompressed radar data on a solid-state drives organized in a RAID-0 array at the full pulse repetition rate up to 5 kHz. These data are stored together with the navigation data (original GPS measurements, integrated trajectories, estimated orientation angles, motion compensation curves, etc.), as well as the pre-filtered range-compressed data and the SAR images formed in real time. Recorded data are used further in our research and development activity on SAR systems.

8. Conclusion

The presented results indicate that some of the essential problems that limited the development of SAR systems for small aircrafts are solved. In particular, the problem of the antenna beam orientation evaluation has been solved by extracting this information from the Doppler shift of the radar echoes. This technique enables to use only a simple GPS receiver to provide a reliable SAR operation. Simultaneously, the problem of the correction of the geometrical distortions in SAR images has been solved via the introduction of a signal processing algorithm, which provides pointing multi-look SAR beams exactly to the nodes of a rectangular grid on the ground plane. The proposed multi-look processing algorithm with extended number of looks has demonstrated a high efficiency for the correction of radiometric errors. The suggested approaches have been successfully implemented in and tested with Ku- and X-band SAR systems deployed on small aircrafts. It should be pointed that these solutions are as well useful for SAR systems deployed on other platforms.

9. Acknowledgment

The authors would like to thank all of their colleagues at the Department of Microwave Electronics, Institute of Radio Astronomy of the National Academy of Sciences of Ukraine for their help and fruitful discussions. In particular, we indebt to Dr. V. V. Vynogradov, Dr. V. A. Volkov, Dr. S. V. Sosnytskiy, Mr. R. V. Kozhyn, Mr. S. S. Sekretarov, Mr. A. Kravtsov, Mr. A. Suvid, and Mr. I. Gorovyi for their essential contributions to the development of practical SAR systems.

10. References

Bamler, R. & Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse Problems*, Vol. 14, pp. R1-R54.

Bezvesilniy, O. O., Dukhopelnykova, I. V., Vynogradov, V. V., & Vavriv, D. M. (2006). Retrieving 3D relief from radar returns with single-antenna, strip-map airborne SAR. *Proceedings of the 6th European Conference on Synthetic Aperture Radar (EUSAR2006)*. 16-18 May 2006, Dresden, Germany. pp. 1-4. (CD-ROM Proceedings).

Bezvesilniy, O. O., Dukhopelnykova, I. V., Vynogradov, V. V. & Vavriv, D. M. (2007). Retrieving 3-D topography by using a single-antenna squint-mode airborne SAR. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 11, pp. 3574-3582.

Bezvesilniy, O. O., Vynogradov, V. V. & Vavriv, D. M. (2008). High-accuracy Doppler measurements for airborne SAR applications. *Proceedings of the 5th European Radar Conference (EuRAD2008)*. 30–31 Oct. 2008, Amsterdam, The Netherlands. pp. 29-32.

Bezvesilniy, O. O., Gorovyi, I. M., Sosnytskiy, S. V., Vynogradov V. V. & Vavriv D. M. (2010a). Multi-look stripmap SAR processing algorithm with built-in correction of geometric distortions. *Proceedings of the 8th European Conference on Synthetic Aperture Radar (EUSAR2010)*. 7-10 June 2010, Aachen, Germany. pp. 712-715.

Bezvesilniy, O. O., Gorovyi I. M., Sosnytskiy, S. V., Vynogradov V.V. & Vavriv D.M. (2010b). Multi-look SAR processing with build-in geometric correction, *Proc. of the 11th Int. Radar Symposium (IRS-2010)*. June 16-18, Vilnius, Lithuania. Vol. 1. pp. 30-33.

Bezvesilniy, O. O., Gorovyi, I. M., Vynogradov V.V. & Vavriv D.M. (2010c). Correction of radiometric errors by multi-look processing with extended number of looks, *Proceedings of the 11th Int. Radar Symposium (IRS-2010)*. June 16-18, Vilnius, Lithuania. Vol. 1. pp. 26-29.

Bezvesilniy, O. O., Gorovyi, I. M., Sosnytskiy, S. V., Vynogradov, V. V. & Vavriv, D. M. (2010d). Improving SAR images: Built-in geometric and multi-look radiometric corrections. *Proceedings of the 7th European Radar Conference (EuRAD2010)*. 30 September - 1 October 2010, Paris, France. pp. 256-259.

Bezvesilniy, O. O., Gorovyi, I. M., Sosnytskiy, S. V., Vynogradov, V. V. & Vavriv, D. M. (2011a). SAR processing algorithm with built-in geometric correction. *Radio Physics and Radio Astronomy*, Vol. 16, No. 1, pp. 98-108.

Bezvesilniy, O. O., Gorovyi, I. M., Vynogradov, V. V. & Vavriv, D. M. (2011b). Multi-look radiometric correction of SAR images. *Radio Physics and Radio Astronomy*, Vol. 16, No. 4, pp. ???-??? (Accepted for publication).

Bezvesilniy, O. O., Gorovyi, I. M., Vynogradov, V. V. & Vavriv, D. M. (2011c). Range-Doppler algorithm with extended number of looks, *Proceedings of the 2011 Microwaves, Radar and Remote Sensing Symposium (MRRS-2011)*. August 25-27, Kiev, Ukraine. pp. 203–206.

Blacknell, D., Freeman, A., Quegan, S., Ward, I. A., Finley, I. P., Oliver, C. J., White, R. G. & J. W. Wood (1989). Geometric accuracy in airborne SAR images. *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 25, No. 2, pp. 241-258.

Buckreuss, S. (1991). Motion errors in an airborne synthetic aperture radar system. *European Transactions on Telecommunications*, Vol. 2, No. 6, pp. 655–664.

Carrara, W. G., Goodman, R. S. & Majewski, R. M. (1995). *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, ISBN 0-89006-728-7.

Cumming, I. G. & Wong, F. H. (2005). *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*, Artech House, ISBN 1-58053-058-3.

Franceschetti, G. & Lanari, R. (1999). *Synthetic Aperture Radar Processing*, CRC Press, ISBN 0-8493-7899-0.

Li, F.-K., Held, D. N., Curlander, J. C. & Wu, C. (1985). Doppler parameter estimation for spaceborne synthetic-aperture radars. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 23, No. 1, pp. 47-56.

Madsen, S. N. (1989). Estimating the Doppler centroid of SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 25, No. 2, pp. 134-140.

Moreira, A. (1991). Improved multilook techniques applied to SAR and SCANSAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 29, No. 4, pp. 529-534.

Oliver, C. J. & Quegan, S. (1998). *Understanding Synthetic Aperture Radar Images*, Artech House, ISBN 0-89006-850-X.

Rosen, P. A., Hensley, S., Joughin, I. R., Li, F.-K., Madsen, S. N., Rodriguez, E. & Goldstein, R. M. (2000). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, Vol. 88, No. 3, pp. 333-382.

Vavriv, D. M., Vynogradov, V. V., Volkov, V. A., Kozhyn, R. V., Bezvesilniy, O. O., Alekseenkov, S. V., Shevchenko, A. V., Belikov, A., Vasilevsky, M.P. & Zaikin D. I. (2006). Cost-effective airborne SAR. *Radio Physics and Radio Astronomy*, Vol. 11, No. 3, pp. 276-297.

Vavriv, D. M. & Bezvesilniy, O. O. (2011a). Developing SAR for small aircrafts in Ukraine. *Proceedings of the 2011 IEEE MTT-S International Microwave Symposium (IMS 2011)*. 5-10 June 2011, Baltimore, USA. pp. 1-4. (CD-ROM Proceedings).

Vavriv, D. M. & Bezvesilniy, O. O. (2011b). Potential of multi-look SAR processing. *Proceedings of the 5th Int. Conference on Recent Advances in Space Technologies (RAST 2011)*. 9-11 June 2011, Istanbul, Turkey. pp. 365-369.

Vavriv, D. M., Bezvesilniy, O. O., Kozhyn, R. V., Vynogradov, V. V., Volkov, V. A. & Sekretarov, S. S. (2011). SAR systems for light-weight aircrafts. *Proceedings of the 2011 Microwaves, Radar and Remote Sensing Symposium (MRRS-2011)*. August 25-27, Kiev, Ukraine. pp. 15-19.

Wehner, D.R. (1995). *High-Resolution Radar (2nd Ed.)*, Artech House, ISBN 0-89006-727-9.

21

Avionics Design for a Sub-Scale Fault-Tolerant Flight Control Test-Bed

Yu Gu[1], Jason Gross[2], Francis Barchesky[3],
Haiyang Chao[4] and Marcello Napolitano[5]
West Virginia University
USA

1. Introduction

The increasingly widespread use of Unmanned Aerial Vehicles (UAVs) has provided researchers with platforms for several different applications:

1. For carrying remote sensing or other scientific payloads. Highly publicized examples of such applications include the forest fire detection effort jointly conducted by NASA Ames research centre and the US Forest Service (Ambrosia et al., 2004), and the mission into the eye of hurricane Ophelia by an Aerosonde® UAV (Cione et al., 2008);
2. For evaluating different sensing and decision-making strategies as an autonomous vehicle. For examples, an obstacle and terrain avoidance experiment was performed at Brigham Young University to navigate a small UAV in the Goshen canyon (Griffiths et al., 2006); an autonomous formation flight experiment was performed at West Virginia University (WVU) with three turbine-powered UAVs (Gu et al., 2009);
3. As a sub-scale test bed to help solving known or potential issues facing full-scale manned aircraft. For example, a series of flight test experiments were performed at Rockwell Collins (Jourdan et al., 2010) with a sub-scale F-18 aircraft to control and recover the aircraft after wing damages. Another example is the X-48B blended wing body aircraft (Liebeck, 2004) jointly developed by Boeing and NASA to investigate new design concepts for future-generation transport aircraft.

Each of these applications poses different requirements on the design of the on-board avionic package. For example, the remote sensing platforms are often tele-operated by a ground pilot or controlled with a Commercial-off-the-Shelf (COTS) or open-source autopilot (Chao et al., 2009). The UAVs for sensing and decision making research often requires a higher level of customization for the avionic system. This can be achieved through either

[1] Research Assistant Professor, Mechanical and Aerospace Engineering (MAE) Department, West Virginia University (WVU), Morgantown, WV 26506, Email: Yu.Gu@mail.wvu.edu;
[2] Ph.D., MAE Dept., WVU, now at Jet Propulsion Laboratory, Pasadena, CA, Email: Jason.Gross@jpl.nasa.gov;
[3] M.S. Student, MAE Dept., WVU, Email: fjbarchesky@gmail.com;
[4] Post-Doctoral Research Fellow, MAE Dept, WVU, Email: Haiyang.Chao@mail.wvu.edu;
[5] Professor, MAE Dept., WVU, Email: Marcello.Napolitano@mail.wvu.edu.

augmenting a COTS autopilot with a dedicated payload computer (Miller et al., 2005), or by having an entirely specialized avionics design (Evans et al., 2001). An alternative approach for smaller UAVs is to instrument an indoor testing environment (How et al., 2008) for measuring aircraft states so that a less complex avionic system could be used on-board the aircraft.

The avionic systems for sub-scale aircraft aimed at improving the safety of full-scale manned aircraft have a different set of design requirements. In addition to providing the standard measurement and control functions, the avionic system also needs to enable the simulation of different aircraft upset or failure conditions. Two general approaches have been used by different research groups. The first approach is to develop a highly realistic experimental environment in simulating a full-scale aircraft operation. For example, the Airborne Subscale Transport Aircraft Research Test bed (AirSTAR) program at NASA Langley research centre uses dynamically scaled airframe equipped with customized avionics for aviation safety research (Jordan et al., 2006) (Murch, 2008). During the research portion of the flight, the aircraft is controlled by a ground research pilot augmented by control algorithms running at a mobile ground station. An alternative approach is to develop a low-cost and expansible aircraft/avionic system for evaluating high-risk flight conditions (Christophersen et al., 2004).

Sub-scale aircraft have played critical complimentary roles to full-scale flight testing programs due to lower risks, costs, and turn-around time. The objective of this chapter is to discuss the specific avionics design requirements for supporting these experiments, and to share the design experience and lessons learned at WVU over the last decade of flight testing research. Specifically, in this chapter, detailed information for a WVU Generation-V (Gen-V) avionic system design is presented, which is based on an innovative approach for integrating both human and autonomous decision-making capabilities. Due to the high risk and uncertain nature of experiments that explore adverse flight conditions, the avionics itself is designed to reduce the risk of a Single Point of Failure (SPOF). This makes it possible to achieve a reliable operation and seamless flight mode switching. The Gen-V avionics design builds upon several earlier generations of WVU avionics that supported a variety of research topics such as aircraft Parameter Identification (PID) (Phillips et al., 2010), formation flight control (Gu et al., 2009), fault-tolerant flight control (Perhinschi et al., 2005), and sensor fusion (Gross et al., 2011).

The rest of the chapter is organized as follows. Section 2 introduces the general design requirements for avionic systems used in fault-tolerant flight control research. Section 3 discusses the overall hardware design architecture and main sub-systems. Section 4 presents the control command signal distribution logic that enables the flexible and reliable transition among different flight modes. Section 5 presents the aircraft on-board software architecture and the real-time Global Positioning System/Inertial Navigation System (GPS/INS) sensor fusion algorithm. Ground and flight testing procedures and results for validating avionics functionalities are discussed in Section 6, and finally, Section 7 concludes the chapter.

2. Avionics design requirements for fault-tolerant flight control research

Fault tolerant flight control research pose special challenges for avionics design due to the complex nature of aviation accidents. The occurrence of aviation accidents can be attributed to many factors, such as weather conditions (e.g. icing or turbulence), pilot errors (e.g. disorientation or mis-judgment), air and ground traffic management errors, and a variety of sub-system failures (e.g. sensor, actuator, or propulsion system failures). Furthermore, the introduction of new technologies in aviation systems poses new threats to the safe operation of an aircraft. For example, modern fly-by-wire flight control systems are known to introduce new failure modes due to their dependence on computers and avionics (Yeh, 1998). Increased automation and flight deck complexity could also potentially degrade situational awareness, and require increased and highly aircraft-specific pilot training. These factors could potentially create new failure scenarios that have not yet been recognized as causes of accidents.

2.1 Research requirements

Due to the complexity of aviation accidents, a multi-functional avionics design is needed to support the fault-tolerant flight control research. The most important requirements for such a design include maintaining accurate and timely measurements of aircraft states, having the ability to emulate various aircraft upset or failure conditions, and providing a flexible interface between humans and automatic control systems. A breakdown of more specific avionics requirements for several aviation safety related research topics is summarized in Table 1.

Research Topic	Specific Avionics Requirements
Aircraft modelling with manually injected manoeuvres	High quality sensor measurements; adequate update rate; monitoring of pilot activities; precise time-alignment of all measured channels.
Aircraft modelling with an On-Board Excitation System (OBES)	Ability to automatically apply pre-specified waveform inputs to control effectors.
Failure emulation	Ability to inject and remove simulated aircraft sub-system failures, such as failures in a particular sensor, actuator, or propulsion unit, or in the control command transmission link.
Fault-tolerant flight control (automatic)	Ability to command and reconfigure individual aircraft control effectors; having low system latency and abundant computational resources.
Fault-tolerant flight control (pilot-in-the-loop)	Ability to augment the pilot command with automatic control algorithms.

Table 1. Design requirements for typical fault-tolerant flight control research topics.

2.2 Operational scenarios

A fundamental difference between operating a sub-scale and a full-scale aircraft is the absence of humans on-board. The removal of the physical presence of human pilots allows the testing of high-risk flight conditions and reduces the cost of the experiment. However, pilots are integral components of modern aviation systems and contributed to 29% of "*fatal accidents involving commercial aircraft, world-wide, from 1950 thru 2009 for which a specific cause is known*" (Planecrashinfo.com, 2011). Pilots are also the ultimate decision-makers on-board; therefore, the evaluations of their response under adverse situations and the detailed understanding of their interaction with the rest of the flight control system play crucial roles in improving aviation safety (NRC, 1997). From this point of view, a realistic fault-tolerant flight testing program should not only take advantage of the low-cost and low-risk features of the sub-scale aircraft, but also to provide a highly relevant operational environment for human pilots. Figure 1 illustrates two potential sub-scale flight testing scenarios for different research topics.

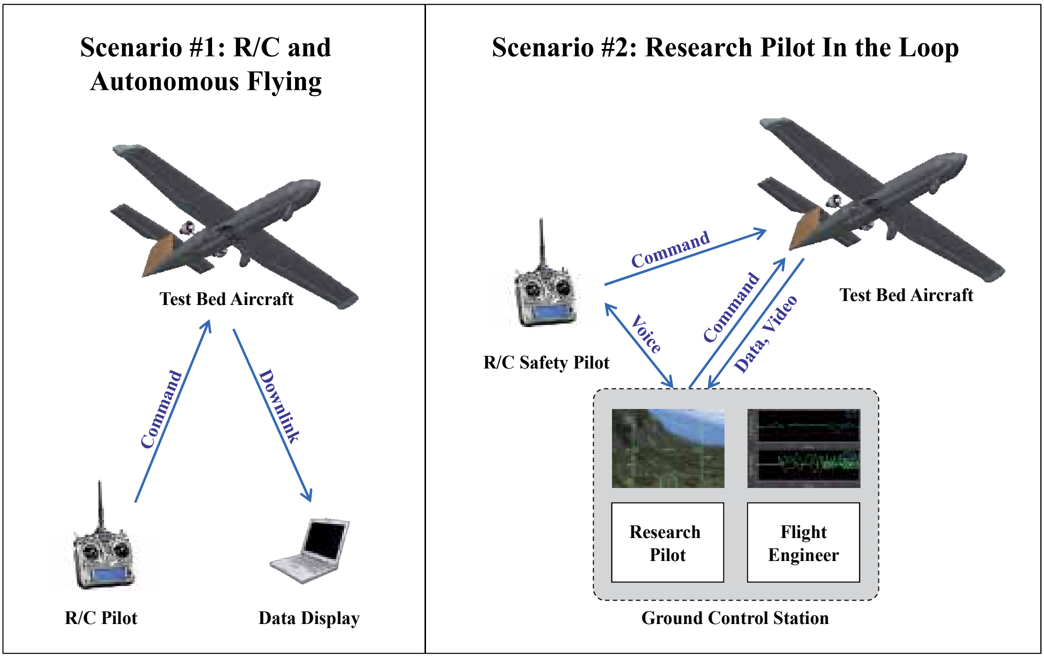


Fig. 1. Two sub-scale aircraft operational scenarios.

Scenario #1 can be used for modelling the aircraft dynamics under different flight conditions and to evaluate automatic Guidance, Navigation, and Control (GNC) algorithms. Within this scenario, a Remote Control (R/C) pilot either directly controls the test bed aircraft or serves as a safety monitor to the on-board flight control system during the test. This scenario provides a simple but reliable method for operating a research aircraft.

Scenario #2 expands upon the first scenario by adding an additional Ground Control Station (GCS), a research pilot, and a flight engineer. The GCS provides a simulated cockpit for the research pilot, who controls the aircraft based on the transmitted flight data and video. This configuration allows the research pilot to have a first-person perspective and enables a fully instrumented flight operation. The role of the flight engineer is to control the configuration of the aircraft by adjusting controller modes/parameters or inject/remove different failure scenarios during the flight. The R/C safety pilot monitors the flight and takes over the aircraft control under emergency situations or during non-research portions of the experiment. Scenario #2 provides additional capabilities for studying the pilot’s role in a flight.

2.3 Operational modes

To support the previously described research topics and the two operational scenarios, the following operational modes are typically required:

1. *Manual Mode I – Direct Vision*. An R/C safety pilot has full authority on all control channels in the basic stick-to-surface format. The pilot should always have the option of switching to this mode instantaneously under any conditions as long as the R/C link is available. This mode can be used for aircraft manual take-off and landing, manual PID manoeuvre injection, as well as emergency recovery from other operational modes;
2. *Manual Mode II – Virtual Flight Display*. A research pilot inside the ground control station has full authority on all control channels;
3. *Fully Autonomous Mode.* The on-board flight control system has full control of the aircraft, while the R/C pilot is only serving as an observer and safety backup;
4. *Partially Autonomous Mode.* A subset of the flight control channels is under autonomous control while other channels are still operated by the ground pilot;
5. *Pilot-In-The-Loop Mode.* The pilot command is supplied as input to a Stability Augmentation System (SAS) or a Control Augmentation System (CAS). This mode allows for studying the interaction between a human pilot and the automatic control system;
6. *Failure Emulation Mode*. A simulated failure condition is induced by the on-board computer to one or multiple control channels, while the remaining channels could be under manual, autonomous, or pilot-in-the-loop control;
7. *Fail-Safe Modes*. In the event that the ground pilot could not maintain manual control of the aircraft due to loss of an R/C link, the avionic system should explore redundant communication links and on-board autonomy to help in regaining the aircraft control or minimize the damage of a potential accident.

2.4 Hardware requirement

Due to the high-risk involved in testing various adverse flight conditions and the need for switching between multiple operational modes, the reliability requirements for the avionics hardware are significantly higher than that of a conventional autopilot system for a similar class of UAV. In other words, the avionic system needs to be fault-tolerant itself, and its design should minimize the risk of a SPOF condition. For example, redundant command and control links are needed in case the primary link is lost or interfered. Additionally, the safety pilot should be able to instantaneously switch back to the manual mode from any other operational mode, even in the event of main computer shutdown or power loss.

Additional requirements to the avionics hardware design typically include low-cost, low-weight, low-power consumption, low Electromagnetic Interference (EMI), configurable and expandable, and user-friendly.

3. WVU avionics architecture and main sub-systems

Based on design requirements outlined in the previous section, a Gen-V avionic system is being developed for a WVU '*Phastball*' sub-scale research aircraft. The '*Phastball*' aircraft has a 2.2 meter wingspan and a 2.2 meter total length. The typical take-off weight is 10.5 Kg with a 3.2 Kg payload capacity. The aircraft is propelled by two brushless electric ducted fans;

each can provide up to 30 N of static thrust. The use of electric propulsion systems simplifies the flight operations and reduces vibrations on the airframe. Additionally, the low time constant associated with an electric ducted fan allows it to be used directly as an actuator or for simulating the dynamics of a slower jet engine. The cruise speed of the '*Phastball*' aircraft is approximately 30 m/s. As a dedicated test-bed for fault-tolerant flight control research, the following nine channels can be independently controlled on the '*Phastball*' aircraft: left/right elevators, left/right ailerons, left/right engines, rudder, nose gear, and longitudinal thrust vectoring.

The avionic system features a flight computer, a nose sensor connection board, a control signal distribution board, a sensor suite, an R/C sub-system, a communication sub-system, a power sub-system, and a set of real-time software. It performs functions such as data acquisition, signal conditioning & distribution, GPS/INS sensor fusion, GNC, failure emulation, aircraft health monitoring, and failsafe functions. Figure 2 shows the '*Phastball*' aircraft along with the main avionics hardware components.

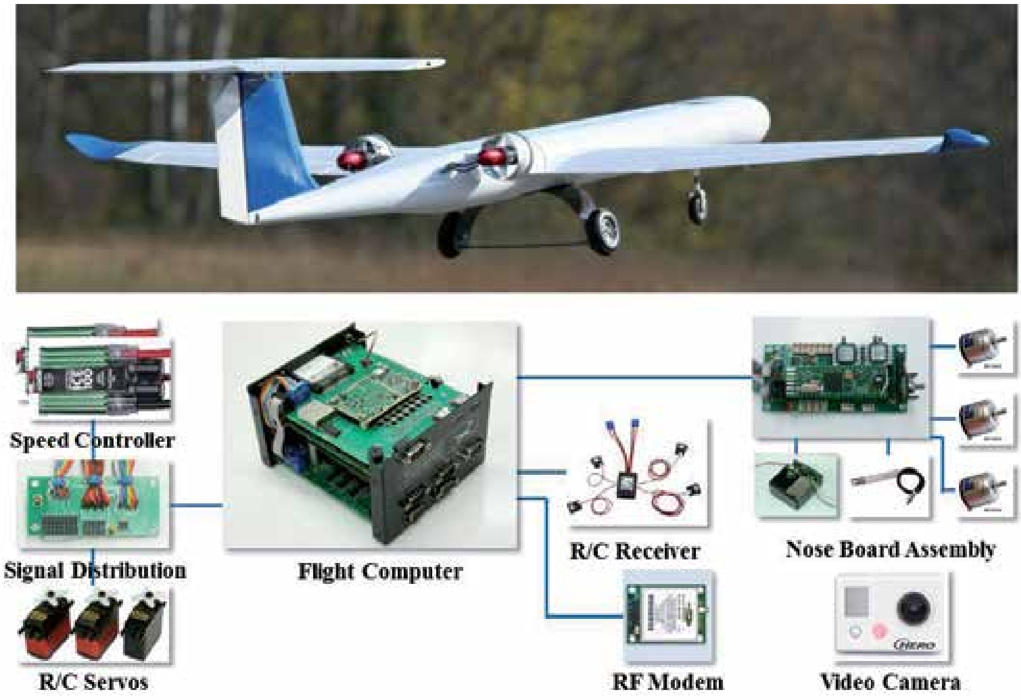


Fig. 2. '*Phastball*' aircraft and main avionics hardware components.

A detailed functioning block diagram for the Gen-V avionics hardware design is provided in Figure 3. The functionality of each main sub-system is described in the following sections.

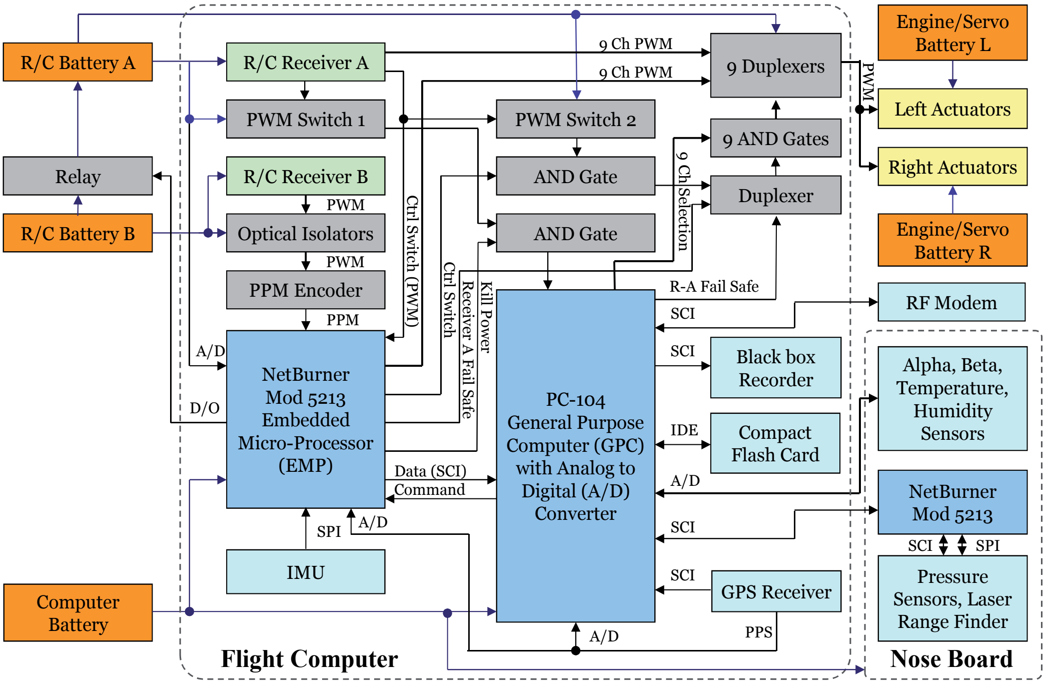


Fig. 3. Functional block diagram for the WVU Gen-V avionics hardware design.

3.1 Flight computer

The Gen-V flight computer integrates the functions of data acquisition, signal conditioning, GPS/INS sensor fusion, failure emulation, automatic control command generation, and control command distribution into a compact package. In terms of hardware, the following main components are included in the flight computer:

1. An Analog Devices® ADIS16405 Inertial Measurement Unit (IMU) that measures the aircraft 3-axis accelerations and 3-axis angular rates. Additionally, it provides readings of the magnetic field for potential use in the navigation filter for an improved aircraft attitude estimation;
2. A Novatel® OEMV-1 GPS receiver that provides aircraft position and velocity measurements. It also provides the precision time information in the form of Pulse Per Second (PPS) signal, which is used to synchronize measurements from different parts of the avionic system;
3. A Netburner® MOD5213 Embedded Micro-Processor (EMP) provides lower level interfaces for measuring human pilot commands in the Pulse-Position Modulation (PPM) format, generating on-board control command in the Pulse-Width Modulation (PWM) format, collecting data from the IMU through a Serial Peripheral Interface (SPI), monitoring battery voltages, and communicating with a general-purpose computer. The EMP also monitors two important PWM signals from the R/C receiver: a *ctrl*-switch and a *kill*-switch. The state of the *ctrl*-switch determines whether the aircraft will be operating in the manual mode or one of the other modes. The *kill*-switch gives pilot the option to power-off the computer during flight if needed for achieving improved

ground-control reliability during the safety critical (such as landing) portion of the flight;

4. Two COTS PWM switches provide independent monitoring of the critical *ctrl*-switch and *kill*-switch;
5. An 800 MHz PC-104+ form factor General-Purpose Computer (GPC) hosts the aircraft on-board software. It also provides additional 16 Analog to Digital Conversion (ADC) channels and 6 Serial Communication Interfaces (SCI) for communicating with the GPS receiver, the EMP, the nose board assembly, and the ground control station;
6. A logic network that distributes control command from both human pilots and automatic control systems to individual actuators based on the selected operational mode of the avionic system;
7. A compact flash memory card storing the operating system, the on-board software, and the collected flight data;
8. A black-box data recorder stores a real-time stream of sensory data, control command, and the avionics health information during the flight.

A detailed description of the aircraft control command generation and distribution is provided in Section 4.

3.2 Sensor suite

In addition to the IMU and the GPS receiver embedded inside the Gen-V flight computer, the '*Phastball*' aircraft is also equipped with three P3America® MP1545A inductive potentiometers for measuring aircraft flow angles, two Sensor Technics® pressure sensors for measuring the dynamic and static pressures, a Measurement Specialities® HTM2500 temperature and relative humidity sensor, and an Opti-Logic® RS400 laser range finder. Additionally, the pilot input, engine operating parameters, and R/C receiver status are also recorded in flight. The aircraft attitude angles are provided with a real-time GPS/INS sensor fusion algorithm, which will be described in Section 5.

3.3 Power system

To reduce SPOF, the arrangement of battery power has been carefully determined. A total of five battery packs are used to power different components of the avionic system. Specifically, an R/C battery-A is connected to R/C receiver-A and an logic network for control command distribution; an R/C battery-B is used to power receiver-B; the computer battery powers EMP, GPC, and all sensing, communication, and data storage devices; engine/servo batteries L and R power the left and right side engines and R/C servos independently. With this configuration, the failure of any given battery would not cause a total loss of aircraft control during the flight. Specifically, if EMP detects that receiver-A battery is low, it activates a relay to tie up R/C batteries A and B so that there would be enough power for a safe landing. If the computer battery loses its power, the logic network powered by receiver-A battery automatically switches to the manual mode and gives the R/C pilot full control authority. If one of the engine/servo batteries fails, the pilot still has independent control for half of the aircraft actuators (propulsion and control surfaces) and would be able to perform a controlled landing.

3.4 Ground control station

The GCS computer collects the aircraft downlink telemetry data, the nose camera video, the weather information, the GPS time/position measurements, the voice communication, as well as inputs from the R/C pilot, the research pilot, and the flight engineer. The data stream is accessible by the research pilot, the flight engineer, and field researchers in near real-time through a local network. For the research pilot station, three displays are provided including an X-Plane® based synthetic-vision primary flight display overlapped with a Heads-Up Display (HUD) that shows the flight parameters and mission constraints, a flight instrumentation display with a navigation window, and a screen showing the real-time flight video transmitted from the aircraft nose camera. The research pilot flies the aircraft through a set of joystick, rudder pedals, and throttle handles. The flight engineer has access to all available flight data and can change the aircraft operational mode or inject/remove failures with or without notifying the research pilot. Figure 4 shows the layout of the GCS vehicle.

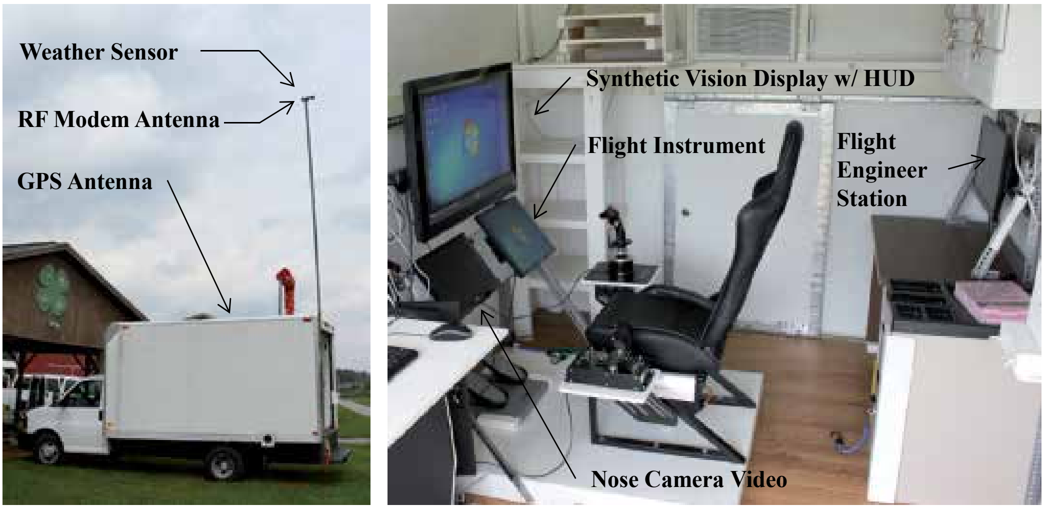


Fig. 4. The exterior (left) and interior (right) of the ground control station vehicle.

The duplex communications between the ground control station and the test bed aircraft is provided with a pair of 900 MHz Freewave ® Radio Frequency (RF) modems. The downlink communication packet contains information about aircraft states and avionics health conditions. The uplink packet integrates both the research pilot control commands and the flight engineer configuration commands. Both the uplink and downlink data are transmitted at a rate of 50Hz.

4. Control command signal distribution

One of the important features of the WVU Gen-V avionics design is its ability to provide a flexible and reliable interface between control commands generated by humans and automatic controllers. This capability is achieved through the interaction of different hardware components and software functions.

4.1 Control command generation

Depending on mission requirements, the aircraft control command could come from several potential sources:

1. *The R/C pilot*. The R/C pilot commands are provided to the flight computer through two redundant R/C receivers (A & B). The antennas of the two receivers are installed at different locations of the aircraft to reduce the likelihood that both are interfered at the same time. Each receiver can operate in either a nominal mode, when maintaining a good reception of the radio signal, or a fail-safe mode, when the communication with the radio transmitter is lost. In the fail-safe mode, each receiver channel output is either independently programmed to a pre-set value, or assigned to latch on to the last received value from the R/C transmitter.
   For R/C receiver-A, 9-channel pilot control commands are sent directly to a duplexer network for later distribution to individual actuators. Two additional channels are used as *ctrl*-switch and *kill*-switch. In order to provide information about both operational modes and the R/C receiver status, the *ctrl*-switch is programmed to have three different output levels: a lower pulse width for '*ctrl*-switch off', a higher pulse width for '*ctrl*-switch on', and a median pulse width indicating the receiver went into a fail-safe mode. A pulse width indicating '*ctrl*-switch on' may trigger a fully autonomous, a partially autonomous, or a pilot-in-the-loop mode depending on additional hardware and software settings.
   The output of receiver-B is first processed with a PPM encoder before being measured with an EMP General-Purpose Timer (GPT). This pilot input is then transmitted to GPC to be used by the flight control system;
2. *The research pilot at GCS*. Commands from the research pilot are transmitted through a pair of RF modems to the flight computer. This signal can be used to control the aircraft directly or indirectly through a SAS or CAS controller;
3. *The Flight Control System (FCS)*. The FCS running inside GPC generates the automatic control command based on sensor feedbacks as well as pilot commands provided through either receiver-B (for the R/C safety pilot) or the RF modems (for the research pilot);
4. *Failure Emulation Software (FES)*. A faulty actuator locked at a given deflection or a failed engine can both be simulated by sending a constant value to the selected control channel. A slower responding engine can be simulated by inserting additional dynamics between the control command and the engine speed controller. A floating control surface can be simulated with the feedback from a local flow indicator. More complicated failure scenarios can also be introduced through exploring feedbacks from various sensors;
5. *On-Board Excitation System (OBES)*. OBES provides specified waveform to be applied on aircraft control actuators. The OBES manoeuver can be either stand-alone or superimposed onto the pilot or controller commands.

The R/C pilot command is in a PWM format recognizable by R/C servos and engine speed controllers. The commands from the research pilot, FCS, FES, and OBES are first integrated (selected or combined) within the on-board software before being converted into a set of PWM signals. Due to the existence of these two parallel streams of PWM commands, there are several layers of checking and signal distribution to ensure the reliability and the flexibility of the transition.

4.2 Command signal distribution

The command signal distribution system manages and distributes the R/C Pilot Control Command (PCC) provided by receiver-A and the on-board Software-generated Control Commands (SCC) to individual control actuators. Based on the operational mode, the SCC can be one of or a combination of the R/C pilot commands provided by receiver-B, research pilot command, and commands from FCS, FES, and OBES.

The *ctrl*-switch, which the R/C pilot can turn on/off at any given time during the operation, plays a central role in determining the operational mode of the system. Specifically, based on measured receiver-B *ctrl*-switch signal, the EMP sends out a logic (high/low) signal indicating the status (on/off) of the *ctrl*-switch. This status indicator meets with the output of PWM switch-2, which measures the receiver-A *ctrl*-switch signal, at an AND gate. The output of the AND gate, which is called as Confirmed Ctrl Switch Signal (CCSS), becomes logic high only if both input signals are high. This provides a cross-check avoiding accidental activation of the on-board control due to either an EMP or PWM switch-2 failure.

If both receiver-A and B are functioning in the normal mode, a low CCSS initiates the logic network to feed the receiver-A pilot command directly to the control actuators for enabling the pilot manual control. The CCSS can only be overridden in the situation that receiver-A is in the fail-safe mode. Under this condition, the avionic system is able to relay the receiver-B output to actuators through EMP and GPC even if the CCSS signal is low. To achieve this capability, a duplexer is used to switch between CCSS and an EMP provided receiver-A fail-safe indicator. The switching signal for the duplexer is generated by the GPC, which provides a second confirmation that receiver-A is in the fail-safe mode.

To further improve the flexibility of the avionic system and for enabling the partially autonomous mode, another level of logic is provided before the SCC reaches an actuator. Specifically, the GPC is sending out a set of 9-channel selection signals through digital output ports. These channel selection signals are then joined with CCSS at nine AND gates to independently control a 9-channel duplexer network with both SCC and PCC as inputs. Within this configuration, if CCSS is low, all channels will be under manual control. If CCSS is high, the on-board software controls any channel with a high channel selection signal with the rest channels being controlled by the R/C pilot.

The configuration of channel selection signals is normally defined prior to flight based on mission requirements. They can also be modified by the GCS flight engineer during the operation through changing uplink communication packets. Additionally, if receiver-A goes into the fail-safe mode the GPC will activate all channel selection signals along with the fail-safe indicator. This allows the pilot command registered from receiver-B to reach actuators, maintaining the R/C pilot control.

The above-mentioned command signal distribution between PCC and SCC relies on a collaboration of both hardware and software functions. For generating SCC, the integration of commands from R/C pilot, research pilot, FCS, FES, and OBES are performed by the GPC software and are determined based on the specific flight mode. To help clarify the command signal distribution process, pseudo-codes for the EMP software and the command signal distribution portion of the GPC software are provided in Figures 5 and 6 respectively.

```
function EMP_software
    while (1)                               // start an infinite loop
        read IMU data;
        read receiver-B PPM signal;
        if (kill_switch = 'on')
            set kill_switch pin high;
        else
            set kill_switch pin low;
        read receiver-A ctrl-switch;
        if (ctrl_switch = 'fail-safe')
            set fail_safe pin high;
        else
            set fail_safe pin low;
        if (ctrl_switch = 'on')
            set ctrl_switch pin high;
        else
            set ctrl_switch pin low;
        read ADC;
        if (Receiver_A_battery = 'low')
            tie receiver batteries A&B;
        send data packet to GPC;
        receive control command packet from GPC;
        generate PWM Signal;
```

Fig. 5. Pseudo-code for the embedded micro-processor software.

```
function GPC_Command_Distribution (CC_RCP, CC_GCS, CC_FCS, CC_FES, CC_OBES, CSD,
flight_mode, ctrl_switch, fail_safe)
    // CC - control command, RCP- R/C pilot, GCS- ground control station,
    // FCS – flight control system, FES – failure emulation software,
    // OBES – on-board excitation system, CSD – channel selection data
    if (ctrl_switch = 'off')
        flight_mode= 'Manual I';
    if (fail_safe = 'on')                   // indicating receiver-A fail safe
        set fail-safe pin high;
        flight_mode= 'Fail Safe';
    else
        set fail-safe pin low;
    switch (flight_mode)
        case 'Fail Safe'                    ctrl_command = CC_RCP;
                                            CSD = 511;        // all 9 channels;
        case 'Manual I'                     ctrl_command = CC_RCP;
                                            CSD = 0;    // no channel;
        case 'Manual II'                    ctrl_command = CC_GCS;
        case 'Autonomous'        ctrl_command = CC_FCS;
        case 'Pilot_in_the_loop'            ctrl_command = CC_FCS;
        // the FCS will have pilot input as input in this case.
        case 'Failure Emulation'            ctrl_command = CC_FES;
        case 'OBES'              ctrl_command = CC_OBES;
        case 'OBES + Manual II'             ctrl_command = CC_GCS+OBES;
        // additional operational modes are available through different
        // combinations of control commands.
    set channel selection digital I/O pins according to CSD;
    send control command packet to EMP;
    return flight_mode;
```

Fig. 6. Pseudo-code for the command signal distribution portion of the GPC software.

5. On-board software

5.1 Operating systems

The use of a general-purpose computer within the avionics design facilitates the use of abundant COTS and open source software products. The on-board Operating System (OS) for GPC is the Linux kernel 2.6.9 patched with Real-Time Application Interface (RTAI) 3.2. An RTAI target was implemented so that Simulink® schemes can be compiled into real-time executable files using the Matlab Real Time Workshop®. The auto-coding capability allows for a rapid integration and testing of algorithms developed by independent researchers.

The NetBurner® MOD5213 EMP uses a µC/OS real-time operating system. The main functionality of the EMP software was outlined in Figure 5.

5.2 GPC software

The GPC software has a modular structure that is first implemented in Simulink® before being compiled into real-time executable files. Each module is either a combination of existing Simulink blocks or a custom S-function written in C language. The modular structure allows for parallel development and debugging, quick and easy configuration for different mission requirements, and intuitive visual interpolation of the software. Additionally, without any modification the same software module can be first simulated in the Simulink® environment before being tested in flight. The main modules of the GPC software and their connectivity are shown in Figure 7.

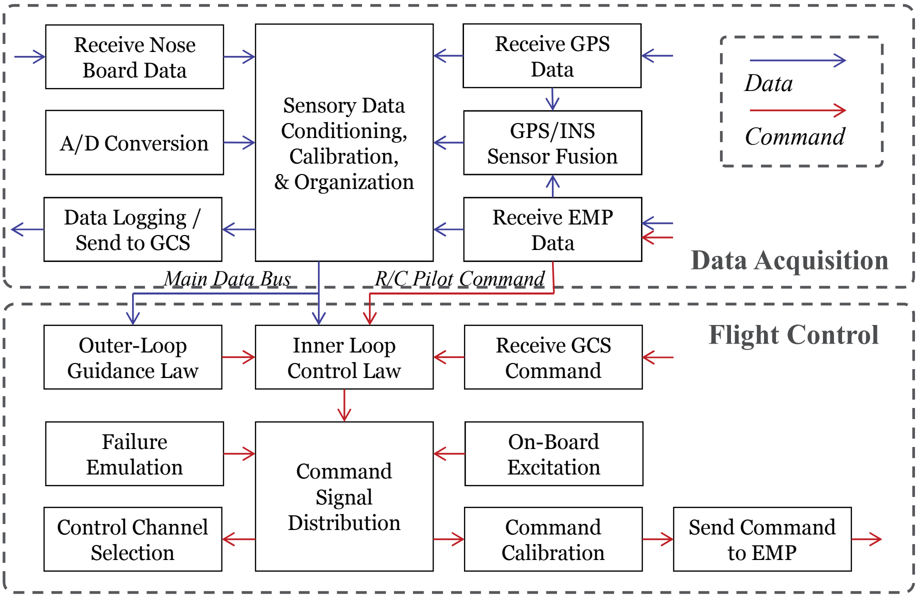


Fig. 7. GPC on-board software architecture.

5.3 GPS/INS sensor fusion

A low-cost INS is regulated with measurements from a GPS receiver to provide navigation solutions to the avionic system. Including a real-time GPS/INS sensor fusion algorithm eliminates the need of heavier and more expensive navigation-grade inertial sensors for a small and low-cost research aircraft.

A 9-state Extended Kalman Filter (EKF) based GPS/INS sensor fusion algorithm is selected for the Gen-V avionics design after a comprehensive comparison study of different sensor fusion formulations and nonlinear filtering algorithms (Rhudy et al., 2011) (Gross et al., 2011). This solution provides a good balance between attitude estimation performance and computational requirements. Within this formulation, the state vector includes the aircraft 3-axis position (*x*, *y*, *z*) and velocity (V_x, V_y, V_z) defined in a Local Cartesian frame (*L*), and aircraft attitude represented by three Euler angles (φ, θ, ψ) defined in the aircraft Body-axis (*B*):

$$\mathbf{x} = \begin{bmatrix} x^L & y^L & z^L & V_x^L & V_y^L & V_z^L & \phi^B & \theta^B & \psi^B \end{bmatrix}^T \tag{1}$$

During the state prediction stage, the inertial measurements in terms of three axis accelerations $(\tilde{a}_x^b = a_x^b + v_{ax}, \tilde{a}_y^b = a_y^b + v_{ay}, \tilde{a}_z^b = a_z^b + v_{az})$, and 3-axis angular rates $(\tilde{p}^b = p^b + v_p, \tilde{q}^b = q^b + v_q, \tilde{r}^b = r^b + v_r)$ are integrated to provide an estimate of the state vector $\mathbf{x}$. Each measurement (e.g. $\tilde{a}_x^b$) is a combination of the true measured parameter (e.g. a_x^b) and an noise term (e.g. v_{ax}). The noise is assumed to be zero mean and normally distributed, with its variance approximated by statistical analyses from static ground tests.

The three position states are predicted through straight forward integration, as represented in discrete-time:

$$\begin{bmatrix} x_{k|k-1}^L \\ y_{k|k-1}^L \\ z_{k|k-1}^L \end{bmatrix} = \begin{bmatrix} x_{k-1|k-1}^L \\ y_{k-1|k-1}^L \\ z_{k-1|k-1}^L \end{bmatrix} + \begin{bmatrix} V_{x\,k-1|k-1}^L \\ V_{y\,k-1|k-1}^L \\ V_{z\,k-1|k-1}^L \end{bmatrix} T_s \tag{2}$$

where *Ts = 0.02 s* is the length of the discrete time step. For velocity prediction, the 3D acceleration measurements are integrated and transformed from the aircraft body-axis (B) to the local Cartesian navigation frame:

$$\begin{bmatrix} V_{x\,k|k-1}^L \\ V_{y\,k|k-1}^L \\ V_{z\,k|k-1}^L \end{bmatrix} = \begin{bmatrix} V_{x\,k-1|k-1}^L \\ V_{y\,k-1|k-1}^L \\ V_{z\,k-1|k-1}^L \end{bmatrix} + DCM(\phi_{k-1|k-1}^B, \theta_{k-1|k-1}^B, \psi_{k-1|k-1}^B) \begin{bmatrix} \tilde{a}_{x\,k}^B \\ \tilde{a}_{y\,k}^B \\ \tilde{a}_{z\,k}^B \end{bmatrix} T_s + \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} T_s \tag{3}$$

where *g* is the earth's gravity, DCM stands for the Direction Cosine Matrix:

$$DCM(\phi,\theta,\psi) = \begin{bmatrix} \mathrm{c}\psi\,\mathrm{c}\theta & -\mathrm{s}\psi\,\mathrm{c}\phi + \mathrm{c}\psi\,\mathrm{s}\theta\,\mathrm{s}\phi & \mathrm{s}\psi\,\mathrm{s}\phi + \mathrm{c}\psi\,\mathrm{s}\theta\,\mathrm{c}\phi \\ \mathrm{s}\psi\,\mathrm{c}\theta & \mathrm{c}\psi\,\mathrm{c}\phi + \mathrm{s}\psi\,\mathrm{s}\theta\,\mathrm{s}\phi & -\mathrm{c}\psi\,\mathrm{s}\phi + \mathrm{s}\psi\,\mathrm{s}\theta\,\mathrm{c}\phi \\ -\mathrm{s}\theta & \mathrm{c}\theta\,\mathrm{s}\phi & \mathrm{c}\theta\,\mathrm{c}\phi \end{bmatrix} \tag{4}$$

where ‘s’ and ‘c’ are abbreviated sine and cosine functions respectively.

The aircraft Euler angles are predicted with the 3-axis angular rate measurements:

$$\begin{bmatrix} \phi^B_{k|k-1} \\ \theta^B_{k|k-1} \\ \psi^B_{k|k-1} \end{bmatrix} = \begin{bmatrix} \phi^B_{k-1|k-1} \\ \theta^B_{k-1|k-1} \\ \psi^B_{k-1|k-1} \end{bmatrix} + \begin{bmatrix} \tilde{p}^B_k + \tilde{q}^B_k \sin\phi^B_{k-1|k-1} \tan\theta^B_{k-1|k-1} + \tilde{r}^B_k \cos\phi^B_{k-1|k-1} \tan\theta^B_{k-1|k-1} \\ \left(\tilde{q}^B_k \cos\phi^B_{k-1|k-1} - \tilde{r}^B_k \sin\phi^B_{k-1|k-1}\right) \\ \left((\tilde{q}^B_k \sin\phi^B_{k-1|k-1} + \tilde{r}^B_k \cos\phi^B_{k-1|k-1})\sec\theta^B_{k-1|k-1}\right) \end{bmatrix} T_s \qquad (5)$$

The nine predicted state variables are then regulated by the GPS position and velocity measurements during the measurement update process with a simple observation equation:

$$\begin{aligned} \mathbf{z}_k = \Big[& \tilde{x}^L_k = x^L_k + v_x \quad \tilde{y}^L_k = y^L_k + v_y \quad \tilde{z}^L_k = z^L_k + v_z \\ & \tilde{V}^L_{x\,k} = V^L_{x\,k} + v_{Vx} \quad \tilde{V}^L_{y\,k} = V^L_{y\,k} + v_{Vy} \quad \tilde{V}^L_{z\,k} = V^L_{z\,k} + v_{Vz} \Big]^T \end{aligned} \qquad (6)$$

The solution of the GPS/INS sensor fusion problem follows the classis EKF approach as outlined in (Simon, 2006). The filter tuning is performed through the selection of the process noise covariance matrix Q and the measurement noise covariance matrix R. Specifically, the process noise is approximated by the sensor-level noise present on the IMU measurement.

$$Q = diag([0,0,0,\sigma^2_{v_{ax}},\sigma^2_{v_{ay}},\sigma^2_{v_{az}},\sigma^2_{v_p},\sigma^2_{v_q},\sigma^2_{v_r}])T_s^2 \qquad (7)$$

where the first three zeros indicate that no uncertainty is associated with Equation (2). Similarly, the variance of the GPS measurement noise calculated with a ground static test is used for providing the R matrix:

$$R = diag([\sigma^2_{v_x},\sigma^2_{v_y},\sigma^2_{v_z},\sigma^2_{v_{Vx}},\sigma^2_{v_{Vy}},\sigma^2_{v_{Vz}}]) \qquad (8)$$

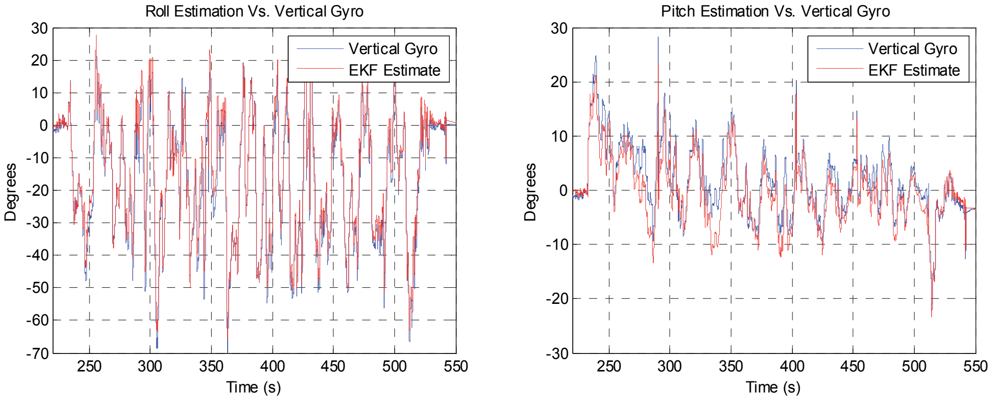


Fig. 8. Validation of the GPS/INS sensor fusion algorithm performance.

The performance and robustness of the attitude estimation algorithm was evaluated against multiple sets of flight data. Within these flights, a Goodrich VG34® mechanical vertical gyroscope was carried on-board to provide independent pitch and roll angle measurements and is used as the reference for evaluating the GPS/INS sensor fusion performance. The VG34 has a self-erection system, and reported accuracy of within 0.25° of true vertical. Figure 8 shows a comparison between the GPS/INS estimates and VG34 measurements on both roll and pitch channels for one of the May 27, 2011 flight tests. The mean absolute error and standard deviation error for roll estimation are 2.64° and 2.29° respectively in this particular flight. The mean absolute error and standard deviation error for pitch estimation are 2.22° and 1.93° respectively.

6. Avionic system testing

Extensive ground and flight testing experiments were performed to verify the functionality and performance of the Gen-V avionics design and to enable different aviation safety related flight experiments.

6.1 Avionics integration

The integration of avionics components into an airframe is constrained by many practical factors, such as aircraft balance, sensor alignment, signal interference, heat-dissipation, vibration damping, and user accessibility. Particularly, a key consideration for the avionic integration is to minimize the EMI effect. Within a sub-scale aircraft, the EMI issue is recurrent due to the close proximity of electrical components within a confined space. The effect of EMI includes reduced sensor measurement quality and disruptions of the command and control link, which could potentially lead to the loss of an aircraft. An integrated approach is used to mitigate the EMI problem. This include careful circuit design to reduce cross-interferences; providing redundancy on safety-critical components; proper shielding of main electronic components and cables; separation of EMI sources from R/C receivers; and reducing the number and length of cables. Once every avionics sub-system is installed, a comprehensive spectrum analysis and ground range tests are performed to identify residual EMI issues. Remaining problems can usually be alleviated through application of additional shielding materials, addition of ferrite chokes on selected cables, or through alternative antenna placements for RF modem and R/C receivers. Finally, a systematic ground range check procedure is performed before each flight to ensure a safe operation.

6.2 Ground testing

The ground testing procedure for the WVU Gen-V avionics system involves the following main categories:

1. *Hardware testing*, which includes the basic conductivity tests, evaluation of system power consumption and heat dissipation, EMI tests, and range tests for the R/C and data links;
2. *Software testing*, which includes the latency measurement of the real-time operating system and profiling the computational resource use by different software components;
3. *Hardware/software integration*, which includes the evaluation of sensor measurement quality, communication dropouts, PWM reading and generating accuracy, control system delay, and the functionality of the flight mode transition logics;

4. *Reliability testing*, which includes a number of duration tests under simulated dynamic operating environments;
5. *Calibration*, which includes the calibration of individual sensors, PWM reading and generating processes, individual control actuators, and pilot input devices such as R/C transmitter and the research pilot control station;
6. *Modelling*, this includes the development of mathematical models for the test-bed aircraft, actuators, propulsion systems, and sensors, as well as the identification of model parameters;
7. *Simulation*, which includes model-based simulation for initial validation of mission-specific research algorithms, and hardware-in-the-loop simulation for evaluating the integration between hardware and software sub-systems.

A flight test is considered after all related ground tests are performed.

6.3 Flight testing

Flight testing provides the final validation of the aircraft and its flight control system. However, it is also well known that experimental flight testing program, either with a full-scale or a sub-scale aircraft, is associated with substantial risks. A general strategy for flight risk mitigation focus on three steps:

1. *Prevent* the aircraft from entering an adverse flight condition;
2. Timely *identification* of the problem when an emergency situation develops;
3. *Recover* the aircraft or minimize its damage during the accident.

An adverse flight condition could be caused by improper/inadequate planning, pilot error, atmospheric condition, and aircraft sub-system (e.g. mechanical, electrical, power, control, and communication) failures. Quite often, an aviation accident has multiple inter-connected contributing factors (Boeing, 2009).

It is worth noting that the general objective of a fault-tolerant flight control research program with a sub-scale aircraft is usually to facilitate the development of the fault prevention, identification, and recovery methods for a full-scale manned aircraft. During flight experiments, the aircraft is often commanded to enter deliberately-planned adverse conditions, while minimizing other flight-associated potential risks. This high level of uncertainty, with both expected and unexpected failure contributing factors, provides valuable experiences and insights for understanding aviation accidents and the unique opportunity to practice and refine risk mitigation approaches.

6.3.1 Risk mitigation and flight testing protocol

Two effective approaches for improving the operational safety of a sub-scale flight testing program are incremental testing and the standardization of flight protocols. The incremental flight testing method utilizes a 'divide and conquer' approach to build-up individual sub-system capabilities and allows them to mature over a series of increasingly complex experiments. Each step should be a logic extension of previous steps, but should also be large enough to ensure a timely completion of the project. For example, an experiment to study the aircraft dynamics at high angle of attack flight conditions could be built upon the following key steps:

1. R/C flights for evaluating aircraft handling quality, stall characteristics, and payload capacity;
2. Data acquisition flights to evaluate avionics measurement quality and GPS/INS sensor fusion algorithm performance;
3. Closed-loop flights with a set of inner-loop control laws stabilizing the aircraft at the trim flight condition;
4. Closed-loop flights around the trim condition with OBES injection;
5. Closed-loop flights at high angle of attack conditions with OBES injection.

The standardization of flight testing protocols reduces human error both before and during the flight. It allows a systematic planning, resource allocation, testing, and inspection during the flight preparation. During the flight, having a standard procedure and flight pattern reduces pilot stress and improve the consistency among flights. Additionally, having an emergency handling procedure reduces the pilot reaction time and avoids making arbitrary decisions under adverse flight conditions. The flight testing protocol builds upon years of flight testing experience and provides a media to store and apply lessons learned from past mistakes.

A flow chart for the flight testing operation procedure developed at WVU is shown in Figure 9. A flight test session starts with a flight planning meeting in the lab discussing mission objectives, test methods, and personal responsibilities. A preliminary flight readiness review is normally performed a day before the flight date, following successful efforts in research algorithm development, ground test, and aircraft inspection.

At the airfield, another round of aircraft inspection and ground tests are performed to ensure that all aircraft sub-systems are operational after ground transportation. This is enforced with a flight preparation check-list, which covers airframe, avionics, R/C system, power system, firmware, research software, communication system, and the ground station. Additionally, the aircraft weight and balance are checked before the first flight of each aircraft. A final flight readiness review is then performed after the checklist is completed. Finally, a pre-flight pilot de-briefing discusses the flight procedures, research manoeuvres, and potential risks of this particular flight.

Once the aircraft is positioned at its starting position on the runway, the propulsion, R/C, and avionics systems are powered following an aircraft start-up procedure. A series of range tests are then performed to evaluate the R/C and data link range. A flight operation checklist is filled to verify the general functionality of the aircraft, such as control surface deflections, propulsion system condition, and R/C system fail-safe settings. A set of 'go/no-go' criteria, which includes wind-speed, wind-direction, communication range, and ground crew readiness, are then evaluated before a final approval of the flight by the flight director.

The flight operation itself follows a set of pre-defined take-off, trim, command hands-off, research, and landing procedures. In case of an emergency, such as a single engine failure, both engine failure, controller failure, actuator failure, aircraft upset condition, or changing weather condition, a set of specific emergency handling procedures are followed to abort the flight and recover the aircraft.

After landing and powering off the aircraft, flight data are downloaded and analysed in the field to provide an initial assessment of data quality and determine any potential issues. A post flight discussion session reviews the flight performance, problems encountered and pilot feedbacks. After returning to the lab, a detailed data analysis is performed, follows by a post-flight meeting to conclude the flight session.

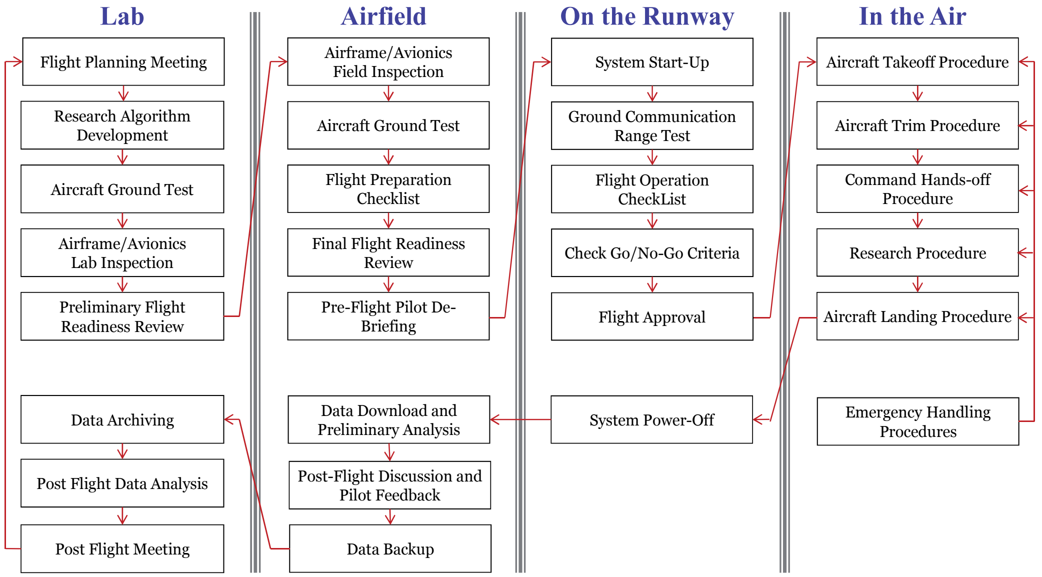

Lab
Flight Planning Meeting
Research Algorithm Development
Aircraft Ground Test
Airframe/Avionics Lab Inspection
Preliminary Flight Readiness Review
Data Archiving
Post Flight Data Analysis
Post Flight Meeting
Airfield
Airframe/Avionics Field Inspection
Aircraft Ground Test
Flight Preparation Checklist
Final Flight Readiness Review
Pre-Flight Pilot De-Briefing
Data Download and Preliminary Analysis
Post-Flight Discussion and Pilot Feedback
Data Backup
On the Runway
System Start-Up
Ground Communication Range Test
Flight Operation CheckList
Check Go/No-Go Criteria
Flight Approval
System Power-Off
In the Air
Aircraft Takeoff Procedure
Aircraft Trim Procedure
Command Hands-off Procedure
Research Procedure
Aircraft Landing Procedure
Emergency Handling Procedures

Fig. 9. WVU flight testing operation protocol.

6.3.2 Flight test examples

Two flight test examples are presented in this section to show the effectiveness of the designed avionics system. The first example is to collect data for identifying mathematical models of the 'Phastball' aircraft under high angle of attack flight conditions. The second example is to evaluate the human pilot performace with delayed control signals.

The objective of the first experiment is to study the aircraft dynamics under high angle of attack conditions. This is particularly important for T-tail aircraft, where the turbulent airflow from the stalled wing can blanket the elevators during a deep stall. For this experiment, the OBES manoeuver is designed with a multi-sine frequency-sweep approach (Klein & Morelli, 2006) to minimize disturbances to the flight condition. Specifically, it composes of six discrete frequency components ranging between 0.2 and 2.2 Hz. During the flight, a set of aircraft inner-loop controllers are activated with the *ctrl*-switch. The inner-loop controllers track zero degree roll angle and 12-degree pitch angle as reference inputs, while holding the throttle positions constant. After 2-seconds into the autonomous flight, a stream of 8-second of OBES manoeuvres are superimposed onto the elevator command generated by the inner-loop controllers. Several flight tests were performed with this configuration. Figure 10 shows a section of data collected from an October, 10, 2011 flight test.

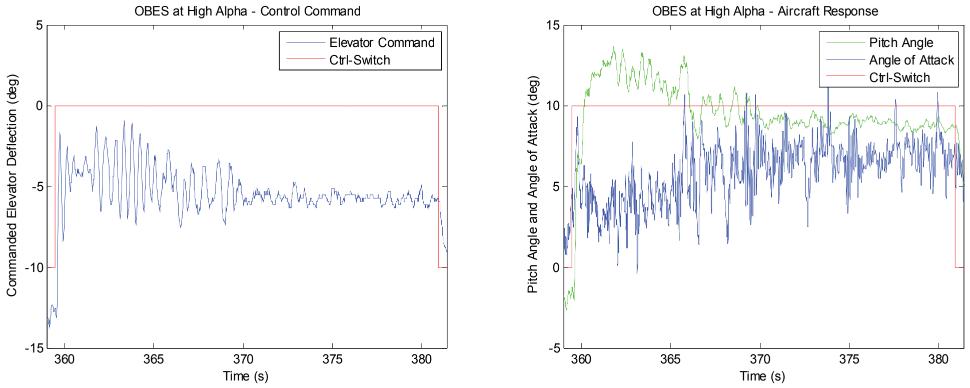


Fig. 10. Elevator control command (left) and aircraft response (right) with OBES manoeuvres at high angle of attack.

Within Figure 10, the red line indicates the turning on/off of the *ctrl*-switch. The angle of attack gradually increases to approximately 7 degrees with the deceleration of the aircraft. Additional flight experiments are planned to investigate higher angles of attack, as well as pre-stall and post-stall flight conditions.

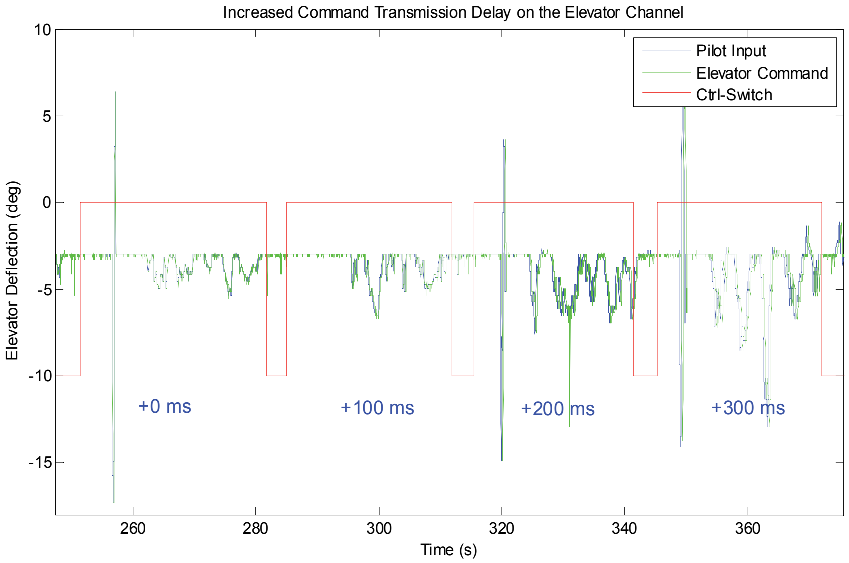


Fig. 11. Pilot elevator command vs. the actual elevator input during a command transmission delay experiment.

The objective of the second experiment is to study how the transmission delay of a fly-by-wire flight control system affects the handling quality of an aircraft. The on-board software is designed to relay the recorded pilot input to actuators with added delay. This occurs whenever the *ctrl*-switch was turned on, and a 100 ms increment is added to the total transmission delay during each *ctrl*-switch activation. The pilot flies the aircraft directly in the '*Manual Mode I*' with the *ctrl*-switch off. During the flight test, the pilot turns on the *ctrl*-switch at the beginning of a straight path. The pilot first injects an elevator doublet, wait until it settles, and then performs a turn manoeuvre. This process repeats multiple times in flight but with increasing transmission delay up to 300ms.

The first experiment of this kind was performed on October 18, 2011. The pilot reported "*…the whole flight was normal, and decay in elevator was negligible…*" during the post-flight discussion. However, flight data collected clearly indicates increased pilot activity with increased command transmission delay, which is shown in Figure 11. A later experiment with a different pilot showed similar results. Additional flight experiments are planned to investigate larger transmission delay, random data dropouts, and flight conditions that require precise control actions.

7. Conclusions

The use of sub-scale research aircraft provides unique opportunities for investigating adverse flight conditions that are too risky or costly to be tested on a full-scale aircraft. It can be considered as an intermediate validation tool between a flight simulator and a full-scale aircraft. It allows the testing of different system design, modelling, control, fault detection, and risk mitigation approaches within a realistic physical environment.

Sub-scale flight testing for fault tolerant flight control research also poses many challenges to the avionic system design: 1) it requires new capabilities for simulating different aircraft upset and failure conditions; and 2) it requires a flexible interface for integrating both human and machine decision-making capabilities; and 3) it needs to be reliable and fault-tolerant to both planned and unexpected failures. The Gen-V avionics system, designed and being developed at WVU meets these complex research requirements, along with strict power, weight, size, and cost limitations. Preliminary flight testing results demonstrate the capability of the proposed avionics design and its flexibility in supporting a variety of research objectives.

8. Acknowledgment

This study was conducted with partial support from NASA grant # NNX07AT53A and grant # NNX10AI14G

9. Appendix A: List of achronoymes

ADC – Analog to Digital Conversion
CAS – Control Augmentation System
CCSS – Confirmed Ctrl Switch Signal
COTS – Commercial-off-the-Shelf
EKF – Extended Kalman Filter
EMI – Electromagnetic Interference
EMP – Embedded Micro-Processor
FCS – Flight Control System
FES – Failure Emulation Software
GCS – Ground Control Station

GNC – Guidance, Navigation, Control
GPC – General-Purpose Computer
GPS – Global Positioning System
GPT – General-Purpose Timer
HUD – Heads-Up Display
IMU – Inertial Measurement Unit
INS – Inertial Navigation System
OBES – On-Board Excitation System
OS – Operating System
PCC – Pilot Control Command
PID – Parameter Identification
PPM – Pulse-Position Modulation
PWM – Pulse-Width Modulation
R/C – Remote Control
RF – Radio Frequency
RTAI – Real-Time Application Interface
SAS – Stability Augmentation System
SCC – Software-generated Control Commands
SCI – Serial Communication Interfaces
SPI – Serial Peripheral Interface
SPOF – Single Point of Failure
UAV – Unmanned Aerial Vehicle
WVU – West Virginia University

10. References

Ambrosia, V.G.; Brass, J.A.; Greenfield, P. & Wegener, S. (2004). *Collaborative Efforts in R&D and Applications of Imaging Wildfires*, US Forest Service. Available: http://geo.arc.nasa.gov/sge/WRAP/projects/docs/RS2004_PAPER.PDF.

Boeing Commercial Airplanes, (2009) Statistical Summary of Commercial Jet Airplane Accidents, World Wide Operations, 1959-2008, Seattle, WA, Available: http://www.boeing.com/news/techissues.

Chao, H.Y.; Jensen, A.M.; Han, Y.; Chen, Y.Q. & McKee, M. (2009). AggieAir: Towards Low-cost Cooperative Multispectral Remote Sensing Using Small Unmanned Aircraft Systems, Chapter, *Advances in Geoscience and Remote Sensing*, Gary Jedlovec, Ed.Vukovar,Croatia:IN-TECH,pp.467–490.

Christophersen, H.B.; Pickell, W.J.; Koller, A.A.; Kannan, S.K & Johnson, E.N. (2004). Small Adaptive Flight Control Systems for UAVs using FPGA/DSP Technology, *Proceedings of the AIAA "Unmanned Unlimited" Technical Conference, Workshop, and Exhibit*, Chicago, IL, September, 2004.

Cione, J.J.; Uhlhorn, E. W.; Cascella, G.; Majumdar, S. J.; Sisko, C.; Carrasco, N.; Powell, M. D.; Bale, P.; Holland, G.; Turlington, P.; Fowler, D.; Landsea, C. W. & Yuhas, C. L. (2008). The First Successful Unmanned Aerial System (UAS) Mission into a Tropical Cyclone (Ophelia 2005), *12th Conference on IOAS-AOLS*, New Orleans, LA, January 2008.

Evans, J.; Inalhan, G.; Jang J.S.; Teo, R. & Tomlin, C.J. (2001). DragonFly: a Versatile UAV Platform for the Advancement of Aircraft Navigation and Control, *Digital Avionics*

Systems, DASC. The 20th Conference, vol.1, pp.1C3/1-1C3/12, Daytona Beach, FL, October, 2001.

Griffiths, S.; Saunders, J.; Curtis, A.; Barber, B.; McLain, T. & Beard, R. (2006). Maximizing Miniature Aerial Vehicles, *IEEE Robotics & Automation Magazine*, vol.13, no.3, pp. 34-43, Sept. 2006.

Gross, J.; Gu, Y.; Rhudy, M.; Gururajan, S. & Napolitano, M.R. (2011). Flight Test Evaluation of Sensor Fusion Algorithms for Attitude Estimation, *IEEE Transactions on Aerospace and Electronic Systems*, In Press, June, 2011.

Gu, Y.; Campa, G.; Seanor, B.; Gururajan, S. & Napolitano, M.R. (2009). Autonomous Formation Flight – Design and Experiments, Chapter, *Aerial Vehicles*, ISBN 978-953-7619-41-1, I-Tech Education and Publishing, Austria, EU, Chapter 12, pp. 233-256.

How, J.P.; Bethke, B.; Frank, A.; Dale, D. & Vian, J. (2008). Real-Time Indoor Autonomous Vehicle Test Environment, *IEEE Control Systems Magzine*, vol.28, no.2, pp.51-64, April, 2008.

Jordan, T. L.; Foster, J. V.; Bailey, R. M.; & Belcastro, C. M. (2006). AirSTAR: A UAV Platform for Flight Dynamics and Control System Testing, *25th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*, San Francisco, CA, June, 2006.

Jourdan, D.B.; Piedmonte,M.D.; Gavrilets,V. & Vos,D.W. (2010). Enhancing UAV Survivability Through Damage Tolerant Control, *AIAA Guidnace Navigation and Control Conference*, Toronto, Ontario, Canada, August, 2010.

Klein, V. & Morelli, E.A. (2006). *Aircraft System Identification – Theory and Practice*, AIAA Education Series, AIAA, Reston, VA.

Liebeck, R.H. (2004). Design of the Blended Wing Body Subsonic Transport, *Journal of Aircraft*, pp. 10-25, Vol. 41, No. 1, January–February, 2004.

Miller, J.A.; Minear, P.D.; Niessner, A.F.; DeLullo, A.M.; Geiger, B.R.; Long, L.L. & Horn, J.F. (2005). Intelligent Unmanned Air Vehicle Flight Systems, *Infotec@AIAA*, Arlington, Virginia, September, 2005.

Murch, A. M. (2008). A Flight Control System Architecture for the NASA AirSTAR Flight Test Infrastructure, *AIAA Guidance, Navigation, and Control Conference*, Honolulu, HI, August, 2008.

NRC (National Research Council) (1997). Aviation Safety And Pilot Control: Under-standing and Preventing Unfavorable Pilot-Vehicle Interactions, *National Academy Press.*

Perhinschi, M.; Napolitano, M.R.; Campa, G.; Seanor, B.; Gururajan, S. & Gu, Y. (2005). Design and Flight Testing of Intelligent Flight Control Laws for the WVU YF-22 Model Aircraft, *AIAA Guidance, Navigation, and Control Conference*, San Francisco, California, August, 2005.

Phillips, K.; Gururajan, S.; Campa, G.; Seanor, B.; Gu, Y. & Napolitano, M.R. (2010). Nonlinear Aircraft Model Identification and Validation for a Fault-Tolerant Flight Control System, *AIAA Atmospheric Flight Mechanics Conference*, Toronto, Ontario, Canada, August 2010.

Planecrashinfo.com (2011). Causes of Fatal Accidents by Decade, Avaliable: http://planecrashinfo.com/cause.htm.

Rhudy, M.; Gu, Y.; Gross, J. & Napolitano, M. R. (2011). Sensitivity Analysis of EKF and UKF in GPS/INS Sensor Fusion, *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, August, 2011.

Simon, D. (2006). *Optimal State Estimation: Kalman, H-Innity, and Nonlinear Approaches*. Wiley & Sons, 1. edition.

Yeh, Y.C. (1998). Design Considerations in Boeing 777 Fly-By-Wire Computers, *Third IEEE International High-Assurance Systems Engineering Symposium*, Washington, DC, November, 1998.

22

Study of Effects of Lightning Strikes to an Aircraft

N.I. Petrov[1], A. Haddad[2], G.N. Petrova[1], H. Griffiths[2] and R.T. Waters[2]
[1]Istra, Moscow region,
[2]Cardiff University,
[1]Russia
[2]United Kingdom

1. Introduction

It is difficult to avoid thunderstorm regions by aircraft, so that on average every commercial airliner is struck by lightning once per year. Defining test and design criteria of aircraft is becoming important since aircraft safety is increasingly dependent on electronic equipment and the development of new materials (carbon composites, etc.) to replace the metallic airframes.

In-flight statistics show that most strikes occurred 3-5 km above sea level, where the temperature is ~ 0°C (Uman & Rakov, 2003; Larsson, 2002). There are two different types of lightning strikes to aircraft. The first type is that the aircraft initiates the lightning discharge when it is found in the intense electric field region of a thundercloud, and the second is the interception by the aircraft of an approaching lightning leader. The mechanism for lightning initiation by aircraft is often explained using the "bidirectional leader" theory (Clifford & Casemir, 1982; Mazur, 1989; Mazur et al., 1990; Mazur & Moreau, 1992), which describes the aircraft-initiated lightning process as a positive leader starting from the aircraft in the direction of the ambient electric field; this is followed, a few milliseconds later, by a negative leader developing in the opposite direction. This order of events is a consequence of the lower electric strength of air in the vicinity of a divergent (anode) field. The ambient thundercloud electric field measured under such conditions is typically in the range 50 - 100 *kV/m* (Marshall & Rust, 1991).

Radome "measles" (coloured spots on the inner radome surface) have been observed in many instances during service (Lalande et al., 1999; Ulmann et al., 1999). Each spot corresponds to a pin hole through the sandwich panel of the radome material. A possible explanation of the origin of these pin holes is that they were caused by breakdown due to double-layer charge accumulation on the radome. However, the physical mechanisms of the occurrence of "measles" are not fully established yet.

The purpose of this chapter was to investigate the physical processes involved in lightning strikes to aircraft and to compare simulation results with other studies involving instrumented aircraft flying in thunderstorms. 3-*D* electric field calculations were performed to determine the field distributions at the nose of aircraft and inside the dielectric radome (nosecone). The influence of the thickness and dielectric constant of the radome wall on the electric field penetration inside the radome was also investigated. The screening effect caused by ice and water layers on the radome wall is demonstrated. A new proposal for radome protection is made possible by the development of strips using materials such as non-linear *ZnO*, which behave as dielectrics under low-field conditions and acquire properties of conductors if the external electric field exceeds the critical value. Experimental tests of the strips on a real aircraft radome were carried out, and the test results reported in this paper.

2. Lightning attachment to aircraft

It was recently reported that about 90% of lightning strikes to aircraft are initiated by the aircraft (Uman & Rakov, 2003). This indicates that the aircraft extremities provide the region of high electric field needed to initiate a lightning discharge by enhancing the ambient electric field. The aircraft geometry and ambient atmospheric conditions are the most important factors in determining the local electric field intensification. Since pressure, absolute humidity and temperature decrease with increasing altitude, the variation of streamer properties with altitude can be inferred from laboratory experiments and incorporated into lightning modelling.

It is inferred from (Petrov & Waters, 1994, 1995) that the electric field needed to initiate a lightning discharge at 4km altitude is only about half of the value at sea level. Calculations show that the required striking distance increases significantly with increasing altitude, causing a corresponding increase in the risk of lightning strikes for aircraft in flight. It is shown, in the following, that ambient electric fields of between 50-80 *kV/m* can initiate positive leaders at the nose of aircraft at such altitudes.

2.1 Aircraft-initiated lightning

Consider the aircraft body as an electrically floating conducting ellipsoid placed in a uniform ambient electric field E_0 (Fig.1). An analytical expression may be obtained for the enhanced electric field in the vicinity of the nose for the case where the major axis is parallel to E_0 (Petrov & Waters, 1994):

$$E(x,a,b) = E_0 \left\{ \begin{array}{l} 1 - \dfrac{ar\tanh(aA^{1/2}/x) - aA^{1/2}/x}{ar\tanh A^{1/2} - A^{1/2}} + \\ \dfrac{A}{(x^2/a^2 + b^2/a^2 - 1)} \dfrac{aA^{1/2}/x}{(ar\tanh A^{1/2} - A^{1/2})} \end{array} \right\} \quad (1)$$

where $A = 1 - b^2/a^2$, a and b are the half-length and half-width of the ellipsoid and $(x - a)$ is the distance from the ellipsoid tip.

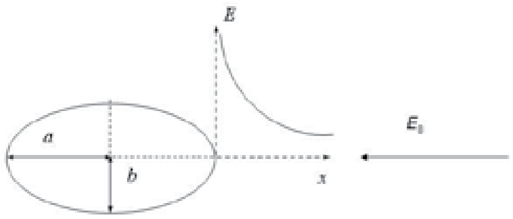


Fig. 1. Aircraft model representation and field intensification.

For given ellipsoid parameters, it is possible to determine the critical value of the ambient electric field which predicts a successful leader development from the aircraft. Using the criteria from (Petrov & Waters, 1994), we find that ambient field magnitudes of $E_{cr} \approx 50$ - 80kV/m (Fig. 2). This is insufficient at sea level to initiate leaders from the aircraft tip. However, at an altitude of 4000m, where the relative air density is around 0.58, triggering of leaders originating from the nose could certainly occur. Ambient fields of 50kV/m agree well with the fields measured inside storm-cloud, consistent with the in-flight measurements of lightning strikes to aircraft (Lalande et al., 1999).

The critical electric field dependence on the half-length of the aircraft, can be approximated with high accuracy using the empirical relationship

$$E_{cr} \cong 570 \cdot a^{-0.68}, \tag{2}$$

where a is in m, and E_{cr} in kV/m.

Similar relationship with slightly different coefficient was obtained in (Petrov & D'Alessandro, 2002) for earthed structures.

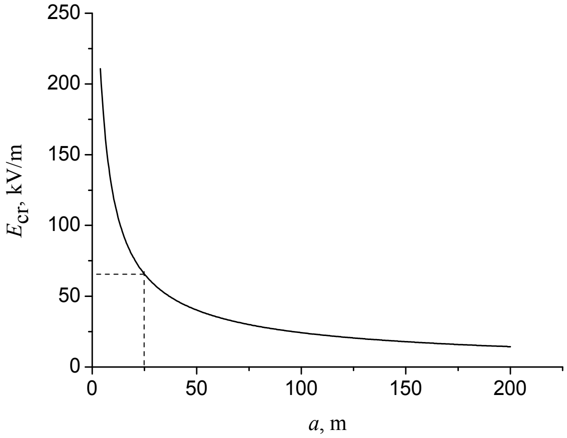


Fig. 2. Critical ambient electric field as a function of aircraft half-length (E_{cr} = 65 kV/m at a=25 m, b=3m.).

2.2 Aircraft-intercepted lightning

An aircraft can, in principle, intercept an approaching lightning leader, although no direct evidence is available. Nevertheless, in this case, the striking distance concept usually used for earthed structures may be applied to estimate the risk factor. The striking distance and the probability of lightning strikes are functions of aircraft geometry and lightning current. Electric field intensification of the field of a nearby lightning leader as a function of the distance from the aircraft tip is presented for different values of lightning peak current in Fig. 3. The aircraft is again modelled as an ellipsoid with half-width of 3m and half-length of 25m. The lightning leader channel is modeled by a charge per length, q, and leader tip charge, Q, at a distance, S, from the aircraft. The values for q and Q correspond to a prospective lightning return stroke current i_0, evaluated from (Petrov & Waters, 1995), i.e.

$$q \approx 0.43 \cdot 10^{-6} i_0^{2/3} \qquad [\mathrm{C/m, A}], \tag{3}$$

Note that there are similar relationships between the leader channel charge and the return stroke current obtained from other models. A review of data concerning this relationship was made in (Cooray et al., 2004).

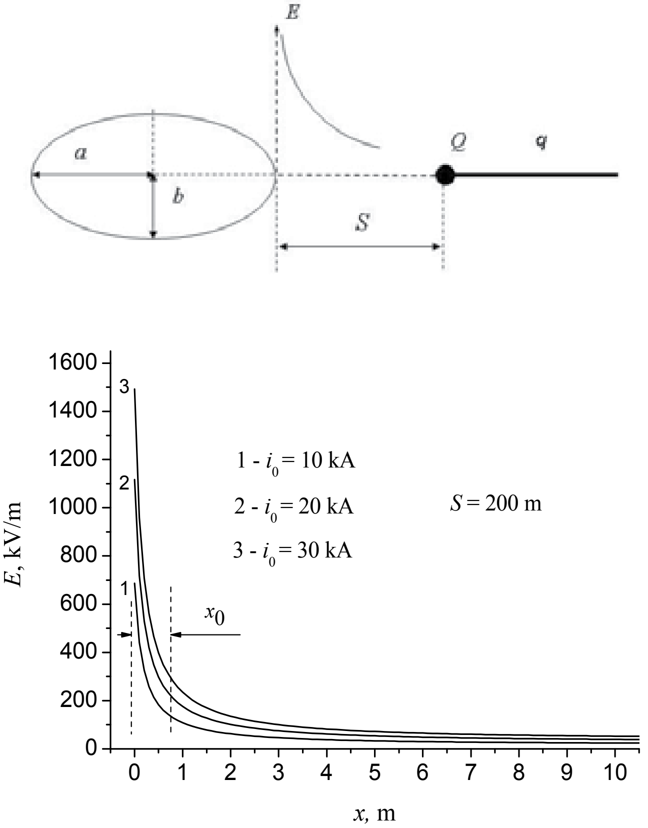


Fig. 3. Electric field intensification as a function of distance from the aircraft tip for different values of lightning peak current.

In Fig. 4, the striking distance of negative lightning to the aircraft as a function of lightning current is presented for different altitudes above sea level. Note, that for positive lightning, these distances are substantially less than those obtained for negative polarity lightning (Petrov & Waters, 1999).

A semi-quantitative estimate of the risk of lightning strike interception by an aircraft can be obtained from the concept of attractive area as used in lightning protection standards for ground structures, which can also be derived from lightning models (Petrov & Waters, 1995). For a grounded structure of the size of a commercial aircraft, the attractive area to a powerful lightning stroke of 100kA is of the order of 0.2km^2. At 4000m altitude, this would increase to 0.6km^2. Then, if the flash activity (cloud-cloud and cloud-ground) is N flashes/km^2/s, the aircraft would be expected to intercept 0.6N flashes/s. Active storms can generate 2 flashes/minute over 10 km^2, which suggests an interception rate of 1 per 500s at the heart of a storm.

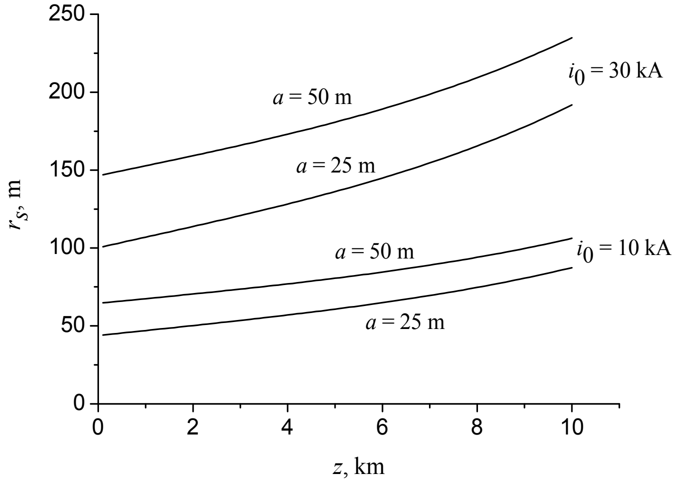


Fig. 4. Lightning interception distances by aircraft of different half-lengths as function of altitude above sea level for lightning peak current values of 10 kA and 30 kA.

3. Electric field around radomes

Radar and communications antennae are usually located at the nose or tail of the aircraft where lightning is most likely to attach. Lightning strikes damage non-metallic radomes, so the diverter strips were developed to mitigate this problem. The diverter strips screen the lightning induced electric fields on the antenna surface, i.e. they move the internal streamer initiation points forward so that strips cause the collapse of electric field inside the radome. Solid strips (permanent conductors) have been used for this purpose. However, they were found to interfere with antenna radiation patterns because they usually extend beyond the antenna. For this reason, segmented diverter strips were developed to reduce the interference effects on antenna radiation (Amason et al., 1975; Plumer & Hoots, 1978). Although they have better electromagnetic transparency for radar, segmented strips need a significant voltage gradient to light up, and their efficiency needs to be further proved.

3.1 Electric field distribution at radome without strips

For a simplified analytical calculation of 3-D electric field, consider a hemi-spherical radome with thickness d placed in uniform field E_0 (Fig. 5). This is equivalent to the floating dielectric hollow sphere (permittivity ε, internal and external radii a and b) placed in the electric field E_0.

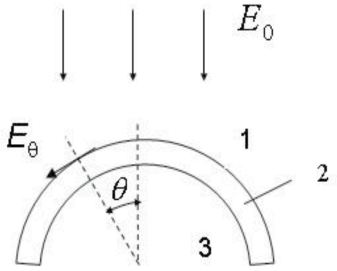


Fig. 5. Simplified model of radome exposed to electric field E_0.

The analytical solution of Laplace's equation for the potentials outside the sphere (Region 1) and inside the sphere (Region 3) can be obtained as:

$$\begin{aligned} \varphi_1 &= -E_0 \cos\theta \left(r - \frac{A}{r^2} \right), \quad [r > b]; \\ \varphi_3 &= -BE_0 r \cos\theta, \qquad [r < a] \end{aligned} \tag{4}$$

and the potential inside the dielectric layer (Region 2)

$$\varphi_2 = -CE_0 \cos\theta \left(r - \frac{D}{r^2} \right), \quad [a < r < b] \tag{5}$$

where *A, B, C, D* are constants determined from the continuity condition for φ and $\varepsilon\partial\varphi/\partial r$ on the boundaries of regions *1-2* and *2-3*. Calculation of these constants leads to the following expressions:

$$A = a^3 \left\{ 1 - \frac{3\left[1 + 2\varepsilon + (\varepsilon - 1)b^3 / a^3\right]}{(\varepsilon + 2)(2\varepsilon + 1) - 2(\varepsilon - 1)^2 b^3 / a^3} \right\},$$

$$B = \frac{9\varepsilon}{(\varepsilon + 2)(2\varepsilon + 1) - 2(\varepsilon - 1)^2 b^3 / a^3},$$

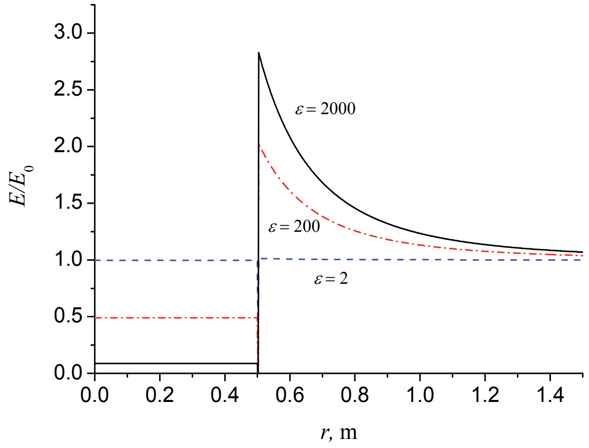
$$C = \frac{3(2\varepsilon + 1)}{(\varepsilon + 2)(2\varepsilon + 1) - 2(\varepsilon - 1)^2 b^3 / a^3}, \; D = -\frac{b^3(\varepsilon - 1)}{2\varepsilon + 1}. \tag{6}$$

For the radial and tangential components of the electric field outside the radome surface, we obtain

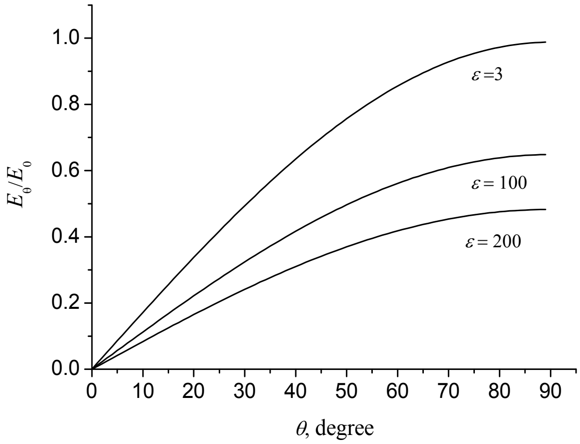
$$E_r = -\frac{\partial\varphi}{\partial r} = E_0 \cos\theta \left(1 + \frac{2A}{r^3} \right), \quad [r > b] \tag{7}$$

$$E_\theta = -\frac{1}{r}\frac{\partial\varphi}{\partial\theta} = -E_0 \sin\theta \left(1 - \frac{A}{r^2} \right), \quad [r > b]$$

In Figs. 6a and 6b, the radial and tangential electric field distributions are presented for radomes having different dielectric constants. It is seen that the screening of the electric field by the radome itself increases when the dielectric constant of the radome material increases.



a. Radial electric field distribution inside and outside the one-layer semi-spherical radome



b. Tangential electric field distribution inside the one-layer semi-spherical radome

Fig. 6. Electric field distribution in the vicinity of a radome.

i. 2-layer radome wall

By analogy, the potentials and electric fields may be obtained for the 2-layer radome wall placed in the field E_0:

$$\varphi_1 = -E_0 \cos\theta\left(r - \frac{A}{r^2}\right), \qquad [r > c]$$

$$\varphi_2 = -CE_0 \cos\theta\left(r - \frac{D}{r^2}\right), \qquad [a < r < c]$$

$$\varphi_3 = -FE_0 \cos\theta\left(r - \frac{G}{r^2}\right), \qquad [b < r < a];$$

$$\varphi_4 = -BE_0 r \cos\theta, \qquad [r < b] \tag{8}$$

$$A = c^3 - C(c^3 - D)\,,\; B = \frac{F(b^3 - G)}{b^3}$$

$$C = \frac{3\varepsilon_1 c^3}{2\varepsilon_1(c^3 - D) + \varepsilon_2(c^3 + 2D)},$$

$$D = \frac{a^3\left[1 - (\varepsilon_2 / \varepsilon_3)(a^3 - G)/(a^3 + 2G)\right]}{1 + 2(\varepsilon_2 / \varepsilon_3)(a^3 - G)/(a^3 + 2G)},$$

$$F = \frac{C(a^3 - D)}{a^3 - G}\,,\; G = \frac{b^3(1 - \varepsilon_3 / \varepsilon_4)}{1 + 2\varepsilon_3 / \varepsilon_4},$$

where $b < a < c$, with b the internal radius of the inner layer, a and c are the internal and external radii of the exterior layer, ε_1, ε_2, ε_3, ε_4 are the dielectric constants of outside medium (air), exterior and interior layers, and inside medium (air), accordingly.

Radial and tangential components of the electric field outside the radome surface are expressed by

$$E_r = -\frac{\partial\varphi}{\partial r} = E_0 \cos\theta\left(1 + \frac{2A}{r^3}\right), \qquad [r > c]$$

$$E_\theta = -\frac{1}{r}\frac{\partial\varphi}{\partial\theta} = -E_0 \sin\theta\left(1 - \frac{A}{r^2}\right), \qquad [r > c] \tag{9}$$

In Fig. 7, the electric field distributions inside and outside the two-layer semi-sphere radome are presented for different values of dielectric constants of layers. It can be seen that the field intensification at the tip of a radome increases with the dielectric constant of the radome layers.

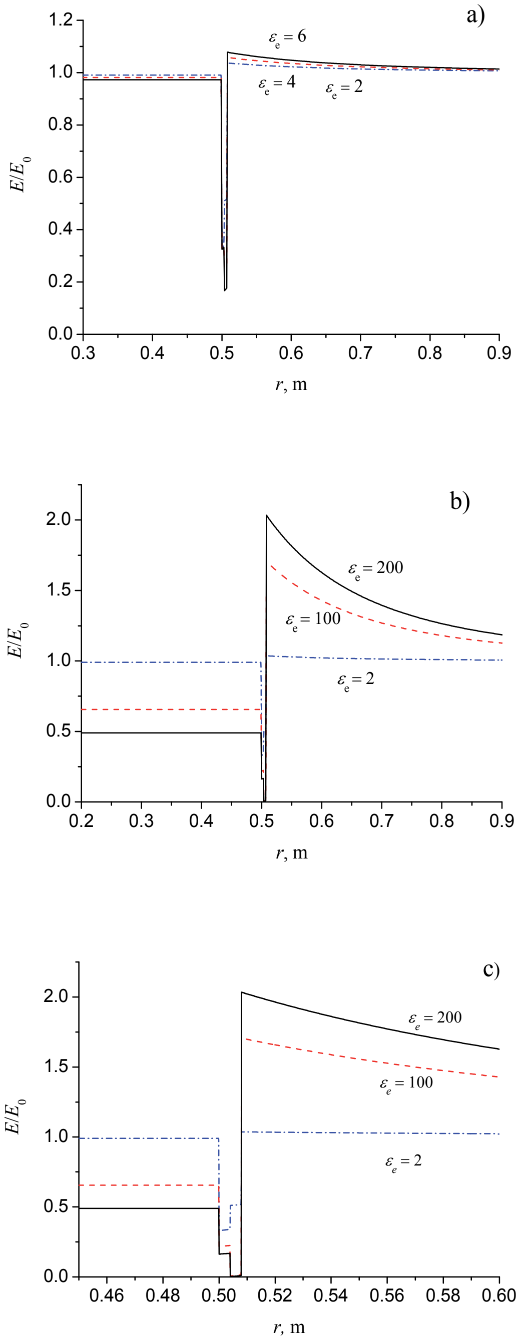


Fig. 7. Electric field distribution inside and outside the two-layer semi-sphere radome: *a*) ε_i=3 for internal layer; ε_{e1} = 2, ε_{e2} = 4, ε_{e3} = 6 for external layer; *b*) ε_i = 3; ε_{e1} = 2, ε_{e2} = 100, ε_{e3}=200; c) expanded scale of *b*).

ii. Effect of ice and water layers

In in-flight environmental conditions, the radome may be covered by ice or water layers. The tests on radomes in rain and icing conditions were conducted recently (Hardwick et al., 1999, 2003), and it was shown that the ice layers increase the light up voltages by a factor 2 to 3.

Calculations of the electric field distributions in the case of ice and water on the radome surface show that radome produces significant shielding effect (Fig. 8). In this case, the lightning leader can be initiated from the radome tip, so the strips will not operate as usual. In Fig. 8, the electric field distributions are presented for different values of permittivity of the radome wall material. The radar is represented by a conducting hemisphere having a radius of 0.2m. Note, that for a wide range of frequencies, the dielectric constants of water and ice are equal to ε_{H2O} = 87.9 and ε_{ice} = 99, respectively (Handbook of Chemistry, 2001).

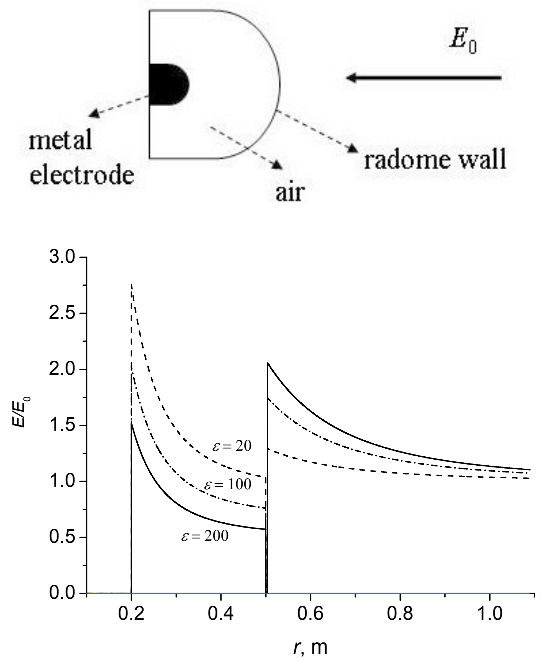
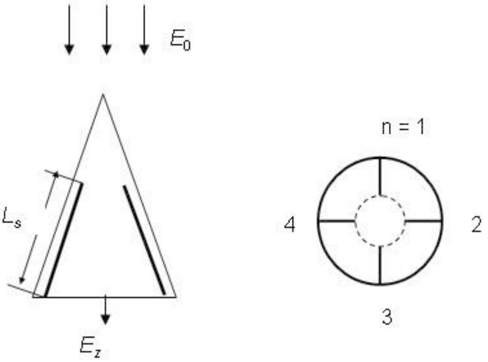


Fig. 8. Electric field distribution inside and outside the hemisphere radome with different dielectric constants of radome material.

3.2 Electric field shielding effect of strips

As was shown above, the electric field inside a dielectric radome is not disturbed significantly by the radome wall itself, so the radome does not produce screening effects. Low-level shielding permits the inception of a discharge from the internal electrode, so the solid strips are usually used to produce the shielding effect. However, high quality shielding has undesirable interference effects on antenna radiation. Therefore, the optimal length and number of strips should be determined. In the following, we consider a conical shaped radome with a base diameter of 0.7m (Fig. 9). For this radome, electric field measurement results at its base were reported (Ulmann et al., 2001; Delannoy et al., 2001), which allows comparison of simulations with experimental data. Here, solid strips were considered as inclined isolated rods in a uniform external electric field, since the analytical expressions for the electric field distribution exist in this case. In Fig. 10, the electric field at the radome base is shown as a function of strip length for different numbers of strips. It can be observed that the electric field at the radome base decreases by 50 % if 6 solid diverter strips of 0.4m length were installed on the radome surface. This is in good agreement with the measurements (Ulmann et al., 2001).



a) side view b) view from top

Fig. 9. A conical radome with conducting solid strips:

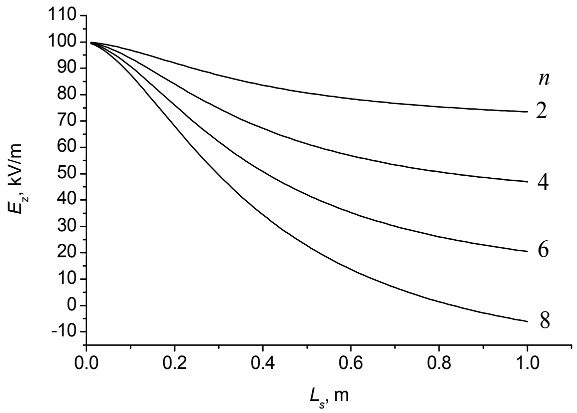


Fig. 10. Calculated electric field at the radome base as a function of strip length for different numbers of strips

4. Laboratory lightning impulse tests

Preliminary tests on a radome, used on a commercial aircraft, with a thickness of ~5 *mm*, a diameter of ~ 1.6 *m* and having six solid strips of 1*m* length were performed in the high voltage laboratory at Cardiff University. Lightning impulses of 1.2/50 shape, positive and negative polarity, were applied to the output electrode (sphere of 10*cm* diameter or rod with spherical end of 1.2*cm* diameter), which was placed at different distances (10-30 *cm)* from the surface of the radome. Breakdown channels were recorded using a video-camera having a picture rate of 50 fps.

4.1 Segmented diverter strips

Tests were also conducted on two commercially available segmented diverter strips of one meter length each. The diverter strips were attached to the aircraft radome surface for testing. It was found that the diverter with smaller buttons (segment diameter 1.524 mm) has higher breakdown voltage.

The segmented diverters had breakdown voltages of 50-60kV while the time to breakdown $t_{\rm br}$ varied between 3 and 7μs. This corresponds to leader velocities v_l in the range 15 to 30 $cm/\mu s$, which is ten times higher than usually registered leader velocities in long air gaps. Dependence of the applied voltage on the polarity is weak, if the rod-type high voltage electrode is placed close (~10-15cm) to the strip end.

Although the segmented strips have good diversion properties, tests have shown problems with multi-impulse lightning strikes. After a number of strikes, damage was observed on the strip buttons (Fig. 11). However, the resistance of the strips after tests was still more than 600 MΩ. This indicates that the discharge current mainly flows not through the strip buttons but in air over the strip.

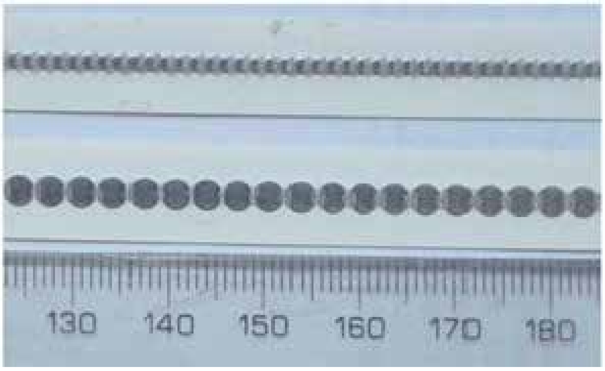


Fig. 11. Segmented diverter strips after the test.

4.2 Isolated multiple-electrode diverter

Isolated rings or disks with diameters of 17*mm* and 20*mm* with a separation of 5-15*cm* were mounted on the radome surface with the help of dielectric tape. These types of strips have several advantages: (a) they have negligible interference effects on antenna radiation due to the small total surface of metal elements and (b) they do not initiate a leader discharge before an approaching lightning leader streamer zone attaches to the radome surface. Both positive and negative polarity impulses of amplitude 200-250 kV were applied to gaps of 10-20cm between the electrode and the radome tip (Fig. 12 and Fig. 13).

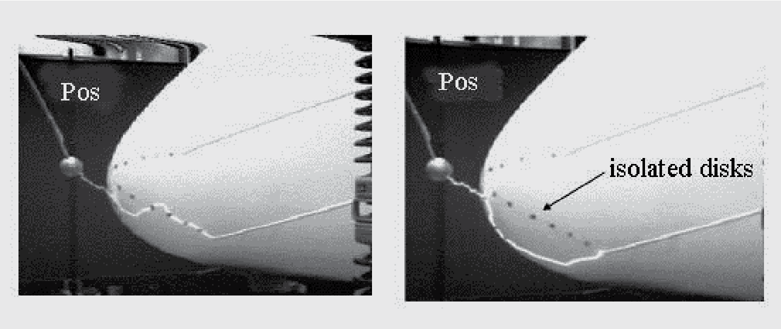


Fig. 12. Influence of isolated disks on the trajectory of breakdown under positive impulse

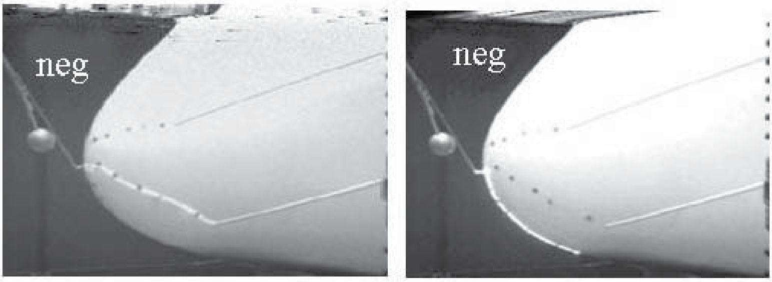


Fig. 13. Influence of isolated disks on the trajectory of breakdown under negative impulse

Breakdown occurs between the electrode and the closest point on the radome surface, propagating further along the radome surface until the end of the solid strip, even if the air gap distance between the electrode and the end of the solid strip is shorter. This indicates that the breakdown voltage along the surface of the radome is lower than the breakdown voltage in an air gap. The time to the surface breakdown was t_{br}~ 30-50 μs depending on the distance between the electrode and the end of the solid strip on the radome wall. This corresponds to a leader velocity v_l ~ 2 $cm/\mu s$, which is usually recorded in long air gaps. In the case of the isolated multiple-electrode diverters, the time to breakdown decreases by a factor of 3-4, i.e. the leader velocity becomes v_l ~ 6-8 $cm/\mu s$. The decrease of the breakdown time indicates that simultaneous development of the discharge in the gaps between different isolated electrodes.

The light up electric field was about 3.3 *kV/cm*, which is close to typical light up voltages with *D* waveform for the segmented strips (Hardwick et al., 1999).

Tests have shown that the isolated electrode strips divert the discharge channel of both polarities. Leaders develop along the surface without any damage to it. For the same applied voltage, the breakdown gap with the isolated multiple-electrode diverter strip can be twice as long as the gap without strip. The diversion ability of isolated multiple-electrode diverter strip is higher for negative polarity discharge than for positive discharge (Figs. 12, 13). This is due to different mechanisms of breakdown for negative and positive polarity discharges (Petrov & Waters, 1999).

4.3 Flashover across the radome wall

The discharge develops along the surface of the radome even if the air gap distance between the output electrode and the termination of the strips is shorter. This indicates that the breakdown voltage along the dielectric surface is lower than the breakdown voltage in air.

A leader discharge can be initiated from the internal radar antenna. In the model, the antenna was represented by a grounded metal hemisphere at the radome base. The leader channel from the antenna was modeled as a metal rod of different lengths connected with the antenna.

The laboratory experiments have shown that both positive and negative polarity discharges can cause a puncture through the radome wall when the internal electrode (antenna) extends beyond the strips and, hence, when it is no longer screened.

In Fig. 14, the flashover path can be seen initially propagating along the surface and then passing through the radome wall to the internal grounded electrode. The distance from the surface puncture point to the grounded outer electrode was only 7.5*cm*. This indicates that a voltage drop of less than 20 *kV* is sufficient to cause a flashover across the radome wall.

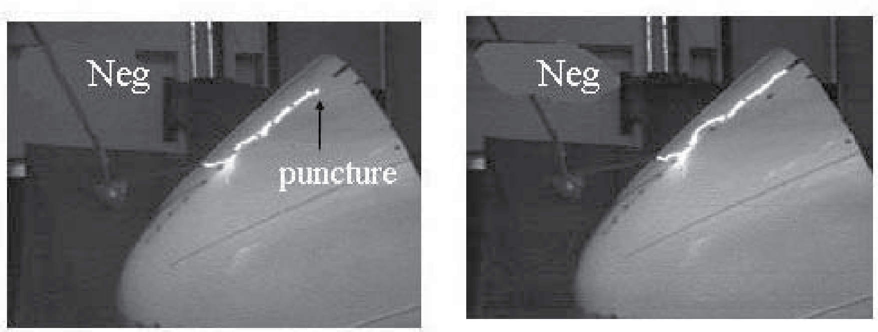


Fig. 14. Puncture through the radome wall with an earth electrode inside the radome.

4.4 Diverter strips with ZnO material

Segmented strips consisting of *ZnO* material between *Al* segments of 3x3 *mm* size were designed and tested (Fig. 15). Experiments have shown that the influence of *ZnO* material on the discharge properties of strips depends on the distances between the segments. Although no significant influence was observed for gaps $d > 10$ *mm*, at $d \sim 1\text{-}3$ *mm*, the influence of *ZnO* material becomes significant. The competitive breakdown tests showed that all discharges pass through the strip consisting of *ZnO* material, which indicates that electric fields created between the segments are sufficient for the ZnO material to become conductive. The breakdown time for these strips is comparable to that of commercial segmented strips. The velocity of leader propagation increases 4-5 times in comparison to the velocity of the surface leader discharge without the strips.

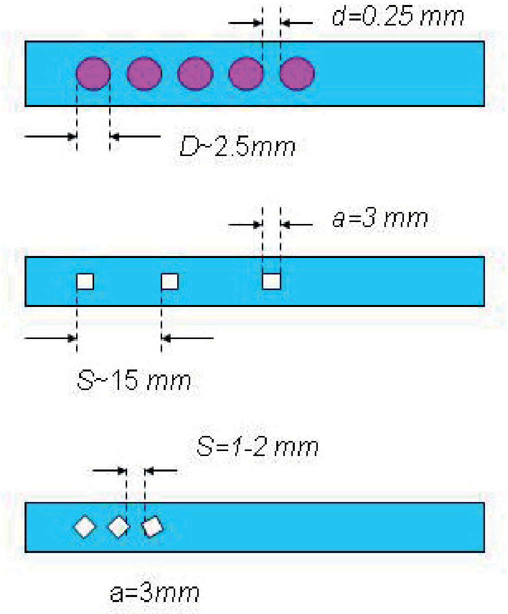


Fig. 15. Designed diverter strips with ZnO material.

5. 3D numerical computation of electric field around radomes

The electric field and potential distributions inside and outside the aircraft radome placed in an external electric field were analyzed using COULOMB software which is based on the boundary element method. This results of this analysis were used to determine the necessary number and length of strips to be utilized to provide the radome with the optimised lightning protection using strips.

A simulation model of the aircraft radome having a hemispherical shape placed in a uniform ambient electric field was used in a plane-plane gap (Fig. 16). The gap length is 5.2 m and the applied voltage is 2 MV. The dielectric hemispherical radome is placed on top of a metal cylinder of 1.5 m length to simulate the end of the fuselage. The hemispherical radome has a radius of 0.5m and a thickness of 4mm and a dielectric constant ε_r = 10. Solid strips of 1cm width and 3mm thickness were considered. The segmented strips have a 5mm diameter and a 3mm thickness of and a gap distance of 1mm. The distance between the radome tip and the upper electrode is 2m. The distance between the bottom of the cylinder and the bottom electrode is 1.2m.

Fig. 17 shows the solid and segmented strips attached to the radome surface. Fig. 18 shows examples of computed voltage contours.

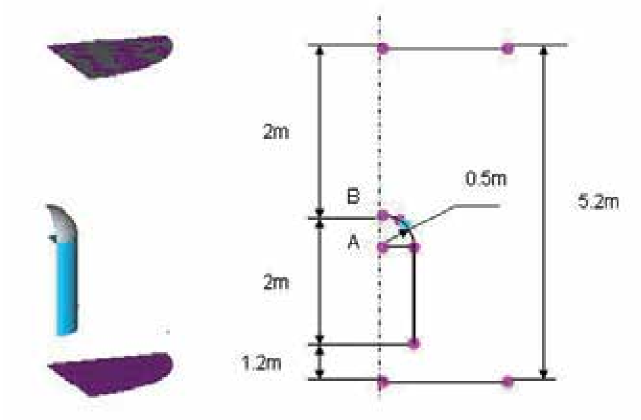


Fig. 16. Model representation: semi-spherical radome.

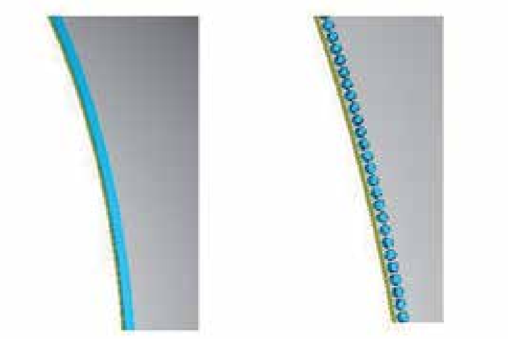


Fig. 17. Modeled solid (thickness: 3mm, width: 10 mm) and segmented (thickness: 3mm, radius: 2.5 mm, gap: 1 mm) strips.

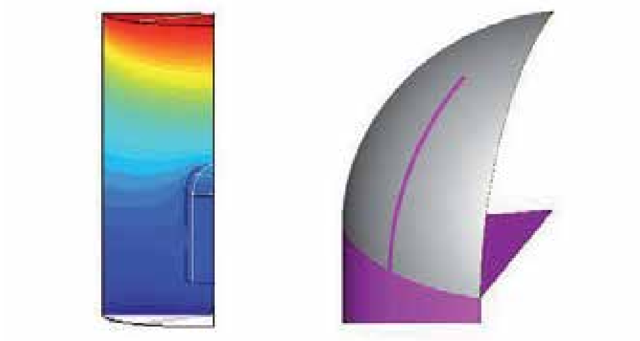


Fig. 18. Voltage contour and section of a radome with a solid strip.

Tables 1 and 2 summarise the computed magnitudes of electric field at the radome base and tip. It can be observed that the shielding effect increases with the length of solid strips and the number of strips. The electric field at the base of the radome is only 50% of the external field if 6 solid strips of 0.5m length are used. Segmented strips do not produce any visible shielding effects.

Detailed analaysis has shown that an increase of the number of solid strips results in a decrease in the electric field at the base of the radome. On the other hand, the electric field was forced out to the frontal area of the radome, so that too strong shielding of the internal electrode can cause undesired field intensification at a radome front. This is a disadvantage with solid strips, in addition to their interference effect on the radiation field from the antenna. In the case of segmented strips, there is no shielding effect. This indicates that there will be no interference effect with the radiation field until the breakdown along the strip takes place, under which condition the strip behaves like a conductor.

	Solid strips					
*N*umber of strips	4		6		8	
Length, m	0.25	0.5	0.25	0.5	0.25	0.5
E(A), *kV/m*	482	335	456	270	435	226
E(B), *kV/m*	493	534	515	588	524	570

Table 1. Electric field magnitudes at radome base (point A in Fig.16) and radome tip (point B in Fig. 16) for different numbers of solid strips.

	Segmented strips					
Nunber of strips	4		6		8	
Length, m	0.25	0.5	0.25	0.5	0.25	0.5
E(A), *kV/m*	517	524	526	508	551	515
E(B), *kV/m*	477	472	481	497	477	475

Table 2. Electric field magnitudes at radome base (point A in Fig. 16) and radome tip (point B in Fig. 16) for different numbers of segmented strips.

6. Discussion

The radome simulations described in this chapter show clearly that the critical electric field magnitude, which is necessary to originate leaders from the aircraft tip, decreases with the aircraft length. The magnitude of the critical electric field decreases from 100 kV/m to 40 kV/m as the aircraft length increases from 20m to 100m. These values are in good agreement with the in-flight measurements of the ambient fields inside storm-cloud (Lalande et al., 1999).

Furthermore, the simulations demonstrated that the electric field inside the radome is not reduced significantly by the radome wall itself, which indicates that the radome does not produce screening effects. This shows that leader can start from the internal electrode (radar antenna) causing flashover across the radome. Therefore, strips to produce the screening effect must be used to avoid the initiation of streamers from the antenna. The lightning strike to the radome does not damage the radome surface if discharges do not occur from the metal parts inside the radome. This points out that the main purpose of the protection system should be the screening (shielding) of the electric field inside the radome. Poor shielding permits the inception of a discharge from the internal electrode, so the solid strips are usually used to produce the shielding effect. However, effective shielding has undesirable interference effects on antenna radiation. Therefore, the optimal length and number of the strips should be determined.

Significant shielding effect is created by water and ice layers on the radome surface. Under these conditions, the lightning leader can be initiated from the radome tip. Note that the dielectric constant values of ice depend on the frequency of the external field or the rate of voltage rise, and these values affect the electric field magnitude. For example, the values of $\varepsilon_{\text{ice}} = 5$ for 1000 $kV/\mu s$ and $\varepsilon_{\text{ice}} = 70$ for 10 $kV/\mu s$ were used in (Hardwick et al., 2003). This work has shown that the ice layer does not screen the high frequency radiation associated with the radar.

In high ambient humidity conditions (>60%), the radome becomes moderately conductive because of humidity absorption at its surface (Ulmann et al., 2001; Delannoy et al., 2001). Although this decreases the internal field due to shielding effect, it also reduces the efficiency of the strips.

Numerical simulations have shown that the shielding effect is produced only by solid strips, there is no practical shielding by segmented strips in the absence of a discharge. It was demonstrated that the field intensification area is forced out from the metal electrode (antenna) surface to the front of the radome, thereby preventing discharge initiation from the antenna. However, too strong shielding of antenna surface by increasing the number and the length of strips can cause the field intensification at the frontal area of the radome which can be sufficient to initiate the discharge. Hence, the shielding of the antenna surface as much as possible is not the best solution to the problem. It is necessary to optimize the electric field distribution with respect to the streamer and leader discharge initiation conditions.

Both the fast and slow waveforms (MIL STD 1757 Waveforms *A* and *D* respectively) are used for testing radomes (Ulmann et al., 1999). Waveform *A* has 1000 $kV/\mu s$ rate of rise, and Waveform *D* has 50-250 μs rise time. It was concluded (Ulmann et al., 1999) that Waveform *D* represents the in-flight environment more accurately than Waveform *A*. For aircraft intercepting approaching leaders, rates of rise of the electric field, dU/dt of 10^8 to 10^{10} $V/m/s$ were estimated (Lalande et al., 1999) at the aircraft. If 1 $MV/\mu s$ (waveform *A*) is applied over a 1m gap, this will give $dU/dt \approx 10^{12}$ $V/m/s$. Hence, the slower voltage Waveform *D* tests might be more appropriate. In our tests, we have $dU/dt \approx U/\tau_f/L \approx 2.8{\cdot}10^5 V/2{\cdot}10^{-6} s/0.7m \approx 2{\cdot}10^{11}$ $V/m/s$. However, the voltage rise time is important when the voltage is applied directly to the strip. If the high-voltage electrode is placed far from the strip, the breakdown process of the strip is determined by the field generated by the ionization front of the discharge, i.e. by the space charge of the streamers. The magnitude of this field is affected by the velocity of the streamer/leader ionization front, but not by the applied voltage waveform.

Besides direct strikes to aircraft radome, the aircraft could be subjected to indirect strikes. Lightning strike entrance and exit points are usually found at sharp structures of the aircraft, around which the electric field enhancement takes place, but also can occur at any part of the aircraft, including the fuselage, stabilisers, antennas, etc. Observations of such strikes were conducted in a laboratory experiments with aircraft models (Chernov et al., 1992; Petrov et al., 1996). It is seen from Fig. 19, that the nose radome can also be exit point of lightning strike depending on the aircraft position with respect to the approaching lightning threat.

It is worth highlighting here that the lightning diverter strips concept could be adapted for use in protection of ground antennas for ultra-high-frequency communications, which are difficult to protect from direct lightning strikes because interference to the radiation field arises when standard air-terminal shielding is installed (Bruel et al., 2004).

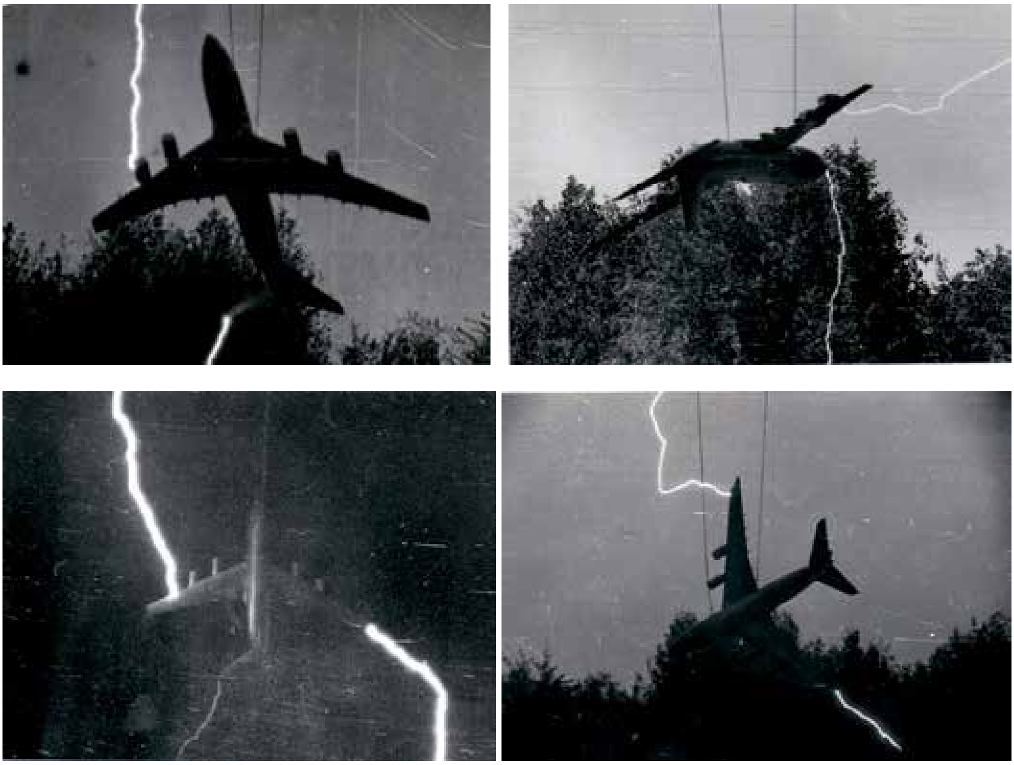


Fig. 19. Laboratory testing of lightning strikes to an aircraft model.

7. Conclusion

Theoretical analysis and numerical simulations together with experimental laboratory tests of lightning discharge interaction with aircraft radome demonstrated the applicability of existing lightning attachment models to create optimal protection systems against lightning strikes.

The following points can be concluded from the analysis:

i. Electric field intensification by aircraft flying at high altitudes exceeds the threshold to initiate the lightning leader (50-100 *kV/m*), this explains why about 90% of lightning strikes to aircraft are initiated by the aircraft.
ii. The shielding effect of dielectric radome material itself is less than 10%, so the lightning leader can be initiated from the radar antenna.
iii. The penetration of the electric field, created by the lightning channel or storm-cloud, into the radome is significantly decreased by ice and/or water layers on the radome surface; however, this may cause also the occurrence of punctures.
iv. Strong diversion effect for the strips comprising isolated metal disks or rings is observed for positive as well as for negative polarity discharges; this type of diverter strip can be used together with the solid strips in order to decrease the interference effect on antenna radiation.
v. Numerical simulations have shown strong radar shielding effects produced by solid strips and no practical shielding by segmented strips in the absence of a discharge.

8. Acknowledgment

N.I.P. and G.N.P. thank colleagues of the High Voltage Group of the Cardiff School of Engineering for hospitality while they worked as guests in their laboratory.

9. References

Amason, M.; et al. (1975). Aircraft application of segmented-strip lightning protection systems, *Proceedings of Conf. on Lightning and Static Electricity*, pp.1-14, London, UK, 1975

Bruel, C.; Barilleau, D. & Rousseau, A. (2004). Application of aircraft lightning protection to radar stations, *Proceedings of 27th Int. Conf. on Lightning Protection*, pp. 975-977, Avignon, France, September 13-16, 2004

Chernov, E.; Lupeiko, A. & Petrov, N. (1992). Repulsion effect in orientation of Lightning discharge. *J.de Phys. III*, Vol. 2, (July 1992), pp. 1359-1365

Clifford, D. & Casemir, H. (1982). Triggered lightning. *IEEE Trans. Electromagnetic Compatibility,* Vol. 21, (January 1982), pp. 112-122, ISSN 0018-9375

Cooray, V.; Rakov, V. & Theethayi, N. (2004). The relationship between the leader charge and the return stroke current – Berger's data revisited, *Proceedings of 27th Int. Conf. on Lightning Protection*, pp. 145-150, Avignon, France, September 13-16, 2004

Delannoy, A.; Bondiou-Clergerie, A.; Lalande, P.; et.al. (2001). New investigations of the mechanisms of lightning strike to radomes Part II: Modeling of the protection efficiency, *Proceedings of Int. Conf. on Lightning and Static Electricity*, paper No 2001-01-2884, Seattle, USA, September 11-13, 2001

Handbook of Chemistry and Physics. (2001). CRC Press, ISBN 0849304822

Hardwick, J.; Plumer, A. & Ulmann, A. (1999). Review of the joint radome programme, *Proceedings of ICOLSE'99*, pp.59-65, ISBN 0768003938, Toulouse, France, June 22-24, 1999

Hardwick, C.; Hawkins, K. & Sanders, M. (2003). Effect of water and icing on segmented diverter strip performance, *Proceedings of ICOLSE'03*, pp. 80.1-80.8, ISBN 1857681525, 9781857681529, Blackpool, UK, September 16-18, 2003

Larsson, A. (2002). The interaction between a lightning flash and an aircraft in flight. *C.R. Physique*, Vol 3, (December 2002), pp. 1423-1444

Lalande, P.; Bondiou-Clergerie, A. & Laroche, P. (1999). Analysis of available in-flight measurements of lightning strikes to aircraft, *Proceedings of ICOLSE'99*, pp.401-408, Toulouse, France, June 22-24, 1999

Mazur, V. (1989). Triggered lightning strikes to aircraft and natural intracloud discharges. *J. Geophys. Res.*, Vol. 94, (March 1989), pp. 3311-3325, ISSN 0148-0227

Mazur, V. (1989). A physical model of lightning initiation on aircraft in thunderstorms. *J. Geophys. Res.,* Vol. 94, (March 1989), pp. 3326-3340, ISSN 0148-0227

Mazur, V.; Fisher, B. & Brown, P. (1990). Multistroke cloud-to-ground strike to the NASA F-106B airplane, *J. Geophysical Research*, Vol. 95, no. D5, (May 1990), pp. 5471-5484, ISSN 0148-0227

Mazur, V. & Moreau, J. (1992). Aircraft-triggered lightning: processes following strike initiation that affect aircraft, *J. Aircr.*, Vol. 29, (August 1992), pp. 575-580, ISSN 0021-8669

Marshall, T. & Rust, W. (1991). Electric field soundings through thunderstorms. *J. Geophys. Res*., Vol. 96(22), (December 1991), pp. 297-306, ISSN 0148-0227

Petrov, N. & Waters, R. (1994). Conductor height and altitude: effect on striking distance, *Proc. Int. Conf. Lightning and Mountains*, pp. 52-57, SEE, Chamonix-Mont-Blanc, June 6-9, 1994.

Petrov, N. & Waters, R. (1995). Determination of the striking distance of lightning to earthed structures, *Proc. R. Soc. Lond. A*, Vol. 450, No. 1940, (September 1995), pp. 589-601, ISSN 1471-2946

Petrov, N.; Avansky, V.; Efimova, N. & Petrova, G. (1996). Experimental and theoretical investigations of the orientation of leader discharge to isolated and earthed objects, Proceedings of 23th Int. Conf. on Lightning Protection, pp.254-259, Vol.1, Firenze, Italy, September 23-27, 1996

Petrov, N. & D'Alessandro, F. (2002). Theoretical analysis of the processes involved in lightning attachment to earthed structures, *J. Phys. D: Appl. Phys.*, Vol. 35, No. 14, (July 2002), pp. 1788-1795, ISSN 0022-3727

Petrov, N. & Waters, R. (1999). Striking distance of Lightning to earthed structures: effect of stroke polarity, *Proc. 11th Int. Symp. on High Voltage Engineering*, pp.220-223, Vol. 2, London, UK, August 23-27, 1999

Plumer, J. & Hoots, L. (1978). Lightning protection with segmented diverters, *Proceedings of IEEE Int. Symp. Electromagnetic Compatibility*, pp.196-203, 1978

Ulmann, A.; Hardwick, J. & Plumer, A. (1999). Laboratory Reproduction of In-Flight Failures of Radomes, *Proceedings of ICOLSE'99*, pp.493-496, ISBN 0768003938, Toulouse,France, June 22-24, 1999

Ulmann, A.; Brechet, P.; Bondiou-Clergerie, A.; et.al. (2001). New investigations of the mechanisms of lightning strike to radomes Part I: Experimental study in high voltage laboratory, *Proceedings of Int. Conf. on Lightning and Static Electricity*, paper No 2001-01-2883, Seattle, USA, September 11-13, 2001

Uman, M. & Rakov, V. (2003). The interaction of lightning with airborne vehicles. *Progress in Aerospace Sciences*, Vol.39, No 1, (January 2003), pp. 61-81, ISSN 0376-0421

Edited by Ramesh K. Agarwal

The book describes the state of the art and latest advancements in technologies for various areas of aircraft systems. In particular it covers wide variety of topics in aircraft structures and advanced materials, control systems, electrical systems, inspection and maintenance, avionics and radar and some miscellaneous topics such as green aviation. The authors are leading experts in their fields. Both the researchers and the students should find the material useful in their work.

ISBN 978-953-51-4345-1

Photo by XavierMarchant / iStock

IntechOpen