Publication of the MIRCE Akademy – 2013

2013 Annals of MIRCE Science

"The goal of a scientist is to uncover new ideas, concepts and tools, practical or theoretical, that extend our understanding of the world around us and enable us to do new things. One must believe in what one is doing and stay the course. Now of course, in science one can ultimately prove the correctness of one's work by appeal to experiment and established theory. But even with this buttressing of one's ideas, acceptance can be a long and difficult road."

Richard F.W. Bader (1931 – 2012), Grand Fellow of the MIRCE Akademy

December 2013, Woodbury Park, Exeter, UK

Publisher:
MIRCE Science Limited
Woodbury Park
Exeter
EX5 1JJ
United Kingdom

Phone: +44 (0) 1395 233 856
Email: quest@mirceakademy.com
Website; www.mirce.com

Editor:
Dr J. Knezevic, President MIRCE Akademy

Editorial Board:
Fellows of the MIRCE Akademy

Content

Mirce-mechanics **5**

On The Non-Existence of Parallel Universes in Science **7**
Richard F. W. Bader, Grand Fellow of the MIRCE Akademy,
Department of Chemistry, McMaster University, Hamilton, ON, Canada

The Two Faces of Chemistry: Can they be reconciled? **15**
Mark E. Eberhart and Travis E. Jones
Molecular Theory Group, Colorado School of Mines,
Golden, Colorado 80401, USA.

Atoms and Molecules in Mirce-mechanics Approach to Functionability, **24**
Dr J. Knezevic, MIRCE Akademy, Exeter, UK

Physics-of-Failure based Reliability Engineering **37**
Elviz George and Michael Pecht
Center for Advanced Life Cycle Engineering (CALCE)
University of Maryland College Park, MD 20742 USA

Mirce-mechanics Analysis of the Impact of Cosmic Phenomena on In-service Reliability **54**
Ian. Zaczyk,
MIRCE Akademy, Exeter, EX5 1JJ, UK

Managing Machine Functionability Using Methods of Complexity Science **68**
George Rzevski
Professor Emeritus, Centre for Complexity Science Applications,
The Open University, UK.
Executive Chairman, Multi-Agent Technology Group, London, UK

Human Effectiveness of Troubleshooting Process in Commercial Aviation **77**
John G. Hessburg, Jezdimir Knezevic
MIRCE Akademy, Exeter, UK

No Fault Found and Air Safety **81**
Christopher J Hockley OBE, CEng MRAeS
Centre for Through-Life Engineering Services,
Cranfield University, Bedford, UK.

Maintenance Axiom of Mirce-mechanics **89**
Dr Jezdimir Knezevic
MIRCE Akademy, Woodbury Park, Exeter, EX5 1JJ, UK

Planning In-service Support **99**
Dr John Crocker
Science Fellow of the MIRCE Akademy, Exeter, UK

The Role of Simplified Technical English in Aviation Maintenance **108**
Orlando Chiarello, Secondo Mona S.p.A., Italy

Call for papers for the Annals of Mirce Akademy 2014 **112**

Mirce-mechanics

According to Einstein **"***Everything that the human race has done and thought is concerned with the satisfaction of felt needs*".

During the history of civilisation, needs for transportation, communication, navigation and many others have been satisfied by human created machines like, trains, aircraft, cars, computers, telephones, radars, radios satellites and so forth. The mechanics of the functioning of machines are well-understood processes, which are predictable by the laws of natural sciences, such as: Newton's laws of motion, Coulomb's law of solid friction, Hook's law of stress and strain, Maxwell's law of electrodynamics, Boltzmann's law of thermodynamics, to name a few.

Machines are constructed by assembling a well-defined number of parts in a precise and preestablished way. As they are functioning in predetermined linear chains of cause and effect, their performance measured through speed, acceleration, power, range, energy usage, capacity and similar is also predictable. The reason for the predictability of the system design-in functionality performance is the fact that they are based on the physical and chemical processes that are characterised by certainty, continuity, reversibility, separability and independence of time, location and humans.

Regarding the long-term satisfaction of human needs, the ability of a machine to function beyond the delivery day is an essential property of its in-service performance. Due to complex interactions between consisting parts and impacts from environment and humans, disturbances of mechanical, electrical, chemical, thermal, radiant and other types are created, some of which cause failures (inability of a system to satisfy felt needs.). To maintain functionality of a machine, actions like servicing, repairs, inspections, replacements and similar, are undertaken by humans. Thus, from the point of view of the ability to function during the in-service life, known as f**unctionability**[1]**,** a machine could be in a functionable or failed state, at any instant of time.

Experience teaches us that unlike quantitative information regarding the design-in functionality performance of a machine that is available on the delivery day, the in-service functionability performance is not. Instead, years later the statistics for various functionability measures become available. The reason for this is the fact that they are emerging properties of the complex interactions between machine in-service processes, which are characterised by uncertainty, discontinuity, irreversibility, inseparability, and dependence of time, location and humans.

To scientifically understand the mechanics of the motion of a machine through functionability states during in-service life and to develop laws and rules that enable predictions of emerging functionability trajectory to be made, at the decision-making stages, Dr Knezevic established the MIRCE Akademy at Woodbury Park, in 1999. Staff, Fellows, Members and students of the Akademy study in-service behaviours of a machine to:

- Determine the patterns of the motion of a machine through functionability states and to measure emerging functionability properties.
- Understand mechanisms of the motion of a machine through functionability states, within the physical scale from 10^{-10} to 10^{10} metre,
- Define the scheme for the prediction of emerging functionability measures for a given: machine in a given in-service conditions.

[1] Knezevic, J., Reliability, Maintainability and Supportability – A probabilistic Approach, Text and Software package, pp. 291, McGraw Hill, London 1993. ISBN 0-07-707691-5

A generated body of scientific knowledge constitutes Mirce-mechanics whose axioms, formulas, methods and rules enable predictions of the emerging functionability trajectory of the future transportation, communication, navigation and many others systems to be made.

On The Non-Existence of Parallel Universes in Science[2]

Richard F. W. Bader, Grand Fellow of the MIRCE Akademy
Department of Chemistry, McMaster University, Hamilton, ON, Canada

Abstract:
Thoughts on the divide that exists in science between those who seek their understanding within a universe wherein the laws of physics apply and those who prefer alternative universes wherein they are suspended or 'bent' to suit preconceived ideas.

Introduction

We are at a cross-roads in science wherein we face a dichotomy in two important areas:: in one we must choose between physics and those who believe the conceptual basis of science to lie beyond physics, a view subscribed to by Professor Hoffmann, a Nobel Prize winner in chemistry, for example. In the other, we must choose between physics and those who challenge the laws of physics to suit their preconceived ideas. The first of these dichotomies has recently been addressed in a paper in the Journal of Physical Chemistry A where one will find a quotation from Hoffmann summarizing his view[1]

Development of the Theory of Atoms in Molecules

These dichotomies of views were bought to the fore with the development of the quantum theory of atoms in molecules (QTAIM). Those who deny the possible existence of a physical basis for the concepts of chemistry are placed directly at odds with QTAIM, whose very existence stems from the discovery that in the observable topology of the electron density, one finds the definitions of atoms, of the bonding between atoms and hence of molecular structure, the conceptual basis of chemistry. By relating these concepts to the electron density, a physically measurable property (that is, the expectation value of a Dirac observable), the theory provides the necessary link for their ultimate quantum definition, one that follows from the extension of a fundamental statement of physics to an atom in a molecule.[2] Feynman[3] and Schwinger[4] demonstrated how the classical action principle embodied in the Lagrangian approach to physics could be generalized to obtain its quantum analogue, an approach suggested by Dirac in 1933.[5] This suggestion led to Feynman's path integral formulation and to Schwinger's principle of stationary action. Schwinger's reformulation of physics, which is a differential statement of Feynman's path integral, combines the principle of least action with Heisenberg's equation of motion for the quantum observables, thus providing "all of physics" in a single statement.[4] It is Schwinger's principle that is most readily extended from its application to an infinite, necessarily closed system, to one with finite boundaries, that is, to an atom in a molecule. An atom in a molecule is necessarily an open system free, to exchange matter and energy with its neighbours. This extension of Schwinger's principle is accomplished by demonstrating that a universal property of the electron density serves as the quantum boundary condition for an open system; that regions

[2] Professor Bader has submitted this paper to the MIRCE Akademy in June 2011 instead of the Grand Fellowship Award acceptance speech, as it was not possible for him to travel to the UK to deliver it, due to medical condition.

of space containing a single nucleus are bounded by surfaces through which there is a zero flux in the gradient vector field of the density. Figure 1 shows the definition of the atomic surfaces in terms of the trajectories traced out by the gradient of the electron density for the formaldehyde molecule, OCH_2. The trajectories defining an interatomic surface terminate a particular kind of critical point, one that is found between all pairs of bonded atoms. Since gradient paths can never cross, the surfaces are necessarily ones of 'zero-flux'. The diagram demonstrates that the critical point serving as the terminus for the trajectories that define an atomic surface, also serves as the origin of a pair of trajectories that terminate at the neighbouring nuclei. They define a line along which the density is a maximum with respect to any neighbouring line – they define a *bond path*. The linked set of bond paths defines a molecular structure, as shown in the same figure. A molecular structure defined in this manner recovers all known 'classical' chemical structures that are denoted by linking atoms with lines representing chemical bonds. There are no exceptions. As nuclei execute spatial excursions, the density may undergo catastrophic, discontinuous changes, causing a change in the molecular structure. These topological changes and the ensuing changes in structure are all predicted and quantified using the mathematics of qualitative dynamics and one obtains a complete theory not only of molecular structure, but of structural stability, as well.[6]

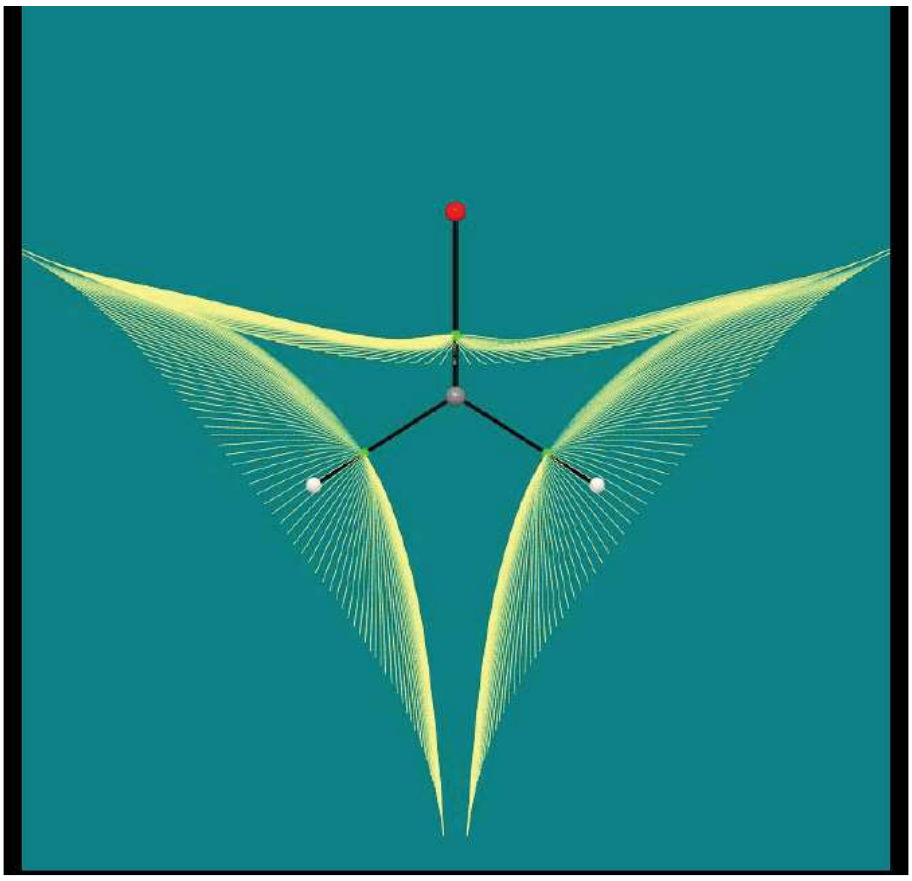

Fig. 1*: Atoms in molecules and the 'bonds' that link them are the natural consequences of the manner in which the density of electronic charge is distributed in real space. The structure imposed on the density by the nuclear attractive force is brought to the fore in terms of the paths traced out by the gradient vectors of the density, as illustrated here for the OCH2 molecule. The trajectories originate and terminate at critical points, points where the gradient vector vanishes. The universal signature of bonding is a consequence of the properties associated with a bond*

critical point, as found and illustrated here, between each pair of bonded nuclei. The surface separating the basins of neighbouring atoms is defined by the set of trajectories that terminate at such a critical point and the bond path that links their nuclei by the unique pair of trajectories that originate there.

Schwinger's principle, as well as yielding Schrödinger's equation, yields Heisenberg's equation of motion for a quantum observable, thus providing one with all of the tools required for the application of physics to any chemical problem. Heisenberg's equation generalized to any system Ω bounded by a 'zero-flux' surface S(**r**,Ω), one that includes the total system, is given in eqn (1).[7,8]

$$d\int_{\Omega}\rho_G(\mathbf{r})/dt = (N/2)\left\{(i/\hbar)\left\langle\Psi\left|[\hat{H},\hat{G}]\right|\Psi\right\rangle_{\Omega}+cc\right\}-(1/2)\left\{\oint dS(\mathbf{r},\Omega)\mathbf{J}_G\bullet\mathbf{n}+cc\right\} + (1/2)\left\{\oint dS(\mathbf{r},\Omega)(\partial S/\partial t)\rho_G(\mathbf{r})\right\} \qquad (1)$$

The electron and vector current densities, the two physical properties that determine the properties of matter are given below in eqns (2) and (3) for the observable $\hat{G}$.

$$\rho_G(\mathbf{r}) = \tfrac{1}{2}\left\{N\int d\tau'(\Psi^*\hat{G}(\mathbf{r})\Psi+(\hat{G}(\mathbf{r})\Psi)^*\Psi)\right\} \qquad (2)$$

$$\mathbf{J}_G(\mathbf{r}) = \tfrac{1}{2}\left\{N\int d\tau'(\Psi^*\hat{G}(\mathbf{r})\Psi+(\hat{G}(\mathbf{r})\Psi)^*\Psi)\right\} \qquad (3)$$

For a closed system, one with infinite boundaries, the surface terms vanish and one is left with the usual statement of the equation of motion relating the time dependence of an observable to its commutator with the Hamiltonian $\hat{H}$. It is the surface flux in the current density of $\hat{G}$ that distinguishes the physics of an open system, a term that persists in a stationary state wherein it describes the instantaneous effect of the surrounding on the system Ω. The subscript Ω on the averaging in the commutator and the prime appearing on d**r** in eqns (2) and (3) denote a summation over spins followed by an integration over all electronic coordinates save those denoted by the position vector **r,** multiplied by N, the number of electrons. This is the same procedure used to obtain the electron density from the product ψ*ψ and its use in the theory of atoms in molecules results in a real-space density representation of all properties. The real-space density is a 'dressed distribution', one that accounts for the corresponding interaction of the property density at some point **r** in space with the remainder of the molecule, a most important result for interpretive understandings of a physical property. Thus one obtains density distributions for all properties, including those that involve two-electron interactions such as the total energy and the electronic potential energy.

Eqn (1) enables one to determine the expectation values of all observables for an atom in a molecule, as for example the Cr atom in the molecule $Cr(CO)_6$ shown in Fig. 2, bounded by its six 'zero-flux' surfaces. This molecule was studied to illustrate the correspondence between the understanding of chemical properties obtained using QTAIM with those

obtained from molecular orbital models.[9] It is most important to note that the calculated atomic contributions to

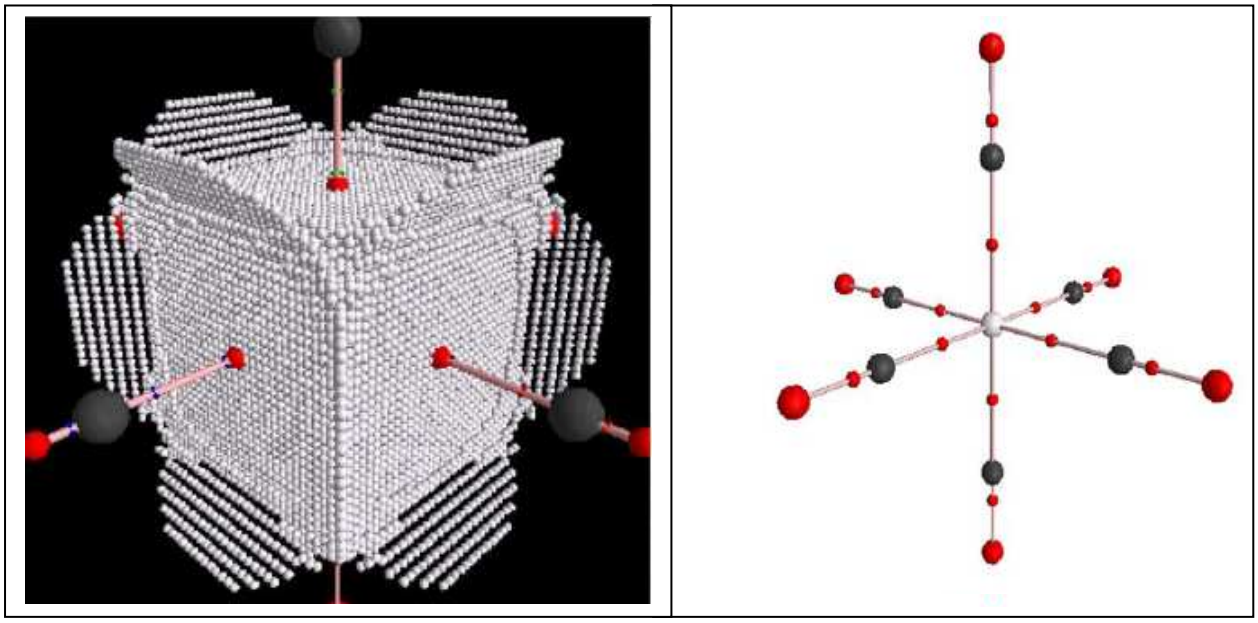

Fig. 2(a) Fig. 2(a)

Fig. 2(a) *Three of the six surfaces bounding a Cr atom in $Cr(CO)_6$, a molecule with D_{6h} symmetry, are shown. All of the properties of this enclosed region are defined by quantum mechanics and they make additive contributions to the properties of the complex. It is shown that the principal source of bonding in $Cr(CO)_6$ is the electrostatic interaction of the Cr atom density with the nuclei of the ligands. The bond CPs are denoted by the red dot embedded in each surface, there being one bond CP in each Cr|C surface. Also shown are the bond paths originating at each of these bond critical points and linking a carbon nucleus, denoted by a black ball, and the further bond path linking each C to an oxygen atom denoted by a red ball.* ***Fig 2(b)*** *The molecular graph of the $Cr(CO)_6$ molecule defined by the bond paths. As in all cases, this structure duplicates the classical structure.*

additive properties agree with those determined experimentally: heats of formation, electric and magnetic susceptibilities for example.[10] All of the nuances of chemical models are recovered by QTAIM including, via the atomic partitioning of the exchange density, a physical measure of the localization/delocalisation of electrons, a property reflected in the Laplacian of the electron density, $\nabla^2\rho(r)$.[11] One now has at hand the physics required to recast of all of chemistry in terms of quantum mechanics.

Why Were the Electron and Current Densities Overlooked?

One can trace the shunning of physics in the period following the advent of Schrödinger's 'wave equation' in 1926[12] to a paper that appeared the following year by Heitler and London[13] wherein they applied the wave equation to the hydrogen diatomic molecule and demonstrated that quantum mechanics could account for 'covalent' bonding, as opposed to ionic bonding, which at the time was readily explained in terms of electrostatics. The Heitler-London (H-L) paper expressed the wave function for H_2 as a combination of two atomic-like

terms: a(1)b(2) assigning electrons 1 and 2 to atoms 'a' and 'b' and the term a(2)b(1) where the electronic positions are exchanged between the atoms. This process was likened to quantum mechanical resonance (although H-L went to some lengths in their paper to counter this analogy[13]) and the interpretation of bonding in terms of the wave function and 'resonance' were to dominate theory from that moment on to the present day, as exemplified in the recent statement: 'which showed that the bond energy in H_2 is due to resonance between the electrons as they exchange positions between the two atoms.'[14]

Where to begin? Schrödinger, in his fourth paper in 1926, wherein he derived the expressions for the electron density, the vector current density and the equation of continuity connecting them, warned against the use of the wave function for other than the determination of the electron and current densities and championed their use in understanding the properties of matter.[15] The wave function is of course, essential to the determination of the eigen- and expectation values of observables and their subsequent use in the theorems of quantum mechanics, the path followed in QTAIM, but his advice that one avoid using ψ directly in the interpretation of physical observations and instead relegate it to the purpose of determining the density $\rho(\mathbf{r})$ was not followed, resulting in a delay in relating the chemical concepts of atoms and bonding to a system's charge distribution. Experimental chemistry is presently explained using many empirical concepts such as electronegativity, Pauli and steric repulsions, covalency and nonbonded interactions, none of which relate directly to quantum mechanical theorems.

It is surprisingly little known that the fundamental role of the density in understanding chemical bonding was first pointed out by London in 1928 in a companion paper to the one he co-authored with Heitler giving the quantum mechanical description of homo-polar bonding.[16] London gave contour diagrams of the density distributions associated with the antisymmetric and symmetric solutions to the Heitler-London (H-L) equations, his diagrams having been recently reproduced.[17] He obtained the densities by integrating $\psi^*\psi(\mathbf{r}_1,\mathbf{r}_2)$ over the coordinates of one of the electrons, employing the definition of the density $\rho(\mathbf{r})$ provided by Schrödinger in the preceding year.[15] London's paper presents the first calculated representations of the electron density. London, in describing the antisymmetric (the lowest excited, unbound state of H_2) and symmetric (the bound, ground state) density distributions so obtained, states: "We see that the densities for the antisymmetric solution are clearly pushed outward, as if they would separate if possible. If we would bring the nuclei closer together, the strangling of the density between the atoms would increase;" "In opposition to this the density for the symmetric state shows the two atoms which are in a state of homo-polar binding. Here the two densities seem to draw closer and become one. *With the help of these figures, one can imagine how in complicated molecules the atoms which form a valence are connected by such a bridge of $\psi^*\psi$-density, while all remaining atoms stay separate.*" His figure for the symmetric state clearly illustrates the build-up of density between the nuclei, a situation commented on by Feynman:[18] "In a H_2 molecule for example, the (H-L) symmetrical solution can easily permit charge concentration between the nuclei and hence it is the only solution which is symmetrical that leads to strong attraction, and the formation of a molecule, as is well known." *London was the first to define a bond path as a 'bridge of density' and to postulate its physical significance in the understanding of bonding.* Unfortunately, this paper went unheeded (and remained so for 80 years)[17] and instead the chemical community embraced the 'mysterious wave function' and resonance.

Ascribing bonding to 'resonance' removes one from the realm of the physics of definable forces - the force exerted on the nuclei, as given by Feynman's electrostatic theorem[18] and

the Ehrenfest force exerted on the electrons.[19] These are the only forces operative in bonding and they are never invoked in present day discussions. Feynman's electrostatic theorem demonstrating that bonding is a result of the accumulation of electron density between the nuclei that exerts an attractive force on the nuclei sufficient to overcome the force of repulsion between them is particularly reviled as being 'too simple'.[20] Instead, one finds statements to the effect "all attempts to explain the chemical bond in terms of the electron density have failed" along with the associated statements that bonding is not the result of the accumulation of the density between the nuclei[21] In brief, all electrostatic interpretations are held suspect when in fact, the potential energy operator in the Hamiltonian consists of the electrostatic interactions between the electrons and the nuclei. The sole attractive interaction in the Hamiltonian, and hence the term responsible for chemical bonding, is the electrostatic interaction between the nuclei and the electrons, one that can be recast in terms of the electron density. This attractive force acts in opposition to the repulsions between the electrons and between the point-like nuclei. While the repulsion between the electrons is a two-electron contribution, that through the Pauli exclusion principle leads to its breakdown into a classical-like Coulomb contribution and the purely quantum mechanical exchange energy, it is describable in terms of electrostatic repulsions using the Ehrenfest theorem which determines the forces acting on the electron density.

The three theorems essential to chemistry and bonding in particular are Feynman's electrostatic theorem governing the forces on the nuclei, Ehrenfest's theorem determining the forces on the electrons, that is, on the electron density, and their unification by Slater's virial theorem which relates the electronic kinetic energy to the total and potential energies.[22] It is the virial of the Ehrenfest force acting on the electrons and of any residual Feynman forces acting on the nuclei that determines a system's virial, its potential energy. Thus through the Ehrenfest and Feynman theorems, one has the tools that are needed to describe the forces acting in a molecule and through the virial theorem, to relate these forces to the molecule's energy and its kinetic and potential contributions. Slater was the first to make extensive use of these theorems in discussions of chemical bonding.[23] He viewed the virial and Feynman theorems as being the two most powerful theorems applicable to molecules and solids.[24] Unfortunately, the first two of these theorems are ignored and the third is contravened in most present discussions of chemistry.

The Feynman, Ehrenfest and virial theorems all lead to the identical cause of chemical bonding: the result of lowering the potential energy of the molecule resulting from the accumulation of electron density between the nuclei.[2,25 9,26] It is difficult to understand how such a view can be questioned in the face of the *observation* that every 'bond' in a classical molecular structure is mirrored by a bond path whose presence denotes an accumulation of density, a 'bridge' of density in London's words, between the bonded nuclei, an observation verified many times over in the quantum theory of atoms in molecules.

Time to Choose Between Single- and Multi-Universes

I believe in the existence of a single universe, one in which the laws of physics apply. I do not believe in the existence of parallel universes wherein the laws are either ignored or 'bent' to accommodate personal points of view. A prime example of a broadly accepted model of bonding that requires the acceptance of alternative universes is the argument that the kinetic energy must decrease in the energy change associated with bond formation.[27,28-30] This view is in direct opposition to the virial theorem that requires the kinetic energy to *increase* by an amount equal to the *decrease* in the total energy, the familiar statement $\Delta T=-\Delta E$. The model

stems from neither physics nor observation but from imaginary steps envisaged in the minds of its proponents. One readily proves, among other things, that the total kinetic energy can decrease only in the presence of an attractive Feynman force acting on the nuclei. While this can and does occur, it does not occur in an equilibrium configuration nor contribute to the equilibrium bonding energy wherein the Feynman forces vanish and the imagined steps require a universe in which the Ehrenfest, Feynman and virial theorems are all suspended.

I wish to distinguish schemes such as this from *models*, which as normally done, spring from attempts to organize and explain *observations*, a most noble and useful cause. Good models are ultimately related to physics and this brings forth the strongest argument against the use of imagined schemes; *they refer to a universe in which the laws of physics are suspended and thus they cannot be related to any observation nor used to make any predictions, the ultimate goal of science.*

The universe in which we live is more exciting and demanding of our abilities than any that one might care to imagine. Most of it, of course remains to be explored, in particular, that portion wherein chemistry resides. The purpose of QTAIM is straightforward: to propose that one maximally embrace observation and physics in the understanding of chemistry, an understanding that has grown from the finding of chemical concepts in the topology of the electron density. It is a goal that is approached more closely each day, as more and more workers apply QTAIM in their research, research that covers all facets of the study of matter at the atomic level.

References

1. Bader, R. F. W. *J. Phys. Chem. A* **2010**, *114*, 7431-7444.
2. Bader, R. F. W. *Atoms in Molecules: a Quantum Theory*; Oxford University Press: Oxford UK, 1990.
3. Feynman, R. P. *Rev. Mod. Phys.* **1948**, *20*, 367-387. 11
4. Schwinger, J. *Phys. Rev.* **1951**, *82*, 914-927.
5. Dirac, P. A. M. *Physik. Zeits. Sowjetunion* **1933**, *3*, 64.
6. Bader, R. F. W.; Nguyen-Dang, T. T.; Tal, Y. *Rep. Prog. Phys.* **1981**, *44*, 893-948.
7. Bader, R. F. W.; Nguyen-Dang, T. T. *Ad. Quantum Chem.* **1981**, *14*, 63-124.
8. Bader, R. F. W. *Phys. Rev.* **1994**, *B 49*, 13348-13356.
9. Cortés-Guzmán, F.; Bader, R. F. W. *Coordination Chem. Rev.* **2005**, *249*, 633-662.
10. Matta, C. F.; Bader, R. F. W. *J. Phys. Chem. A* **2006**, *110*, 6365-6371.
11. Bader, R. F. W.; Heard, G. L. *J. Chem. Phys.* **1999**, *111*, 8789-8798.
12. Schrödinger, E. *Ann. D. Phys.* **1926**, *79*, 361.
13. Heitler, W.; London, F. *Z. Physik* **1927**, *44*, 455.
14. Shaik, S. *J. Comp. Chem.* **2007**, *28*, 51-61.
15. Schrödinger, E. *Ann. D. Physik* **1926**, *81*, 109.
16. London, F. *Z Physics* **1928**, *46*, 455. From the English translation by Hinne Hettema, in *Quantum Chemistry: Classic Scientific Papers*, World Scientific, Hong Kong, 2000.
17. Bader. R. F. W. *J. Phys. Chem.* **2009**, *A 113*, 10391-10396.
18. Feynman, R. P. *Phys. Rev.* **1939**, *56*, 340-343.
19. Ehrenfest, P. *Z. Phys.* **1927**, *45*, 455.
20. Gleick, J. *Genius: The Life and Science of Richard Feynman*; Vintage Books: New York, 1992, page 90.
21. Frenking, G. *Angew. Chem. Int. Ed.* **2003**, *42*, 143-147.

Two quotations from this paper make the stance of the wave function approach clear. "...the authors should be reminded of the ground breaking publication of Heitler and London in 1927, which showed that the fundamental basis of the chemical bond is the resonance of the wave function and it cannot be explained simply by the electron density." "Since (that publication) we know that it is only the wavefunction that gives an explanation of the chemical bond, whereas all attempts to explain the chemical bond in terms of the electron density have failed."

22. Slater, J. C. *J. Chem. Phys.* **1933**, *1*, 687. 12
23. Slater, J. C. *Quantum Theory of Molecules and Solids. I*; McGraw-Hill Book Co. Inc.: New York, 1963.
24. Slater, J. C. *J. Chem. Phys.* **1972**, *57*, 2389.
25. Bader, R. F. W.; Hernández-Trujillo, J.; Cortés-Guzmán, F. *J. Comput. Chem.* **2006**, *28*, 4-14.
26. Bader, R. F. W. *J. Mol. Struct. (THEOCHEM)* **2010**, *943*, 2-18.
27. Hellmann, H. *Einführung in Die Quantumchemie*; Deauticke: Vienna, 1937.
28. Ruedenberg, K. *Rev. Mod. Phys.* **1962**, *34*, 326-376.
29. Ruedenberg, K.; Schmidt, M. W. *J. Phys. Chem.* **2009**, *113*, 1954-1968.
30. Kutzelnigg, W. *Angew. Chem. Int. Ed* **1973**, *12*, 546-562.

The Two Faces of Chemistry: Can they be reconciled?

Mark E. Eberhart and Travis E. Jones
Molecular Theory Group, Colorado School of Mines, Golden, Colorado 80401, USA.

Abstract:

Shortly before his death, Richard Bader commented on the dichotomy that exists within chemistry and between chemists (Bader 2011). We believe that the dichotomy results from different goals and objectives inherent in the chemical disciplines. At one extreme are designers who synthesize new molecules with interesting properties. For these chemists, the rationale underpinning molecular synthesis is far less important than the end product—the molecules themselves. At the other extreme are the chemists who seek a fundamental understanding of molecular properties. We suggest tat the Quantum Theory of Atoms in Molecules, by virtue of the rich hierarchical structure inherent in the theory, offers a bridge through which to unite these two groups. However, if there is to be reconciliation, it falls to the theorists to develop "quantum mechanically" correct tools and concepts useful to the synthetic and applied chemist.

1. INTRODUCTION

Chemistry is a discipline of two faces, one applied and the other theoretical. The applied face focuses on the design and synthesis of molecules and solids, while the theoretical face looks for explanations of a molecule or solid's properties. At first blush, this observation may not seem to warrant note, yet it does set chemistry apart from its sister sciences. The applied and theoretical components of physics, for example, have been subsumed into separate areas of study and specialization. Classical physics is largely concerned with the theories of statics, dynamics, electricity, magnetism, fluid dynamics, and so forth, while the application of these theories forms the basis of civil, mechanical, electrical, and aeronautical engineering.

The formal recognition of the applied and theoretical components of science as distinct is to some degree driven by the different values of designers and theoreticians. From these values, vocabulary arises that is nuanced to the specific disciplines; the physicist and the civil engineer may both use the word mass, but the image evoked by the word is likely to be quite different. For chemists, however, where there has been no clear separation of the science into its design and theoretical parts, our value differences can result in confusion, misunderstanding, and controversy. But from this controversy there also comes an opportunity to align theory to the needs of the applied chemist.

This is exactly the situation that presents itself to the community advancing the Quantum Theory of Atoms in Molecules (QTAIM) (Bader, 1990). That there is controversy and misunderstanding is clearly evidenced by the recent article in which Richard Bader (2011) "presents thoughts on the divide that exists in chemistry between those who seek their understanding within a universe wherein the laws of physics apply and those who prefer alternative universes wherein the laws are suspended or bent to suit preconceived ideas." Before leaping headlong into this controversy, however, it is worth taking a deeper look at the value difference that place the laws of physics in opposition to the ideas of some chemists.

Dirac (1929) was probably the first to articulate the values of the modern quantum chemist when he wrote,

> *The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.*

One can only imagine what Dirac would consider to be "too much computation," nonetheless, for nearly 80 years the emphasis has been to exploit growing compute power to implement increasingly complex but computationally feasible approximate methods that allow the properties of molecules and solids to be calculated. For many theoreticians, the ability to calculate a property from first principles serves as a complete explanation of its origins.

On the other hand, the designer is interested in identifying molecules with a given set of properties. As has been argued before (Eberhart, 2002), this ability is conferred by what are known as structure-property relationships, which express the control that structure plays in mediating properties.

The search for structure-property relationships is an essential component of scientific inquiry and usually begins with a precise hierarchical description of structure, where each level is characterized by a different length scale (Smith, 1981). However, Cohen (1976) has argued that sometimes representations of structure are mere constructs that allow one to rationalize properties. Cohen called structure-property relationships of this type, "reciprocity relationships," which abound in the contemporary picture of molecular structure.

Known to every first year chemistry student is the common classification of molecules and solids as ionic, covalent, Van der Waals, or metallic. These classifications grew exclusively from the desire to see structure as the origin of a particular property. It was Arrhenius' (ca. 1885) need to explain the property of electrical conductivity of some solutions that gave rise to the ion and the description of a crystal as ionic. Lewis (1916) originated the covalent bond as a way to explain the existence of binding forces in nonionic molecules. The need to explain the formation of condensed phases by molecules whose atoms possessed a full octet of electrons (an impossibility according to the Lewis picture) led to the Van der Waals bond (ca. 1922). The metallic bond (ca. 1925) grew from the need to explain the differences in conductivity between nonionic solids. All of these representations of the bond were devised before the discovery of quantum mechanics. However, they were so effective in explaining the chemical phenomena of interest, and so useful in the synthesis of new molecules, that during the first half of the twentieth century they became the principal descriptors of molecular structure and the basis for many, if not most, of the structure property relationships used by the applied molecular scientist. Hence, many of these structure property relationships cannot be reconciled with quantum mechanics, i.e. "wherein the laws of physics apply."

Simply stated, if the goal is to bring chemistry in its entirety under the umbrella of physics, it is unlikely that providing a quantum mechanically correct description of molecular structure will be sufficient. To achieve this goal, one must build an entirely new set of quantum mechanically correct structure-property relationships that are as useful as the reciprocity relationships that they will replace.

The Hohenberg–Kohn theorem (Hohenberg and Kohn, 1964) provides insight as to the form that some of these relationships may take. It posits that all ground-state molecular properties

are a consequence of the charge density. And, of course, QTAIM provides an extremely rich—quantum mechanically correct—description of charge density. It seems only reasonable, therefore, to seek relationships between charge density, as descried by QTAIM and, at a minimum, ground state properties. And as Cyril Smith (1981) has argued, this search should begin with a precise hierarchical description of structure.

2. THE QTAIM STRUCTURAL HIERARCHY

Though little noted, intrinsic to QTAIM is a structural hierarchy derived from volumes bounded by surfaces on which the gradient of the charge density vanishes at every point. Such surfaces are referred to as zero flux surfaces or ZFSs. By virtue of being bounded by ZFSs, these volumes are quantum observables, that is, they are characterized by properties that are in principle measurable (Bader, 1990). Though there are infinitely many such volumes, the initial formulation of QTAIM was concerned only with those where the bounding ZFSs did not intersect an atomic nucleus. These distinct volumes, called Bader atoms or sometimes atomic basins, partition the charge density of a molecule or solid into space filling regions each of which encloses a single nucleon—hence the name, "atom."

QTAIM's structural hierarchy is a consequence of the topological connections between Bader atoms, which are determined by the charge density's rank 3 critical points, CPs. These are the places where the charge density, a three-dimensional scalar field, achieves extreme values in all directions. As with all 3D scalar fields, the charge density possesses at most four kinds of CP: local minima, local maxima, and two types of saddle points. These CPs are denoted by an index, which is the number of principal positive curvatures minus the number of principal negative curvatures. For example, at a minimum, the curvature in all three orthogonal directions is positive; therefore it is called a (3, +3) CP. The first number is simply the number of dimensions of the space and the second is the net number of positive curvatures. A maximum is denoted by (3, -3), because all three curvatures are negative. A saddle point with two of the three curvatures negative is denoted (3, -1), while the other saddle point is a (3, +1) CP.

The charge density at the atomic nucleus is always a maximum a (3,+3) CP (within the approximate Coulomb Hamiltonian the charge density at a nucleus is cusp with the curvatures undefined), hence it is also called a nuclear CP. The other CPs, which must be present in a molecular system, sit on the ZFSs bounding the Bader atoms and mediate their connectivity (Bader, 1990; Zou and Bader, 1994). The simplest topological connection results from a shared (3, -1) CP between two Bader atoms, and is indicative of a charge density ridge originating at the (3, -1) CP and terminating at the nuclear CPs. In essence, this charge density ridge possesses the topological properties imagined for the chemical bond, which motivated studies showing the presence of such a ridge between atoms that conventional wisdom assumed to be bound. Accordingly, this ridge is descriptively referred to as a bond path and the accompanying (3, -1) CP as a bond CP. Thus, through QTAIM, a structure was associated with the strictly heuristic concept of the bond as a simple link.

Other types of CPs have been correlated with other features of molecular connectivity. A (3, +1) CP is required at the center of ring structures (rings of bond paths). Accordingly, it is designated a ring CP. Cage structures must enclose a single (3, +3) CP and are given the name cage CPs. In these cases, the Bader atoms of a ring or cage, share ring and cage CPs with the other atoms of the ring or cage, which is indicative of the more complex nature of the topological connections between these atoms.

Though only recently noted (Eberhart, 2001; Jones and Eberhart, 2009, 2010), CPs, bond paths, and the ZFSs of Bader atoms are elements drawn from the larger set of extremal points, lines, and surfaces. These extremals are generically referred to as ridges and valleys. Incorporating all members of this set into QTAIM provides a more robust, elegant, and unified topological theory of molecular structure.

In 2D, a ridge is a familiar topographic feature, the path (gradient path) connecting mountain passes to neighboring peaks, for example. There is only one such gradient path, and it is a path of locally least steep ascent terminating at the local maximum. Consequently, it is an extremum with respect to all neighboring paths. Similarly, a valley is an extreme gradient path connecting a saddle point to a local minimum, and because valleys and ridges differ only by the sign of the curvatures along the path, both are often referred to as "ridges." In 3D fields (the electron charge density), ridges are the points, gradient paths, and zero flux surfaces that are extreme with respect to all neighboring points, gradient paths, and zero flux surfaces respectively. They are denoted by an index, $n - d$, where n is the dimensionality of the space and d is the number of principal directions in which the charge density is extremal (Eberly et al., 1994). Thus, a 0-ridge is nothing more than one of the four types of critical points. A 1-ridge is an extremal gradient path, of which the bond path is an example. And a 2-ridge is an extremal gradient surface, of which the ZFSs bounding Bader atoms are examples.

For an extended systems[3] there will always be 4 kinds of charge density CPs (0-ridges), six kinds of 1-ridges, and 4 kinds of 2-ridges. The 1-ridges pairwise connect the four kinds of critical points and the 2-ridges are surfaces containing three distinct CPs. The ridge structure forms a set of space filling volumes homeomorphic to a tetrahedron. Coincident with the four vertices of each tetrahedron is a nuclear, bond, ring and cage CP, respectively. The six edges of the tetrahedron are 1-ridges, and the 4 faces are 2-ridges, Figure 1.

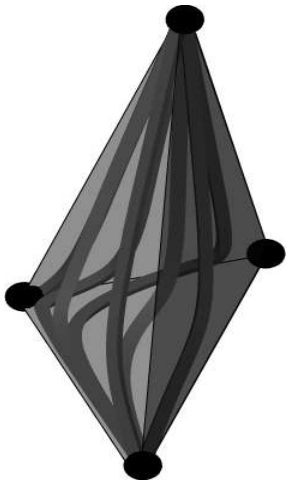

FIG. 1. *An irreducible bundle of fcc Cu. The CPs are shown as spheres: nuclear CP top, ring CP middle left, bond CP middle right, and cage CP bottom. The dark rods are gradient paths (differential gradient bundles) originating at the cage CP and terminating at the atom CP. The gradient paths in the IB are confined to the tetrahedral volume. The faces of this*

[3] *All examples are drawn from extended systems where the ridge structure is somewhat easier to visualize than it is in finite molecules. In solids all possible ridges are present and all gradient paths are of finite length. Though only examples drawn from extended systems will be given here, the arguments are generalizable to open systems, i.e. molecules and surfaces Jones and Eberhart (2010).*)

tetrahedron are 2-ridges each containing three CPs.

The resulting tetrahedra are simplices, which means that they are the most basic unit of charge density retaining local topology. Simplices may be glued together to form a simplical complex that is homeomorphic to the charge density topology of any molecular system. Accordingly, these simplices have been designated irreducible bundles, IBs, where bundle is used to evoke an image of a bundle of gradient paths.

By way of illustration, consider an fcc copper crystal. The topology of this structure is distinguished by five symmetry-unique CPs: a ring CP, a bond CP, a nuclear CP, and two cage CPs–one each at the center of the tetrahedral and octahedral holes. These CPs give rise to two distinct irreducible bundles, shown in Figure 2. The first has as its vertices the nuclear, ring, bond, and cage CP from the tetrahedral hole and the second the nuclear, ring, bond, and cage CP from the octahedral hole. Since the two IBs have three vertices in common, they must share a tetrahedral face, i.e. a 2-ridge. Together these two IBs form the symmetry unique wedge of the fcc crystal structure, which, under the operations of the fcc space group, will generate an extended simplicial complex whose physical realization is the full charge density of fcc copper.

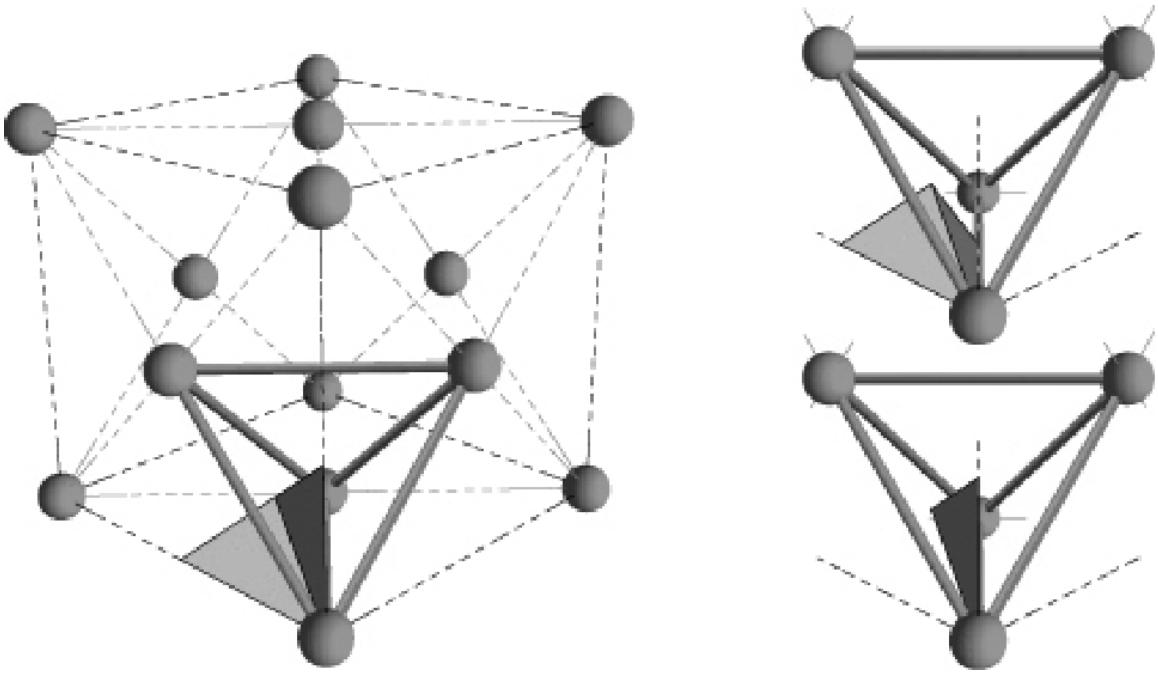

FIG. 2. *The two symmetry unique simplicies, IBs, of the fcc crystal structure. The IBs are shown to the right as shaded polyhedra relative to the fcc tetrahedral hole. Top right is the IB containing the cage CP located in the octahedral hole while bottom right is the IB containing the cage CP from the tetrahedral hole. These two IBs may be glued together to form the symmetry unique wedge of the fcc structure (left), which, under the operations of the fcc space group will generate the charge density of the full fcc crystal.*

From a molecule or solid's simplicial complexes one can construct subcomplexes, called d-skeletons, which recover the charge density at various topological levels. In particular, the 0-skelton of the simplicial complex is the set of all of its CPs. Its 1-skeleton consists of all 1-ridges and is called the underlying graph of the complex. The 2-skeleton is the set of all 2-ridges, and the 3-skeleton is the full simplicial complex. A molecule or solids underlying graph will contain as a subset the molecular graph common to modern depictions of

molecules. However, the molecular graph depicts only bond paths, i.e. the connections between atoms, while the underlying graph (1-skeleton) depicts the full set of 1-ridges and captures the topology of the atomic connections, providing a more complete representation of bonding than the traditional picture of a bond as a simple connection.

In addition to the extended simplicial complexes, local structures can be generated by gluing together a finite number IBs. The most interesting of these are given through the union of IBs sharing a single CP. The union of all IBs sharing the same nuclear CP will generate Bader atoms. The union of all IBs sharing the same cage CP will yield the repulsive basin first noted by Pendaś, Costales, and Luaña (1997). Additionally, one can construct the union of all IBs sharing the same ring point. Finally, there is the union of all IBs sharing the same bond CP. This volume will contain a single bond critical point and its associated bond path and is referred to as a bond bundle, again to stress the fact that as a bundle of gradient paths the volume is bounded by ZFSs and hence has well defined properties, for instance an energy.

To illustrate, consider the B2 ordered intermetallic of NiAl, Figure 3. There are two symmetry unique bond CPs in this structure. One is located on the interatomic axis joining aluminum atoms to their eight nearest-neighbor nickel atoms. The other bond CP is on the interatomic axis joining nickel atoms to their six second-neighbor nickel atoms. Consequently, there are fourteen bond paths terminating at each nickel nuclear CP and eight terminating at each aluminum nuclear CP. These two types of bond CP necessitate two different bond bundles, which are shown in Figure 3. The Ni-Ni second neighbor bond bundle is constructed from the union of sixteen IBs of one type and the Ni-Al bond bundle from the union of six IBs of two types.

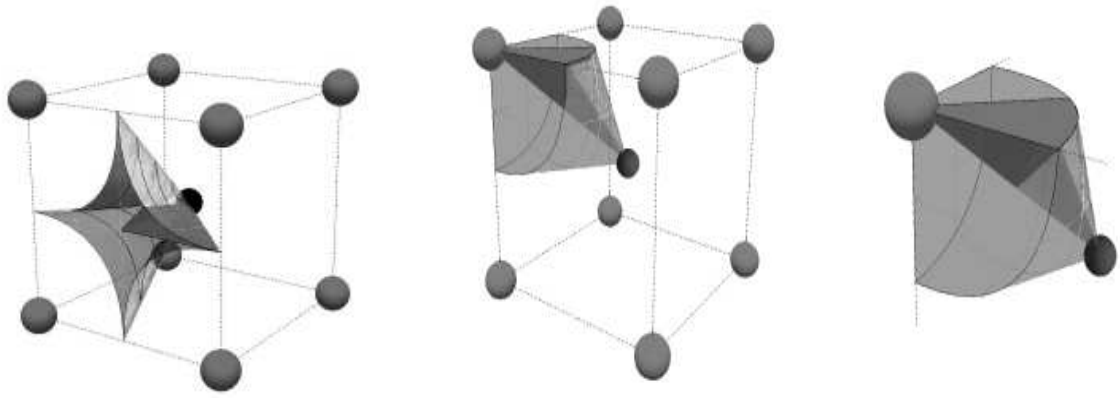

FIG. 3. *The irreducible and bond bundles of the B2 structure. The Al atoms are located at the corners of the cube and the Ni atom at its body center. Shown (left) is half of the Ni-Ni second neighbor bond bundle resulting from the union of the 16 IBs (darker wedge). The first neighbor bond bundle and the IBs from which it is formed are shown on the right.*

Because a bond-bundle is a simplicial complex, it also may be represented in terms of a skeletal structure. Its underlying graph (1-skeleton) provides a compact representation of the connection between atoms.

While the IB is the basic structural element reflecting the connections between atoms, it too has a structure built from volumes bounded by ZFSs. To help with visualization, begin by noting that sufficiently close to a nuclear CP all charge density gradient vectors are radial and of the same magnitude. As a result, it is always possible to find a spherical isosurface, *S*, of

radius *dR* centered on each nuclear CP. In the conventional spherical coordinate system, every point on such a sphere may be specified in terms of a polar and an azimuthal angle, $S(\theta,\varphi)$. Through each point of this surface there corresponds a gradient path, G, that originates from any number of cage CPs and terminates at the nuclear CP at the center of S. Hence, each gradient path may be specified by a pair of coordinates, $G(\theta, \varphi)$. An IB will intersect S to produce a triangle. The set of points on this triangle are denoted as $S_{IB}(\theta,\varphi)$. The gradient paths terminating at the nuclear CP and passing through the points interior to $S_{IB}(\theta,\varphi)$ will originate at the same cage CP and together form a compact open set that is the interior of the IB. Obviously, 2-ridges form the boundary of this open set and its closure is the IB.

Imagine covering $S_{IB}(\theta,\varphi)$ with a set of nonintersecting differential elements of area *dA*. The gradient paths passing through the points comprising each of these area elements gives rise to a family of differential volume elements bounded by ZFSs. These differential gradient bundles, $dG(\theta,\varphi)$, (see Figure 1) are the smallest structures admitted by QTAIM that, in principle, possess measurable properties, and for each nucleon, are parameterized by θ and φ only. The properties of larger structures, such as IBs or Bader atoms, are found by integrating over the properties of the differential gradient bundles. But more significantly, the underlying value of a property derives from a 2D property distribution that is itself well-defined within the QTAIM formalism.

3. DISCUSSION

The existence of QTAIM specific 2D property distributions sheds light on what has been one of the controversial aspects of QTAIM, the assumption that a bond path and CP is indicative of a bonding (energy lowering) interaction. Bader and Preston (1969) have argued that the nature of the bonding interaction derives from the relative curvatures of the charge density parallel and perpendicular to a bond path and not simply from the existence of a bond CP. In support of this argument, Bader and Preston investigated the charge density of dimers including He_2, which served as an example of a "unbounded" interaction, along with several "covalent" dimers such as H_2. Similarly, Figure 4 shows the charge density and gradient paths for "unbounded" Ar_2 and "covalent" N_2 molecules. Note that the family of gradient paths for these molecules is quite distinct. For Ar_2 the gradient paths are atomic like—radial—except in the immediate neighborhood of the 2-ridge that is the ZFS of the Ar Bader atoms. In contrast, for N_2, the gradient path curvature is distributed along a greater length of the path, and particularly, much closer to the nucleus. But significantly, all gradient paths, and hence all differential gradient bundles, sample both the region parallel and perpendicular to the bond path. Values of charge density properties at a point, or along a path other than a differential gradient bundle, have no physical significance within the QTAIM formalism. The implication that a bond CP or a bond path carries information about the "bond" is misguided and contrary to QTAIM formalism.

As an illustration of this point, consider the 2D charge density distribution of N_2 and Ar_2. This distribution is found by integrating the charge density within the differential gradient bundles. In the case of dimers, these are volumes produced by rotating about the internuclear axis a wedge produced by two gradient paths contained in the same internuclear plane and separated by an apex angle of $d\theta$. Figure 4 shows two such wedges for both N_2 and Ar_2, one wedge contains the bond path and the other is perpendicular to the bond path. Consistent with our calculations, and as is apparent, in the case of N_2 it is the wedge containing the bond path with the greatest charge density. On the other hand, for Ar_2, the charge density in the wedge perpendicular to the bond path contains the greatest charge density—though the difference is

small and falls within the range of computational error. Still, the difference in the distribution is telling and provides a quantitative distinction between a “Van der Waals” and a “covalent” bond.

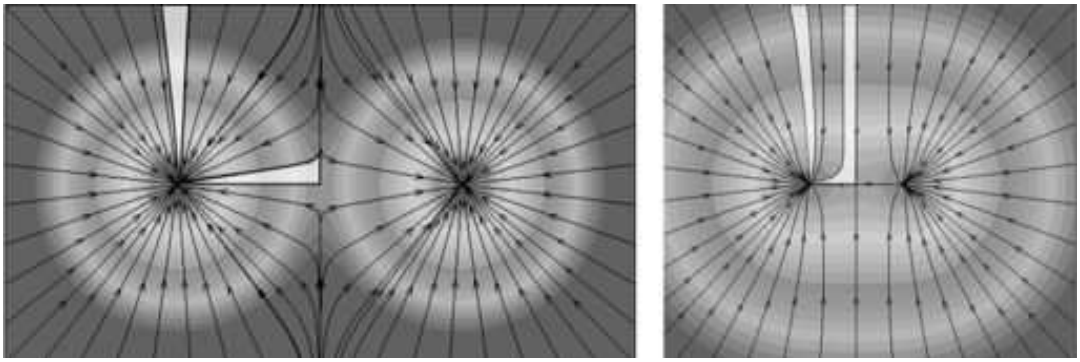

FIG. 4. *The gradient paths of Ar_2 (left) and N_2 (right) superimposed on a charge density contour plot. As discussed in the text, the number of electrons contained in the volumes produced by revolving the white “wedges” about the internuclear axis will give a QTAIM consistent representation of the electron distribution.*

Obviously, the 2D property distributions provides a more detailed and complete descriptions of atom-atom interactions than does the current vocabulary, e.g. ionic, covalent, etc. However, it seems reasonable that these common classes of interactions will manifest distribution similarities. What remains then, is to divine the metrics that will be most useful in characterizing the structure of the family of differential gradient bundles and demonstrating that there are relationships between these structural metrics and the property distributions.

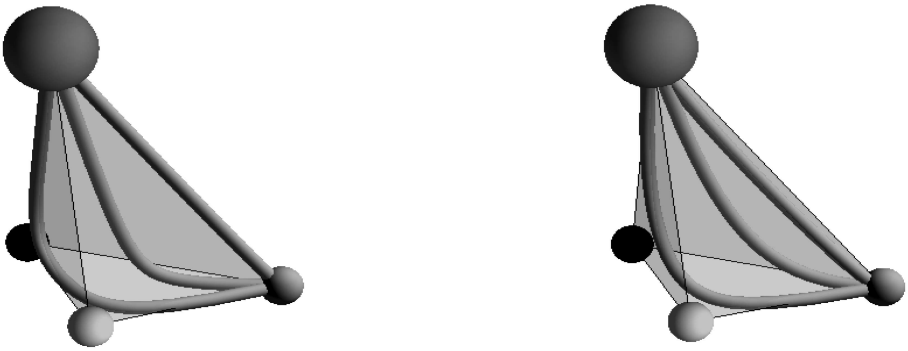

FIG. 5. Representative gradient paths of two IBs from the octahedral hole of fcc elements. For each IB the nuclear CP is at the top, the cage CP to the right, the ring CP is center foreground, and the bond CP on the left. The structure of the family of gradient paths is fully determined by specifying their curvature and torsion at each point in the IB. As in the case of the dimers, we find examples where the curvature and torsion are confined to the regions close to the Bader atom boundaries, e.g. Ar (left) and those where the curvature and torsion are distributed along the gradient path length e.g. Rh (right)

As a first step in developing these relationships, we will hazard a guess as to the structural metrics that might characterize the family of differential gradient bundles by noting that the members of this family are basically space curve, each of which is fully specified by its curvature and torsion at each point along the curve. Inspections of Figures 4 and 5 drive the

point home that these structural features are sensitive to the classes of atom-atom interactions. However, building the convincing case that there are indeed relationships be- tween these structural metrics is clearly beyond the scope of this, or any single investigation, and we leave it as a clarion call to the QTAIM community to confirm the existence of these relationships.

REFERENCES

Bader, R. F. W., *Atoms in Molecules. A Quantum Theory* (Clarendon Press: Oxford, UK, 1990).

Bader, R. F. W., "On the non-existence of parallel universes in chemistry," Found Chem **13**, 11–37 (2011).

Bader, R. F. W. and Preston, H. J. T., "The kinetic energy of molecular charge distributions and molecular stability," International Journal of Quantum Chemistry **3**, 327–347 (1969).

Cohen, M., "Unknowables in the essence of materials science and engineering," Mat. Sci. Eng. 25, 3–4 (1976).

Dirac, P. A. M., "Quantum mechanics of many-electron systems," Proc. R. Soc., London **A123**, 714–733 (1929).

Eberhart, M., "A quantum description of the chemical bond," Phil. Mag. B **81**, 721–729(2001).

Eberhart, M. E., "Quantum mechanics and molecular design in the twenty first century," Found Chem **4**, 201–211 (2002).

Eberly, D., Gardner, R., Morse, B., Pizer, S., and Scharlach, C., "Ridges for image analysis,"J. of Math Imaging Vis **4**, 353–373 (1994).

Hohenberg, P. and Kohn, W., "Inhomogeneous electron gas," Phys. Rev. **136**, B864–B871(1964).

Jones, T. and Eberhart, M., "The bond bundle in open systems," Int. J. Quant. Chem. **110**,1500–1505 (2010).

Jones, T. E. and Eberhart, M. E., "The irreducible bundle: Further structure in the kinetic energy distribution," J. Chem. Phys. **130**, 204108 (2009).

Lewis, G., "The atom and the molecule," J. Amer. Chem. Soc. **38**, 761–785 (1916).

Pendas, A. M., Costales, A., and Luana, V., "Ions in crystals: The topology of the electron density in ionic materials. I. fundamentals," Phys. Rev. B **55**, 4275–4284 (1997).

Smith, C. S., Search for Structure: Selected Essays on Science, Art and History (MIT Press, USA, 1981).

Zou, P. F. and Bader, R. F. W., "A topological definition of a Wigner–Seitz cell and the atomic scattering factor," Acta Crystallogr., Sect. A **50**, 714–725 (1994).

Atoms and Molecules in Mirce-mechanics Approach to Functionability

Dr J. Knezevic
MIRCE Akademy, Exeter, EX5 1JJ, UK.

Abstract

Although functionability properties of machines are defined through probability characteristics, like reliability, availability and similar, the full understanding of them is only possible by observing, analysing and understanding of the physical mechanisms that generate negative functionability events. As the scientific understanding of the mechanisms that generate functionability phenomena, in Mirce-mechanics, is based on the fundamental understandings of the physical properties of atoms and molecules. The understanding and prediction of the properties of matter at the atomic level represents one of the great achievements of twentieth-century science. As matter is composed of atoms, this paper starts with its property and the manner in which the atomic elements are arranged. Electron density describes the distribution of the electronic charge throughout real space resulting from the attractive forces generated by nuclei. It is a measurable property that determines the appearance and form of matter. The theory developed to describe the behaviour of electrons, atoms and molecules differs radically from known Newtonian physics, which governs the motions of macroscopic bodies and the physical events of our everyday experiences. That new theory, which is able to account for all observable behaviour of matter, was named quantum mechanics. Thus, this paper presents the quantum theory approach to atoms in molecules, QTAIM, which is based on the revolutionary approached pioneered by Professor Richard F.W Bader (1931-2012)

1. Introduction

It is commonly accepted that reliability is defined as a probability that a system will maintain a required function during a stated period of time. As a probability cannot be seen or measured directly there seems to be a certain fundamental difficulty in understanding and interpreting statistical and probability functions in real life. This is because physical characteristics of a system like the weight, temperature, volume and similar have a clear and measurable meaning. However, the concepts of probability, and hence reliability, is an abstract property of a system that obtains a physical meaning only when behaviour of a large sample of systems is considered. Hence, understanding of reliability is reduced to the physical observation and analysis of system failures, which are observable and measurable physical characteristics.
According to the Mirce-mechanics, system failures are negative functionability events that cause transition of a system from positive to negative functionability state [1] due to some of the following reasons, or combinations of them:

a) Built-in design errors (incorrect selection of materials, stresses shapes, etc)
b) Production errors (human errors in assembly, delivery and installation tasks)
c) Irreversible changes in the condition of components with time due to wear, fatigue, creep, corrosion, and similar degradation processes
d) Imposition of external stresses resulting from collisions, harsh landings, extreme weather conditions, etc
e) Human errors in execution of maintenance tasks
f) Human errors in execution of in-service support tasks

At the MIRCE Akademy a large number of negative functionability events and associated phenomena have been observed and analysed to understand the physical mechanisms that leads to their occurrences.

Consequently, systematic studies are applied to understand phenomena that cause thermal aging, thermal buckling, photo-chemical degradation, reduction in dielectric strength, evaporation, metal fatigue, actinic degradation, photo-oxidation, swelling/ shrinking, degradation of optical qualities, fogging, photochemical decomposition of paint, blistering, warping, thermal stress, breakdown of lubrication film, increased structural loads, shift in the centre of gravity, jammed control surfaces, attenuation of energy, clutter echoes, blocking of air intakes, decreased lift and increased drag, unequal loading, removal of coating protection, pitting, roughening of the surface, acid reactions, leakage currents, promotion of mould growth, reduction of heat transfer, caking and drying, premature cracking, hot spots creation, erosion, bleaching preservatives, abrasive wear, corrosion, alkaline reactions and similar.

For years, research studies, international conferences, summer schools and other events have been organised in order to understand just a physical scale at which failure phenomena should be studied and understood. In order to understand the motion of negative functionability events it is necessary to understand the physical mechanisms of that motion. That represented a real challenge, as the answers to the question "what are physical and chemical processes that lead to the occurrence of negative functionability events" have to be provided. Without accurate answers to those questions the prediction of their future occurrences is not possible, and without ability to predict the future, the use of the word science becomes inappropriate.

After a numerous discussions, studies and trials, it has been concluded that any serious studies in this direction, from Mirce-mechanics point of view, have to be based between the following two boundaries:

- the "bottom end" of the physical world, which is at the level of the atoms and molecules that exists in the region of 10^{-10} of a metre [3],
- the "top end" of the physical world, which is at the level of the solar system that stretches in the physical scale around 10^{+10} of a metre. [4]

This range is the minimum sufficient "physical scale" which enables scientific understanding of relationships between system functionability processes and system functionability events.

As matter is composed of atoms, its property is a consequence of the manner in which the atomic elements are arranged into molecules. Consequently, the main objective of this paper is to argue that the scientific approach to functionability is the only way forward for all members of the reliability and availability community who wish to make accurate predictions that will be confirmed during the operational processes of the future systems. For that to happen scientific understanding of mechanisms that generate negative functionability phenomena is required. This paper advocates that research of this nature must start with the understanding of the properties of atoms and their bonding to form molecules, in order for negative functionability events to be understood. Then and only then, accurate and meaningful reliability and availability predictions become possible, which finally leads to the reduction of the probability of the occurrence of negative functionability events during the life of a system.

2. Electronic Structure of an Atom

The understanding and prediction of the properties of matter at the atomic level represents one of the great achievements of twentieth-century science. As matter is composed of atoms, this paper starts with its property and the manner in which the atomic elements are arranged. Electron density describes the distribution of the electronic charge throughout real space resulting from the attractive forces generated by nuclei. It is a measurable property that determines the appearance and form of matter. The theory developed to describe the behaviour of electrons, atoms and molecules differs radically from known Newtonian physics, which governs the motions of macroscopic bodies and the physical events of our everyday experiences. That new theory, which is able to account for all observable behaviour of matter, was named quantum mechanics.

The proper formulation of quantum mechanics and its application to a specific problem requires a rather elaborate mathematical framework, as do proper statements and applications of Newtonian physics. Hence, principles of quantum mechanics and its basic concepts are used for the studies of the motion and relationship between atoms in molecules. Thus, the physical laws governing the behaviour of electrons and their arrangements, when bound to nuclei, to form atoms and molecules have been discovered, and termed the electronic structure of the atom or molecule. Furthermore, understanding of the relationship between the electronic structure of an atom and its physical properties enables understanding of the change of electronic structure during a chemical reaction, where only the number and arrangement of the electrons are changed while the nucleus remaining unaltered. Thus, the unchanging charge of the atomic nucleus is responsible for retaining the atom's chemical identity through any chemical reaction. Therefore, for the purpose of understanding the chemical properties and behaviour of atoms, the nucleus may be regarded as simply a point charge of constant magnitude for a given element, giving rise to a central field of force that binds the electrons to the atom. [2]

3. Atomic Phenomena

Rutherford's nuclear model for the atom set the stage for the understanding of the structure of atoms and the forces holding them together. From Rutherford's alpha-scattering experiments it was clear that the atom consisted of a positively-charged nucleus with negatively-charged electrons arranged in some fashion around it, the electrons occupying a volume of space many times larger than that occupied by the nucleus. The diameters of nuclei fall in the range of $l\times10^{-14}$ to 1×10^{-15} m, while the diameter of an atom is typically of the order of magnitude of 1×10^{-10} m. The forces responsible for binding the atom, and in fact all matter apart from the nuclei themselves, are electrostatic in origin: the positively charged nucleus attracts the negatively charged electrons. There are attendant magnetic forces that arise from the motions of the charged particles. These magnetic forces give rise to many important physical phenomena, but they are smaller in magnitude than are the electrostatic forces and they are not responsible for the binding found in matter.

4. Single Atom Structure

When a stable atom is formed, the electron is attracted to the nucleus, and as the radius is less than infinity, the energy has a negative sign, which implies that it must be supplied to the system if the electron is to overcome the attractive force of the nucleus and escape from the atom.

The motion of the electron is not free. The electron is bound to the atom by the attractive force of the nucleus and consequently quantum mechanics predicts that the total energy of the electron is quantised and equal to $E_n = \left(-2\pi^2 me^4 Z^2\right)/\left(n^2 h^2\right)$, $n = 1, 2, 3, ...$, where m is the mass of the electron, e is the magnitude of the electronic charge, n is a quantum number, h is Planck's constant and Z is the atomic number, which is the number of positive charges in the nucleus.

Since the motion of the electron occurs in three dimensions it is correct to anticipate three quantum numbers for the hydrogen atom. However, as the energy depends only on the quantum number n it is called the principal quantum number. When n is equal to infinity and energy is equal to zero the electron is free of the attractive force of the nucleus. The average distance between the nucleus and the electron increases as the energy or the value of n increases. Hence, energy must be supplied to pull the electron away from the nucleus.

4.1 The Probability Distributions for the Hydrogen Atom

To what extent is possible to pinpoint the position of an electron when it is bound to an atom? An order of magnitude for the answer to this question could be obtained by applying the uncertainty principle $\Delta x \Delta p = h$ to estimate Δx. The value of Δx represents the minimum uncertainty in our knowledge of the position of the electron. The momentum of an electron in an atom is of the order of magnitude of 9×10^{-19} g cm/sec. The uncertainty in the momentum Δp must necessarily be of the same order of magnitude. According to Bader [2] the answer is $\Delta x = 7 \times 10^{-27} / 9 \times 10^{-19} \approx 10^{-8}$ cm.

The uncertainty in the position of the electron is of the same order of magnitude as the diameter of the atom itself. As long as the electron is bound to the atom, it is not possible to say much more about its position than that it is in the atom. This fact invalidated all models of the atom that describe the electron, as a particle following a definite trajectory or orbit.

Energy and one or more wave functions could be obtained for every value of n, the principal quantum number, by solving Schrödinger's equation for the hydrogen atom. Knowledge of the wave functions, or probability amplitudes, allows calculation of the probability distributions for the electron in any given quantum level. When n=1, the wave function and the derived probability function are independent of direction and depend only on the distance between the electron and the nucleus.

An experiment designed to detect the position of the electron with an uncertainty much less than the diameter of the atom itself, when repeated a large number of times, shown that the electron is detected close to the nucleus most frequently and the probability of observing it at some distance from the nucleus decreases rapidly with increasing radial distance [2]. The atom was ionised in making each of these observations because the energy of the photons with a wavelength much less than 10^{-8} cm is greater than the amount of energy required to ionise the hydrogen atom. If light with a wavelength comparable to the diameter of the atom was used in the experiment, then the electron would not have been excited and the knowledge of its position would have been correspondingly less precise.

When the electron is in a definite energy level, the electron density distributions describe the manner in which the total electronic charge is distributed in space. The electron density is

expressed in terms of the number of electronic charges per unit volume of space, e-/V. The volume *V* is usually expressed in atomic units of length cubed, and one atomic unit of electron density is then e^-/a_0^3. To give an idea of the order of magnitude of an atomic density unit, 1 au of charge density $e^-/a_0^3 = 6.7$ electronic charges per 10^{-8} cm^3. That is, a cube with a length of 0.52917×10^{-8} cm, if uniformly filled with an electronic charge density of 1 au, would contain 6.7 electronic charges.

The important point of the above discussion is that both the angular momentum and the energy of an atom remain constant if the atom is left undisturbed. Any physical quantity that is constant in a classical system is both conserved and quantised in a quantum mechanical system. Thus both the energy and the angular momentum are quantised for an atom.

Since an electron may exhibit a magnetic moment even when it does not possess orbital angular momentum, it must possess some internal motion. This motion is known as the electron spin and it is treated in quantum mechanics as another kind of angular momentum. Experimentally, however, all that is known is that the electron possesses an intrinsic magnetic moment. The remarkable feature of this intrinsic magnetic moment is that its magnitude and the number of components along a given axis are fixed. A given electron may exhibit only one of two possible components; it may be aligned with the field or against it. Hence only one quantum number is required to describe completely the spin properties of a single electron.

Finally, a total of four quantum numbers is required to specify completely the state of an electron when it is bound to an atom. The quantum numbers *n*, *l* and *m* determine its energy, orbital angular momentum and its component of orbital angular momentum. The fourth quantum number, the spin quantum number, summarises all that can be known about the spin angular momentum of the electron.

5. Pauli Exclusion Principle

The study of the magnetic properties of the ground and excited states of helium is sufficient to point out a general principle. For the ground state of helium, in which both electrons are in the same atomic orbital, only the non-magnetic form exists. This would imply that when two electrons are in the same atomic orbital their spins must be paired, that is, one up and one down. This is an experimental fact because helium is never found to be magnetic when it is in its electronic ground state. When the electrons are in different orbitals, then it is again an experimental fact that their spins may now be either paired or unpaired. This lead to the creation of the Pauli Exclusion Principle that states: no two electrons in the same atom may have all four quantum numbers the same. The Pauli principle cannot be derived from, nor is it predicted by, quantum mechanics. It is a law of nature that must be taken into account along with quantum mechanics if the properties of matter are to be correctly described. The concept of atomic orbitals, as derived from quantum mechanics, together with the Pauli Exclusion Principle that limits the occupation of a given orbital, provides an understanding of the electronic structure of many-electron atoms. [2]

The concept of atomic orbitals in conjunction with the Pauli principle has indeed predicted a periodicity in the electronic structures of the elements. The form of this periodicity replicates exactly that one found in the Mendeleev's periodic table of the elements in which the periodicity is founded on the observed chemical and physical properties of the elements.

The diameter of an atom is difficult to define precisely as the density distribution tails off at large distances. However, there is a limit as to how close two atoms can be pushed together in a solid material. The size of the atom in general decreases as the number of electrons in the quantum shell increases. This observation, which at first sight might appear surprising, finds a ready explanation through the concept of an effective nuclear charge.

The electric field and hence the attractive force exerted by the nucleus on an electron in the outer quantum shell is reduced because of the screening effect of the other electrons which are present in the atom. An outer electron does not penetrate to any great extent the tightly bound density distribution of the inner shell electrons. Hence, each inner electron, which is an electron with an *n* value less than the *n* value of the electron in question, reduces the value of the nuclear charge experienced by the outer electron by almost one unit. The remaining outer electrons on the other hand are, on the average, all at the same distance away from the nucleus as is the electron under consideration. Consequently each outer electron screens considerably less than one nuclear charge from the other outer electrons. Thus the higher the ratio of outer shell to inner shell electrons, the larger is the "effective nuclear charge" which is experienced by an electron in the outer shell. [2]

6. Chemical Implications of Effective Nuclear Charge

The effective nuclear charge is a minimum for the group I elements in any given row of the periodic table. Therefore, it requires less energy to remove an outer electron from one of these elements than from any other element in the periodic table. The strong reducing ability of these elements is readily accounted for. The variation in the relative reducing power of the elements across a given period or within a given group is determined by the variation in the effective nuclear charge. The ability of the elements in a given row of the periodic table to act as reducing agents should undergo a continuous decrease from group I to group VII, since the effective nuclear charge increases across a given row. Similarly, the reducing ability should increase down a given column (group) in the table since the effective nuclear charge decreases as the principal quantum number is increased. Anticipating the fact that electrons can be transferred from one atom (the reducing agent) to another (the oxidizing agent) during a chemical reaction, it is expected that the elements to the left of the periodic table to exhibit a strong tendency to form positively charged ions.

The ability of the elements to act as oxidising agents should parallel directly the variations in the effective nuclear charge. Thus the oxidising ability should increase across a given row (from group I to group VII) and decreases down a given family. These trends are, of course, just the opposite of those noted for the reducing ability. The reducing ability should vary inversely with the ionisation potential, and the oxidising ability should vary directly with the electron affinity. The elements in groups VI and VII should exhibit a strong tendency for accepting electrons in chemical reactions to form negatively charged ions. For example, Francium, which possesses a single outer electron in the 7*s* orbital, is the strongest chemical reducing agent and fluorine, with an orbital vacancy in the 2*p* subshell is the strongest oxidizing agent. These are only a few examples of how knowledge of the electronic structure of atoms may be used to understand and correlate a large amount of chemical information that certainly has significant impact of the failure mechanisms affecting the reliability of systems and their components. [2]

It should be pointed out that chemistry is a study of very complex interactions and the few simple concepts advanced here cannot begin to account for the incredible variety of phenomena actually observed.

7. The Chemical Bond

With understanding of the electronic structure of atoms, only briefly summarised above, it is possible to understand the existence of molecules. Clearly, the force that binds the atoms together to form a molecule is, as in the atomic case, the electrostatic force of attraction between the nuclei and electrons. In a molecule, however, a force of repulsion between the nuclei in addition to that between the electrons is encounter. To account for the existence of molecules it is necessary to account for the predominance of the attractive interactions, which could be shown in terms of the energy of a molecule, relative to the energies of the constituent atoms, and in terms of the forces acting on the nuclei in a molecule.

In order to determine what attractive and repulsive interactions are possible in a molecule, an instantaneous configuration of the nuclei and electrons in a hydrogen molecule is considered, as shown in Figure 1.When the two atoms are initially far apart, the distance *R* is very large, the only potential interactions are the attraction of nucleus A for electron number (1) and the attraction of nucleus B for electron number (2). When *R* is comparable to the diameter of an atom (A and B are close enough to form a molecule) then new interactions appear. Nucleus A now attracts electron (2) as well as (1) and similarly nucleus B attracts electron (1) as well as (2). The dashed lines represent the repulsive interactions between like charges and the solid lines indicate the attractive interactions between opposite charges.

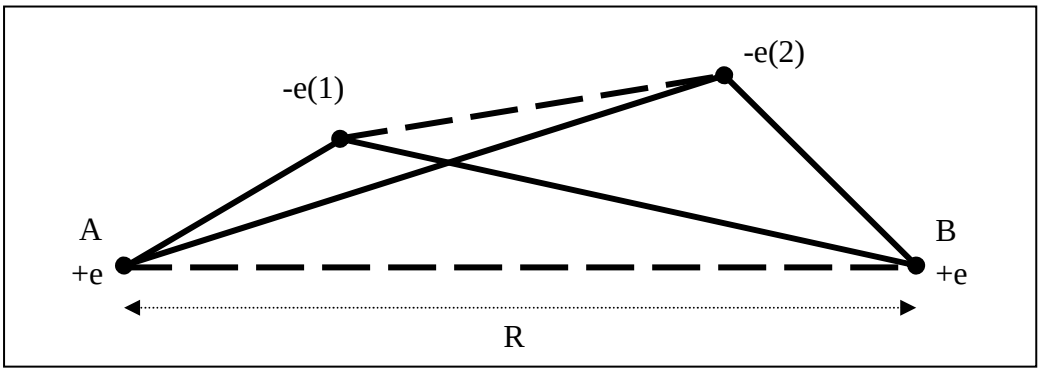

Figure 1: One possible set of the instantaneous relative positions of the Electrons and nuclei in a hydrogen molecule. [2]

The number of attractive interactions has been doubled from what it was when the atoms were far apart. However, the reduction in *R* introduces two repulsive interactions as well, namely the two electrons now repel one another as do the two nuclei. If the two atoms are to remain together to form a molecule, the attractive interactions must exceed the repulsive ones. It is clear from Figure 1 that the new attractive interactions, nucleus A attracting electron (2) and nucleus B attracting electron (1), is large only if there is a high probability of both electrons being found in the region between the nuclei. When the average potential energy is calculated by quantum mechanics, the attractive interactions are found to predominate over the repulsive ones because quantum mechanics does indeed predict a high probability for each electron being in the region between the nuclei. This general consideration of the energy demonstrates that electron density must be concentrated between the nuclei if a stable molecule is to be formed, for only in this way can the attractive interaction be maximised.

In the atomic case it is possible to fix the position of the nucleus in space and consider only the motion of the electrons relative to the nucleus. However, in molecules, the nuclei may also change positions relative to one another. This additional movement can be neglected as the nuclei are very massive compared to the electrons and their average velocities are consequently much smaller than those possessed by the electrons. In a classical picture of the molecule we would see a slow, lumbering motion of the nuclei accompanied by a very rapid motion of the electrons. The physical implication of this large disparity in the two sets of velocities is that the electrons can immediately adjust to any change in the position of the nuclei. The positions of the nuclei determine the potential field in which the electrons move. However, as the nuclei change their positions and hence the potential field, the electrons can immediately adjust to the new positions. Thus the motion of the electrons is determined by where the nuclei are but not by how fast the nuclei are moving.

For a given distance between the nuclei it is possible to determine the energy, the wave function and the electron density distribution of the electrons, the nuclei being held in fixed positions. Then the distance between the nuclei is changed to a new value, and the calculation of the energy, wave function and electron density distribution of the electrons is performed again. This process, repeated for every possible internuclear distance, allows understanding of how the energy of the electrons changes as the distance between the nuclei is changed. However, for the purpose on this analysis only the motion of the electrons and hold the nuclei stationary at some particular value for the internuclear distance R, could be considered.

The energy of the electrons in a molecule is quantised, as it is in atoms. When the nuclei are held stationary at some fixed value of *R*, there are a number of allowed energy levels for the electrons. There are, however, no simple expressions for the energy levels of a molecule in terms of a set of quantum numbers, such as was the case with the hydrogen atom. As in the case of atoms, there is a wave function that governs the motion of all the electrons for each of the allowed energy levels. Each wave function again determines the manner in which the electronic charge is distributed in three-dimensional space.

7.1 An Electrostatic Interpretation of the Chemical Bond

In the light of the above discussion of a molecular electron density distribution, a molecule may be regarded as two or more nuclei imbedded in a rigid three-dimensional distribution of negative charge. There is a theorem of quantum mechanics that states that the force acting on a nucleus in a molecule may be determined by the methods of classical electrostatics. The nuclei in a molecule repel one another, since they are of like charge. This repulsive force, if unbalanced, would push the nuclei apart and the molecule would separate into atoms. In a stable molecule, however, an attractive force exerted by the negatively charged electron density distribution balances the nuclear force of repulsion. The usefulness of this approach lies in the fact that the stability of molecules in terms of the classical concept of a balance between the electrostatic forces of attraction and repulsion could be considered.

A chemical bond is thus the result of the accumulation of negative charge density in the region between the nuclei to an extent sufficient to balance the nuclear forces of repulsion. [2] This corresponds to a state of electrostatic equilibrium, as the net force acting on each nucleus, is zero for this one particular value of the internuclear distance. If the distance between the nuclei is increased from the equilibrium value, the nuclear force of repulsion is

decreased. At the same time the force of attraction exerted by the electron density distribution is increased as the binding region is increased in size. Thus, when the radial distance is increased from its equilibrium value there are net forces of attraction acting on the nuclei that pull the two nuclei together again. A definite force would have to be applied to overcome the force of attraction exerted by the electron density distribution and separate the molecule into atoms. Similarly, if the value of radial distance is decreased from its equilibrium value, the force of nuclear repulsion is increased over its equilibrium value. At the same time, the attractive force exerted by the electron density is decreased, because the binding region is decreased in size. In this case there is a net force of repulsion pushing the two nuclei apart and back to their equilibrium separation. There is thus one value of radial distance for which the forces on the nuclei are zero and the whole molecule is in a state of electrostatic equilibrium. [2]

This is an important result as it shows that the density distribution in a molecule cannot be considered as the simple sum of the two atomic charge densities. The overlap of rigid atomic densities does not place sufficient charge density in the binding region to overcome the nuclear force of repulsion. Hence it is reasonable to conclude that the original atomic charge distributions must be distorted in the formation of a molecule, and the distortion is such that charge density is concentrated in the binding region between the nuclei. A quantum mechanical calculation predicts this very result. The calculation shows that there is a continuous distortion of the original atomic density distributions, a distortion that increases as the internuclear distance decreases.

The changes in the original atomic density distributions caused by the formation of the chemical bond may be isolated and studied directly by the construction of a density difference distribution. Such a distribution is obtained by subtracting the density obtained from the overlap of the undistorted atomic densities separated by a radial distance, from the molecular charge distribution evaluated at the same value. Wherever this density difference is positive in value it means that the electron density in the molecule is greater than that obtained from the simple overlap of the original atomic densities. Where the density difference is negative, it means that there is less density at this point in space in the molecule than in the distribution obtained from the overlap of the original atomic distributions. Such a density difference map thus provides a detailed picture of the net reorganisation of the charge density of the separated atoms accompanying the formation of a molecule. This just proves that the density distribution resulting from the overlap of the undistorted atomic densities does not place sufficient charge density in the binding region to balance the forces of nuclear repulsion. The regions of charge increase in the density difference maps are, therefore, the regions to which charge is transferred relative to the separated atoms to obtain a state of electrostatic equilibrium and hence a chemical bond. From this point of view a density difference map provides a picture of the "bond density."

7.2 The Effect of the Pauli Principle on Chemical Binding

The Pauli Exclusion Principle plays as important a role in the understanding of the electronic structure of molecules as it does in the case of atoms. The end result of the Pauli principle is to limit the amount of electronic charge density that can be placed at any one point in space. For example, the Pauli principle prevents the 1*s* orbital in an atom from containing more than two electrons. Since the 1*s* orbital places most of its charge density in regions close to the nucleus, the Pauli principle, by limiting the occupation of the 1*s* orbital, limits the amount of

density close to the nucleus. Any remaining electrons must be placed in orbitals that concentrate their charge density further from the nucleus. [2]

It is proven that the reason the electron doesn't fall onto the nucleus is because it must possess kinetic energy if Heisenberg's uncertainty principle is not to be violated. This is one reason why matter doesn't collapse. The Pauli principle is equally important in this regard. The electron density of the outer electrons in an atom cannot collapse and move closer to the nucleus since it can do so only if the electrons occupy an orbital with a lower *n* value. If, however, the inner orbital contains two electrons, then the Pauli principle states that the collapse cannot occur. The Pauli principle demands that when two electrons are placed in the same orbital their spins must be paired. What restriction is placed on the spins of the electrons during the formation of a molecule, when two orbitals, each on a different atom, overlap one another? To address this question a hydrogen molecule that consists of two hydrogen atoms is considered, where atom A has the configuration $1s^1$ and atom B has the configuration $1s^1$. Even when the atoms approach very close to one another the Pauli principle would be satisfied as the spins of the two electrons are opposed. This is the situation that has been assumed in all discussions of the hydrogen molecule.

However, what would occur if two hydrogen atoms approached one another and both had the same configuration and spin, say $1s^1$? When two atoms are relatively close together the electrons become indistinguishable. It is no longer possible to say which electron is associated with which atom as both electrons move in the vicinity of both nuclei. Indeed this is the effect which gives rise to the chemical bond. In so far as the region around each atom to be governed by its own atomic orbital can be considered, distorted as it may be, two electrons with the same spin are not be able to concentrate their density in the binding region. This region is common to the orbitals on both atoms, and since the electrons possess the same spin they cannot both be there simultaneously. In the region of greatest overlap of the orbitals, the binding region, the presence of one electron tends to exclude the presence of the other if their spins are parallel. Hence, instead of density accumulating in the binding region as two atoms approach, electron density is removed from this region and placed in the antibinding region behind each nucleus where the overlap of the orbitals is much smaller. Thus, the approach of two hydrogen atoms with parallel spins does not result in the formation of a stable molecule. This repulsive state of the hydrogen molecule, in which both electrons have the same spin and atomic orbital quantum numbers, can be detected spectroscopically. [2]

Consequently, the general requirements for the formation of a chemical bond can be formulated. Electron density must be accumulated in the region between the nuclei to an extent greater than that obtained by allowing the original atomic density distributions to overlap. In general, the increase in charge density necessary to balance the nuclear force of repulsion requires the presence of two electrons.

In the atomic orbital approximation we picture the bond as resulting from the overlap of two distorted atomic orbitals, one centred on each nucleus. When the orbitals overlap, both electrons may move in the field of either nuclear charge as the electrons may now exchange orbitals. Finally, the pair of electrons must possess opposed spins. When their spins are parallel, the charge density from each electron is accumulated in the antibinding rather than in the binding region.

7.3 Classification of Chemical Bonds

To make a quantitative assessment of the type of binding present in a particular molecule it is necessary to have a measure of the extent of charge transfer present in the molecule relative to the charge distributions of the separated atoms. This information is contained in the density difference or bond density distribution, the distribution obtained by subtracting the atomic densities from the molecular charge distribution. Such a distribution provides a detailed measure of the net reorganisation of the charge densities of the separated atoms accompanying the formation of the molecule.

The density distribution resulting from the overlap of the undistorted atomic densities (the distribution which is subtracted from the molecular distribution) does not place sufficient charge density in the binding region to balance the nuclear forces of repulsion. The regions of charge increase in a bond density map are, therefore, the regions to which charge is transferred relative to the separated atoms to obtain a state of electrostatic equilibrium and hence a chemical bond. Thus, it is possible to use the location of this charge increase relative to the positions of the nuclei to characterise the bond and to obtain an explanation for its electrostatic stability.

A bond is classified as covalent when the bond density distribution indicates that the charge increase responsible for the binding of the nuclei is shared by both nuclei. It is not necessary for covalent binding that the density increase in the binding region be shared equally. It is possible to encounter molecules with different nuclei, in which the net force binding the nuclei is exerted by a density increase that, while shared, is not shared equally between the two nuclei. [2]

The charge distribution of a molecule with an ionic bond is characterised not only by the transfer of electronic charge from one atom to another, but also by a polarisation of each of the resulting ions in a direction counter to the transfer of charge. [2]

In a covalent bond the increase in charge density that binds both nuclei is shared between them, while in an ionic bond the forces exerted by the charge density localised on a single nucleus bind both nuclei. It must be stressed that there is no fundamental difference between the forces responsible for a covalent or an ionic bond, as they are electrostatic in both cases.

7.4 Interaction Between Molecules

The properties of matter observed on the macroscopic level are determined by the properties of the constituent molecules and the interactions between them. The polar or non-polar character of a molecule is clearly important in determining the nature of its interactions with other molecules. There are relatively strong forces of attraction acting between molecules with large dipole moments. To a first approximation, the energy of interaction between dipolar molecules can be considered as completely electrostatic in origin, the negative end of one molecule attracting the positive end of another.

The presence of intermolecular forces accounts for the existence of solids and liquids. A molecule in a condensed phase is in a region of low potential energy, as a result of the attractive forces that the neighbouring molecules exert on it. By supplying energy in the form of heat, a molecule in a solid or liquid phase can acquire sufficient kinetic energy to overcome the potential energy of attraction and escape into the vapour phase. The pressure of the vapour in equilibrium with a solid or liquid, at a given temperature, provides a measure of

the tendency of a molecule in a condensed phase to escape into the vapour (the larger the vapour pressure, the greater the escaping tendency). The average kinetic energy of the molecule in the vapour is directly proportional to the absolute temperature. Thus the observation of a large vapour pressure at a low temperature implies that relatively little kinetic energy is required to overcome the potential interactions between the molecules in the condensed phase.

7.5 Polyatomic Molecules

The concept of a molecular orbital is readily extended to provide a description of the electronic structure of a polyatomic molecule. Indeed molecular orbital theory forms the basis for most of the quantitative theoretical investigations of the properties of large molecules.

In general a molecular orbital in a polyatomic system extends over all the nuclei in a molecule are essential, if the spatial properties of the orbitals are to be understood and predicted. An analysis of the molecular orbitals for the water molecule provides a good introduction to the way in which the symmetry of a molecule determines the forms of the molecular orbitals in a polyatomic system.

8. Conclusion

The main objective of this paper was to present Mirce-mechanics approach to Reliability, one that is based on the laws of science. This approach is in direct agreement with the observed with occurrence of negative functionability phenomena resulting from physical processes like corrosion, fatigue, creep, wear and similar.

Finally, it is essential to distinguish the scientific formulation of the motion of functionability through the life of a system, contained in Mirce-mechanics and presented in this paper, from the best “industrial practices” approach that is based on reliability models of systems that are created to demonstrate the contractual compliance of the legally binding acquisition processes. As science is the proved model of reality that is confirmed through observation, the summary message of this paper to reliability professionals is to move from the universe in which the laws of science are suspended to the universe that is based on the laws of science that govern behaviour of atoms in molecules of the matter, in order for their predictions to become future realities.

9. Acknowledgment

This paper is dedicated to the memories of Richard F.W. Bader, Emeritus Professor of McMaster University in Canada, (1931-2012). He revolutionised chemistry by introducing quantum mechanical approach its physical understanding that led to the creation of the atoms in molecules theory [6]. Most material in this paper is based on the knowledge that Richard has accumulated during an extremely successful career and which he shared so enthusiastically with all of those who were able to embrace it. Professor Bader was a Grand Fellow of the MIRCE Akademy and during the last two years of his life he has actively supported the development of Mirce-mechanics.

10. References

[1] Knezevic J., Reliability, Maintainability and Supportability – A Probabilistic Approach, with Probchar Software. pp. 292, McGraw Hill, UK, 1993.

[2] Bader, R.F.W., An Introduction to the Electronic Structure of Atoms and Molecules, pp 240, Clark Irvin, 1970.

[3] Knezevic, J., Physical Scale of Mirce-mechanics, Lecture Notice, Master Diploma Programme, MIRCE Akademy, Woodbury Park, Exeter, UK, 2009.

[4] Knezevic, J., **Functionability in Motion**, Proceedings 10th International Conference on Dependability and Quality, DQM Institute, 2010, Belgrade, Serbia.

[5] Knezevic, J., **Scientific Scale of Reliability**, Proceedings of International Conference on Reliability, Safety and Hazard, Bhabha Atomic Research Centre, 2010, Mumbai, India.

[6] Bader, R.F.W., Atoms in Molecules: a Quantum Theory, Oxford University Press, Oxford, UK, 1990.

Physics-of-Failure based Reliability Engineering

Elviz George and Michael Pecht

Center for Advanced Life Cycle Engineering (CALCE)
University of Maryland
College Park MD 20742 USA

Abstract

With increased use, complexity, and miniaturization of electronics in various applications, there is a need to improve the reliability of electronic products swiftly and cost-effectively. The shift from the use of traditional constant failure rate approaches to estimate the reliability of electronic products was enabled by physics-of-failure based reliability approaches. In a physics-of-failure based reliability assessment, the failure mechanisms that degrade products and ultimately produce failures are identified, based on the hardware configuration and anticipated life-cycle profile. Physics-of-failure models associated with specific failure mechanisms are utilized to provide a statistical distribution of the time-to-failure for a particular failure mechanism and site. Physics-of-failure based reliability assessment is integrated into the product development process to aid in design-for-reliability, stress test selection, product qualification, product screening, and prognostics and health management. This paper describes various steps involved in the physics-of-failure based reliability of electronic products and its integration into the product development process.

1. Introduction

Reliability is the ability of a product to perform its intended function within specified performance limits without failure, for a specified time in its life cycle environment. Development of a reliable product is a consequence of conscious, systematic, and rigorous efforts conducted throughout the entire life cycle of the product—from design to scrap. Early approaches to reliability prediction relied on statistical analysis of field data, which was characterized by a wide spectrum of applications and therefore large variation in environments. Historically, reliability modeling and prediction can be divided into two eras. The first era, lasting from World War II until the 1980s [1], was based on the exponential or constant failure rate model [2][3]. A drawback to this approach is that there is no physical justification for using a constant rate failure model where the failure distribution of random events is mathematically represented by a bottom-up statistical method [4]. With the introduction of integrated circuits during the 1980s, evidence suggested that the constant failure rate model was no longer applicable [1]. Based on their work to provide guidelines to update MIL-HDBK-217 in 1987-1990, Center for Advanced Life Cycle Engineering suggested that the constant failure rate model should not be used [5]. Responding to the increasing criticism of the constant failure rate model, Secretary of Defense William Perry issued a memorandum to eliminate the use of the MIL-HDBK-217 in 1994 [1][6]. This event marked the beginning of the second era, in which the physics-of-failure (PoF) approach dominated reliability modeling. In the PoF approach, the root causes of individual failure mechanisms are studied and lifetimes are predicted for each failure mechanism.

Electronics are rapidly becoming smaller and more complex. New failure modes are in the rise due to the higher stress-strength ratios associated with miniaturization and due to the advent of new materials and manufacturing processes [7]. Therefore, new reliability methodologies are needed to accurately assess the reliability of components during the development process. Debates [8][9][10][11] questioning the appropriateness of reliability modeling procedures based on chance failures contributed to the need for an approach that overcomes the pitfalls of the constant failure rate approach. A quantitative understanding and modeling of the relevant failure mechanisms guides product design, manufacturing and testing.

Reliability is critical in the competitive global business environment, where time-to-market is a key metric for success. As a result, companies must ensure product reliability, but at the same time they cannot afford to design and test their products' reliability over a long period of time due to time-to-market pressure. Reliability prediction methodologies can provide a quick estimate of the reliability of products. PoF is the most promising approach among the reliability prediction methods discussed in the literature [12][2][13][14][15][16].

Quantitative and qualitative comparisons between the PoF and constant failure rate models have been published [7]. Mortin [17] compared the failure rates in a dual in-line package using a constant failure rate model and failure rate specific electromigration model. Results showed that the constant failure rate model overestimated or underestimated the true instantaneous hazard rate. This inaccuracy may result in a significant cost impact on logistics support and maintenance requirements. Mortin recommended the shift from constant failure rate to instantaneous failure rate based on root-cause failure mechanisms to improve reliability assessment. Analysis of test and field data by Weil [18] showed that the results predicted for plastic encapsulated microcircuits (PEMs) using constant failure rate (based on MIL-HDBK-217) was misleading and inaccurate. Wei recommended the use of experimental data and methods instead of MIL-HDBK-217. Cushing [19] compared the constant failure rate and PoF approaches from model development to final cost analysis. Cushing showed that PoF approach proactively incorporated reliability into the design by establishing a scientific basis for evaluating new materials and technologies. Foucher et al. [4] assessed the ability of bottom-up statistical methods, top-down similarity analysis methods, and bottom-up physics-of-failure methods to estimate whether a reliability requirement is achievable, achieve a reliable manufacturing process, assess potential warranty risks, and provide inputs to safety analysis. Foucher et al. recommended the combined use of statistical and PoF methods for accurate prediction of lifetime. Matic and Sruk [7] showed that PoF-based approach is better than constant failure rate approaches to improve the reliability of a product during the design stage.

This paper discusses various steps involved in the PoF-based reliability assessment of electronic products. An introduction to the PoF-based approach along with a step-by-step discussion on how this methodology is conducted is provided in section 2. Examples of the incorporation of PoF-based reliability assessment methods into the product development process are also listed in section 2. Examples of application of PoF approach are provided in section 3.

2. Physics-of-failure based reliability assessment

PoF enables the design and development of reliable products based on knowledge of the failure mechanisms [20]. PoF can be used in reliability engineering applications including design-for-reliability (DfR), reliability prediction, test planning, and prognostics. To be used in the design and test of a product, PoF requires that the product and its anticipated life-

cycle loading profile be sufficiently defined to identify potential failure sites and failure mechanisms. Failure mechanisms are the processes by which physical, chemical, mechanical, or any other process induces product failures [23].

Numerical and/or analytical models that are based on the failure mechanisms are used to make PoF-based reliability predictions. Physics-of-failure based models, commonly referred to as PoF models, provide a myriad of stress-time relationships that describe the failure mechanisms [20][21]. Lee et al. [24] has provided a comprehensive review of the existing PoF models for solder interconnects. PoF can be broadly classified as strain-based, energy based or damage-based. A general procedure to choose the appropriate PoF model based on the package type and environmental conditions was also illustrated. In general, inputs to these failure models include the specific product geometry, material information, and stress information. The stress information needs to include stress levels, as well as the duration or frequency of the application of stress. The time-to-failure predicted by a failure model generally represents the time to a specific percentage of failure, depending on how a model is developed and validated. Since the inputs to all the models have known or expected levels of uncertainty, simulation of these uncertainties allows the development of a series of possible times-to-failure and the statistical distribution that represents the failure probability over time. Using these distribution parameters, a confidence interval can be associated with the estimated times-to-failure and other reliability parameters [22].

The process of reliability assessment can be divided into four steps, as discussed by Snook [26]: 1) understand the application environmental extremes, life cycle environment, and life/reliability requirements; 2) ensure that the assembly can function under the application environment; 3) ensure that the component construction and assembly technologies have the capability to meet the life requirement subjected to the life cycle environment, and 4) assess the probability of failure within the desired life, and review design/manufacturing methods to minimize this probability. The process for PoF-based reliability assessment is shown in Figure 1. In the PoF-based reliability assessment, the failure sites and mechanisms are identified based on the product design and life cycle loading conditions.

The potential failure mechanisms are then ranked based on the probability of the occurrence, detection, and severity. PoF models are utilized to predict the time to failure, followed by verification of the time to failure by physical testing. Physical testing also verifies the failure sites and mechanisms thereby providing confidence on the PoF model and methodology. Identification of the critical parameters that influence failures allows reliability engineers to modify the design and use conditions to improve product reliability.

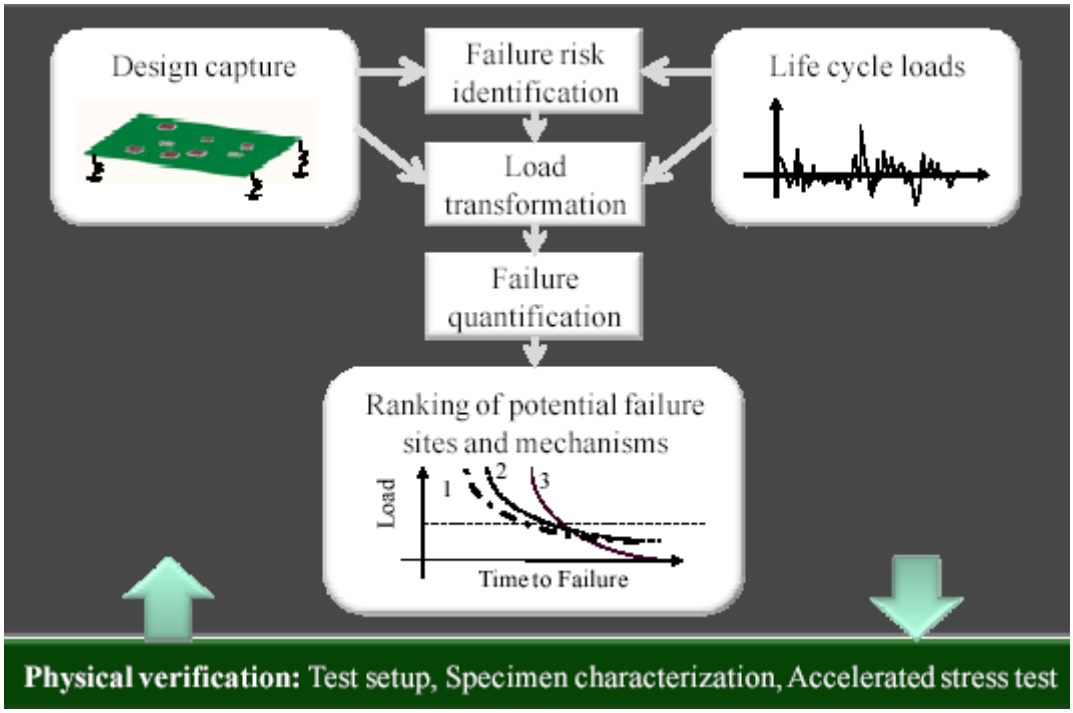

Figure 1: PoF-based reliability assessment

A PoF methodology for reliability engineering can be described in the flowchart shown in Figure 2. The inputs to the PoF methodology include hardware configuration, life-cycle profile, and the appropriate PoF models, as described in Section 2.1. Based on the inputs, a failure modes, mechanisms, and effects analysis (FMMEA) and stress analysis is carried out, as shown in Section 2.2. The PoF analysis gives the probability of failure, failure time and the corresponding sensitivity of design parameters. The outputs from the PoF methodology can be utilized to improve the design, provide virtual qualification, and so on, as described in Section 3.

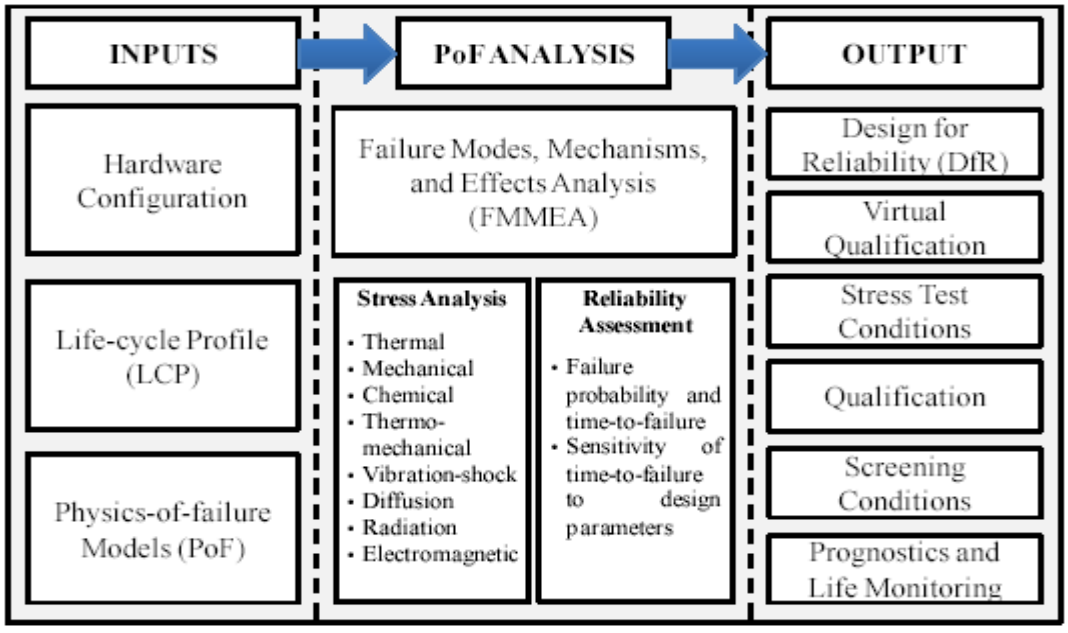

Figure 2: Physics-of-failure methodology.

2.1 Inputs

To conduct a PoF-based reliability assessment, the design and life-cycle profile (LCP) of the product under analysis must be defined. The inputs to the PoF-based reliability assessment include definition of hardware configuration and life cycle profile condition. The PoF model corresponding to a specific failure mechanism should also be known to predict the time to failure.

2.1.1 Hardware configuration

As discussed by Tilgner [27], the design of a product identifies the materials, components, manufacturing processes, and connectivity (including the physical and functional relations between the subassemblies) of the product. The hardware configuration of the product describes the design of the components and subassemblies and the product architecture. It may also include the effects of the manufacturing processes on the final product in the form of tolerances on the dimensions and material properties.

For electronic products, designs will likely include electronic components, connectors, and enclosures. An example of an electronic component is a plastic encapsulated semiconductor chip, which provides signal communication paths to the subassemblies. Electronic components are generally mounted on a printed circuit board that includes the laminate materials, layer stacks, metallization and plating materials and thicknesses [27]. Examples of connectors include solder joints or wire bonds.

The stresses at potential failure sites within the product due to external and internal loads are dependent on the materials used to construct a product. To determine the effect of material type on the stresses, the physical properties of material used are needed as inputs to PoF-based failure models. For example, a failure in a solder joint may be driven by stresses arising from repeated temperature excursions through a fatigue failure mechanism. In this case, the coefficients of thermal expansion of the materials are needed to determine the cyclic shear strain range of the solder joint due to thermo-mechanical fatigue.

During manufacturing, a product is subjected to multiple processes that apply stresses on materials, resulting in residual stresses in the final product. The manufacturing process may even modify some of the materials' properties. For example, the reflow and wave soldering processes can change the thermo-physical properties of a printed circuit board. Material property inputs to PoF-based assessments must reflect the properties and their tolerances after completion of the manufacturing process.

2.1.2 Life-cycle profile (LCP)

Environmental and operational loads applied to a product during its life cycle cause a product to degrade and fail. In PoF-based reliability assessment, every load that can degrade and cause a product to fail product should be characterized. A product may experience loads at different stages of its life cycle, such as manufacturing, assembly, testing, rework, storage, transportation, handling, operation, repair, and maintenance. The types of loads experienced include thermal (e.g., steady-state temperature, temperature ranges, temperature cycles, temperature gradients), mechanical (e.g., pressure levels, pressure gradients, vibrations, shock loads, acoustic levels), chemical (e.g., aggressive or inert environments, ozone, pollution humidity levels, contamination, fuel spills), physical (e.g., radiation, electromagnetic interference, high or low air pressure), and operational loading conditions (e.g., power, power surge, heat dissipation, current, voltage spikes). These loads, in various

combinations, can influence the reliability of a product. The extent and rate of product degradation depend on the nature, magnitude, and duration of exposure to such loads [28][29].

Definition and characterization of the life cycle conditions should take into account the nature of user pattern. For example, desktop computers are typically designed for office or home environments. However, the operational profile of each unit will depend on user behavior. Some users may shut down their computers after each use, others may shut it down only once at the end of the day, and other users may keep their computers powered on all the time. Thus, the temperature profile experienced by each product, and hence, its degradation due to thermal loads, will be different.

A life-cycle profile (LCP) is a time history of events and conditions to which a product is exposed. The LCP should include the various phases that a product will encounter in its life, such as handling, shipping, and storage prior to use. A life-cycle profile is used to quantify the types and levels of loads on a product and identify potential failure mechanisms that can be active in the product. For example, the frequency of use may determine the number of temperature cycles to which a product is exposed on a daily basis. During transportation, the product may be subjected to substantial vibration loading. In storage, elevated humidity levels may be present. During operation, an electrical bias may exist between isolated conductors. The levels of temperature, vibration, humidity, and voltage potential may be inputs to determine thermo-mechanical stress, mechanical strain, moisture content, and dendritic growth, respectively.

Since a product may experience numerous loads simultaneously, it is necessary to identify the critical loads that are applied to the product. Some of the loads will play dominant roles in activating and accelerating the failure of the product, while other loads can be ignored. For example, radiation can often be ignored for ground-based electronic products, since radiation levels are often too low to affect the function of a product or cause any damage. Usually only the dominant loads are considered to reduce the amount of calculation.

2.1.3Physics-of-failure (PoF) models

Failures are broadly categorized by the nature of the loads that trigger or accelerate the failure mechanisms, as overstress (i.e., based on stress strength interference) and wear-out (i.e., based on damage accumulation). Overstress and wear-out failures generally result from irreversible material damage; however, some overstress failures can be caused by reversible material damage (e.g., elastic deformation) [25]. Examples of overstress and wear-out failure types for electronic assemblies are given in Table 1.

Failure models are used to assess failure propensity. There are two categories of models: empirical and PoF. PoF methodology requires inputs from PoF-based models. PoF models, such as Engelmaier model, the Coffin Mason model, and the Basquin model, exist for predicting the time-to-failure of design elements. All failure models have associated assumptions that can be utilized to determine the applicability (or lack thereof) of failure models to a specific design and loading conditions [27].

Each potential failure mechanism is represented by one or more of the PoF models. For electronic products, there is a plethora of PoF models describing the behaviors of components such as printed circuit boards, interconnections, and metallizations under various

conditions (including temperature cycling, vibration, humidity, and corrosion) [21][24][34][35][36]. Commonly used PoF models include the strain-range-based model [30], which describes temperature-cycle-induced solder interconnect fatigue; the Black's model [31], which describe the electromigration in semiconductor device metallization; the Fowler-Nordheim model [32], which describes time dependent dielectric breakdown due to tunneling in gate oxide devices; and the Rudra model [33], which describes conductive filament formation in a printed circuit board.

A model should provide repeatable results, be sensitive to the variables and interactions that are causing degradation and failures, and predict the behavior of the product over the entire domain of its operational environment. PoF models also allow for the development of accelerated tests and acceleration factors. If no models are available, or if the models are found to be inapplicable to specific failure sites and loads, then new models can be developed using controlled experiments that identify the design and environmental factors governing failure and the mathematical relationships linking those factors to the time-to-failure. New failure mechanisms or variations of known failure mechanisms in products usually arise with the introduction of new materials and technologies. As a result, research into the failure mechanisms of new materials and technologies is critical to evaluating products' life expectancies.

2.2 PoF Analysis

The steps involved in PoF analysis include FMMEA, stress analysis, and reliability assessment. Hardware configuration and product design are provided as inputs to the reliability assessment process in conjunction with PoF models.

2.2.1 Failure modes, mechanisms, and effects analysis (FMMEA)

FMMEA utilizes the basic steps in developing a traditional failure modes and effects analysis (FMEA), in combination with knowledge of PoF to identify failure sites. FMMEA then uses a life-cycle profile to identify the active stresses and select the potential failure mechanisms. Knowledge of load type, level, and frequency is used to prioritize failure mechanisms according to their severity and likelihood of occurrence. Figure 3 is a schematic of the FMMEA methodology.

The first step in the FMMEA process is to define the system to be analyzed. A system is a composite of subsystems or levels that are integrated to achieve a specific objective. A failure mode is the effect by which a failure is observed. For the subsystems that have been identified, possible failure modes for each subsystem are listed. For example, in a solder joint, the potential failure modes are electrical opens or intermittent changes in resistance. In the absence of information on potential failure modes, potential failure modes may be identified using stress analysis, accelerated tests to failure (e.g., HALT), past experience, or engineering judgment [38]. One potential failure mode for one component can be the cause of a different potential failure mode in a higher or lower level subsystem or system.

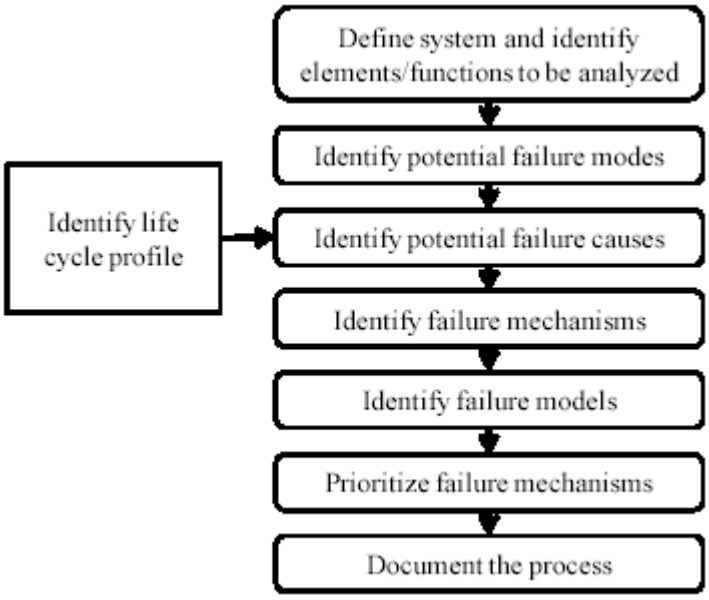

Figure 3: FMMEA methodology.

Potential failure mechanisms for a product are determined based on known failure mechanisms, functional sites, and materials in a product, as well as the anticipated stresses arising in the product [27][28][29]. FMMEA prioritizes failure mechanisms based on their probability of occurrence and severity. The life-cycle profile (LCP) is used to evaluate the failure susceptibility. If certain environmental and operating conditions are nonexistent or generate a low stress level that is below the trigger conditions for a failure mechanism, the failure mechanisms that are exclusively dependent on those environmental and operating conditions are considered to be low occurrence. The quality of materials used also affects the probability of occurrence for a failure mechanism. For example, printed circuit boards (PCBs) made with low levels of hollow glass fibers and high adhesion strength between fiber bundles and epoxies will have a lower level of occurrence of conductive filament formation failure than PCBs with higher levels of glass fibers and low bonding strength. Severity of a specific failure mechanism is defined with regards to a particular application condition. For example, a particular failure mechanism may result in a minor change to an electrical parameter at a specific site in the assembly, whereas the same failure mechanism may shut down the system at another site. The effect of the latter could be catastrophic; therefore, the severity of the same failure mechanism will be higher based on the application. Each critical failure mechanism may have one or more associated causes, sites, and modes in an FMMEA result.

2.2.2 Stress analysis

The identification of failure modes, sites, and mechanisms in the FMMEA process allows for the quantification of time to and probability of failure. However, to evaluate the failures using failure models, it is often necessary to evaluate the stress levels acting at the failure sites. Stress analysis is used to convert the loads on the system to stresses at the potential failure sites. The product's life cycle loads and design information are inputs to the stress analysis process. Stress levels within a product over its anticipated life cycle may be evaluated through physical testing or simulation techniques.

For electronic products, thermal analysis, thermo-mechanical analysis, and vibration analysis are often performed to determine the inputs for the relevant failure models [27]. Thermal

analysis determines the temperature distribution within a product. For example, in a printed circuit assembly (PCA), thermal analysis provides the temperature profile on each layer of laminate and each electronic component. Thermal analysis solves heat transfer equations for the relevant and applicable methods of heat transfer and may target transient or steady state temperatures. In most cases, steady-state thermal analysis results provide sufficient inputs for the identified failure models. For example, a temperature cycle is usually defined as occurring between a power-on operating state and a power-off state. Here, a steady-state thermal analysis may be used to establish the power-on state. Cases where transient behavior alters the damage accumulation rate or failure mechanism, transient behavior should also be analyzed. For example, if there is a delay in temperature changes between connected structures, the connection point may be subjected to higher stress than that estimated by a steady-state evaluation. Thermal analysis problem can be solved by finite difference, finite element, resistance network, and bulk analysis [37]. Vibration analysis determines the response of a PCA to the random oscillating motion of the system that contains the PCA. To calculate the natural frequency of the PCA, boundary conditions are assigned based on an understanding of the mounting conditions of the board in hardware configurations. First-order approximations can provided a quick estimate of the natural frequency and first order response of a PCA. However, finite element modeling is employed to evaluate a wide range of boundary conditions.

The properties of component materials and their interactions may change in response to stress caused by physical or chemical processes. Examples of such processes include diffusion and reaction between materials; fatiguing of materials; and transport, accumulation, and trapping of charges driven by electrical fields [20]. Since these processes may act as failure mechanisms that degrade a product, these property changes have to be taken into account for stress analysis.

The simulation results of stress analysis may be verified and updated through experimental tests. For example, the natural frequency of a PCB can be experimentally determined using a strain gauge or accelerometer on a PCB, attaching the PCB to a dynamic shaker, and measuring the response of the PCB to a known input.

2.2.3 Reliability assessment

With the understanding of failure mechanisms, failure models, stress inputs, failure modes, and failure sites, the reliability of a product can be quantified in terms of time-to-failure of the identified failure sites with a certain probability. By providing the stress at the potential failure sites as input to the PoF models for the potential failure mechanisms, the time-to-failure and failure probability are calculated. Thus, the potential failure mechanisms associated with failure sites can be ranked according to their time-to-failure, failure probability, and severity. The ranked failure mechanisms associated with failure sites can be used for reliability assessment [1][29].

For most products, the life-cycle profile consists of multiple loading conditions. The damage accumulated under multiple loading conditions is estimated using Miner's linear damage accumulation rule. The time-to-failure for a specific failure mechanism, referred to as damage ratio, is defined as the ratio of exposure time under a particular stress condition to the time-to-failure for that stress condition. If the exposure time is equivalent to the time-to-failure, then the ratio would equal unity. Assuming linear damage accumulation, the damage ratios for the same failure site and mechanism could be added over multiple stress conditions. When the summation of damage ratios over multiple stress conditions equals unity, failure

would occur at the site. For the same site and the same failure mechanism and for fixed-duration load events, a specific damage ratio can be determined. For example, dropping a handheld device from a certain height may result in a loss of 10% of the life of a solder interconnect. In this case, each drop will result in some increase in damage to the solder interconnect. For repetitive events, a damage rate may be established by using the appropriate failure models to estimate the number of events required to produce a failure [28][29].

Due to variations in manufacturing processes, all dimensional and material properties are distributed around a nominal value. The same is true for environmental loads. PoF-based reliability assessment allows for utilization of these natural variations in reliability assessment and provides the time-to-failure data as a distribution for each failure site and failure mechanism. With known time-to-failure distribution on each site, reliability can be evaluated by different metrics such as failure rate, warranty return rate, or mean-time-to-failure (MTTF).

The failure mechanisms, modes, and sites are ranked in terms of severity based on the anticipated life-cycle profile and the function of the product. Higher-risk failure mechanisms are given a higher severity when the product design is updated, screening conditions are determined, accelerated testing plans are designed, and remaining life predictions are conducted.

PoF models also allow sensitivity analysis of time-to-failure to material properties, geometries, and life-cycle profiles. Sensitivity analysis would identify the parameters that affect the time-to-failure of the product and thereby improve design by addressing critical design parameters.

2.3 Output

The output from PoF-based reliability assessment process can be used to support various product development activities including design for reliability, stress test conditions, product qualification, screening conditions, accelerated testing, and prognostics implementation. Higher-risk failure mechanisms and failure sites are given a high priority in these activities.

2.3.1 Design-for-reliability

Design-for-reliability (DfR) is a systematic, streamlined, and concurrent approach where reliability is woven into the total development cycle. The PoF-based reliability assessment process can examine the effect of variations in design parameters on a product's reliability. Hence, the PoF approach can be used to examine the impact of proposed design changes, as well as to compare competing designs. PoF-based reliability assessments may be performed on multiple product designs and manufacturing variables to compare their estimated reliability. The sensitivity analysis in PoF assessment also provides a way to obtain reliability as a function of product attributes. The ranking of failure mechanisms in the PoF approach allows designers to concentrate on locations and mechanisms that are most likely to cause product failure. The PoF methodology may be used to guide design improvement by identifying the drivers for the dominant failure mechanisms during each life-cycle phase. Design trade-offs can then be evaluated by determining the sensitivity of the dominant degradation mechanisms to the drivers of the mechanisms.

Knowledge of PoF will assist in the design, plan and implementation of the testing. The critical parameters that affect the reliability of the product can be identified and quantified by sensitivity analysis of PoF models, thereby conducting a more effective design of experiment (DOE). PoF-based reliability assessment can also provide a basis to conduct environmental

stress management [38]. Stress management solutions can be evaluated in terms of their effectiveness to reduce loads that may result in failure and increase the useful life of a product. For example, the response of a circuit card to various levels of vibration could be used to determine the amount of damping or the number of support locations required to adequately protect the circuit card. Derating is another form of stress management of electronic components to reduce their electrical and thermal stress levels with the goal of increasing the times-to-failure.

2.3.2 Virtual qualification

Virtual qualification (VQ, also called simulation-assisted reliability assessment) is the assessment of life expectancy under anticipated life cycle loading conditions. The assessment of reliability goals under anticipated lifecycle profiles is based on material properties, geometries, and operating characteristics. The physical hardware is modeled in simulation software and the failure probability and distribution is estimated using PoF models [39] [40] [41]. Application of VQ in the design stage enables improvement in product design and comparison of competing designs [23] [42]. VQ uses computer-aided engineering software to qualify components and systems based on analysis of the susceptibility of the designs to failure. The reliability assessment tool assesses the DfRs in the environments present in the life-cycle profile using a database of validated PoF models. VQ calculates times-to- failure for the mechanisms that cause failures and evaluates the effects of variables in product design, manufacturing, and loading conditions on reliability. VQ also aids in the selection of design parameters and components by providing information on their impact on reliability and facilitates the selection of cost-effective test methods for validating reliability assessment and design [43].

2.3.3 Stress test conditions

Stress is usually applied to ruggedize the design and manufacturing process of a product during systematic step-stress testing, increase the stress margins for reliability enhancement testing or to verify in-service reliability by accelerated life tests in the laboratory. The stress conditions are determined from the life-cycle profile of the product, FMMEA, experience with previous similar products, application requirements, and the time-to-market [26]. Critical failure mechanisms identified by FMMEA that are associated with failure sites and failure modes will be considered in the selection of stress conditions.

Accelerated stress testing is used to precipitate failures in products—either prototypes or manufactured products. Accelerated stress testing allows verification of design goals over a shorter time period, thereby achieving faster design modification. However, the value of accelerated tests can be realized only if there is an associated predictive capacity regarding the effect of stress tests on the failure mechanisms and modes.

For obtaining useful information from accelerated stress testing, a correlation (defined as acceleration factor) must be established between the accelerated stress test condition and the field condition [44]. PoF-based reliability assessment provides the acceleration factor by identifying the failure mechanisms and appropriate failure models that relate the stress test to field conditions. Stress levels for accelerated testing should be selected to ensure that the stress will induce the same failure mechanism under accelerated conditions as it would under field conditions without introducing any new failure mechanisms. PoF-based reliability assessment can also be used to establish a test plan. This test plan will include the test stress types; stress levels; duration of the test; definitions of failure; expected times of failures; and failure mechanisms, sites, and modes. In addition, if the accelerated tests are meant to be

qualification or acceptance tests, then acceptance criteria (e.g., number of samples, number of allowable failures, and minimum time-to-failure) for such decisions are also identified in the plan [44][45].

2.3.4 Qualification

Qualification is carried out to verify whether a product meets the reliability and quality requirements of its intended application. Qualification includes verification of the product's function, performance, and reliability. PoF can be used in the qualification of products to define the qualification requirements, especially the qualification test conditions, based on the understanding of potential failure modes and mechanisms. Failure analysis is also performed, when applicable, to verify the potential failure modes and failure mechanisms. The qualification test conditions and stress levels are determined from the life-cycle profile of the product, FMMEA, experience from previous similar products, application requirements, and time-to-market. Target values and tolerances have to be met by the product or product elements under the test conditions selected with respect to the life cycle profile [26]. The critical failure mechanisms identified by FMMEA associated with the failure sites and failure modes will be considered in the qualification tests. When these failure mechanisms are combined with the reliability requirements, the targeted failure mechanisms in the qualification tests can be determined. The qualification test stress level should be selected so that the stress will induce the same failure mechanism as it would under operating conditions without introducing any new failure mechanisms. If the purpose of the qualification test is to test the samples to failure and make predictions of the product life, the acceleration factor between the qualification test results and the actual life under application conditions should be estimated. Stress levels and test time are considered based on the time-to-market. High stress levels need less time for testing, while a low stress level may need a longer test time. However, stress levels beyond a critical limit may induce unexpected failure mechanisms. So, the tradeoffs between the time-to-market and stress level selections have to be examined.

2.3.5 Screening conditions

Screening eliminates defective products with a certain level of tolerance after manufacturing. During screening all manufactured products are subjected to a set of defined stress tests to expose defective products. Due to the application of a defined set of stresses, screening induces a certain amount of damage on the products. To be effective, the screening conditions must not introduce defects to the products or reduce their life expectancy below acceptable levels.

PoF allows for the identification of design parameters that affect product life and the stress conditions that can be used to precipitate failures through an understanding of the failure mechanisms. PoF is used to define stress conditions that will produce failure in defective products with design parameters below a desired threshold [46]. PoF also identifies the damage induced on products subjected to a screen in order to determine if the screen will result in the compromised reliability of the surviving products.

2.3.6 Prognostics and health management (PHM)

Prognostics and health management (PHM) is the process of predicting the remaining useful life (RUL) of a product by assessing the extent of degradation from its expected state of health. PoF-based PHM permits in-situ assessment of system reliability under field conditions. Sensor data is used by PoF models to enable in-situ assessment of the deviation or degradation of a product from the expected normal operating condition (i.e., the system's

"health") and the prediction of the expected state of reliability. PoF-based PHM identifies the components that are critical to system reliability, provides estimates of damage for given loading conditions and failure mechanisms, enables remaining life estimation for different loading conditions, and obtain remaining useful life even in non-operating conditions. PoF-based PHM provides previously unknown information on life-cycle environmental and operational conditions, prevents premature failures, and gives information on remaining useful life of electronic products and systems. Schematic of a PoF-based PHM methodology is provided in Figure 4.

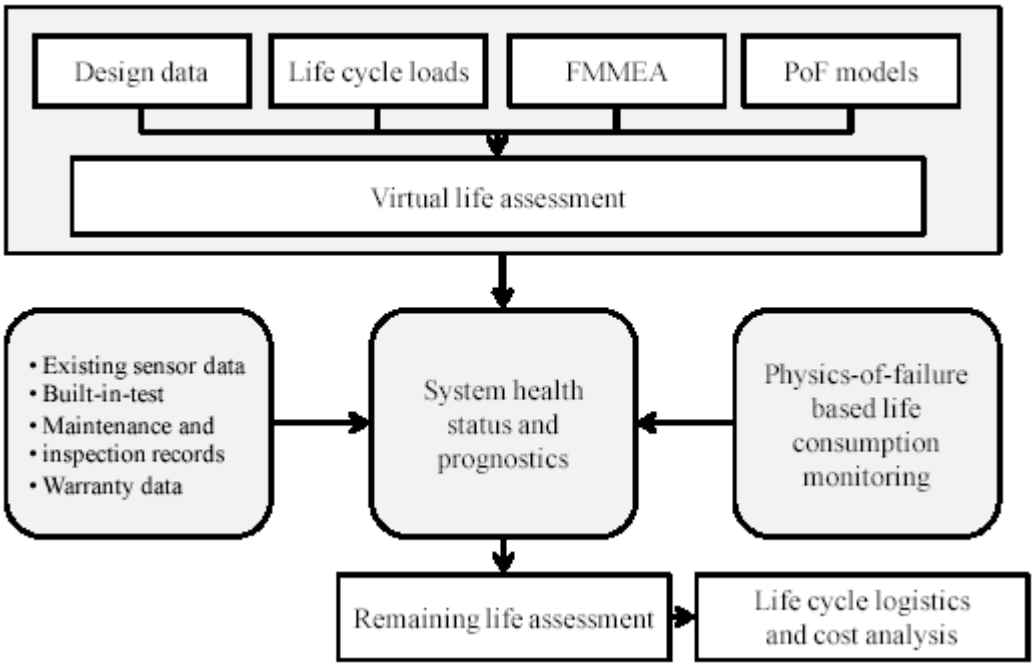

Figure 4: PoF based PHM methodology [5]

3. Physics-of-failure cases

PoF-based reliability assessment requires the documentation of product attributes, life-cycle loads, failure mechanisms and associated PoF models. While the implementation of PoF may appear challenging, a structured approach will allow the reuse of information and streamline the information gathering during product development. Cases of successful implementation of PoF-based reliability assessment are discussed in this section.

The semiconductor industry has been successfully using PoF for semiconductor product qualification and reliability evaluation. Standards have been developed for qualification testing to address the different failure mechanisms under different conditions based on application conditions. Failure mechanisms have been identified, and failure models have been developed through research efforts for various electronic products. For example, time dependent dielectric breakdown due to tunnelling was identified as a failure mechanism in gate oxide devices and a model was developed by Fowler-Nordheim [32].

The Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland has been applying PoF-based reliability assessment techniques on various electronic products for over two decades. PoF has been applied to evaluate the reliability of different

packaging and assembly levels of electronic products, including wire bonding [47], electronic devices [48][49], flip-chip bonding [50], board laminates [51], printed board assemblies with a complex layout [26], and cruise control modules for automobiles [52]. Computer software tools based on PoF have been developed by CALCE to simplify the reliability assessment process. George at al. [53] demonstrated the PoF-based virtual qualification of a communication hardware device using the CALCE software. Guidelines have been developed for PoF-based accelerated testing to address electronics reliability [54]. J. Gu et al. [55] successfully used the PoF approach as the basis for PHM for an electronic card assembly. Ramakrishnan and Pecht [56] also performed remaining useful life estimation based on a PoF analysis of printed circuit board assemblies under application conditions. In addition to printed circuit board applications, PoF-based approaches have been developed for LEDs, cooling fans, and so on. Oh et al. [57] developed a PoF based PHM approach to predict the reliability of brushless direct current fans used in the thermal management of electronic equipments. A step-by-step PoF approach, including FMMEA, prioritization of potential failure mechanisms, and estimation of life expectancy was provided. Fan et al. [58] presented a PoF based PHM approach for high-power white LEDs. The knowledge of LED's life-cycle loading, materials and geometries was utilized to conduct FMMEA followed by the application of appropriate PoF models to predict life. Cheng et al. [59] provided the considerations for sensor selection for PHM applications including the parameters to be measured, performance needs, electrical and physical attributes, reliability, and cost of the sensor system.

The U.S. Army Material Systems Analysis Activity (AMSAA) provides PoF analysis, guidance, and support to the U.S. Army, DoD, and defense contractors to reduce costs and increase system readiness. AMSAA assesses and applies PoF tools and methodologies to army systems, assist research institutions in the development and validation of new PoF tools and methodologies, and develop approaches for probabilistic PoF and PoF-based prognostics.

4. Conclusions

Physics-of-failure based methodology is a more realistic and accurate approach to assess reliability than constant failure rate approaches due to identification of the failure mechanisms that degrade a product and result in its failure based on its design and life-cycle profile. PoF ranks potential failure mechanisms and helps in design for reliability, selection of stress test conditions, product qualification, screening, and prognostics and health management. PoF is widely accepted as a reliability assessment methodology by industry, U.S., government organizations, and international standards. Qualification standards based on PoF approaches are widely used by the electronics industry during product development and evaluation processes.

5. References

[1] J. Bernstein, M. Gurfinkel, X. Li, J. Walters, Y. Shapira, and M. Talmor, "Electronic circuit reliability modeling," Microelectronics Reliability, Vol. 46, pp. 1957-1979, 2006.
[2] J. Mclinn, "Constant failure rate - a paradigm in transition", Quality Reliability Engineering International, Vol. 6, pp. 237-241, 1990.
[3] J. Bowles, "A survey of reliability prediction procedures for microelectronic devices," IEEE Transactions on Reliability, Vol. 41, No. 1, pp. 2-12, 1992.

[4] B. Foucher, J. Boulli, B. Meslet, and D. Das, "A review of reliability prediction methods for electronic devices," Microelectronics Reliability, Vol. 42, No. 8, pp. 1155-1162, 2002.
[5] M. Pecht and J. Gu, "Physics-of-failure-based prognostics for electronic products", Transactions of the Institute of Measurement and Control, Vol. 31, No. ¾, pp. 309-322, 2009.
[6] W. Perry, "Specifications and standards - a new way of doing business", U.S. Department of Defense Policy Memorandum, 1994.
[7] Z. Matic and V. Sruk, "The physics-of-failure approach in reliability engineering," Proceedings of the International Conference on Information Technology Interfaces, Cavtat, Croatia, pp. 745-750, June 23-26,2008.
[8] A. Goel and R. Graves, "Electronic system reliability: collating prediction models," IEEE Transaction on Device and Materials Reliability, Vol. 6, No. 2, pp. 258-265, 2006.
[9] I. Knowles, "Is it time for a new approach?" IEEE Transactions on Reliability, Vol. 42, No. 1, pp. 2-3, 1993.
[10] M. Pecht, "Why the traditional reliability prediction models do not work - Is there an alternative?", Electronic Cooling, Vol. 2, No. 1, pp. 10-12, 1996.
[11] M. Talmor and S. Arueti, "Reliability prediction: The turn-over point," Proceedings of Annual Reliability and Maintainability Symposium, Philadelphia, PA, pp. 254-262, January 13-16, 1997.
[12] P. O'Connor, "Reliability prediction for microelectronic systems", Reliability Engineering, Vol. 10, No. 3, pp. 129-140, 1985.
[13] C. Leonard, "How failure prediction methodology affects electronic equipment design," Quality Reliability International, Vol. 6, No. 4, pp. 243-249, 1993.
[14] M. Pecht and F. Nash, "Predicting the reliability of electronic equipment," Proceedings of IEEE, Vol. 82, No. 7, pp. 992-1004, 1994.
[15] T. Stadterman, M. Cushing, B. Hum, A. Malhorta, and M. Pecht, "The transition from statistical-field failure based models to physics-of-failure based models for reliability assessment of electronic packages," Proceedings of InterPACK, Lahaina, Maui, HI, pp. 619-625, 1995.
[16] B. Johnson and L. Gullo, "Improvements in reliability assessment and prediction methodology," Proceedings of the Annual Reliability Maintainability Symposium (RAMS), pp. 181-187, 2000.
[17] D. Mortin, J. Krolewski, and M. Cushing, "Consideration of component failure mechanisms in the reliability assessment of electronic equipment - Addressing the constant failure rate assumption," Proceedings of Annual Reliability and Maintainability Symposium, Washington, DC, pp. 54-59, January 16-19, 1995.
[18] L. Weil, M. Pecht, and E. Hakim, "Reliability evaluation of plastic encapsulated parts," IEEE Transactions on Reliability, Vol. 42, No. 4, pp. 536-540, 1993.
[19] M. Cushing, D. Mortin, T. Stadterman, and A. Malhorta, "Comparison of electronics-reliability assessment approaches," IEEE Transactions on Reliability, Vol. 42, No. 4, pp.542-546, 1993.
[20] JEDEC, JEP 148, "Reliability qualification of semiconductor devices based on physics of failure risk and opportunity assessment", April 2008.
[21] A. Dasgupta and M. Pecht, "Failure mechanisms and damage models", IEEE Transactions on Reliability, Vol. 40, No. 5, pp. 531-536, 1991.
[22] D. Ryu and S. Chang, "Novel concepts for reliability technology,", Microelectronics Reliability, Vol. 45, pp. 611-622, 2005.
[23] J. Hu, D. Barker, A. Dasgupta, and A. Arora, "The role of failure mechanism identification in accelerated testing," Journal of the Institute of Environmental Sciences, Vol. 36, No. 4, pp. 39-45, 1993.

[24] W. Lee, L. Nguyen, and G. Selvaduray, "Solder joint fatigue models: Review and applicability to chip scale packages," Microelectronics Reliability, Vol. 40, pp. 231-244, 2000.
[25] M. Osterman and T. Stadterman, "Reliability and performance of advanced PWB assemblies", High Performance Printed Circuit Boards, McGraw-Hill, New York, 1999.
[26] I. Snook, J. Marshall, and R. Newman, "Physics of failure as an integrated part of design for reliability", Proceedings of the IEEE Annual Reliability and Maintainability Symposium, pp. 46-54, 2003.
[27] R. Tilgner, "Physics of failure for interconnect structures: an essay," Microsystem Technology, Vol. 15, pp. 129-138, 2009.
[28] M. Pecht and A. Dasgupta, "Physics of failure: An approach to reliable product development," Proceedings of the International Integrated Reliability Workshop, Lake Tahoe, CA, pp. 1-4, Oct. 22-25, 1995.
[29] U. Kumar and A. Dasgupta, "Guidelines for physics-of-failure based accelerated stress testing," Proceedings of the Annual Reliability and Maintainability Symposium, pp. 345-357, 1998.
[30] M. Osterman, A. Dasgupta, and B. Han, "A strain range based model for life assessment of Pb-free SAC solder interconnects," Proceedings of the 56th Electronic Component and Technology Conference, pp. 884-890, May 30-June 2, 2006.
[31] J. Clement, "Electromigration modeling for integrated circuit interconnect reliability analysis," IEEE Transactions on Device and Materials Reliability, Vol. 1, No. 1, pp. 33-42, March 2001.
[32] J. Lee, I. Chen, and C. Hu, "Modeling and characterization of gate oxide reliability," IEEE Transactions on Electron Device, Vol. 35, No. 22, pp. 2268-2278, 1988.
[33] B. Rudra and D. Jennings, "Tutorial: Failure-mechanism models for conductive filament formation," IEEE Transactions on Reliability, Vol. 43, No.3, pp.354-60, 2004.
[34] M. Pecht, R. Radojcic, and G. Rao, "Guidebook for managing silicon chip reliability," CRC Press, Boca Raton, FL, 1999.
[35] P. Lall, M. Pecht, and E. Hakim, "The influence of temperature on microelectronic device reliability", CRC Press, Boca Raton, FL, 1997.
[36] J. Li and A. Dasgupta, "Failure mechanism models for material aging due to inter-diffusion", IEEE Transactions on Reliability, Vol. 43, No. 1, pp. 2-10, March 1994.
[37] M. Pecht, R. Agarwal, P. McCluskey, T. Dishongh, S. Javadpour, and R. Mahajan, "Electronic packaging materials and their properties," CRC Press, Boca Raton, FL, 1999.
[38] R. Bauernschub and P. Lall, "Addressing defect related reliability and screening levels through physics-of- failure analysis," American Society of Mechanical Engineering, Advances in Electronic Packaging, Vol. 10-12, pp. 635-645, 1995.
[39] J. Cunningham, R. Valentin, C. Hillman, A. Dasgupta, and M. Osterman, "A demonstration of virtual qualification for the design of electronic hardware", Proceedings of the Institute of Environmental Sciences and Technology Meeting, April 24, 2001.
[40] M. Cushing, D. Mortin, T. Stadterman, and A. Malhotra, "Comparison of electronics-reliability assessment approaches," IEEE Transactions on Reliability, Vol. 42, No. 4, pp. 542-546, December 1993.
[41] T. Larson and J. Newel, "Test philosophies for the new millennium," Journal of the Institute of Environmental Sciences, Vol. 40, No. 3, pp. 22-27, 1997.
[42] H. Caruso and A. Dasgupta, "A fundamental overview of analytical accelerated testing models," Journal of the Institute of Environmental Sciences, Vol. 41, No. 1, pp. 16-30, 1998.
[43] P. McCluskey, M. Pecht, and S. Azarm, "Reducing time-to-market using virtual qualification," Proceedings of the Institute of Environmental Sciences Conference, pp. 148-152, 1997.

[44] T. Rothman, A. Dasgupta, and M. Binder, "Physics-of-failure case study for accelerated testing for electronic packaging," Institute of Environmental Sciences Proceedings, Annual Technical meeting, Product Reliability, pp.63-71, 1995.
[45] M. Cushing, J. Evans, and R. Bauernschub, "Physics of failure (POF) approach to addressing device reliability in accelerated testing of MCM's," Proceedings of the IEEE Multi Chip Module Conference, pp.14-25, 1993.
[46] A. Foucher, J. Tomas, F. Mounsi, and M. Jerremias, "Life margin assessment with physics of failure tools applications for BGA packages," Microelectronics Reliability, Vol. 46, No. 5-6, pp. 1013-1018, 2006.
[47] M. Pecht, A. Dasgupta, and P. Lall, "A failure prediction model for wire bonds," ISHM Proceedings, Baltimore, MD, October 17-19, 1989.
[48] P. Lall, M. Pecht, and M. Cushing, "A physics-of-failure (PoF) approach to addressing device reliability in accelerated testing," 5th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis, Glasgow, Scotland, October 4-7, 1994.
[49] M. Osterman, "A physics of failure approach to component placement," Journal of Electronic Packaging, Vol. 114, pp. 305-309, 1992.
[50] C. Pusarla, A. Dasgupta, M. Pecht, and A. Christou, "A physics-of-failure design philosophy applied to flip-chip bonds, ,Microlectronics International, No. 36, 1995.
[51] M. Pecht, and A. Dasgupta," Physics-of-failure: An approach to reliable product development," Journal of the Institute of Environmental Sciences, Vol. 38, pp. 30-34, 1995.
[52] K. Kimseng, M. Hoit, N. Tiwari, and M. Pecht, "Physics-of-failure assessment of a cruise control module", Microelectronics Reliability, Vol. 39, pp. 1423-1444, 1999.
[53] E. George, D. Das, M. Osterman, and M. Pecht, "Physics of failure based virtual testing of communication hardware", ASME International Mechanical Engineering Congress and Exposition, Lake Buena Vista, Florida, US, 13-19 November 2009.
[54] K. Upadhyayula, and A. Dasgupta, "Physics-of-failure guidelines for accelerated qualification of electronic systems," Quality and Reliability Engineering International, Vol. 14, No. 6, pp. 433-447, 1998.
[55] J. Gu and M. Pecht, "Prognostics and health management using physics of failure," 54th annual Reliability and Maintainability Symposium (RAMS), Las Vegas, Nevada, January 2008.
[56] A. Ramakrishnan and M. Pecht, "A life consumption monitoring methodology for electronic systems," IEEE Transactions on Components and Packaging Technologies, Vol. 26, No. 3, pp. 625-634, September, 2003.
[57] H. Oh, M. Azarian, M. Pecht, C. White, R. Sohaney, and E. Rhem, "Physics-of-failure approach for fan PHM in electronics applications", Prognostics and Health Management Conference, Macao, January 12-14, 2010.
[58] J. Fan, K. Yung, and M. Pecht, "Physics-of-failure-based prognostics and health management for high-power white light-emitting diode lighting", IEEE Transactions on Device and Materials Reliability, Vol. 11, No. 3, pp. 407-416, September, 2011.
[59] S. Cheng, M. Azarian, and M. Pecht, "Sensor systems for prognostics and health management", Sensors, Vol. 10, pp. 5774-5797, June 2010.

Mirce-mechanics Analysis of the Impact of Cosmic Phenomena on In-service Reliability

Ian. Zaczyk,
Doctoral Diploma Student, MIRCE Akademy, Exeter, EX5 1JJ, UK

Abstract

The main objective of this paper is to argue that the scientific approach to functionability is the only way forward for the engineering community if accurate predictions regarding occurrences of negative functionability events are to be made, which are to be confirmed during the operational processes of the future man made, managed and maintained systems. Hence, science based understanding of the mechanisms that cause occurrences of functionability events generated by the surrounding natural environment are required. Then and only then, accurate and meaningful functionability predictions become possible, which will ultimately lead to the reduction of the probability of the occurrence of failure events during the life of man made, managed and maintained systems. This paper focuses on the scientific understandings of the relevant cosmic phenomena on the in-service reliability of systems, as conducted within Mirce-mechanics principles.

1. Introduction

Analysis of the events that caused the blackout on 13 March 1989 in Quebec confirmed that magnetic storms affect power system behaviour. Mainly, they cause transformer saturation, which reduces or distorts voltage. Power supply systems with long lines and static compensators are particularly sensitive to such natural phenomena. Quebec utility's experts noted a correlation between the exceptional intensity of the magnetic storm and the tripping of several static compensators, at Chibougamau and La Verendrye substations. Immediately after this event took place records show voltage oscillations and power-swings increase until the lines from James Bay failed. Within seconds, the whole grid lost functionability (ability to function). This negative functionability event was caused by the strongest magnetic storm ever recorded at this location. The storm, which resulted from a solar flare, tripped five lines from James Bay and caused a generation loss of 9,450 MW. With a load of some 21,350 MW at that moment, the system was unable to withstand this sudden loss and failed to function within seconds. The system-wide blackout resulted in a loss of some 19,400 MW in Quebec and 1,325 MW of exports. An additional load of 625 MW was also being exported from generating stations isolated from the Hydro-Quebec system.

Restoration of functionability took more than nine hours. This can be explained by the fact that some of the essential equipment, particularly on the James Bay transmission network, was made unavailable by the blackout. Generation from isolated stations normally intended for export was made available to Quebec's needs and the utility purchased electricity from Ontario. By noon, the entire generating and transmission system was back in service, although 17 percent of Quebec customers were still without electricity. In fact, several distribution-system failures occurred because of the high demand typical of Monday mornings, combined with the jump in heating load after several hours without power.

On the other side of the scale spectrum, atmospheric radiation causes daily concerns regarding the functionability of avionics equipment, particularly for those systems that are considered safety critical. The trend with each new generation of avionics system is to use

increasing quantities of semiconductor memories and other complex devices that are susceptible to failures induced by ionising radiation from the following two main sources: cosmic rays from space. and alpha particles from radioactive impurities in the device itself. The interaction of this radiation can result in either a transient 'soft error' effect such as a bit flip in memory or a voltage transient in logic, alternatively a 'hard error' can be induced resulting in permanent damage such as the burn out of a transistor. These functionability effects caused by a single radiation event are collectively termed as Single Event Effects (SEEs).

If device memory cells used for flight safety or mission critical functions are affected the concern is that the loss of key system functionality due to corrupted data could cause a flight safety or mission critical failure. The ability to predict and quantify the rate of occurrence of erroneous data bits in memories or voltage transients in logic is one of the key objectives in the field of avionics SEEs research. Baumann [1] stated that: "Left unchallenged, soft errors have the potential for inducing the highest failure rate of all other reliability mechanisms combined"

The main challenge in both examples given, as in all cases regarding the operation of man made and maintained systems, is the true understanding of the impact of the environment that surrounds them. The functionability of their operation is influenced by a multitude of different factors extending from the Earth's atmosphere to the far reaches of space beyond our own galaxy. In order to determine the probabilities of occurrence and the resultant impact of functionability events on a system a full awareness of the dynamic nature of the environmental phenomena is required. To identify the causes of functionability events the mechanisms that cause those physical phenomena must first be understood.

Consequently, the main objective of this paper is to argue that the scientific approach to functionability is the only way forward for all members of the reliability community who wish to make accurate predictions regarding occurrences of negative functionability events, which will be confirmed during the operational processes of the future systems, is required. For that to happen scientific understanding of functionability phenomena is required. This paper advocates that research of this nature must include the understanding of the cosmic phenomena, in order for occurrence of functionability events to be understood. Then and only then, accurate and meaningful functionability predictions become possible, which will ultimately lead to the reduction of the probability of the occurrence of failure events during the life of man made, managed and maintained systems.

2. Scientific Principles of Mirce-mechanics

Mirce-mechanics is a new scientific theory, developed at the MIRCE Akademy by Dr. J. Knezevic, that aims to scientifically understand the physical causes and human actions that shape the motion of functionability through lives of man made, managed and maintained systems. [2] For years, research studies, international conferences, summer schools and other events have been organised in order to understand just a physical scale at which failure phenomena should be studied and understood. In order to understand the motion of functionability events it is necessary to understand the physical mechanisms that cause their occurrences. That represented a real challenge, as the answers to the question "what are physical and chemical processes that lead to the occurrence of given functionability events" have to be provided. Without accurate answers to those questions the prediction of their

future occurrences is not possible, and without ability to predict the future, the use of the word science becomes inappropriate.

After a numerous discussions, studies and trials, it has been concluded that any serious studies in this direction, from Mirce-mechanics point of view, have to be based between the following two boundaries:

- the "bottom end" of the physical world, which is at the level of the atoms and molecules that exists in the region of 10^{-10} of a metre [3],
- the "top end" of the physical world, which is at the level of the solar system that stretches in the physical scale around 10^{+10} of a metre. [4]

This range is the minimum sufficient "physical scale" which enables scientific understanding of relationships between system life processes and system failure events.

One of the interacting factors from the physical world that directly impacts the functionability trajectory of man made systems are cosmic phenomena, as illustrated by the examples given above. This paper therefore considers major causes of cosmic phenomena from the physical world that can influence system functionability from functionability point of view.

Using the scientific principles of Mirce-mechanics the primary goal of this paper is to present the dynamic nature of the cosmic environment and the mechanisms that cause occurrences of negative functionability events. To achieve this goal, the paper examines the nature of the cosmic phenomena to understand the mechanisms of their occurrences as well as their possible impacts on systems functionability.

2. Atmospheric Radiation

In the natural environment there are two fundamental radiation particles that can cause transient errors in electronic devices, which can be classified into the following three groups:

- High-energy cosmic ray neutrons.
- Thermal or low energy cosmic ray neutrons.
- Low energy alpha particles emitted from within the semiconductor device and packaging materials.

Each of these particle categories is different in terms of flux, energy level, charge or composition, but in essence a single particle of any of the above forms could result in a soft error if it deposits sufficient charge within the susceptible volume of a device.

3. Cosmic Rays

Cosmic rays are individual energetic particles that originate from a variety of energetic sources ranging from our Sun to supernovas and other phenomena in distant galaxies all the way out to the edge of the visible universe. The majority of energetic particles however come from our galaxy with only the most energetic particles believed to have originated from extra-galactic sources. Although the term cosmic ray is commonly used, this term is misleading because no cohesive ray or beam actually exists. Cosmic rays are in fact independent energetic particles that travel at approximately 87% of the speed of light.

Victor Hess first discovered cosmic rays in 1912, when he discovered the fourfold increase in ionisation rates as he ascended to altitude in a balloon. From this experiment he concluded that
"The results of my observation are best explained by the assumption that a radiation of very great penetrating power enters our atmosphere from above. In 1936 he was awarded the Nobel Prize in Physics for this discovery, although the term 'cosmic rays' is actually credited to a fellow scientist, R.A Millikan in 1925.

The majority of cosmic rays consist of the nuclei of atoms (atoms stripped of their outer electrons) ranging from the lightest elements in the periodic table to the heaviest. In terms of composition about 90% of the nuclei are hydrogen, therefore just single protons, 9% are helium, alpha particles with the remaining 1% a mix of heavier element nuclei, high energy electrons, positrons and other sub-atomic particles.

Cosmic rays must not be confused with gamma rays (high energy photons) that constitute the most energetic form of electromagnetic radiation. However there is a component of cosmic rays, < 0.1% which consists of gamma ray photons produced after high energy particle collisions with matter.

Within the atmosphere the three most important parameters used to define the variability of the particle flux at a specific location are altitude, latitude and energy. Within the field of cosmic ray physics altitude is expressed in terms of atmospheric depth, which is the mass thickness per unit of area in the Earth's atmosphere. At sea level this is approximately 1033 g/cm^2 of oxygen and nitrogen and reduces as the altitude increases. Atmospheric depth is the key-determining factor in the particle flux for a specific point in the atmosphere. For example at an altitude of 3000m the flux of neutrons within the atmospheric cascade is around 10 times greater than at sea level.

Energy is usually shown as the flux per unit of energy called the differential flux, and geographic latitude is expressed in terms of the geomagnetic field strength expressed in units of GeV and also referred to as a locations geomagnetic rigidity or cut-off.

Cosmic rays can be broadly divided into two main categories, primary cosmic rays and secondary cosmic rays. Primary cosmic rays are particles accelerated at astrophysical sources and generally do not penetrate the Earth's atmosphere. Primary cosmic rays are composed from a mixture of different energetic particles that can be categorised based on origin and energy level into the groups listed below in order of descending particle energy:

- Extra galactic cosmic rays.
- Galactic cosmic rays,
- Solar cosmic rays,
- Anomalous cosmic rays.

Secondary cosmic rays are created when primary cosmic rays collide with particles and break into lighter nuclei in a process known as cosmic ray spallation. Cosmic ray spallation is a naturally occurring form of nuclear fission and nucleosynthesis. Spallation can also occur with the dust and gas that inhabits the interstellar medium. However the resultant products from these interactions are not relevant to the avionics radiation environment.

As cosmic ray particles are charged, magnetic fields in space will bend their motion paths. Due to the impact of magnetic fields, cosmic ray particles are incident on the Earth from all directions and as a consequence it is impossible to retrace their trajectories to determine their point of origin. However, the trajectory of a gamma ray photon is a straight line, due to their neutral charge. This makes it possible to retrace the trajectories of gamma rays to discover their source.

3.1 Extra galactic and galactic cosmic rays

Extra galactic cosmic rays originating from outside our galaxy and galactic cosmic rays from within bombard the top of the Earth's atmosphere with a low but continuous flux of protons and heavy ions. The majority of energetic particles are accelerated from within our galaxy but external to the solar system. Cosmic ray particles from extra galactic and galactic sources are typically highly energetic and arrive at the Earth with an approximate flux rate of between 2 to 4 cm^{-2} s^{-1}.

3.2 Solar cosmic rays

Solar cosmic rays, also termed Solar Energetic Particles, SEPs or Solar Proton Events SPEs, are produced by highly energetic processes that occur on or close to the Sun's surface. Unlike galactic cosmic rays that arrive at the Earth with an almost steady constant flux, the occurrence of solar particles is not only irregular but also highly variable in terms of flux rate. Typically most solar protons arriving from the Sun lack the energy level required to penetrate the Earth's magnetic field.

Solar cosmic rays consist of heavy ions and protons with a less energetic spectrum than galactic cosmic rays. In comparison to the maximum energy possessed by galactic cosmic ray protons of 10^{21}eV, the solar proton peak energy of about 20 GeV is many orders of magnitude smaller.

In the case of very powerful flux ejections, SPEs manifest as Ground Level Enhancements or Events, GLEs, on the Earth's surface and typically last between 20 minutes to a few days dependent on the originating solar mechanism. SPEs can therefore be categorised as either an impulsive event linked to solar flares or gradual events linked to coronal mass ejections, CMEs.. The main concern however regarding SPEs are the significant neutron flux enhancements generated at aircraft altitudes particularly at high geographic latitudes where the Earth's level of magnetic shielding is reduced.

During the Sun's eleven year solar cycle the flux of solar particles incident upon the Earth's upper atmosphere can increase by a million fold during a GLE relative to the level at a quiescent period close to or at the solar minimum. In contrast the difference between the flux rates between solar minimum and solar maximum, whilst still significant, are less dramatic than the sporadic peak flux rates caused by the most energetic SPEs., as shown in Table 1.

Energy Range	Solar Maximum (Particles : $cm^{-2}s^{-1}$)	Solar Minimum (Particles : $cm^{-2}s^{-1}$)
Above 30 MeV	3×10^{2}	2×10^{-2}
Above 100 MeV	20	2×10^{-3}

Table 1: Mean integral solar cosmic ray flux at solar minimum and maximum

GLEs in general occur 1 to 3 years after a solar maximum and to date since 1942 in total 63 of them have been observed. Over a longer period analysis of nitrate spikes obtained from polar ice cores indicate 154 large SPEs have occurred in the last 450 years,. These powerful and evidently rare events are believed to be caused by the most energetic solar flares rather than CMEs.

In terms of energy levels SPEs typically range from 10 MeV to 100 MeV although protons up to 20 GeV travelling at near relativistic speeds can be discharged from the Sun during extremely energetic events. The proton energy level determines the speed and hence the arrival time of incident protons. At 1 MeV, protons arrive in 2.9 hrs but at 1 GeV the arrival time is reduced to just 9.5 minutes.

Anomalous cosmic rays

Anomalous cosmic rays are the final component of primary cosmic rays and possess energy levels significantly lower than any other type of cosmic ray, typically less than ~10 MeV. They are created when electrically neutral atoms enter the heliosheath of the Sun's solar wind, become ionised and are then accelerated by the termination shock. The termination shock region forms the inner edge of the heliosheath where the solar wind becomes subsonic. This region varies between 75 and 100 AU (1 AU is a unit of length approximately equal to the semi-major axis of Earth's orbit around the Sun) from the Earth.

4. Energy and origins of cosmic rays

The kinetic energy possessed by cosmic rays particles are measured in terms of electron volts, eV. One electron volt is defined as the energy gained when an electron is accelerated through a potential difference of 1 volt. The energy levels of cosmic ray charged particles range from a few billion eV to more than 10^{20} eV. Consequently units of MeV for mega electrons volts or GeV giga-electron volts are generally used to quantify the voltage levels.

The energy spectrum of cosmic rays that is represented by a power-law function over an expansive range of energies, 10^9 eV to over 10^{20} eV, as shown in Figure 1. The energy spectrum for cosmic rays is relatively featureless except for the break points traditionally referred to as the 'Knee' and 'Ankle'. The 'Knee' point is located around the energy level $3x10^{15}$ eV and the 'Ankle' around $3x10^{18}$ eV.

To clearly portray the difference between the incident cosmic ray flux of particles with energies of 10^{15} eV, 10^{18} eV and 10^{20}eV consider that at 10^{15} eV, one particle is incident per m^2 every year, at 10^{18} eV, one particle is incident per km^2 every year but at 10^{20} eV one particle is only incident per km^2 once every century. At the energy level of 10^{20} eV galactic cosmic rays are equivalent in kinetic energy to a tennis ball travelling at 340 mph. Considering the diameter of a proton is 1.5×10^{-15} m and a tennis ball is 13 orders of magnitude bigger at 0.065 m this is a considerable amount of energy packed into a very small volume.

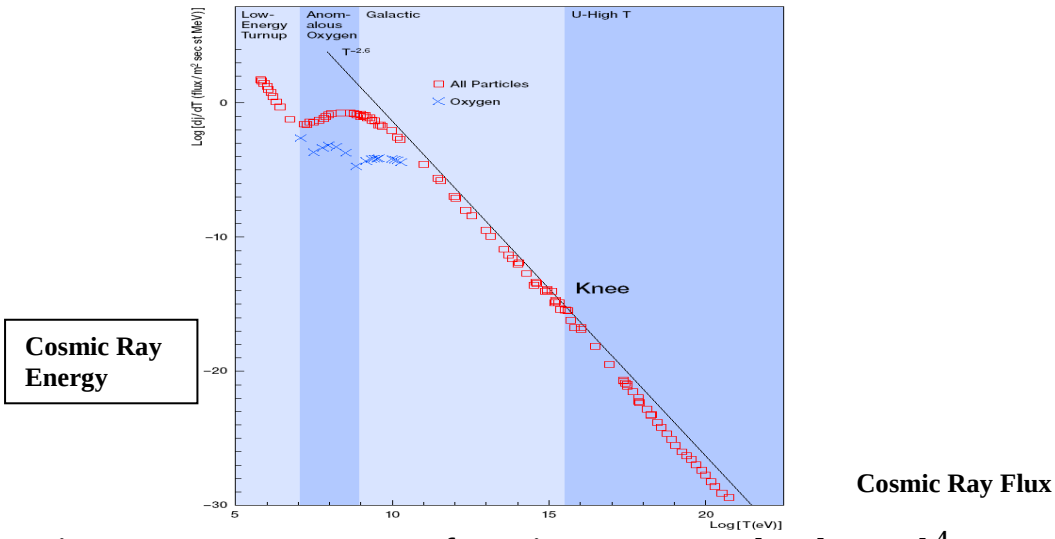

Figure 1 – Energy spectrum of cosmic rays measured at the Earth[4]

It is postulated that the 'Knee' and 'Ankle' in addition to other less significant break points in the energy spectrum are a function of the origin, acceleration and propagation mechanisms of cosmic rays. The 'Knee' point may also reflect the gradual transition in particle composition as the energy level increases. The acceleration of cosmic rays with energies below the 'knee' can be attributed to the interaction of cosmic ray charged particles within the magnetic fields generated by the Sun, solar wind and in the remnants of supernova explosions in our own galaxy, the Milky Way.

For energies above the knee, it is believed that multiple "bounces" off turbulent magnetic fields generated by supernova shock waves could account for energies up to the "Ankle". But beyond this energy level there is no scientific consensus on the acceleration mechanism or origin of cosmic rays with these extremely high energy levels. A range of space phenomena exist that could potentially generate the tremendous energies required to accelerate particles to these ultra high energy levels. Candidate sources as proposed by current astrophysics research are as follows:

- Cores of active galactic nuclei: galaxies that exhibit a substantial release of energy from their core that exceeds the radiation produced from the rest of the entire galaxy. Quasars are a form of distant active galactic nuclei.

- Powerful radio galaxies: type of active galaxy that emits radio waves from its central core.

- Cosmic strings : Theoretical one-dimensional topical defect in the fabric of space-time.

A summary of cosmic ray types, origin and acceleration mechanism ranked by particle energy level is shown in Table 2.

[4] "Energy spectrum of cosmic rays measured at the Earth". Figure retrieved June 24 2007 from "Cosmic Rays" Spatium, published by the International Space Science Institute, No 11,Nov 2003. http://www.issibern.ch/PDF-Files/Spatium_11.pdf

Energy Level (eV)	Cosmic Ray Type, Origin and Acceleration Process
$E < 1 \times 10^9$ eV	**Anomalous cosmic rays:** Possess energies in the region of 10 MeV. Solar cosmic rays are typically below 1 GeV.
Below the Knee $E < 3x10^{15}$ eV	**Galactic cosmic rays:** Galactic Origin, acceleration in magnetic fields of the Sun, solar wind and in shocks waves of supernova remnants.
Above the Knee $3x10^{15} \leq E \leq 10^{18}$ eV	**Galactic Cosmic Rays:** Galactic origin. Secondary acceleration of galactic cosmic rays.
Above $E \geq \approx 10^{18}$ eV	**Extra galactic cosmic rays:** Acceleration in active galactic nuclei, powerful radio galaxies or cosmic strings.

Table 2 – Categories of cosmic ray particles[5]

Modulation of cosmic rays

The intensity of the secondary cosmic ray flux in the atmosphere is not constant because it is influenced by a plethora of solar and terrestrial based mechanisms. The objective of this paper is to provide a summary of these physical processes detailing the magnitude and periodicity of each effect without providing an in-depth description of the physics involved which is outside the scope of it. Thus, the most significant solar and terrestrial based modulating mechanisms is listed in Table 4.

Type of Change	**Magnitude of influence (% Sea level Flux Intensity Variation)**	**Origin of Influence**	**Physical Nature**
Period: 11 year solar cycle	Up to 30%	Solar	Solar modulation of the Earth's magnetosphere reducing the incident flux of Galactic cosmic rays. Resultant 30% reduction in the flux of sea level cosmic rays. - Discussed further in this section -
Period: 27 day	< 2%	Solar & Interplanetary Magnetic Field	Variability in the structure of the IMF or solar wind.
Impulsive – Solar Energetic Particles	1 to 300%	Solar	Potentially dramatic increase of secondary cosmic rays resulting in a Ground Level Enhancement or Event, (GLE) induced by a solar particle event.
Impulsive – Forbush decrease	Up to 30%	Solar	Reduction in Galactic Cosmic rays due to a solar interplanetary shock disrupting the Earth's magnetosphere and creating a condition on Earth known as a geomagnetic storm. This investigated has the affect of temporarily increasing the shielding effect of the Earth's magnetosphere. Decreases usually occur over several hours.

[5] "Categories of cosmic Ray particles". Based on a table retrieved June 24 2007 from "Cosmic Rays" Spatium, published by the International Space Science Institute, No 11,Nov 2003. http://www.issibern.ch/PDF-Files/Spatium_11.pdf

Impulsive – Forbush increase	< 2 %	Solar	Small increase due to a build up of galactic cosmic rays on the bow wave of an interplanetary shock.
Periodic - Seasonal	< 1%	Terrestrial	Seasonal changes in the Earth's atmospheric structure that results in a deviation between the absorption rates of cascade particles.
Periodic - Diurnal	< 1%	Terrestrial	Variation in the Earth's atmospheric structure between day and night that results in a deviation between the absorption rates of cascade particles.
Impulsive - Increase during a geomagnetic storm	Up to 10%	Terrestrial	Reduction in geomagnetic rigidity due to the influence of a geomagnetic storm on the Earth's magnetosphere.

Table 3 – Modulation of Cosmic Rays[6]

Variations in the flux of primary cosmic rays takes place extra-terrestrially prior to cascade creation at the top of the atmosphere and also to secondary cosmic rays within the atmosphere itself. The main source of extra-terrestrial modulation is the Sun that is responsible for periodic and sudden impulsive changes in the flux of primary cosmic rays.
Periodic changes in intensity are caused as a result of the Sun's rotation or solar cycle whereas impulsive random events are initiated by solar flares and coronal mass ejections. Primary cosmic rays are modulated by the Sun's magnetic field that is carried out into the Solar System by the Sun's solar wind. This extension of the Sun's magnetic field is known as the Interplanetary Magnetic Field or IMF that acts on the Earth's magnetic field or magnetosphere compressing one side and stretching the other. The Earth's magnetosphere is composed of electrons and free ions held in place by magnetic and electric forces which behave like a filter for particles with an incident energy below approximately 10 GeV.

These periodic changes to the shape of the Earth's magnetosphere results in an increasing and decreasing flux of galactic comic ray radiation in anti-correlation with the Sun's 11 year solar cycle. During an active Sun the shielding effect of the magnetosphere is increased, reducing the net terrestrial level flux by around 30% in comparison to a quiescent Sun.

Terrestrial changes in intensity are produced by small periodic changes in the structure of the atmosphere and impulsive terrestrial variations are once again caused by events on the Sun.

The concept of Space Weather

Space weather is the term used to describe conditions on the Sun and in the Earth's magnetosphere and atmosphere that can impact either the functionability of man-made systems or human health.

The Sun has a major influence on the radiation environment at aircraft altitudes and on the Earth's surface. This section will review the impact of space weather on the avionics radiation environment and discuss each of the components that make up a solar storm.

[6] "Modulation of cosmic rays", Based on two tables presented in lecture notes fall term 2003,"Heliospheric Physics and Cosmic Rays", Chapter 9, " Variations of Cosmic Ray Intensity ", prepared by Kalevi Mursula and Ilya Usoskin, University of Oulu.

The three constituent elements of a solar storm and their resultant space weather manifestations are shown in Figure 2. The largest solar storms typically generate all three components whereas less powerful storms may not.

Solar Storm Components	Space Weather Effects
Solar Flares	Intense EM Burst
Solar Photon Event	Ground level Events
Coronal Mass Ejection	Geomagnetic Storm

Figure 2 – Space Weather Constituents

Solar flares are magnetically initiated explosions that occur at or near the surface of the Sun that release intense bursts of electro-magnetic radiation in the form of x-rays, ultraviolet and radio emissions that can cause disruptions to the Earth's ionosphere leading to radio and communications interference.

Coronal mass ejections are huge clouds of charged plasma containing particles of low to medium energy levels thrown into space by the Sun. Upon reaching the Earth the charged plasma cloud depresses the Earth's geomagnetic field, producing a disturbance known as a geomagnetic storm. A storms severity is related to the size of the CME and the magnetic orientation between the Earth's and plasma clouds, magnetic fields. Geomagnetic storms are also responsible for a diversity of effects on the Earth ranging from electrical power blackouts to human affects such as heart attacks and strokes.

Finally to provide an appreciation of the temporal characteristics of the Sun's effects on the radiation environment, the differences between the arrival times of each solar storm component will be addressed. Hence:

X-Rays and radio waves travel from the Sun at the same speed as visible light, hence they take approximately 8 minutes to reach Earth. T

The speed of protons during SPEs is dependent on energy level and therefore typically take between 15 minutes to a few hrs to generate atmospheric and ground level particle enhancements.

The solar plasma cloud of CMEs takes between 2 and 4 days to impact the Earth's geomagnetic field and generate a geomagnetic storm that may take several days or even weeks to recover.

Geomagnetic Rigidity

The Earth's magnetic field or magnetosphere is the first line of protection against energetic primary cosmic rays from space and is composed of electrons plus free ions held in place by magnetic and electric forces. This magnetic field surrounding the Earth acts on incoming charged particles like a shield directing particles below a threshold energy level along the magnetic lines of force towards the Polar Regions.

As a result, for each point in the magnetosphere there exists a minimum energy level for a particle with a vertical trajectory to create cascade of particles that will reach sea level. This

energy level is defined as a point's geomagnetic rigidity or cut-off. For particles with a non-vertical trajectory a higher energy level is required for the same location.

Due to the nature and shape of the Earth's magnetosphere the values of geomagnetic cut-off value vary significantly with different latitudes, highest at the equator, approximately 15 GeV, reducing to less than 1 GeV at the poles. Cut-off values also vary with longitude but this affect is much less pronounced than the latitude variation

Secondary Cosmic Rays

Secondary cosmic rays are produced when primary cosmic rays interact with oxygen and nitrogen atoms in the upper atmosphere creating a chain reaction cascade of secondary particles that increases rapidly as the particles move down through the atmosphere. At an altitude of approximately 60,000ft (20 km) known as the Pfotzer point the maximum flux of particles is reached due to the rate of particle absorption exceeding the rate of particle spallation. The small fraction of particles that propagate to the Earth's surface are termed terrestrial cosmic rays and are largely the product of sixth and seventh order primary cosmic ray spallations.

As a general guide the incident primary cosmic ray flux at the top of the atmosphere is about 3 particles per cm^2 per second increasing to a secondary flux maximum of approximately 10 particles per cm^2 at the Pfotzer point before reducing to fewer than 0.1 particles per cm^2 at sea level.

When a highly energetic primary particle at the top of the atmosphere collides with the nucleus of an oxygen or nitrogen atom, it reacts with the strong interaction to create an atmospheric particle cascade consisting of three main components, electromagnetic or "soft", meson or "hard" and nucleonic.

The "soft" electromagnetic component is composed of electrons, positrons and photons that have a stable lifetime and the "hard" component made up from muons and pions that have a very short lifetime, decaying within approximately 2 μs and 26 ns respectively. As a result pions will not reach ground level due to their extremely short lifetime but will decay mainly to muons, the most abundant particle at sea level.

Protons and neutrons constituent the nucleonic component and each interact differently in the atmosphere. Both particles will lose energy through nuclear disintegrations after colliding with atmospheric nuclei but as a charged particle, protons also lose energy to electrons in the atmosphere whereas neutrons that carry no charge do not. This characteristic makes neutrons very penetrating through all forms of material.

In physics there are four discrete fundamental forces, strong, electromagnetic, weak and gravitation that govern the interactions of all matter. A fundamental force describes the type of mechanism and behaviour of particles with each other that cannot be described in terms of another fundamental force. The main fundamental force that controls the propagation and interaction of cascade particles through the atmosphere is the strong interaction, although there are other weaker interactions that also take place.

Each type of particle within a cascade will interact differently with other particles dependent on its inherent properties of mass, life and fundamental interaction type. Table 4 details the characteristic properties of each particle type grouped by cascade component.

Cascade Component	Particle	Interaction Type			Mass (MeV)	Lifetime
		Electro-magnetic	Strong	Weak		
Electro-magnetic	Electrons	✓			0.5	Stable
	Photons	✓			0	Stable
Meson	Pions	✓	✓		≈134	≈26 ns
	Muons	✓		✓	≈106	≈2 µs
Nucleonic	Neutrons		✓		940	12 Min
	Protons	✓	✓		938	Stable

Table 4 – Secondary cosmic ray particles[7]

The resultant distribution of each particle type at a specific atmospheric depth is therefore determined by the complex collisions, interactions and particle decay processes as the cascade moves down through the atmosphere. Within Table 5 the composite particles protons, neutrons and pions, within the class of particles known as hadrons, all interact via the strong interaction and will consequently lose energy much more rapidly than particles that are only acted upon by the electromagnetic and weak forces.

As a result hadrons will reach a maximum flux at the Pfotzer point then continue to lose energy via multiple nuclear collisions until ground level. In contrast the particles without the strong interaction, electrons, photons and muons will relinquish energy to atmospheric electrons much more gradually.

Another attribute of a particle cascade is its shape which can be described as a set of concentric cones, with different spatial widths that defines the particle envelope of each type of cascade component. The inner cone will consist of the heaviest particles the nucleons, followed by pions and muons with the lightest and easiest scattered electromagnetic components spread out the widest.

The absolute width of each cone is dependent on the energy of the incident particle, the higher the incident energy the greater the size of each component of the cascade.

High Energy Cosmic Ray Neutrons

As neutrons posses no charge they are very penetrating and in most cases pass straight through a material completely unhindered. For example 140cm of concrete only attenuates the neutron flux by 50% [?]. Neutrons therefore can only cause ionisation within a silicon semiconductor through indirect processes whereas charged particles can interact directly with the silicon. The Linear Energy Transfer, LET, of silicon reaction products caused by an incident high energy neutron is also much higher than the LET of an incident alpha particle.

This also means that soft error effects such as MBU and SEL are generally caused only by high energy neutron impacts because the LET threshold of approximately 16 fC/µm needed to induce these failure mechanisms cannot be generated by alpha particles,(fC - units denote

[7] "Secondary cosmic ray particles" derived from a table retrieved September 02 2008 from " Terrestrial cosmic ray intensities" by J. F. Ziegler, http://www.research.ibm.com/journal/rd/421/ziegler.html

10^{-15} coulombs). As a result incident neutrons pose a much greater upset risk to semiconductors than alpha particles.

Thermal Neutrons

High-energy neutrons lose energy in collisions with atomic nuclei and disperse throughout the aircraft reaching an energy level where they are in thermal equilibrium with the local environment. At normal room temperature this equates to a kinetic energy of approximately 0.025 eV. For the purposes of this dissertation any low energy neutron of less than 1eV will be classified as a thermal neutron.

In comparison with the atmospheric thermal neutron flux the flux level inside the aircraft is increased by about an order of magnitude and varies dependent on location due to the different composition and distribution of materials.

Low Energy Alpha Particles

An alpha particle is a doubly ionised helium atom consisting of two neutrons and two protons, which can also be described as a helium atom, which has been stripped of its electrons. When an alpha particle travels through a material it will lose kinetic energy primarily through interactions with the materials electrons, leaving a trail of atoms with ' kicked out ' orbital valence electrons. This process is called ionisation, which can be described as the physical process of converting an atom or molecule, into a positively or negatively charged state by either adding or removing charged particles. The resulting atom is then referred to as an ion, or more specifically a cation if positively charged or a anion if negatively charged.

Low energy alpha particles are emitted from the decay of trace radioactive materials in semi-conductor device and packing materials. The most common source of radioactive impurities is naturally occurring uranium-238, uranium-235 and thorium- 232. Within a material these impurities are typically evenly distributed and emit alpha particles at specific discrete energy levels, resulting in a characteristically broad energy spectrum between a range of 4 to 9 MeV.

The distance an alpha particle travels in a material before it is stopped, referred to as its 'range' is therefore determined by the energy of the incident particle and the physical properties of the material, principally density. In silicon, alpha particles with an energy of 10 MeV, only have a range of < 100 μM due to their relatively large atomic size. As a result of this short range of travel within a material and the ability of surrounding structures to easily shield out external sources of alpha particles only alpha particles actually emitted from the device itself and its packaging materials should be investigated as a potential upset threat.

As a result alpha particle induced soft errors, have a much smaller significance than high energy or thermal neutrons, due to the improved purity and alpha particle screening measures now employed by component manufactures. The emphasis of the research in this dissertation will therefore focus on quantifying the influence of cosmic radiation particles on avionics. High energy and thermal neutron flux rates are highly dependent on many factors, such as: time of day, date, altitude and geographic location, whereas the alpha particle flux is solely dependent on the concentration and position of impurities within the device and package.

Conclusion

The main objective of this paper is to show that the Mirce-mechanics scientific approach to understanding of the mechanisms that cause occurrences of functionability events of the surrounding natural environment is required for the accurate predictions regarding occurrences of negative functionability events are to be made. Then and only then, the reduction of the probability of the occurrence of failure events during the life of man made, managed and maintained systems could be achieved. This paper focuses on the scientific understandings of the physical mechanisms originated by the cosmic phenomena.

As science is the proved model of reality that is confirmed through observation, the summary message of this paper to reliability professionals is to move from then universe in which the laws of science are suspended to the universe that is based on the laws of science in order for their predictions to become future realities.

References

Baumann, R., " Radiation-induced soft errors in advanced semiconductor technologies, " IEEE Transactions on Device and Materials Reliability, vol 5, No 3, pp. 305–316, Sept. 2005.

Knezevic, J. Physical Scale of Mirce-Mechanics," Lecture Notice, Master Diploma Programme, MIRCE Akademy, Woodbury Park, Exeter, UK, 2009.

Knezevic, J/, Atoms and Molecules in Mirce-mechanics Approach to Reliability, SRESA Journal of Life Cycle Reliability and Safety Engineering, Vol 1, Issue 1, pp 15-25, Mumbai, India, 2012. ISSN-22500820

Knezevic, J., Functionability in Motion, Proceedings 10th International Conference on Dependability and Quality, DQM Institute, 2010, Belgrade, Serbia.

Zaczyk. I, "Analysis of the Influence of Atmospheric Radiation Induced Single Event Effects on Avionics Failures", Master Dissertation, MIRCE Akademy, Exeter, UK, 2010.

C. H. Tsao, R. Silberberg, and J. R. Letaw, "Cosmic ray heavy ions at and above 40,000 feet, "IEEE Trans. Nucl. Sci., vol. 31, pp. 1066–1068, Dec 1984.

R. Silberberg, C. H. Tsao, and J. R Letaw, "Neutron generated single event upsets," IEEE Trans. Nucl. Sci., vol. 31, pp. 1183–1185, Dec 1984.

"Cosmic Rays" Spatium, published by the International Space Science Institute, No 11, Nov 2003. **http://www.issibern.ch/PDF-Files/Spatium_11.pdf**

Marusek, J., "Solar storm threat analysis, " Impact 2007, Bloomfield, Indiana 47424.

Lei. F., Clucas. S., Dyer. C., "An atmospheric radiation model based on response matrices generated by detailed Monte Carlo simulations of cosmic ray interactions, " IEEE Trans. Nucl. Sci, vol. 51, no. 6, pp. 3442–3451, Dec. 2004, Part II.

Mursula, K., Usoskin, I., Chapter 4 Lecture notes fall term 2003, "Heliospheric Physics and Cosmic Rays ",prepared by University of Oulu., retrieved September 02 2008.

Managing Machine Functionability using Methods of Complexity Science

George Rzevski
Professor Emeritus, Centre for Complexity Science Applications, The Open University, UK.
Executive Chairman, Multi-Agent Technology Group, London, UK

Abstract

The implication of Mirce-mechanics is that functionability trajectory patterns, driven by failure and repair events, are complex and therefore must be managed using methods of Complexity Science. Traditional methods for reliability and availability assessment and for planning and scheduling of in-service activities, spare parts supply and logistics are inadequate, as shown in [4], [5]. The new approach described in this paper, has been developed taking into account uniqueness of individual machine failure and repair patterns, component failure interdependencies and dependence of failure and repair patterns on changing operating and in-service conditions. To facilitate understanding of principles of the new approach an outline of the concept of Complexity is provided and references given to key developments in Complexity Science. In order to follow the author's arguments the readers will have to adjust their mindsets so that they can clearly distinguish between problems that can be solved using classical, Newtonian science and those that require a completely new approach – the Complexity Thinking.

Introduction

The paper describes a new research project undertaken by the Multi-Agent Technology Group, London in collaboration with MIRCE Akademy, with the objectives:

- To construct a large-scale software model of an individual machine life as a function of component failures and repairs
- To simulate machine life under different operational and servicing conditions, which affect component failures and repairs
- To develop an intelligent management system, which would ensure high levels of machine functionability (the capability to function)

Results of this research are applicable to all large engineering systems, including, machine (civil and military), vehicles (Formula 1 cars, tanks and heavy trucks) and communication equipment (civil and military).
The modeling, simulation and management methodology has been developed by the author based on his long-term research into methods for modeling, simulating and managing complex systems and processes [1], [2], [3].
Failure analysis adopted in this project is based on Mirce-mechanics, as published by Knezevic [4], [5], which has established that:

1. Each machine has a unique individual failure pattern, which depends on how it is used, how it is maintained and how it was assembled.

2. Machine failures are interdependent; a failure of one component may cause failures of others depending on proximity of locations and functional links.

3. Component failures are unpredictable but not random; patterns of failures can be discovered from failure data; these patterns are not permanent – they change as operating conditions of the machine change.

These findings clearly show that current reliability assessment methods are inadequate and therefore a new approach is required.

Consider the following:

- If each individual machine has a unique life, is it correct to assess reliability and organize identical maintenance for all machines manufactured to the same design?
- If failures are interdependent, is it correct to calculate failure rates ignoring spatial interdependence?
- If a pattern of failures change in time does it make sense to prescribe maintenance procedures for the whole life of the machine?

The potencial readers of this paper are managers and engineers engaged in engineering systems design, reliability and availability assessment, servicing, spare parts supply and logistics. It will be of interest to those working for manufacturers or operating organizations, as well as those engaged in academic study of reliability and logistics.

Why a New Approach?

The implication of Mirce-mechanics is that failure and repair patterns are complex and therefore must be managed using methods of Complexity Science, as defined in the next section.

Traditional methods for reliability and availability assessment and for planning and scheduling of in-service activities, spare parts supply and logistics are inadequate, as shown in [4], [5].

The new approach described in this paper, has been developed taking into account uniqueness of individual machine failure and repair patterns, component failure interdependencies and dependence of failure and repair patterns on changing operating and in-service conditions. To facilitate understanding of principles of the new approach an outline of the concept of Complexity is provided and references given to key developments in Complexity Science.

In order to follow the author's arguments the readers will have to adjust their mindsets so that they can clearly distinguish between problems that can be solved using classical, Newtonian science and those that require a completely new approach – the Complexity Thinking.

The Concept of Complexity

A system is complex if it consists of a large variety of autonomous components engaged in interaction. The global behavior of such a system is unpredictable but not random – it *emerges* from interconnected local behaviors of constituent components and follows discernible patterns.

Complex systems are nonlinear - a small disturbance may cause large changes in their global behavior, whilst large disturbances may be unnoticeable.
The following three paragraphs from Wikipedia are a good introduction to the concept of complexity.

“Complexity has always been a part of our environment, and therefore many scientific fields have dealt with complex systems and phenomena. Indeed, some would say that only what is somehow complex – what displays variation without being random – is worthy of interest.

The use of the term complex is often confused with the term complicated. To understand the differences, it is best to examine the roots of the two words. “Complicated” uses the Latin ending “plic” that means, “to fold” while “complex” uses the “plex” that means, “to weave.” Thus, a complicated structure is one that is folded with hidden facets and stuffed into a smaller space. On the other hand, a complex structure uses interwoven components that introduce mutual dependencies and produce more than a sum of the parts. This means that complex is the opposite of independent, while complicated is the opposite of simple.
While this has led some fields to come up with specific definitions of complexity, there is a more recent movement to regroup observations from different fields to study complexity in itself, whether it appears in anthills, human brains, or stock markets.”

Let us consider this difference between Complex and Complicated in more details.
A machine, when it functions as specified, is a Complicated rather than a Complex System because the interactions between its components are well defined and predictable. The overall behavior of a functioning machine, which is a result of a very large number of predictable interactions between its components, is also predictable.
However, when we consider functionability of a machine during its life, then we have to take into account its failure and repair patterns, which are both unpredictable because they comprise unpredictable failures and unpredictable failure interactions as well as unpredictable repair patterns. Therefore the life of a machine, and its functionability, is a Complex System.

It is important to note that with time the frequency of occurrence of unpredictable events that affect functionability of machines tend to increase. This tendency is particularly evident in warfare, but it is also present in civil applications. There is evidence that as evolution of the Universe takes its course, the complexity of ecological, economic, business and social systems increases.

The often-quoted statement made by Stephen Hawking at the end of the 20th century best illustrates the importance of the increasing complexity of our environment: "*I think the next century will be the century of complexity*."

In a classification of systems according to their predictability, complex systems are between random and deterministic systems, as shown in Fig. 1.

RANDOM	COMPLEX	DETERMINISTIC
Uncertainty = 1	1 > Uncertainty > 0	Uncertainty = 0

Components have full autonomy	Components (called agents) have partial autonomy	Components have no autonomy
Disorganised	Self-organising, Evolving	Organised
Unpredictable behaviour	Emergent behaviour	Predictable behaviour

Figure 1. Complexity and Uncertainty

Physics, often referred to as the queen of sciences, and was for centuries preoccupied with studies of systems in equilibrium, such as the movement of solid bodies and fluids. The laws governing the behavior of such systems are valid at any place and any time and are reversible. The elegance and power of these "natural laws" led many scientists to believe that, with the advancement of science, it will be possible to reduce the understanding of all systems to a similar set of simple and logical laws. This notion is known as Reductionism. Recent research, following the work of Prigogine, has shown that, to the contrary, in physics like in all other branches of science, majority of interesting phenomena are complex, consisting of richly interlinked elements, irreversibly co-evolving with their environments, and are not reducible to simple laws.

The Science of Complexity

The Nobel Laureate in chemistry, 1977, Ilya Prigogine is generally recognized as the father of Complexity Science. In his books "Is Future Given?" [6] and "The End of Certainty: Time, Chaos and the new Laws of Nature" [7], he contrasts his discoveries with notions of classical science as follows.

Classical science, as represented by Newtonian dynamics, is deterministic. The world is, it always was, and always will be, the same; just as argued by the classical Greek philosopher Parmenides. All natural laws are permanent and reversible and independent of time and space. The key scientific concepts are equilibrium, stability and predictability.
Prigogine argued that this view of the world in which we live is rather restrictive. It is valid only for a narrow domain of the world, such as Mechanics. Newtonian science does not concern itself with major domains of interest - what we need is the science that can explain behavior of systems consisting of physical, chemical, biological and socio-technical systems. A broader perspective is required to incorporate into the science readily observable phenomena of change, evolution and unpredictability of events that may or may not occur.

According to Prigogine, Future is not Given, it is constructed in front of our eyes by billions of actions and decisions made by natural or artificial Agents operating in the Universe. The Universe is perpetually changing and constituent systems co-evolve affecting each other.
The idea of the world in perpetual change is by no means new. Heraclitus (535 – 475 bc) believed that "No man ever steps in the same river twice" because between two attempts both the man and the river changed.

The Seven Criteria of Complexity

For the purposes of my research and development I define complexity using the following seven criteria [1]:

1. CONNECTIVITY - A system consists of a large number of diverse components, referred to as Agents, which are richly interconnected.
2. AUTONOMY - Agents are not centrally controlled; they have a degree of autonomy but their behaviour is always subject to certain laws, rules or norms.
3. EMERGENCE - Global behaviour of a complex system emerges from the interaction of agents and is therefore unpredictable but not random; it generally follows discernable patterns.
4. NONEQUILIBRIUM - Global behaviour of a complex system is "far from equilibrium" because frequent occurrences of disruptive events do not allow the system to return to the equilibrium between two disruptive events.
5. NONLINEARITY - Relations between agents are nonlinear, which occasionally causes an insignificant input to be amplified into an extreme event (butterfly effect).
6. SELF-ORGANISATION - A system is capable of self-organizing in response to disruptive events, a feature termed Adaptability. Self-organisation may also be initiated autonomously by the system in response to a perceived need, a feature termed Creativity.
7. CO-EVOLUTION - A system irreversibly coevolves with its environment.

From above discussions it should be obvious why classical Newtonian science is perfectly adequate for designing and constructing machines and yet completely inadequate if applied to the prediction of the functionability trajectory, of the very same machine.

Managing Complexity

If we accept that "to control" means to specify a desirable behavior of a system and to steer the system towards achieving it, then complex systems cannot be controlled. The very concept of "emergent behavior" precludes controllability.
Fortunately we are not helpless in the presence of complexity, the complexity science gives us tools for analyzing complexity issues, for planning how to react to unpredictable disruptive events, for designing adaptability into social, business or technological processes at hand, and for "tuning" complexity using experimentally derived heuristics.
Let us call these activities "Managing Complexity".
In this project we consider a particular complex system, Machine Life, and apply our experience in managing complexity to Managing Machine Life.

Managing Machine Life

Managing Machine Life amounts to Managing Functionability, that is, maintaining the capability of a system to function under conditions of uncertainty at the specified level.
In practical terms it means achieving the repair of any failure, whenever and wherever it occurred, within specified time and within specified cost.
For effective Functionability Management we must (a) have clearly specified measures of Functionability; (b) conduct a thorough Failure Analysis, which will yield the probable failure pattern for the system under investigation; and (c) design and run in real time an effective Functionability Management System capable of maintaining the specified level of Functionability under conditions of uncertainty.

Functionability Measures

The key measures of Functionability are *cost* and *time* required returning the system after a failure to the functioning state. These two criteria are used to determine performance of Functionability Management Systems. Clearly there is always a trade-off between the cost of functionability and cost of performance of machine. The appropriate balance is found at the point that offers the maximum Value for the machine user.

Failure and Repair Analysis

The objective of Failure and Repair Analysis is to discover the probable failure and repair pattern of a system.
As an example, let us consider Failure and Repair Analysis of an individual machine within its particular operating and servicing environment using seven criteria of complexity.

Connectivity

First of all, machine consists of a very large number of components that can fail and whose failures are interdependent due to their functional or proximity relations. The objective of Failure and Repair Analysis is to establish, using Functionability Mechanics, all possible failures and failure interdependencies, as well as conditions under which they are likely to occur.

Secondly, machine is operated and serviced by a large number of personnel many of whom cooperate and occasionally compete with each other. The objective of Failure and Repair Analysis is to establish human factor effects on failures, on failure dependencies and on repairs.

Finally, failures are dependent on external factors such as prevailing climate and other environmental conditions (e.g., snow, ice, or sand) at airports. The objective of Failure and Repair Analysis is to establish how these external factors affect failures and failure dependencies.

Autonomy

Autonomy is the capability of making decisions without being instructed, under given constraints, or within given rules. Physical components of machine have no functional autonomy but they appear to have pseudo autonomy of failure (because they fail without being instructed to do so); intelligent components (e.g., software) may have a limited autonomy whilst personnel operating and servicing machine may have a considerable autonomy. The objective of Failure and Repair Analysis is to establish autonomy of every Functionability component and how this autonomy affects failures and repairs.

Emergence

The actual machine failure pattern emerges from the unpredictable occurrences of individual failures and is obviously unpredictable. The objective of Failure and Repair Analysis is to establish *probable* patterns of failures and repairs for each individual machine.

Nonequilibrium

An analogy can be construed between a dynamic system that may have no time to return to equilibrium between two disruptive events and machine functionability, which may not return to the original value after a failure is repaired. The objective of

Failure Analysis is to establish effects of this feature on the occurrence of failures and on activities of repair.

Nonlinearity

A relatively insignificant failure (such as a tire damage) may cause a catastrophic event (such as fuel leakage, fire and machine explosion). The objective of Failure Analysis is to identify possible catastrophic consequences of failures and to pint out how to prevent them.

Self-Organization

In reaction to a failure some machine subsystems are capable of autonomously reconfiguring with a view to reducing or eliminating failure consequences. The objective of Failure and Repair Analysis is to identify self-organizing activities and their effects on failures and repairs. The classic example of self-organization is the "fail-safe" capabilities of some systems.

Co-Evolution

Since the occurrence of machine failures depends of operating and servicing conditions, the failure and repair patterns will change as these conditions change. In other words, failure and repair patterns of an individual machine and its functionability co-evolve with its operational and servicing environment.

Tools are required for simulating the occurrence of failures under varying conditions, taking into account interdependence, co-evolution, autonomy, behavior far from equilibrium, nonlinearity, self-organization and emergence, to yield probable failure patterns.

Functionability Management Systems

A machine in each in-service environment (war, peace, training and other scenarios) should have a customized Functionability Management System, designed to optimize Functionability of the machine concerned.
To illustrate what needs to be done to achieve the effective Functionability Management let us continue considering in-service life of a machine.
To design an in-service system of a machine, which will ensure maximum functionability, which means: minimum repair time within specified cost constraint for a fleet of machines each with a different, evolving probable failure pattern it is necessary to determine:

- Skills and numbers of servicing personnel and their locations
- The range and quantity of spare parts and their locations
- Mode of transport and types and quantity of transportation resources

The design can be done using trial-and-error method supported by design simulation tools capable of conducting experiments with alternative in-service configurations and different probable failure patterns.
In order to achieve desired functionability level, the machine in-service system must be capable of operating in real time, i.e., capable of deciding on and delivering servicing personnel and spare parts to the servicing location within specified time.

Tools for Functionability Management

To manage functionability of a large fleet of machines in real time under conditions of unforeseen failures, stretched servicing capacities, delays in supply chains and limited availability of spare parts, there is a need for powerful, intelligent software tools capable of autonomous and rapid decision making. The problem is *too complex* for conventional planners and schedulers.
The most appropriate technology is Ontology-Driven Multi-Agent Software.

Multi-Agent Software

The author and his team have worked on the development and delivery of advanced agent-based technology and applications for the last 15 years. A very large number of our multi-agent applications are currently deployed in businesses, administrations, space exploration and healthcare. Few papers describing special features of technology are included in references [8], [9], [10], [11], [12], [13], [14].
Here are some simple explanations. Software Agents are small computer programs that are capable of accomplishing their goals under conditions of uncertainty through the interaction with other intelligent agents or humans; an agent is created when needed and it achieves its goals by:

- Analysing its current task
- Composing messages (for other agents or humans)
- Sending messages to selected correspondents
- Receiving messages (from other agents or humans)
- Interpreting received messages
- Deciding how to react to received messages
- Acting upon its decision

When an agent accomplishes its mission it is destroyed.
AGENTS are organised in SWARMS, which may contain thousands of agents that exchange messages among themselves with a view to solving a complex problem such as clustering of large number of data elements. Agent actions are driven by the occurrence of events such as the arrival of a new data element. Decisions how to react to an event are made by negotiation among agents affected by the occurrence of that event.
Each agent acts upon local information and is aiming at achieving own task and to accommodate needs of other agents. The overall solution emerges from the agent interaction. Agents are guided by the problem domain knowledge collected in a Knowledge Base, which among other elements, contains rules that constrain freedom of agent behaviour and yet leave agents sufficient freedom to act creatively.

References

1. Rzevski, G., "A practical Methodology for Managing Complexity". *Emergence: Complexity & Organization – An International Transdisciplinary Journal of Complex Social Systems.* Volume 13, Nos. 1-2, 2011, ISSN 1521-3250, pp. 38-56.
2. Rzevski, G., "Using Tools of Complexity Science to Diagnose the Current Financial Crisis". ISSN 8756-6990, *Optoelectronics, Instrumentation and Data Processing*, 2010, Vol.46, No. 2.
3. Rzevski, G., "Using Complexity Science Framework and Multi-Agent Technology in Design". In Alexiou, K., Johnson, J., Zamenopoulos, T. (eds.), *Embracing Complexity in Design*, Routledge, 2010, ISBN 978-0-415-49700-8, pp 61-72.

4. Knezevic, J., “Reliability, Maintainability and Supportability: A probabilistic Approach”. McGraw-Hill, 1993
5. Knezevic, J., “Systems Maintainability: Analysis, Engineering and Management” Springer, 1997.
6. Prigogine, I. (2003). “Is Future Given?” ISBN 9789812385086.
7. Prigogine, I. (1997). “The End of Certainty: Time, Chaos and the new Laws of Nature”, ISBN 9780684837055.
8. Rzevski, G., “Modelling Large Complex Systems Using Multi-Agent Technology”. Kyoto Conf. August 2012.
9. Madsen, B., Rzevski, G., Skobelev, P., Tsarev, A. “Real-time Multi-Agent Forecasting & Replenishment Solution for LEGO Branded Retail Outlets”. Kyoto Conf. August 2012.
10. Rzevski, G. “A New Direction of Research into Artificial Intelligence”, Invited Paper, *The Annual Conference of Sri Lankan Association for Artificial Intelligence*, Colombo, 2008.
11. Rzevski, G., Skobelev, P., “Emergent Intelligence in Large Scale Multi-Agent Systems”. *International Journal of Education and Information Technology*, Issue 2, Volume 1, 2007, pp 64-71.
12. Rzevski, G., Skobelev, P. Andreev, V. “MagentaToolkit: A Set of Multi-Agent Tools for Developing Adaptive Real-Time Applications”. *Lecture Notes in Computer Science, Volume 4659, Holonic and Multi-Agent Systems for Manufacturing. Third International Conference on Industrial Applications of Holonic and Multi-Agent Systems, HoloMAS 2007*, Regensburg, Germany, Springer, ISBN 978-3-540-74478-8, pp 303-314.
13. Glaschenko, A., Ivaschenko, A., Rzevski, G., Skobelev, P. “Multi-Agent Real Time Scheduling System for Taxi Companies”. *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary. ISBN: 978-0-9817381-6-1, pp 29-35.
14. Andreev, S., Rzevski, G., Shveykin, P., Skobelev, P., Yankov, I. “Multi-Agent Scheduler for Rent-A-Car Companies”. **”.** *Lecture Notes in Computer Science, Volume 5696, Holonic and Multi-Agent Systems for Manufacturing: Forth International Conference on Industrial Applications of Holonic and Multi-Agent Systems, HoloMAS 2009, Linz, Austria.* Springer, ISBN 978-3-540-74478-8, pp. 305-314.

Human Effectiveness of Troubleshooting Process in Commercial Aviation

J.G. Hessburg, J. Knezevic, MIRCE Akademy, Exeter, UK

Abstract

As with anything involving human factors, everybody has good days and bad ones from the point of view of troubleshooting. Additionally, the short turnaround times during the operating day do not allow for in-depth analysis before taking action to resolve an elusive system problem. It gets down to a business decision whether to delay this and successive flights in order to go through detailed troubleshooting (which will probably be unsuccessful anyway), or take a chance of incurring an NFF charge by replacing the most likely part. In almost every case, it is more cost-effective to pay the NFF. Then there are those systems with "built-in-test" assistance, which is often erroneous and/or misleading. The needle of truth may be buried in a haystack of other messages. These factors also add to the NFF rate. The main causes of imperfect troubleshooting are addressed in this paper, based on, over 40 years, of the author's experience in commercial aviation.

1. Introduction

In 1996 Boeing stated that they believed there was a 40% rate of incorrect parts removal from the airframe. In 2007 Aviation Week reported that avionics constitutes 75% of No Fault Found, NFF, occurrences in the airline industry.

As with anything involving human factors, everybody has good troubleshooting days and bad ones. Additionally, the short turnaround times during the operating day do not allow for in-depth analysis before taking action to resolve an elusive system problem. It gets down to a business decision whether to delay this and successive flights in order to go through detailed troubleshooting (which will probably be unsuccessful anyway), or take a chance of incurring an No fault Found NFF charge by replacing the most likely part. In almost every case, it is more cost-effective to pay the NFF.

Then there are those systems with "built-in-test" assistance, which is often erroneous and/or misleading. The needle of truth may be buried in a haystack of other messages. These factors also add to the NFF rate.

2. Imperfect bench check

The biggest contributor to NFF is the component bench check. Rather than being the precise measurement of component functionality, the bench check does not mimic the component's complete operation in the system. It is nearly impossible to check everything it does in every situation while interfacing with every associated system part. So if the test approaches 90 percent of system operational functions, it will leave a 10 percent margin of functional failures that can occur without any means to identify on the bench.

Furthermore, the conditions of the bench test are very much different that the operational environment of the component. I'm reminded of a control panel that would not exhibit failure until it was actually fastened into a rack. The standard bench test did not direct installation of the component as it was fitted into the aircraft, so it became a rogue unit. A rogue unit is a component that repeatedly experiences short service periods, exhibiting the same system fault each time, whose failure cannot be detected by standard bench test or overhaul procedures, and whose replacement on the aircraft resolves the systems fault. In essence, it is a defective unit that never gets fixed, no matter what is done to identify and resolve the problem.

The most insidious thing about rogue units is that there is a "natural selection" phenomenon that ensures they will displace the serviceable spares. It's pretty much like a Darwin thing, but instead of survival of the fittest, it's more like survival of the worst. The gathering place for these rogue units is the spare pool-after all, being defective, they won't stay on the aircraft, so where else can they go? Furthermore, if they are not recognised and resolved, the spare pool will become more and more polluted until there are no spares that are naturally serviceable.

When rogue units develop in an airline's inventory, the NFF rate skyrockets. The way if happens follows: During the troubleshooting of an elusive problem, the most likely part is installed, and it is a rogue unit pulled from the spare pool. The system complaint continues because a defective part has been introduced. Logical troubleshooting trees don't direct replacement of the same part again, so the other related components are replace systematically, to no avail. All these related parts score NFF in the shop, because there was nothing wrong with them in the first place. When the rogue unit is finally replaced with a "good" spare, it goes to the repair facility to score NFF, then back to stock waiting to wreak havoc again.

Let's make this nightmare worse: Think of what happens if, during the course of replacing the related parts, another rogue unit was installed- now there are two bad parts causing system problems (the originally installed rogue unit plus this new one). This troubleshooting nightmare will go on for a very long time, because it defies logic that there could be two defective parts causing a single problem.

Again this goes back to the inability of the shop to perform a bench check that mimics the complete operating functions and environmental characteristics of the component when it is installed. It is not an intentional situation, but a fact of life.

A few more facts of life: Every component has the opportunity to become a rogue unit. I've even seen rogue antennas! Also, there will always be rogue units, because over the course of time new and bizarre failures will occur in that unchecked region of the test.

The key to recognising rogue units is to track each component by serial number, showing the date installed and removed the aircraft on which it was installed, the flight hours and cycles accrued hours since overhaul, and a reason for removal code. The removal codes should be very simple-if hundreds of codes are offered, human factors will reduce it to only a handful anyway.

Additionally, the aircraft history needs to be archived and easily retrieved in order to determine the exact system effect of the component's failure and whether its replacement

resolved the problem. Finally, it would be good to have the shop history in a database in order to determine the work performed during previous shop visits.

If you don't track by serial number, disregard this whole section. You will never be able to identify rogue units, or determine component work scope, aircraft maintenance program shortcomings, or whether an aircraft or component reliability modification is really needed. Your best course of action would be to send everybody to brainwashing sessions so they believe they work at "Utopia Airlines," and the OEMs really know everything about their components and what is best for the airline operator.

3. What's it good for?

In a quick review, there are circumstances driving NFF that are completely out of control of the line maintenance process; there are times when generating an NFF is an excellent business decision and the component bench check is not complete. All facts considered, it is obviously a poor measurement of line maintenance effectiveness. This is not to say it is completely worthless. A high NFF rate should prompt an investigation to see what is driving it-whether it is line, the shop, or both.

4. Measuring Troubleshooting Effectiveness

The only good measurement of the effectiveness of both line maintenance and the shop is the aircraft system itself. When a part is replaced that fixes the problem on the aircraft, the evaluations can begin. The line maintenance process can be evaluated by how long it took to arrive at the resolution, as well as the shop's ability to identify that component's failure and repair it.

When the line process appears lacking, the next step is to determine whether it is because the problem was nearly impossible to troubleshoot, the built-in test lied, more analytic training is needed, or the system is just one of those prone to business decisions to keep the aircraft moving.

If the shop arrives at NFF for a component that finally fixed the system problem, then special techniques need to be employed to duplicate the problem. Obviously the test is lacking, so new experimentation must be utilised to duplicate the system operation and environment.

A word of caution regarding this experiment: If the unit doesn't get hot or freeze in normal operation, don't expose it to those extremes. If it doesn't vibrate like a paint shaker, don't do it. If it is failing in the normal parameters of its operation, duplicate those. Any extremes would be considered performance of highly accelerated life testing, which is intended to develop premature failure of subassemblies within the component. When using these desperate measures you'll find a problem, but the rogue failure will probably still be there.

The bottom line is that there is no one number that tells the whole story. NFF, mean time between unscheduled removals (MTBUR), mean time between failures (MTBF), or any averaged value will most often mislead rather than be an absolute guide.

5. Conclusion

No fault Found is here to stay. The only way to get a grip of it, at "Real Life Airlines", is to:

- Gather aircraft history;
- Track component installations and removals as described above
- Develop methodology to link these data during analysis.

The closing statement from jack Hessburg is:

"*Do not test a device unless you duplicate the environment in which it is tested. We have these remarkable test benches that don't shake, rattle or roll; don't go to 41,000 ft; don't have coffee spilled on them and the list is endless"*

6. Acknowledgement

The main part of this paper is based on the original research performed by Jack Hessburg, and communicated to the staff and students of the MIRCE Akademy through numerous lectures and discussions that has taken place since 1999. However, the final touches were made during our brief discussion on the subject in July this year when I visited him at Seattle Post Acute Care Hospital. Sadly, Jack passed away in August 2013.

7. References

Hessburg, J.G., Air Carrier MRO Handbook, pp 400, McGraw-Hill, New York, 2001, ISBN 0-07-136133-2

Knezevic, J., Systems Maintainability, Analysis, Engineering and Management, Chapman & Hall, pp 400, 1997, UK. ISBN 0-412-80270-8

Hockley, C.J., Disruption of Supportability Mechanics and Through Life Engineering Services by the Problem of No Fault Found (NFF). Paper presented at the 21st International MIRCE Symposium, 5-7 Dec 2011, Exeter, UK.

No Fault Found and Air Safety

Christopher J Hockley OBE, CEng MRAeS,
Centre for Through-Life Engineering Services, Cranfield University, Bedford, UK.

Abstract

*There is a view that has been expressed in some organizations that No Fault Found, NFF, is not an air safety issue. Consequently the occurrence of NFF and the rates for a particular fleet do not get the attention that they deserve in these organisations. In this paper it is shown that there is a distinct similarity between maintenance errors that could cause accidents and NFF causes and their impact on air safety. It is concluded that NFF needs a higher profile and the acknowledgement that it certainly **is** an air safety issue.*

What is No fault Found (NFF)?

In considering whether NFF is an air safety issue it must first be established what is meant by NFF. In simple terms when a fault occurs and the cause cannot be found or duplicated when the diagnosis is carried out, and the system is then tested and passed serviceable; this is recorded as NFF. Subsequently, however, the same fault may re-occur in the next or a subsequent mission. Individual items, components and Line Replaceable Units (LRUs) may then be removed as part of the diagnosis and replaced with a known serviceable item and the system tested satisfactorily; however, now another NFF may occur when the item is tested further down the support chain at another maintenance facility and no fault is apparent. All these instances incur costs of one sort or another, such as the nugatory maintenance performed with no fault confirmed or with transport and handling costs throughout the support chain. This drives up the cost of Through-life Engineering Services and Support and solutions must be found if these damaging in-service costs are to be reduced throughout equipment's life.

There are many and varied causes for NFF ranging from organisational, procedural, process and even behavioural issues, to the more obvious original design faults that do not cope with the current operational conditions and changes in usage. These are just some of the situations that contribute to NFF and the costs can be huge. The onus for solutions though surely lies with engineers and the maintenance organisations.

What is Maintenance?

We are all familiar with the concept of maintenance and understand it is all about inspecting and servicing equipment to ensure they are able to be put back into service in a fit condition to last until the next maintenance intervention. Most people also associate maintenance with finding failures and repairing them. Indeed definitions in the air environment indicate that there is a very necessary and vital flight safety link with maintenance and the need for it to be undertaken; definitions will cite the need for maintenance to ensure or restore "aircraft integrity". Consequently the view expressed by Jack Hessburg[8] should certainly be

[8] Jack Hessburg (1934 -2013) was Boeing Chief Mechanic responsibility for all maintenance design during development of the Boeing 777.

accepted which is that maintenance is therefore “nothing more than the management of failures. (Hessburg J, 2013)

It is important to note of course that the management of failures is driven by the consequences of the occurrence of failures. These are:

- The impact on safety
- The impact on operational availability

Both are vital and the impact on safety receives huge attention and rightly so. The impact on availability, however receives more attention in commercial aviation where delay and cancellation cost money and reputation. In military aviation it is, however, beginning to receive more attention as resources and numbers of aircraft are reduced and more commercial ways are found to provide support and availability. Yet the whole process of the management of failures and the need for maintenance is different between military aircraft and civilian aircraft. Whilst both will consider the impact on safety in the same way, the impact on availability will be largely economic for the airline industry but will be driven by the need for a battle-winning edge in the military. This also produces subtly different cultures and behaviours between these two groups. The need to achieve dispatch reliability for an airline will be paramount as the economic consequences of delays, or worse, cancellations, can be very damaging. Consequently, the maintenance staff will do almost anything to achieve the minimum delay when faced with a failure “at the gate.” The culture in many airlines is one that minimises delays at the gate and if this means changing three boxes rather than carrying out the diagnostics to find the root cause of the failure and the exact box at fault, then three boxes will be changed. In peace-time operations this culture in the military would be unusual and particularly now that so many civilian companies are providing the support. In actual operations where battle-winning availability is vital, then the same culture may well pervade.

A second factor is also at play here. Civilian airliners are built to fail-safe principles where every system and part of the design is meticulously analysed for the consequences of it failing. Should that possibility happen, there must be an alternative load-path or alternate system to provide redundancy. Military aircraft, however, are built to safe-life principles where maintenance is a key factor in providing the early warning of failure before it is catastrophic.

Various maintenance techniques are increasingly used and incorporated into the design of both military and civilian aircraft to provide maintenance assistance. Condition monitoring in its widest sense in aircraft such as the Boeing 777, uses a huge amount of condition monitoring of all forms to monitor the deterioration of systems and components. By using spare capacity or redundancy, the need for urgent maintenance is avoided. The Aircraft Integrity Monitoring System (AIMS) continually monitors and informs the maintenance staff of both impending and actual failures. The necessary maintenance can then be programmed at a convenient time for maintenance staff with the right skills, the right test equipment and the right spares. Built in Test (BIT) and Built-in Test Equipment (BITE) are part of the whole AIMS system and contribute to this management of failures or impending failures.

Operational Pressure

The pressure in commercial operations on maintenance staff is often overwhelming. Delays, cancellations and lack of availability not only mean lost revenue but have a knock-on effect in customer perception. Reputation is hard won but all too easily lost if delays or cancellations occur. Delays and cancellations between 2003 – 2013 for US domestic carriers, averaged more than 21%. (US DOT and BTS) Whilst some of these are due to uncontrollable issues such as weather or air traffic controls, a great many are because of faults or maintenance delays.[9] The pressure on maintenance staff then becomes extreme, yet safety is still paramount. In that case the easiest solution to a fault or failure will be taken, perhaps without time for proper diagnosis. If the system can be re-set and tests satisfactorily, the fault is no longer present! Yet it may need vibration or temperature whilst airborne to provide the conditions when it will fail again. Operational pressure might also suggest that changing three boxes will solve the failure and so it does, but now two of the boxes will prove to be NFF when tested further down the support chain. In some cases, speed and operational imperatives will have masked the failure which may then re-occur at an inappropriate moment during the next flight. The integrity of maintenance staff is all that stands in the way of whether a fault or failure is solved in the most effective way. There will surely be some occasions when speed and operational pressure win and a dormant fault remains on the aircraft, or in the removed component. The operational pressure is created of course by the organisation and the humans who manage the organisation. There are also human factors at work within the maintenance organisation that relies on the maintenance personnel to undertake work and these human factors must also be understood.

The Human Factors Contribution

When humans are involved, errors can occur for any number of reasons. The Civil Aviation Authority (CAA) goes further stating:

> *"It is an unequivocal fact that whenever men and women are involved in an activity, human error will occur at some point"* (CAA, 2002)

Firstly maintenance errors cost lives and secondly maintenance errors cost money; maintenance errors also cost a Company its reputation though. In fact errors merely keep lawyers in business and ultimately generate more and more regulations. Maintenance errors can be thought of as resulting from what can be described as "The Error Chain". Simple errors often combine to create a catastrophe; by themselves they would not be a problem but the combination becomes serious. The Error Chain can cost a Company millions in re-work and lost revenue and invites unwelcome attention from regulators.

Examples of errors are:

- Incorrect installation of components
- Fitting the wrong part

[9] The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers. Summary information is provided on the number of on-time, delayed and cancelled flights - On time 78.02%, Delay – 20.03%, Cancellation – 1.72% between June 03 and Oct 13. For purposes of this report, a flight is considered delayed if it arrived at (or departed) the gate 15 minutes or more after the scheduled arrival (departure) time as reflected in the Computerized Reservation System. The information is based on data submitted by reporting carriers which number between 14 and 19 over the period.

- Electrical wiring discrepancies
- Loose objects left
- Inadequate lubrication
- Access panels, fairings or cowlings not secured
- Fuel or oil caps not secured
- Safety or gear pins not removed before aircraft departure

Maintenance technicians work in all sorts of environments, often extremely challenging ones, to deliver the outputs that are required. The performance of those maintenance tasks is affected and interfered with by many things, yet the technician will be coping by using both sub-conscious and conscious approaches to deliver the desired performance. The sub-conscious will be delivered as automatic or emotional actions, whereas the conscious approach will be delivered with logical and rational activities. The conscious actions include activities delivered according to rules and procedures, or based on experience and knowledge. The maintenance activities delivered with a sub-conscious approach will include those activities done automatically without thinking and could involve fast reaction and perhaps repetitive activities. As long ago as 1994, after a spate of accidents, the airline industry identified 12 human factors that degrade people's ability to perform effectively and safely, which could lead to maintenance errors; they have been christened the Dirty Dozen. They are well known in the commercial airline industry and feature prominently in maintenance training courses. They are:

- Stress
- Fatigue
- Lack of Communication
- Lack of Assertiveness
- Complacency
- Distraction
- Pressure
- Lack of Resources
- Lack of Knowledge
- Lack of Awareness
- Norms (where incorrect procedures or quick fixes become the normal way of working)
- Lack of Teamwork

Any one of these factors, or a combination of them, can result in a maintenance error or the failure to detect a fault. It is this latter point that is often dismissed or not considered and where the connection with NFF can be critical in its impact on air safety. Maintenance errors are usually obvious and can be traced to one or more of the "Dirty Dozen". The failure to locate or find a fault does not usually have such an obvious cause and is usually not considered a maintenance error. Yet if the dirty dozen is considered in the context of fault finding and achieving diagnostic success, many of the dirty dozen will actually cause a NFF to be registered. In that case, NFF resolution must surely be given the same prominence as the Dirty Dozen!

Table 1 – The Dirty Dozen and NFF

Dirty Dozen Factors	Contribute or Cause NFF?	Comment
Stress	Yes	Stress affects concentration and clear think for successful diagnosis
Fatigue	Yes	Fatigue will hamper ability to diagnose cause of fault
Lack of Communication	Yes	With rushed or poor communication. Poor briefing and description of fault symptoms often leads to NFF
Lack of Assertiveness	Yes	When directed by supervisor to a specific course of action Technician fails to question the course of action he has been directed to which results in NFF.
Complacency	Yes	Action is to perform the usually accepted solution which may result in temporary fix of intermittent faults or connector faults
Distraction	Possibly	Technician may miss out elements of diagnosis due to distraction and thus not find the fault
Pressure	Yes	Pressure may involve changing three items in order to make sure the cause of the fault is covered. This subsequently creates a NFF further down the support chain.
Lack of Resources	Yes	Inadequate resources will hamper diagnosis e.g. unsuitable test equipment my be used or lesser skilled technicians
Lack of Knowledge	Yes	Inadequate training will cause poor diagnosis
Lack of Awareness	Possibly	Similar to lack or poor training and lack of awareness of best diagnostic process.
Norms (where incorrect procedures or quick fixes become the normal way of working)	Yes	Some “norms” will have become the usual “fix” for particular faults and will have become the first fix to be tried because it usually clears the fault. Intermittent faults or connector faults will be temporarily rectified this way.
Lack of Teamwork	Possibly	The inability of a team to work successfully together may result in NFF as a way of shortening the maintenance time so that the team has the least time working together.

It can be seen that the human factors issues that cause maintenance errors and possible safety issues or accidents, are also the same factors that can contribute to NFF. It is therefore logical to conclude that there is a strong link, a cause and effect even, to the fact that NFF is also an air safety issue.

Diagnostic Maintenance Success

Having made the link therefore with maintenance errors, it is worth looking at maintenance help. Where does the technician get help? In modern aircraft it is increasingly from the On-board Maintenance System (OMS) or the Aircraft Integrity Management System (AIMS). The OMS on the Boeing 777 provides direct computer access to many of the maintenance functions on the aircraft. It consists of a central maintenance computer that takes inputs from condition monitoring systems and BIT. There are direct access points for a maintenance engineer to plug in a terminal around the aircraft.

However, BIT and BITE have their own inherent problems. Bit and BITE have become central to the diagnosis of faults, yet they have their own level of reliability built upon the ever increasing level of complexity of the systems they are monitoring. There are subtle relationships between systems that need to be understood by the designer of the BIT and BITE. More and more parameters can be monitored and so the complexity and difficulty of producing reliable test routines continually increases. Unfortunately what is needed for success here is a logical method for effective fault consolidation. If BITE falsely identifies component failures that do not exist, components may be designated as faulty when they are not. Perhaps the fault has in fact been caused by another component that feeds data into the first one - an example of what is known as cascading faults. Complex digital circuits are extremely sensitive to power surges and transient voltages which cause the monitoring circuits to register a fault. When a reset or a test fails to reproduce the fault, a NFF is generated and the BIT/BITE starts to get a poor reputation for identifying spurious faults that cannot be reproduced. As aircraft design and the OMS has been developed, the danger for the maintenance organization is an overload of data. There can be in excess of 100 BITE messages describing the condition of one system such as the landing gear. Has this helped the engineer with his diagnostics? Now he has too many options and may take the path of least resistance, especially if operational pressure demands that there is too little time to diagnose the fault more carefully. If the human factors contribution is added in to the ever more complex problem of achieving maintenance diagnostic success, there is a huge potential for NFF to be recorded and an error chain to be created.

A Case Study: No Fault Found (NFF) Certainly *IS* an Air Safety Issue

Examples that would appear to have serious flight safety implications such as faults with the Merlin radio in Afghanistan, where transmit/receive faults were often not obvious to the pilots but also could not be replicated on the ground. However, a recent Air Accident Investigation Bulletin (AAIB) reported an incident on the 11th Sept 2010 which very nearly led to a crash by a Dash-8 Q400 aircraft (G-JECF). (AAIB, Bulletin 6/2012)

During approach the aircraft experienced a failure of the number 1 Input Output Processor (IOP 1). The flight crew became distracted with this failure and were unaware that the altitude select mode of the flight director had become disengaged and that the aircraft had descended below its cleared altitude. Descent continued until, alerted by an Enhanced Proximity Ground Warning System (EPGWS) warning, the pilots climbed the aircraft and re-established the glidepath. The maintenance action following the incident recorded NFF with the relevant circuit breaker being reset and system tested satisfactorily. The aircraft was released for service with a request for further reports from the aircrew.

The subsequent detailed AAIB investigation found that the IOP 1 failure was caused by intermittent electrical contact arising from cracked solder on two pins of a transformer on the IOP power supply module. This IOP fault happened on this aircraft no less than 8 times between 22nd August and 8th October. In each case the fault had been recorded as NFF with various maintenance actions completed such as swapping with the number 2 IOP. Indeed after the first swap on 20 Sep, it was then number 2 IOP being recorded as faulty. Yet it was not then until the 8 Oct that the faulty serial numbered item was removed and replaced and sent to the OEM for investigation.

It was established that extensive tests were needed by the OEM to finally reproduce the fault on this IOP that was subsequently proved to have an intermittent fault caused by cracks in the solder of some surface mounted components on one of the electronic boards. IOP failures were a common occurrence but were often tested satisfactorily on the ground or tested serviceable by resetting a circuit breaker or re-installing the processor.

Removals were not common due to the high rate of NFF with only 20% of IOP failures being confirmed. Even those returned to the OEM produced a number that were NFF. Consequently, the Company has instituted a procedure where serial numbers need to be tracked more carefully with linkage to reported faults. In order to reduce the risk further of IOP units with intermittent faults being declared serviceable and subsequently fitted to aircraft, the following Safety Recommendation was made:

> **Safety Recommendation 2012-019**
> *It is recommended that Thales Aerospace review the Input Output Processor test procedures to improve the detection of intermittent failures of the ERACLE power supply module in order to reduce the number of faulty units being returned to service.*

Hopefully, this particular NFF history has now been solved and will now not contribute to an accident. How many other NFFs with other operators though are just waiting to contribute to an accident?

Summary

It is clear that NFF is a serious problem to the airline industry in particular as it affects aircraft availability and causes delays and cancellations, all of which have a damaging effect to airline revenue and reputation. Airlines thus treat NFF in a number of ways; many will accept high NFF rates if their delays and cancellations are minimised, reputation and revenue is paramount; others may hide or be unaware of the problem. There are many causes though of NFF and it has been shown that there is a huge similarity between the human factors that cause maintenance errors and those that cause or contribute to NFF. Yet maintenance errors, described as the “Dirty Dozen”, have received a great deal of publicity as they have been accepted as being the factors that contribute to maintenance errors that cause aircraft accidents. The link between NFF and aircraft safety is, however, yet to be fully understood and accepted. If there is such similarity between NFF causes and the causes and impact of maintenance errors, it is only time before an accident and loss of life can be directly linked to the occurrence of NFF.

References:

AAIB Bulletin: 6/2012 G-JECF EW/C2010/09/04 available at
http://www.aaib.gov.uk/publications/bulletins/june_2012/dhc_8_402_dash_8__g_jecf.cfm

CAA, 2002. Safety Regulation Group (2002) CAP 715. An introduction to aircraft maintenance engineering HF for JAR 66 CAA 2002.

Hessburg J, 2013, Functionability Management – A tribute to Jack Hessburg at the 23rd MIRCE international Symposium, 3-5 Dec 2013, The MIRCE Akademy, Exeter, UK. http://www.mirceakademy.com/uploads/MIRCE-Symposium-Programme-2013-last.pdf

U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS http://www.transtats.bts.gov/ot_delay/ot_delaycause1.asp accessed 20 Dec 2013

Maintenance Axiom of Mirce-mechanics

Dr Jezdimir Knezevic
MIRCE Akademy, Woodbury Park, Exeter, EX5 1JJ, UK

Abstract

Reliability of maintenance process is quantifiable in statistical terms related to the occurrences of maintenance faults and errors. However as statistics does not study the causes of statistical behaviour, full understanding of the reliability of maintenance is only possible by understanding physical causes and mechanisms that lead to the occurrence of maintenance faults during the maintenance process. Based on the analysis of tens of thousands of maintenance tasks in defence, aerospace, transportation (including Formula 1 Grand Prix racing), communication and other industries the author has formulated the Maintenance Axiom of Mirce-mechanics, which is: The probability of faulty execution of any maintenance task is greater then zero. This axiom has a profound impact on all aspects of the life on any maintainable system, such as: reliability, availability, safety, cost, effectiveness and many others, on one hand, and associated processes like: manufacturing, operation, logistics support, on the other.

1. Introduction

Although the word reliability is used daily in many different contexts, from the reliability of products to the reliability of people or even relationships among nations, the concept and meaning of reliability differs a lot. In the Oxford English dictionary it is defined as "a particular property inherent in a body or substance". For the purpose of reliability in maintenance, that particular inherent property is "faultless execution of the maintenance task". Hence, the reliability of maintenance could be defined as a probability that maintenance tasks will be completed without any faults, resulting from the maintenance process.

However, as a probability cannot be seen or measured directly, there seems to be a certain fundamental difficulty in understanding and interpreting statistical and probability functions in real life. This is because physical characteristics of a system like the weight, temperature, volume and similar have a clear and measurable meaning. However, the concepts of probability, and hence reliability of maintenance, is an abstract property of a maintenance system that obtains a physical meaning only when behaviour of a large number of maintenance tasks is considered. Hence, understanding the reliability of maintenance is reduced to the physical observation and analysis of faulty maintenance tasks, which are observable and measurable physical quantities.

At the MIRCE Akademy a large number of faulty maintenance tasks like overhauls, tests, inspections, visual checks, scheduled maintenance tasks, repairs, replacements, no fault found, non distractive tests, examinations and many others, have been observed and analysed in order to understand the mechanisms of the motion of a system through the maintenance process, resulting from the execution of maintenance tasks.

Based on the analysis of tens of thousands of maintenance tasks in defence, aerospace, transportation (including Formula 1 Grand Prix racing), communication and other industries

the author has formulated the Maintenance Axiom of Mirce-mechanics, which is presented in this paper. This axiom has a profound impact on all aspects of the life on any maintainable system, such as: reliability, availability, safety, cost, effectiveness and many others, on one hand, and associated processes like: manufacturing, operation, logistics support, on the other.

2. Concept of Maintenance Task

According to Ben-Daya at el 2009, "A Maintenance task is a set of activities that need to be performed in a specified manner, usually by humans, for functionality of the item/system to be maintained."

In accordance to the Maintenance Program Development Document MSG-3, revision 2, published in 1993 [2], maintenance tasks could be categorised in the following categories:

- **Servicing:** replenishment of consumable fluids, cleaning, washing, painting, etc.,
- **Lubrication:** installing or replenishing lubricant
- **Inspection:** Examination of an item against a defined physical standard
- **General Visual Inspection** performed to detect obvious unsatisfactory conditions. It may require the removal of panels and access doors, work stands, ladders, and may be required to gain access.
- **Detailed Visual Inspection** consists of intensive visual search for evidence of any irregularity. Inspection aids, like mirrors, special lighting, hand lens, boroscopes, etc. are usually required. Surface cleaning may be required, as well as elaborate access procedure
- **Special Visual Inspection:** an intensive examination of specific area using special inspection equipment such as radiography, thermography, dye penetrant, eddies current, high power magnification or other NDT. Elaborate access and detailed disassembly may be required.
- **Check:** a qualitative or quantitative assessment of function
- **Examination:** a quantitative assessment of one/more functions on an item to determine if it performs within acceptable limits.
- **Operational:** a qualitative assessment to determine if an item is fulfilling its intended function. It does not require quantitative tolerances.
- **Restoration:** perform to return an item to a specific standard. This may involve cleaning, repair, replacement or overhaul.
- **Discard:** removal of an item from service.

It is necessary to stress that some resources are needed to facilitate the successful completion of the maintenance task. As the main function of these resources is to facilitate the maintenance process they will be called maintenance resources. The resources needed for the successful completion of every maintenance task, could be grouped into following categories:

- **Maintenance Personnel, MP**: a generic name for trained and qualified humans required for the installation, checkout, handling, and sustaining maintenance of the item/system and its associated test and support equipment are included in this category.
- **Maintenance Material, MM**: a generic name which includes all spares, repair items, consumables, lubricants, special supplies, and related inventories needed for the execution of a maintenance task;

- **Maintenance Test and Support Equipment, MTE:** a generic name for all tools, special condition monitoring equipment, diagnostic and check-out equipment, metrology and calibration equipment, maintenance stands and servicing and handling equipment required for the execution of a maintenance task
- **Maintenance Facilities, MF**: a generic name for all facilities needed for completion of maintenance tasks, such as buildings, portable repair shops, inspection pits, dry dock, housing, maintenance shops, calibration laboratories, and special repair and overhaul facilities
- **Maintenance Data, MD**: a generic name for all necessary technical information required for check-out procedures, maintenance inspection and calibration procedures, overhaul procedures, modification instructions, facilities information, drawings and specifications that are necessary in the performance of system maintenance functions.

At the same time it is important to stress that each task is performed in a specific working environment that could make a significant impact on the reliability of the execution of each task. The main environmental factors could be grouped as follows:

- Space impediment (which reflects the obstructions imposed on maintenance personnel during the task execution which requires them to operate in restricted positions)
- Climatic conditions (rain/snow, solar radiation, humidity, temperature, and similar situations, which could make a significant impact on the task completion.
- Platform on which maintenance task is performed (board of the ship/submarine, space vehicle, off-shore platform and similar).

3. Measures of Maintenance Task

Like all other physical phenomena that have to be measured in order to be understood, a maintenance task has to be measured. For that purpose, Knezevic 1997 created the concept of the Duration of a Maintenance Task, DMT, and associated measures of a maintenance task, thus:

Maintainability Function, denoted as M(t), represents the probability that the maintenance task considered will be successfully completed before or at the specified moment of elapsed time t, thus:

$$M(t) = P(DMT \leq t) \qquad 1.$$

Percentual Duration of Maintenance Task, DMT_p, represents the duration of a maintenance task by which a given percentage of maintenance tasks considered would be successfully completed. It is the abscissa of the point whose coordinate presents a given percentage of task completion. Mathematically, DMT_p can be represented as:

$$DMT_p = t, \text{ for which, } M(t) = P(DMT \leq t) = p \qquad 2.$$

The most frequently used is DMT_p measure is DMT_{90} time which presents the duration of the restoration time by which 90 percent of maintenance trials will be completed, thus:

$$DMT_{90} = t, \text{ for which, } M(t) = P(DMT \leq t) = 0.9$$

Expected Duration of a Maintenance Task, denoted as MDMT, represents the expectation of the random variable DMT, which can be used for calculating this characteristic of a maintenance process, thus:

$$E(DMT) = MDMT = \int_0^\infty [1 - M(t)]dt \qquad 3.$$

Methods for numerical evaluation of maintenance task measures, in accordance with Mirce-mechanics, are presented by Ben-Daya 2009, for both:

- Statistical inference method for evaluation of empirical data (parametric and distribution approach), used at the in-service stages of a system life.
- Probabilistic predictive method based on Maintainability Block Diagrams used for engineering prediction at the design stages of a system life.

It is necessary to stress that both approaches, statistical and probabilistic, are related to the maintenance tasks that are completed without faults or errors during their executions.

4. Concept of Faulty Maintenance Task

Numerous industrial and personal accidents have been maintenance related. It means that as result of an inherent fault or a failure that took place during the maintenance process, the maintained system experienced in-service failure. Hence, all maintenance tasks of this type, in Mirce-mechanics, are named Faulty Maintenance Task, FMT.

Some ideas of the type, scale and frequencies of the maintenance induced failures and their consequences in commercial aviation, among many sources, can be obtained from the National Transportation Safety Board of the USA, and Civil Aviation Authorities of the UK report published on 12 August 2002, where, among others, the following events are recorded:

- May 25, 2002. **China Airlines** B747-200. Structural failure at top of climb to cruise altitude resulted in a crash into Taiwan Strait, due to repair of previous tail strike used steel doubler that are prohibited by structural repair manual. Toll: 225 killed.
- Aug. 24, 2001. **Air Transat** A330. Improper engine repair caused by leak from cracked fuel line resulted in duel engine flameout at cruise over Atlantic. Aircraft glided 135 miles to emergency landing in Azores. No serious injuries.
- April 26, 2001. **Emery Worldwide Airlines** DC-8-71F. Left main landing gear would not extend for landing. Cause was failure of maintenance to install the correct hydraulic landing gear extension component and the failure of inspection to comply with post-maintenance test procedures. No injuries.
- March 20, 2001. **Lufthansa** A320. Cross-connected pins reversed the polarity of captain's side stick. Post-maintenance functional checks failed to detect the crossed connection. Aircraft ended up in 21º left bank, almost hitting the ground. Co-pilot switched his side-stick to priority and recovered the aircraft. No injuries.
- Feb. 16, 2000. **Emery Worldwide Airlines** DC-8-71F. Crashed attempting to return to Rancho Cordova, California. Cause was improperly installed right elevator control. Toll: 3 crew killed.
- Jan. 31, 2000. **Alaska Airlines** MD-83. Crashed in Pacific Ocean near Port Hueneme due to loss of horizontal stabilizer caused by the maintainer failure to lubricate jackscrew assembly that controls pitch trim. Toll: all 88 aboard killed.
- Jan 21, 1998. **Continental Express** ATR-42. Fire in right engine during landing, due to improper overhaul of lugholes in the fuel/oil heat exchanger. No serious injuries.

- Sept. 27, 1997. **Continental Airlines** B737. Separation of aileron bus cable forced the crew to return to the airport shortly after takeoff. Separation was caused by wear in the cable and inadequate inspection of it. No serious injuries.
- March 18, 1997. **Continental Airlines** DC-9-32. Failure of maintenance personnel to perform a proper inspection of the combustion chamber outer case, allowing a detectable crack to grow to a length at which the case ruptured, causing uncontained failure of right engine. No injuries.
- Nov. 1996. A320 (operator unknown). Both fan cowl doors detached from No. 1 engine during rotation. Doors had been closed but not latched during maintenance. According to AAIB, "Similar incidents have occurred on at least seven other occasions."
- July 17, 1996. **TWA** Flight 800, B747. Fuel/air explosion due to inadequate maintenance on an aging fleet and noncompliant parts. Toll: all 230 passengers and crew killed.
- July 6, 1996. **Delta Air Lines** MD-88. Uncontained engine failure on takeoff due to inadequate parts cleaning, drying, processing and handling. Toll: 2 passengers killed, 2 passengers seriously injured.
- June 8, 1995. **ValuJet Airlines** DC-9-32. Maintenance technicians failed to perform a proper inspection of the 7th stage high compression disk, allowing a detectable crack to grow to a length at which it ruptured. Toll: 1 crew seriously injured.
- Feb. 1995. **British Midland** B737-400. Oil pressure lost on both engines. Covers had not been replaced from borescope inspection the previous night, resulting in loss of almost all oil from both engines during flight. Diverted and landed safely. No injuries.
- March 1, 1994. **Northwest Airlines** B747. Narita, lower forward engine cowling dragged along runway. During maintenance, the No. 1 pylon diagonal brace primary retainer had been removed but not reinstalled. No injuries
- Aug. 1993. **Excalibur Airways** A320. Un-commanded roll in first flight after flap change. Returned to land safely at Gatwick. Lack of adequate briefing on status of spoilers (in maintenance mode) during shift change. Locked spoiler not detected during standard pilot functional checks. No injuries.
- Sept. 11, 1991. **Continental Express Airlines**, EMB-120. Horizontal stabilizer separated from fuselage during flight because maintenance personnel failed to install 47 screw fasteners. Toll: all 14 passengers and crew killed.
- Aug. 21, 1990. **United Airlines** B737. Flashlight left by maintenance, sandwiched between cargo floor and landing gear retract/extend linkage, causing the crew to make a gear up landing. Toll: No injuries.
- July 22, 1990. **USAir** B737. Fuel pump control failure due to improper machining. No injuries.
- June 1990. **British Airways** BAC1-11. Captain sucked halfway out of windscreen, which blew out under effects of cabin pressure, as 84 of 90 securing bolts were smaller than the specified diameter. Toll: 1 serious injury.
- Aug. 12, 1985. **Japan Air Lines** B-747SR. Improper repair of aft pressure bulkhead led to sudden decompression in flight that damaged hydraulic systems and vertical fin. Aircraft struck Mt. Ogura. Toll: 520 passengers and crew killed; 4 surviving passengers injured.

- May 1983. **Eastern Airlines** L-1011. Loss of all power from improperly fitted O-ring seals Aircraft landed on one engine. No injuries.
- May 25, 1979. **American Airlines**, DC-10. Separation of No. 1 engine and pylon assembly on takeoff at Chicago's O'Hare. Toll: all 298 passengers and crew plus 2 killed and 2 seriously injured on the ground.

5. Causes of Faulty Maintenance Tasks

The human constituents of a maintenance system, either as decision maker or as a task executor, bear the ultimate responsibility for recognising, interpreting, compensating for, and correcting or mitigating the consequences of deficiencies and faults of a maintenance process. Thus "human error" and "judgement error" are terms found frequently in reports on system failures. In many cases, however, system failures can be attributed to the poor design of tasks for human capabilities, to defective interfaces between task performers and equipment subsystems, to inadequate training, to poorly conceived operating or maintenance procedures, or to other situational factors. In-depth investigation and analysis of accidents and injuries provide proof that often inadequate attention is given to the physical understanding of the causes of faulty maintenance tasks as well as the mechanisms of their occurrences.

For example, according to a Pratt & Whitney survey the major causes for the 120 in-flight engine shutdowns on Boeing 747 aircraft's, due to human errors in maintenance, were mainly due to:

- Incomplete installation (33%)
- Damaged on installation (14.5%)
- Improper installation (11%)
- Equipment not installed or missing (11%)
- Foreign object damage (6.5%)
- Improper fault isolation, inspection (6%)
- Equipment not activated or deactivated (4%)

Some of the other related causes are:

- Complex maintenance related tasks
- Time pressure for delivering the aircraft
- Fatigue of the maintenance personnel
- Maintenance procedures not followed accordingly
- Usage of outdated maintenance manuals

Numerous examples of a similar nature regarding the maintenance processes for other industries and applications can be easily found in literature. For example, during the 1999 British Grand Prix F1 race, McLaren-Mercedes driver Mika Hakkinen experienced two successive FMTs within 3 laps of racing, which ended in his retirement from the race. The first FMT took place, in lap 25, during a scheduled maintenance task for refuelling and tyre change. After the leaving the pit lane the driver felt "uneasy" regarding the rear wheel, so he decided to immediately return to the pit. After the replacement of the rear tyres, the maintenance crew released him to the track. A few seconds later the rear wheel separated from the car and Hakkinen continued driving towards the pit on three wheels. By the time he

arrived back to McLaren pit, the race was at the lap 28. The sequence and the duration of all pit stops, in seconds, are shown in Figure. 1.

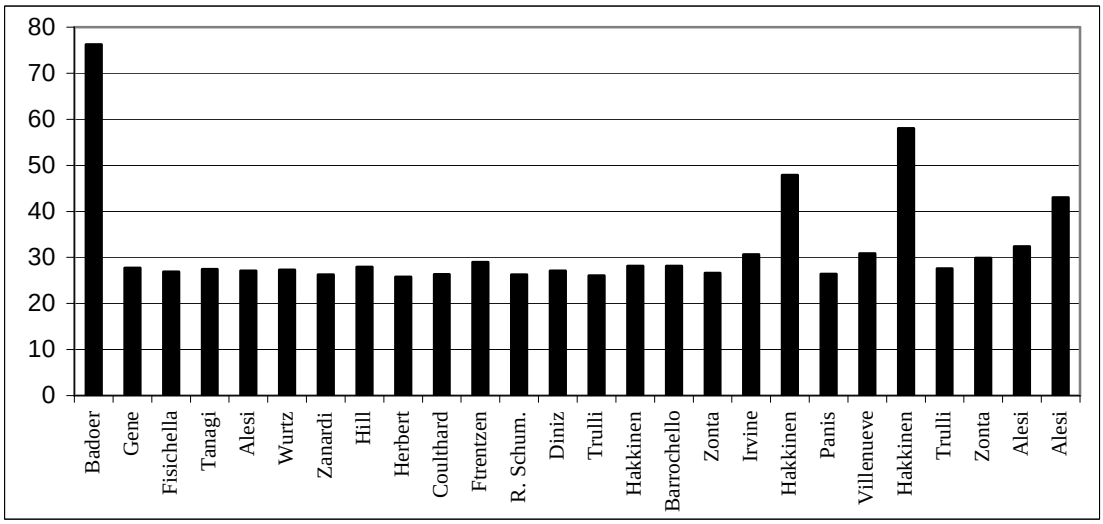

Figure 1: The sequence and the duration of pit stops of the 1999 British Grand Prix

6. Mirce-mechanics Maintenance Axiom

Analysis of over tens of thousands of maintenance tasks in defence, aerospace, transportation (including Formula 1 Grand Prix racing), communication and other industries, some of which are presented in this paper, the author of the paper has formulated the Maintenance Axiom of Mirce-mechanics, which is:

The probability of faulty execution of any maintenance task is greater then zero.

This axiom has a profound impact on all aspects of the life on any maintainable system, such as: reliability, availability, safety, cost, effectiveness and many others, on one hand, and associated processes like: manufacturing, operation, logistics support, on the other.
The impact of the axiom on studies, analysis and classifications of maintenance tasks is briefly presented below.

7. Reliability Based Maintenance Tasks

The direct consequence of the Mirce-mechanics Axiom is additional classification of the types of maintenance tasks. Based on the information presented above it could be concluded that, from point of view of reliability, each physically observable maintenance task could be categorised as:

Successful Maintenance Task, SMT, where all maintenance activities have been completed successfully in the first attempt.

Faulty Maintenance Task, FMT, where all maintenance activities have not been completed successfully in the first attempt.

Consequently, each maintenance task performed has some probability functions attached to it. Generally speaking there is unlimited number of possible function related to reliability of the execution of maintenance tasks. This is source of uncertainty regarding maintenance process

on one hand and uncertainty of system in-service reliability, cost and effectiveness, on the other. It is through observations, science based analysis and added assumptions that one of them gets adopted for a practical applications.

8. Faulty Maintenance Tasks

By studying numerous maintenance processes regarding the execution of maintenance tasks, in respect to the possibility of maintenance faults to be detected during the maintenance process, it is possible to group them into following two categories:

Detectable Faulty Maintenance Tasks, FMT_D are those where the faulty activities could be detected during the execution of consisting activities or at the end of that task and corrective action taken.

Non-Detectable Faulty Maintenance Tasks, FMT_{ND} where faults and errors induced during the maintenance process could not be detected and it is left to the operational process to detect maintenance faults and to deal with its consequences.

8.1 Detectable Faulty Maintenance Tasks

By studying numerous maintenance processes regarding the execution of maintenance tasks, in respect to the ability of maintenance faults to be detected during the maintenance process, it is possible to group them into following two categories:

Detectable Faulty Maintenance Task Type 1, FMT_{DT1}
Detectable Faulty Maintenance Task Type 2, , FMT_{DT2}

The main characteristics of both types of tasks will be briefly presented below.

8.2 Detectable Faulty Maintenance Task Type 1

These are maintenance tasks where at the end of each maintenance activity it is possible to detect maintenance induced faults and it is possible to correct these by redoing the activity or taking appropriate action until the fault is eliminated. Hence, each faulty activity has a chance of being detected and the opportunity exists for its recovery by redoing it (fully or partially). Fault elimination might take additional maintenance time and resources.

Diagrammatic representation of the sequence of maintenance activities for a , FMT_{DT1} is shown in the figure below:

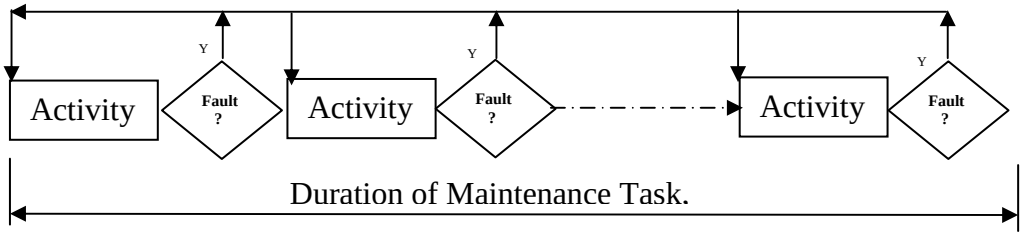

Figure 2: Detectable Faulty Maintenance Task Type 1

8.3 Detectable Faulty Maintenance Task Type 2

These are maintenance tasks where at the end of the maintenance task it is possible to detect maintenance induced faults and it is possible to correct these by redoing the task or taking appropriate actions until the fault is eliminated. Hence, each faulty task has a chance of being detected and the opportunity exists for its recovery by redoing it (fully or partially). Fault elimination might take additional maintenance time and resources.

Diagrammatic representation of the sequence of maintenance activities for a , FMT_{DT2} is shown in the figure below:

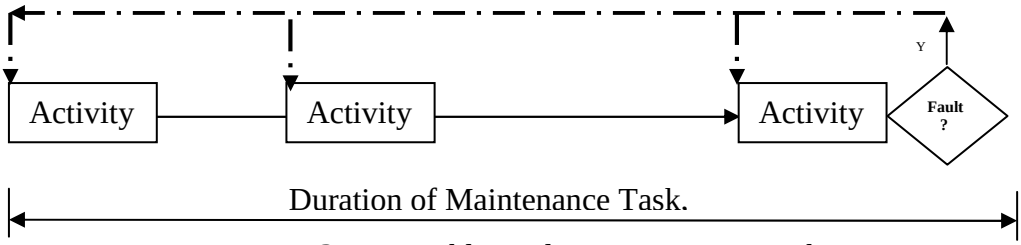

Figure 3: Detectable Faulty Maintenance Task Type 2

9. Reliability Based Classification of Maintenance Tasks

Over the years maintenance tasks have been classified in accordance to several different criteria. However, as result of the Mirce-mechanics Axiom presented in this paper, an additional classification of maintenance task has emerged.

Classification of maintenance tasks, in respect to the reliability of their execution, is shown in Figure 4.

Maintenance Task
- Successful maintenance Task
- **Faulty Maintenance Task**
  - Detectable Faulty Maintenance Task
  - **Non-Detectable Faulty Maintenance Task**
    - Detectable Faulty Maintenance Task Type 1
    - Detectable Faulty Maintenance Task Type 2

Figure 4: Reliability Based Classification of Maintenance Tasks

It is necessary to stress that the measures of maintenance tasks defined by equations 1, 2 and 3, are still valid. However, the “mechanism” for incorporation of the impact of extra time required for the rectifications of faulty maintenance activities or tasks has to be determined, which is beyond the scope of this paper.

10. Conclusion

The main objective of this paper was to demonstrate that although reliability of maintenance is intuitively associated with well-executed maintenance tasks, it can only be physically observed through the occurrences of faulty or erroneous maintenance tasks. As the research performed by the author at the MIRCE Akademy in association with Fellows, Master and Doctoral students has clearly demonstrated continuous occurrence of faulty maintenance tasks the author has formulated the Maintenance Axiom of Mirce-mechanics: *The probability of faulty execution of any maintenance task is greater then zero.* As this axiom has a profound impact on all aspects of the life on any maintainable system, such as: reliability, availability, safety, cost, effectiveness and many others, on one hand, and associated processes like: manufacturing, operation, logistics support, on the other, the categorisation of maintenance tasks, based on reliability of their execution, has been presented.

Finally it is necessary to stress that, in some situations; practicing engineers and managers, for the purpose of the specific project or contract could decide to ignore this axiom. However, making an executive decision to ignore the axiom for practical reasons does not mean that the faults cannot take place. Hence, the Mirce-mechanics Axiom presented is the difference between scientifically accurate and practically acceptable accuracy of prediction methods for the reliability of the maintenance process.

11. References

[1] Ben-Daya, Duffuaa, Raouf, Knezevic, Ait-Kadi, Handbook of Maintenance Management and Engineering, Springer, 2009. ISBN 978-1-84882-471-3

[2] Maintenance Program Development Document MSG-3, revision 2, Air Transport Association of America, 1993.

[3] Knezevic, J., Systems Maintainability, Analysis, Engineering and Management, Chapman & Hall, pp 400, 1997, UK. ISBN 0-412-80270-8

[4] Knezevic, J., Effective Analysis of Existing Maintainability Data, Journal of Communications in Reliability, Maintainability and Supportability, Vol. 2, No. 1 pp. 18-22, SAE, (January 1995).

Planning In-service Support

John Crocker
Science Fellow of the MIRCE Akademy, UK

Abstract

A major part of the cost of operating a fleet of high-valued, repairable assets (HVRA) is in keeping them operational. HVRAs generally only produce a return on investment when they are in use, performing the task[s] for which they were designed. Commercial aircraft, for example, usually earn revenue by carrying freight (passengers are often referred to as "self-loading" freight). Few aircraft generate income by sitting on tarmac, at least, not until they become museum exhibits. Maintaining these assets in a state of functioning can often cost the operators considerably more than the initial purchase price.

Having the right spares in the right place at the right time is obviously important but significant savings can also be made by performing the right maintenance at the right time provided, of course, that it in no way compromises safety or any other legal requirements.

Introduction

Consider a fleet of a number of similar high-valued, repairable systems operating over a period of from twenty to thirty years. Possibly as much as ten years before their entry into service (EIS) date, the operator will need to consider which platforms, what configurations and how many will be required to serve the particular operations envisioned for the given period. For a mature operator, this will probably be to replace some, or all, of the current fleet plus a number to allow for a potential expansion of services. It will also be assumed that the operators will expect to get a high level of utilisation out of the fleet and will be subject to some form of penalty either tangible, in the form of fines or, intangible such as loss of reputation if service levels fail to meet expectation.

Given that the system is high-valued then it is likely to comprise a large number of sub-systems each of which is critical to the successful operation of the whole. This means that although the system may not actually stop functioning instantaneously if one of these fails, there will be a requirement to apply invasive maintenance to restore the system to its full functionability as soon as is practically possible. Although this sounds rather complicated, imagine the system is an airliner and a subsystem has failed. There will almost always be a number of other subsystems able to meet the demand for the duration of the current flight but the aircraft will be grounded at the end of the current flight until the appropriate maintenance has been completed. In exceptional circumstances, the flight may have to be aborted with the aircraft being diverted to the nearest suitable airport

In general, every operator will want to keep as many of the fleet in a state of functioning as possible for as much of the time as possible. One way of ensuring that all of the fleet are functionable for all of the time is to not operate any of them. In fact, this is the only way. The higher the utilisation of the systems, the more likely they are to fail, or at least require invasive maintenance but the relationship between usage and the need for maintenance, in general, will not be a linear one. Similarly, the more maintenance that is needed, the more spares will be required to achieve desired service levels but, again this is not a simple relationship.

In this paper we look at some of the ways that we can use to predict when maintenance will be required and what can be done to better manage this.

Causes of failure

Components are designed to operate within an envelope or set of constraints explicitly or implicitly. Failure to adequately take these into account, at the design stage, may lead to unacceptably low levels of reliability or too many failures too soon. In general, such failures are likely to be blamed on poor design and will often result in claims under the terms of the warranty. In some cases, action may need to be taken to redesign the components, change the way they are manufactured or possibly the way they are stored or handled.

If the system has been operated outside of its design envelope failures are again likely to occur earlier than expected. This category will also include failures due to not maintaining the correct levels of fluids (e.g. coolant, lubricant, fuel and hydraulic fluid). It would also include instances of when hatches have been left unsecured, the wrong fasteners have been used or incorrectly fitted (e.g. over-tightened or cross-threaded) or parts have been fitted the wrong way round.

Although the above types of failure can happen, it is hoped that the designers will make every effort to eliminate them and that adequate procedures are in place to minimise the risks once the system has gone into service.

Failures can occur as a result of external factors. In the aerospace world, these are generally referred to as "foreign object damage" or "FOD". By and large, these can be considered as truly random events – the bird almost certainly does not know the age of the engine into which it gets ingested, the forklift truck driver does not deliberately choose the oldest or newest aircraft to back into. Because the failure is unrelated to the age of the component, it is not unreasonable to model these using a [negative] exponential distribution (unless the data suggests otherwise, in which case, it is very likely that there is a fault somewhere in the data-recording process).

This does not mean that FOD cannot be prevented only that premature replacement of components will almost certainly have no effect on the number of failures over time.
The final category of failure is those caused by usage and exposure to their natural environment. These will, almost always be age-related where "age" may be measured in calendar time, operating time (how long the system is actually in use) or accumulation of stress, say. If the correct unit of aging is used then these failures will generally be adequately modelled using a Weibull distribution with a shape parameter not less than 3. There is some evidence to suggest that the shape parameter may be influenced by how consistent the materials and manufacturing processes are. The use of high-grade material and manufacturing to very tight tolerances will tend to push the value of the shape parameter higher which also means that there will be less variation in the times-to-failure.
At this point it is probably worth mentioning some of the problems associated with "reliability measures". Design engineers will generally be aware of the mechanisms that may cause a component to fail and will, as far as possible, minimise the effects of these mechanisms but with all designs, the end result is a compromise across many conflicting factors. Unfortunately, no matter how good the designers are, they will not be able to design a system to meet a given level of reliability nor will they be able to determine what the level will be. That said, it is not unreasonable to expect the latest system to be no less reliable than

its predecessors – cars, television sets, washing machines, aircraft are all far more reliable now than they were 10, 20, 50 or 100 years ago.

The second problem with measuring reliability is that when a component fails, in service, the actual cause will often not be known. There may be several mechanisms, e.g. erosion, corrosion, fatigue, thermal and mechanical stress acting on the given component at different levels over its life. As a result, failures are more likely to be recorded against symptoms (or modes) rather than mechanisms (or causes). These may include failure to start, smoke, oil smell, vibration, rattle, puncture, over-heating, etc. It is not uncommon for such modes to be the result of several components working below their optimum so there may not actually be a single cause. In gas-turbine engines, a build-up of dirt, soot and other substances on the blades will reduce their efficiency, as will dents and other imperfections. Eventually, this will lead to having to run the engine at temperatures for which it was not designed so action will need to be taken.

A third factor is that as components become increasingly more reliable so the number of failures over time reduces. Unless there is a major problem, it can be many years before there is sufficient in-service failure data to be able to derive [time-to-failure] distribution parameters with any confidence. In addition, health monitoring and pre-emptive replacements of components further reduce the in-service time-to-failure (ttf) data.

Types of Maintenance

It is convenient to consider three types of maintenance: corrective, preventative and opportunistic.

Corrective Maintenance

This is done following an actual failure for whatever reason. Very often it will comprise the minimum amount of work and parts replacement necessary to return the system to a serviceable condition. However, dependent on the age of the parts within the system and possibly the age of the system itself, it may be decided to do additional work.

Secondary Damage

When a component fails, there is a certain risk of it causing damage to other components. For example, if the cam-belt in a petrol engine breaks, there is a high risk that damage will be caused to camshaft, valves, pistons and cylinder linings and possibly other parts as well. This collateral, caused or secondary damage would not have happened if the primary cause had not occurred and there is a reasonably high probability that none of these parts would have needed to be replaced had it not done so. It is generally a relatively straightforward task to identify which parts are likely to be damaged given the primary cause.

Preventative Maintenance

Preventative maintenance can take many forms ranging from checking fluid levels through to a complete overhaul or refurbishment. It is usually triggered by the system or some component thereof reaching a certain age. The age limit, sometimes referred to as life or time expiry (often shorted to “timex” or” lifex”) may be measured in calendar time, usage (e.g. flying hours), cycles (e.g. number of missions, flights, sorties) or [stress] cycles calculated by applying complex algorithms to measured parameters.

Preventative maintenance may be performed for one of two reasons: safety and economics. If the failure of a component could result in catastrophic consequences (i.e. put lives at risk)

with a certain probability and that probability is demonstrably related to the age/usage of the component then one way of reducing the risk is to replace the component before it reaches an age at which the probability of failure during the next mission is unacceptably high. Examples of this can be found in the nuclear power industry and in aircraft (both civil and military).

Occasionally, components are subject to a "routine inspection" policy. This usually, but not invariably, is due to components of the given type having failed earlier than expected once the system has entered into service. If the consequence of failure of these parts is of a high enough level of severity and the probability of failure is also unacceptably high based on the [small] sample of in-service failures then it may be considered expedient to introduce an inspection policy to manage this risk until such time as the cause can be found or the extra operational experience reduces the probability of failure to acceptable levels.

Opportunistic Maintenance

The third category covers the repair or replacement of components that have been rejected during the recovery of the system for one of the two previous reasons. In particular, it is concerned with components that were not implicated in the primary reason for taking the system out of service.

Secondary Inspection or Found Damage

During the recovery of the system or one of its sub-systems, the opportunity is often taken to inspect those parts which are exposed but which would otherwise be difficult to see. For example, if your car's engine has a leaking head gasket (something that used to happen quite often a few decades ago), it is necessary to remove the [cylinder] head which exposes the inlet and exhaust valves, the tops of the pistons and the cylinder linings. It is quite rare to find all of the exhaust valves in a healthy condition so it is invariably necessary to replace one, or more of these along with decoking the tops of the pistons and possibly regrinding some of the valves that are starting to show signs of wear.

The subtle point which distinguishes these rejections from the secondary damage ones is that their unsatisfactory state was not as a direct consequence of the primary failure. A secondary point is that had the system not required maintenance at that time, these components rejected during inspection would almost certainly not have caused an imminent system failure although it is, of course, hoped that by replacing them during this opportunity it will delay the time to next failure.

Age-related removals

With life-limited parts (see above under preventative maintenance), the amount of life remaining can be calculated at the time the system is taken out of service. If this is less than the "Minimum Issue [Service] Life" (MISL) sometimes referred to as the "stub life" then the part will be rejected and replaced. Very often, such a part will be placed into "quarantine" to be released if the "Release Life" is subsequently increased.

[When a part is identified as a potential life-limited component, the normal practice would be to run a sample under test conditions until they fail or until a certain number of them fail. From these recorded times to failure usually measured in stress cycles it is then possible to work out (based on certain assumptions and within certain confidence levels) the parameters of the time to failure distribution. From this it is again possible to calculate (given certain assumptions and confidence levels) the "Predicted Safe Cyclic Life". If this is within

acceptable limits (i.e. not too low) then a “Release life” will be assigned (typically this is initially set to around half the PSCL). There the then follows a complicated procedure to raise the Release Life in a series of steps up to the PSCL. If the value is too low or the conditions are not met to incrementally increase the Release Life then the only course of action is to redesign the component in the hope that the new design will be that much more reliable and hence meet the required level.]

The MISL is set based on purely economic grounds – allowing a component to exceed the MISL does not place the system under undue risk. The point is that by replacing lifed parts at this time, the system (or sub-system) will not need to be taken out of service for scheduled maintenance so soon.

For components subject to a routine inspection regime, it may again be expedient to bring forward such inspections as the opportunity allows. Normally this would only happen if it was necessary to take the system out of action for a significant period to perform the inspection.
If the predominant failure mechanisms or modes are age-related such that the probability of failing within a given period increases as the age of the component increases then it may be cost-effective to replace such a component while the parent system or sub-system is in maintenance if the age of the component has exceeded some arbitrary value. This age-limit is sometimes referred to as a “soft-life”.

Health monitoring

For some types of equipment installed in certain systems, health monitoring can be an effective means of preventing failures. Essentially, this works by fitting the equipment with a number of probes to measure and record changes in a number of the physical properties of the equipment, such as temperatures, pressures and vibrations (both frequency and amplitude). This is based on the premise that if a component is about to fail, it will undergo some form of physical change that will be reflected in changes to one, or more of the monitored parameters which will show up as anomalies prior to failure with sufficient time to take evasive action in a relatively planned manner.

Levels of Maintenance

Just as there are different categories of maintenance so we can also consider different levels. In general, it is sufficient to consider two levels: repair and recondition (also referred to as overhaul or refurbishment). At the system (or sub-system) level, a repair is usually regarded as the minimum amount of work necessary to restore the system (sub-system) to a state of functioning. At component level, it is often used to describe work done on the old/rejected component to restore it to a satisfactory condition (e.g. regrinding the valves).
When a component fails, is rejected because it is in an unsatisfactory condition or is removed for health monitoring reasons, it may be possible to repair the component to a state which gives it a new lease of life. It may not be possible to restore it to a same-as-new condition but it might still be better than same-as-old. For some components, it may only be possible/practical/safe to repair them a limited number of times after which they would have to be reconditioned/replaced.

Very often, components are redesigned possibly because it has been decided to out-source their manufacture or conversely, bring it back in-house. Sometimes it is because it has proved necessary to change the material from which they are made, sometimes it is the manufacturing process for example using lasers to cut holes rather than conventional drill

bits. Other reasons include improving its reliability or to allow an increase in the performance. Whatever the reason, the new standard of component may be introduced as and when the opportunity arises or if the system is recovered at a certain location or "retrospectively" in which case all system containing the old standard will be taken out of service as soon as practical and the old standard component replaced a new one – this generally only happens if there is a safety issue relating to the old one.

The Model

For simple systems, it may be possible to use some form of Markov model by assuming times-to-failure can be adequately modelled using a [negative] exponential distribution, the inverse of which is the Poisson distribution. As a first order approximation, such a model can give quite acceptable results provided one is not interested in managing arisings (component rejections) as indicated in the "opportunistic maintenance" section above.
If times-to-failure are modelled using an exponential [distribution] then replacing components pre-emptively to avoid unplanned system downtime will simply not work. The unique feature of this distribution is that the probability a component will fail in the next increment of time is completely independent of its current age. To put it mathematically, $h(t) = \lambda$ for $t \geq 0$ where λ is a constant and h(t) is the hazard function for the given distribution. In fact, λ is the mean time to failure or between failures (MTTF or MTBF).

If using a Markov model is not an option (because the system is too complicated and some of the times-to-failure are best modelled using a distribution for which the hazard function is a function of the age of the component (e.g. the Weibull distribution with shape parameter greater than 1(i.e. $\beta>1$)) then probably the best option is to use a discrete-event simulation (DES) model.

It has been said that a model should be as complicated as necessary but as simple as possible. Although this is a good maxim to work to, it is not always very easy to apply. There is a caveat that should also be applied to this maxim: the influence of anything that is not included explicitly in a model cannot be assessed using that model.

At this stage it might be appropriate to consider a particular example. The gas-turbine engines of a commercial airliner provide an excellent example. They are complex and expensive to make, maintain and support but can be considered pretty much in isolation from the rest of the aircraft (or platform in which they are installed). Typically, they will be maintained and supported by the manufactured possibly via what may be referred to as a "total care package".

Case Study

An aircraft has been described as 4 million bits flying in close formation. If the skin of the aircraft is aluminium [alloy] then a large percentage of these bits will be rivets so although there may be 4 million bits, the number of different types of component will, in general, be significantly less than this. In general, rivets and fasteners of all types are probably not worth modelling – their failure will generally be very rare, it is quite likely to go unnoticed for years, it will, in any case be unlikely to be critical and in most cases these components can be regarded as disposable and, as such, be treated as "consumables". In practice, it is normal to replace any fastener that has been removed during maintenance of an aircraft with a new one. For systems that are less safety-critical, such as cars or washing machines, say, fasteners are likely to be re-used many times and are most commonly only replaced if they have had to be destroyed when being removed (e.g. drilled out) or go astray.

A gas-turbine engine is a subsystem of the propulsion system which is in turn a subsystem of the aircraft (which can also be regarded as a subsystem of the fleet). Although it may contain around 30,000 bits, made up of maybe 3,000 different parts, there are really only a small number (around 100) parts whose failure is likely to result in the immediate need for invasive maintenance. Perhaps, not surprisingly, these tend to be amongst the most expensive parts in the engine so are perhaps ideal candidates to model in some detail. Many of these parts are so reliable that they are very unlikely to fail during the normal life of the engine so can effectively be disregarded. Others, however, can, and often do fail for several different reasons. A small number fall into the category of being "safety-critical" which basically means that if they were to fail, there is an unacceptably high risk that they could cause catastrophic failure.

This last group of parts are something of a problem. If a turbine disc breaks when the engine is running at full thrust, the fragments have sufficient energy to penetrate 12 feet of reinforced concrete so containment is not really an option. Redundancy, even if it were feasible, is irrelevant; indeed it would simply add to the problem. In order to manage the risk, these components are given a "life-limit" or "hard life". This is the age beyond which the risk of failure is considered to be unacceptably high. Note this is based on the premise that the primary failure mechanisms of these components are strongly age-related, i.e. the probability of failing in the next unit of time increases with the age of the component. Experience over the past 50 years, or so, has indicated that the times to failure when measured in stress cycles can be adequately described using either a log-normal distribution (with a very small variance) or, more commonly, a Weibull distribution with a shape parameter significantly greater than 1 (typically above 4.5). Determining the "hard life" is a complex process incorporating engineering knowledge, accelerated testing and in-service data but the results are borne out in practice – at the time of writing the author can only recall Rolls-Royce aero-engines having suffered one uncontained failure in over 10 years and that is at a rate of well over 1 million engine flying hours a year. (Note: there is absolutely no point in giving a component a "hard life" if the primary cause of failure is due to external factors such as FOD, poor handling or incorrect installation.)
In the above brief explanation, it was mentioned that the hard life is given in [stress] cycles as opposed to hours. A stress cycle for civil engines (i.e. ones installed in commercial airliners) is generally taken to be one cycle per flight – i.e. from take-off to landing. For military aircraft, it is a great deal more complicated and its measurement relies on a complicated algorithm that converts in-flight data of spool speeds, temperatures, pressures, vibration levels and a number of other factors. In reality, the number of stress cycles any component will experience during any given flight is a random variable dependent on a number of factors ranging from the purpose of the flight to the ambient conditions to the mood and skills of the pilot. Since most of these factors are unpredictable and, more to the point, outside of the control of the operator or maintainer, using a simple cyclic exchange rate for each type of sortie / mission / flight is about the best we can do.

This group of parts (the lifed or life-limited ones) generate most of the "planned" engine removals. With engine health monitoring (EHM) actual component failure has become something of a rarity so very few engine removals are truly unplanned, however, EHM can only predict a failure a relatively short time before it would have happened so for convenience these engine removals will still be referred to as "unplanned".

The need for maintenance on an engine (or an engine removal, euphemistically referred to as an "arising") is driven by the aging process – flying. What maintenance is done on the engine after it has been removed from the aircraft will depend on many factors.
Predicting when components will be rejected, for whatever cause, is only part of the process. True this drives when maintenance will need to be done and to a certain extent what will need to be done during that maintenance but, what is done during maintenance will affect when future maintenance will be needed. It is this that adds particular value to a) collecting good failure data and b) using a simulation model with age-related failure data.
For a component with a significant probability of failing due to some form of aging process, replacing it before it fails will increase the **expected** time to next failure for that component. (Note: there is no guarantee that it will increase the actual time to failure.) For an aero-engine taking it off the aircraft and sending into maintenance is actually quite an expensive process so it is unlikely to be cost-effective to remove an engine simply to replace an old, but otherwise healthy component. If the engine has already been removed and is in maintenance, then the marginal cost of replacing an old component will generally be very small so it may prove cost-effective to do so if this is going to extend the expected time to next engine removal. This is basically the same argument as used to justify applying minimum issue lives.

For a new engine (or any other system), the expected time to the first failure can be calculated if the time to failure distributions for all of the causes for all of the components are known. It is not a particularly simple calculation and a much more convenient statistic to work with is the "characteristic life". This is the age by which one would expect just over 63% of failures to have occurred. It is totally independent of the shape. It also happens to be the age at which the cumulative hazard function (H(t) or Z(t)) is equal to 1.
The cumulative hazard function (CHF) is both interesting and useful. If the failure mechanisms are competing, in the sense that failure of the system will occur as soon as failure occurs in any one of the components of that system then the CHF for the system is simply the sum of the CHF for each failure mechanism of each component in the system. We can now define the characteristic life of the system as the value of T (i.e. the time from new) for which the CHF of the system is equal to 1 (which we will call CLS). And, by definition, it is also the age by which we can expect the first failure of the system to have occurred.

What also makes this function useful is that for a component which has survived to time T, the conditional CHF (H(t+T|T)) is simply the CDF from new minus the CDF up to its present age (i.e. $H(t+T|T) = H(t+T) - H(T)$). Thus, if we wish to rebuild the system (following maintenance) to have an expected time to next failure of a certain value ($L \leq CLS$, say) we can calculate the conditional CHF for the system as the sum of the conditional CHF for each failure mechanism (given the age of each component).

If this sum exceeds 1, then the expected time to next failure is less than the required time (L). We know how much each failure mechanism contributes to this sum and we can also calculate by how much this sum would be reduced if we replaced any given component with a new one so we can calculate the marginal benefit of replacing any given component with a new one ($MB = H(L)+H(T)-H(L+T)$). It then becomes a matter of simple arithmetic to calculate which subset of components to replace in order to restore the system such that its expected time to next failure is greater than a given time.

As to whether this is worth doing will clearly depend on the economics – the relative costs of replacing parts versus the expected revenue. Using a simulation model, which incorporates the above decision process will allow the users to determine the expected costs and benefits for any given value of L (which can be defined as the "target build life"). Running the model for a range of values of L will allow the users to determine the optimum target build life for any given fleet.

Conclusions

For systems whose primary causes for the need for unplanned maintenance are due to aging processes, simulation models can provide a distinct advantage over the tradition Poisson process or Markovian models. Not only will these models tend to produce a more accurate prediction of when failures can be expected but they can also provide a very useful tool for determining what opportunistic maintenance should be done during each maintenance activity, whether unplanned or planned.

These benefits do not, however, come free. In order to run a simulation model of this sort, it is necessary to collect in-service data which gives the age at the time of failure and the cause of that failure. At present, most data recording systems (sometimes referred to as FRACAS) may record the time to failure accurately but seldom record the true cause of the failure, rather they record these events by the symptoms. For example, typical "causes" are given as degradation, half-height failure, root failure, oil smell, cabin smoke or [loss of] TGT margin. None of these explain why the component or system failed; they only give the external symptoms. We may be able to infer from "root failure" or "half-height failure" that this was due to excessive vibration at certain frequencies, "creep", thermal fatigue or metal fatigue but for the others it is like saying that the cause of death was due to the fact that the person stopped breathing.

Over the past 70 years since gas-turbine engines were first used to power aircraft, the reliability of their parts has improved phenomenally. A civil engine installed on a large commercial airline flying "long-haul" say, between London and Los Angeles, can be expected to stay on wing without any invasive maintenance for over 20,000 flying hours (the equivalent of around 10 million miles or 5 billion passenger miles). Unfortunately, from a modeller's point of view, this is disastrous as it means there simply will not be sufficient in-service failure data to be able to obtain estimates of the time-to-failure distributions with any confidence, even if the right data is recorded accurately. Much of the modelling which is currently done therefore relies to a certain extent on "engineering judgment". Accelerated life testing would help improve the statistical accuracy of this data but such testing is both very time-consuming and expensive so is rarely performed on non-safety-critical components. Indeed, the only time it might be done for these parts is likely to be if there is an unexpectedly large number of failures and, in the vast number of cases, this will be due to either a design fault or a problem during manufacture.

If aero-engines can be better protected from extrinsic failures, it is quite possible that there will come a time when they will be regarded as another item that comes under the category of "fit and forget". Until that time, however, simulation modelling is still likely to provide the best means of managing in-service unreliability.

The Role of Simplified Technical English in Aviation Maintenance

Orlando Chiarello, Secondo Mona S.p.A., Italy

Abstract

The role that Simplified Technical English, STE, plays in aviation maintenance has been investigated and discussed in this paper. A brief history of the development of the ASD-STE100 specification [1] within the aviation industry, together with its range of application, an overview of its principles, structure and rules are presented. The primary objective of STE is the removal of linguistic barriers in the continuous attempt for correct understanding of the instructions by the operators, the improvement of flight safety and reliability.

Introduction

Since the first man-made flight in 1903, achieved by a machine made by "bicycle experts", flying machines have become significantly complicated. Consequently, achieving the required level of in-service reliability and functionability has become increasingly more and more difficult. The advanced scientific and technological developments built into flying machines has had to be communicated to thousands and thousands of operational, maintenance and support personnel all over the world, who deliver in-service reliability and functionability of aircraft on a daily basis.

The designers of flying machines communicate with operators and maintainers through technical documentations. Operational and maintenance manuals are generated by manufactures that understand very well the science and technology embedded in their machines. However, operators and maintainers, who are neither scientists nor engineers, have to understand information that the design community is trying to communicate to them. This process was less challenging in the time of the Wright brothers when the associated documentation consisted of several pages, but the corresponding technical documentation in the current, web-oriented "Interactive Electronic Technical Publications (IETP)" represents a real challenge. [2]

The international language of the aviation industry, and all technical documentation is English, but for 80% of operational, maintenance and support personnel in the aviation industry English is not the native language. Consequently, the problem arises daily when in-service personnel, the majority of whom have a limited knowledge of the English language, are trained to make use of the relevant technical documentation and there are endless opportunities for misinterpretation. Like other living languages used around the world, many English words, phrases, and expressions have several meanings, which can be very confusing and potentially dangerous.

For example, for the English word "lift" there are 13 different meanings if it is used as a verb and 8 different meanings if it is used as a noun and there are many words in the same category.

The Need for Improving Technical Communication

The following sentence has been found in one of the existing maintenance manuals: "Round the edges of the round cap. If it then turns round and round, as it circles round the casing, another round of tests is required." This example shows that the word "round" is used with

different meanings and roles (as a verb, as a noun, as an adjective) and clearly illustrates the need for significant improvement of the accuracy of the communication in aviation maintenance.

However, it is necessary to stress that the above example does not mean that the previous maintenance texts were not technically or linguistically correct, but accurate text and correct presentation are not always sufficient. The selection of the English words was (and is) not easy and some users still make mistakes with texts that authors think are written in good English.

The aviation industry is not the only one experiencing problems with technical communication. The first attempt to create a Controlled Language was made by Professor Charles Ogden who created the Basic English [3] in 1930s. In the 1970's the American Company Caterpillar created the Caterpillar Fundamental English, CFE [4] to improve technical communication with their customers and consequently improve the reliability of their products. Similar attempts for the creation of Controlled Languages could be found also in other industries, but none of them became the world standard until the early 1980s when AECMA (The European Association of Aerospace Industries) developed the AECMA Simplified English to help users of the English-language maintenance documentation understand what they read. It was initially applicable to commercial aviation, but became also a requirement for defense projects including land and sea vehicles.

The Development of AECMA Simplified Technical English

In 1979, the Association of European Airlines (AEA) thought that a simplified language applicable to the aircraft maintenance documentation was necessary. The new Controlled Language, a Simplified form of English, should satisfy the following requirements: a small number of words, words with defined meanings, words with defined parts of speech and a simplified structure.

AECMA accepted the task of finding the solution, and for this purpose it formed the Simplified English Working Group (SEWG). After the initial analysis of many maintenance texts, the SEWG defined a set of writing rules, a controlled dictionary and examples. Also, the Aerospace Industries Association (AIA) of America sent their representative in the SEWG.

The result of such extensive work was the AECMA Simplified English (SE) Guide which was first released in 1986 and soon became a mandatory language requirement of the e major specifications for writing maintenance manuals, ATA 100 (now ATA iSpec 2200) [5], and AECMA 1000D (now S1000D) [6].

STE was therefore created to produce documentation that must be:

1. Accurate = conforming to a standard with total accuracy
2. Complete = having all necessary information
3. Relevant = having connection with the subject
4. Concise = express much in few words
5. Convincing = causing to believe the truth
6. Meaningful = having a meaning for the purpose
7. Unambiguous = causing no doubts and leading to only one conclusion.

In 2005, AECMA became ASD, "The AeroSpace and Defence Industries Association of Europe" and the Simplified English Guide became an official Specification, ASD-STE100 (STE), with the word "technical" added to its name. The European Community, in 2006, trademarked STE.

Today STE is maintained by a dedicated group of 16 members, the ASD Simplified Technical English Maintenance Group (STEMG), consisting of representatives from ASD member countries and non-ASD member-countries and in addition, there are associate members representing the customers (Airlines and Military Organizations).

Although the STE structure is stable and consolidated, the language has to be kept in line with the technology evolution and amended on the basis of the continuous and important feedback received from the users.

On 15th January 2013, the STEMG released Issue 6 of the specification, which represents a further step of continuous improvement.

Principles and structure of STE

The STE Specification has the following two parts:

- A set of writing rules
- A controlled vocabulary.

The writing rules (approximately 60) cover aspects of grammar and style. They regulate the use of words, layout and sentence length. The STE writing rules are not meant to replace the English Grammar nor to diminish it. STE is for the benefit of the readers that should, in fact, receive a text in simple and correct English without the necessity of knowing STE. However, writing in STE correctly is not an easy task. It requires a good command of the language and a high level of professionalism from the author's side.

The dictionary (approximately 860 approved keywords) specifies the general words that can be used. The approved words were chosen for their simplicity and ease of recognition.
In general, there is only "one word for one meaning", and "one part of speech for one word". In addition to the specified general vocabulary, STE accepts the use of company-specific or project-oriented words (identified as Technical Names and Technical Verbs), providing that they obey their STE specific rules and fit into one of the specific categories listed in the STE specification.

STE as an important resource for Aviation safety and reliability

While new technologies in aviation are introduced to increase the reliability of systems, human errors committed during their operation and maintenance can become the primary cause of safety and reliability reduction events. In aviation, the concepts of "Human Factors" are directly connected to maintenance. During maintenance, a maintainer can do an incorrect task that can cause a malfunction of components and systems. This malfunction can have important effects on Flight Safety and aircraft reliability. The language used in the maintenance documentation and procedures is important for the correct execution of maintenance tasks. It is more important when the maintainers are relying on procedures not written in their native language and they do not have a high knowledge of English.

Applications of Simplified Technical English in Mirce-mechanics

Quality of maintenance process is quantifiable in statistical terms related to the occurrences of maintenance faults and errors. However, as statistics do not study the causes of statistical behaviour, full understanding of the quality of maintenance is only possible by understanding physical causes and mechanisms that lead to the occurrence of maintenance faults during the maintenance process. Based on the analysis of tens of thousands of maintenance tasks in defence, aerospace, nuclear, transportation, Formula 1 Grand Prix racing, chemical, and other industries, Dr. Jezdimir Knezevic has formulated the second Axiom of Mirce-mechanics, which states that "The probability of faulty execution of any maintenance task is greater than zero." [7]

As on one hand this axiom has a profound impact on all aspects of the life on any maintainable system, such as: reliability, availability, safety, cost, effectiveness and many others, and on the other on associated processes like: manufacturing, operation and logistics support, a continuous effort is made in the application of the Simplified Technical English to the principles and methods of the Mirce-mechanics.

The main objective of this research project is the reduction of human errors in operation, maintenance and support processes through the creation of an STE-based structure for technical communication between design-in community and in-service community, which will consequently improve the safety and reliability of flying machines and flights.

References

[1] Aerospace and Defence Industries Associations of Europe (2013), ASD Simplified Technical English, Specification ASD-STE100.
[2] Chiarello, O., Impact of Accuracy of Technical Communication on the Motion of Functionability, Proceedings of the 1st World Congress of Mirce-mechanics, 28-30 May 2912, Exeter, UK.
[3] Ogden, C. K. (1930), BASIC English – A general introduction with rules and grammar, London, Kegan Paul, Trench, Trubner & Co. Ltd.
[4] Verbeke C. A. (1973), "Caterpillar Fundamental English", in Training and Development Journal, 27, 2, 36-40.
[5] Air Transport Association of America (2011), ATA iSpec 2200, Information Standards for Aviation Maintenance.
[6] Aerospace and Defence Industries Associations of Europe (2009), S1000D – International specification for technical publications utilizing a common source database.
[7] Knezevic, J., Quality of Maintenance – Mirce-mechanics axiom. Journal of Quality in Maintenance Engineering, Vol. 18, No 2, pp 216-226, Emerald Group Publishing Ltd. UK.

Price:

General Public:	**£75.00**
MIRCE Akademy Fellows:	**£50.00**
MIRCE Akademy Members:	**No charge**

To Purchase the Annals of Mirce-mechanics:
Please end an email to: quest@mirceakademy.com

To become a Member of the MIRCE Akademy please follow the link:
http://www.mirceakademy.com/index.php?page=Membership

Call for Papers..Call for Papers..Call for.

Annals of Mirce-mechanics – 2014

Published by the MIRCE Akademy to facilitate exchanges of knowledge and experience between: scientific, engineering and management professionals, which are interested in Mirce-mechanics.

The Annals welcomes the following types of original contributions:

- Presentations of the research results related to all aspects of Mirce-mechanics
- Applications of existing Mirce-mechanics knowledge
- Observational knowledge that could be beneficial for further developments of Mirce-mechanics
- Reports, book reviews and short news that are of a general benefit to Mirce-mechanics

Potential authors please see: Guidance for Authors at http://www.mirceakademy.com

Papers..Call for Papers..Call for Papers.