Life Cycle Reliability and Safety Engineering
Vol.2 Issue 2 (2013) 01-10

Impact of Cosmic Radiation Induced Single Event Effects on Avionics Reliability and Safety

I. Zaczyk, J. Knezevic
MIRCE Akademy, Exeter, EX5 1JJ, United Kingdom
jk@mirceakademy.com

Abstract

Due to the rapid advances in electronics technology and the unrelenting demand for increased avionics functionality in the competitive commercial aircraft industry, the complexity of avionics systems has risen exponentially. As a consequence, ever more advanced microprocessor and memory semiconductor devices are being used that exhibit an increased susceptibility to cosmic phenomena.

Single Event Effects (SEEs) have been the primary radiation concern for avionics since the late 1980's when the phenomenon, which had previously only been observed in orbiting satellites, also began to appear in aircraft electronic systems.

This paper presents the research results obtained by applying the principles of Mirce Mechanics to the scientific understanding of the physical mechanisms that lead to the occurrence of the Single Event Upset (SEU), which is the principal SEE affecting avionic devices. It is caused when a sole incident particle creates a charge disturbance of sufficient magnitude in a memory cell, flip-flop, latch or register to reverse or flip its currently stored data state. Alternatively, in logic or support circuitry a transient voltage pulse can be generated that dependent on the right conditions can propagate through the logic of the device and become latched into a memory cell. Voltage spikes on power supply lines and noise can also cause transient errors

Keywords: Single event upset, Fault tolerant design

1. Introduction

"Left unchallenged, soft errors have the potential for inducing the highest failure rate of all other reliability mechanisms combined." R. Baumann [1]

Mirce Mechanics is a scientific theory that studies the motion of functionability (ability to deliver functionality) through the life of a system. Accordingly a life of a system is a time dependent sequence of positive and negative functionability events that cause the change of its functionability states. One of the most challenging tasks of Mirce Mechanics is the understanding of the physical mechanisms that cause occurrences of negative functionability events, which signify the transition of the system from the positive functionability state (state in which it is able to deliver functionality) to the negative functionability state (state in which it is not able to deliver functionality) [2].

Mirce Mechanics recognise the following three categories of the causes of negative functionability events: atomic, environmental and human. Hence, this paper will examine the overall impact of cosmic rays on the in-service reliability of manmade systems, with a particular focus on the effects on avionics systems.

As the reliance on avionics systems within aircraft increases so do concerns regarding the safety and reliability of these systems, particularly for those systems, which are considered to be 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 negative functionability events induced by ionising radiation from two main sources:

- Cosmic rays from space.
- 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

negative functionability event 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 negative functionability event. 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.

In order to determine the probabilities of occurrence and the resultant impact on a systems safety and reliability a fully comprehensive understanding of the generation, behaviour and the interactions between the relevant physical phenomena must first be understood. Then and only then, can accurate and meaningful reliability and safety predictions become possible, enabling the ultimate goal of reducing the probability of negative functionability event occurrences during the life of manmade, managed and maintained systems. [3]

2. Single Event Effects in Avionics

Single Event Effects (SEEs) have been the primary radiation concern for avionics since the late 1980's when the phenomenon, which had previously only been observed in orbiting satellites, also began to appear in aircraft electronic systems.

The principal SEE affecting avionic devices is the Single Event Upset (SEU) caused when a sole incident particle creates a charge disturbance of sufficient magnitude in a memory cell, flip-flop, latch or register to reverse or flip its currently stored data state. Alternatively, in logic or support circuitry a transient voltage pulse can be generated that dependent on the right conditions can propagate through the logic of the device and become latched into a memory cell. Voltage spikes on power supply lines and noise can also cause transient errors; however appropriate shielding and filtering design measures can suppress these types of disturbances.

The primary sources of radiation are high energy cosmic particles, low energy (thermal) neutrons and low energy alpha particles emitted from device and packaging contaminants.

SEEs can be classified into two main categories, soft errors and hard errors. This distinction is made to clarify the difference between soft errors, which are transient non-destructive errors that can be cleared by resetting the device or writing new data to the upset cell and hard errors that are permanent and potentially destructive. Table 1 summarises the types of SEEs currently affecting electronic devices operating in the avionics environment.

Table 1: Types of Single Event Effects

<table>
<tr><th colspan="3">Single Event Effects</th></tr>
<tr><th colspan="2">Hard Errors</th><th>Soft Errors</th></tr>
<tr><td rowspan="2">SEFI, Single Event Functional Interrupt</td><td rowspan="2">SEL, Single Event Latch Ups</td><td colspan="2">Single Event Upsets</td></tr>
<tr><td>Single BIT Upset</td><td>Multiple BIT Upsets</td></tr>
</table>

The use of the term 'soft error' can be ambiguous and is often used interchangeably with SEU. To make a clear distinction between the terms SEU and soft error within this paper the following definitions will apply:

- Single Event Upset: An incorrect data value held at any storage node.
- Soft Error: A soft error only occurs when the corrupted data state of a node is interpreted by the system as valid data. This term therefore encompasses all of the soft forms of SEE: SEU, MBU and SET which have the potential to produce a soft error.

Radiation can affect electronic devices as the consequence of a single energetic particle strike, termed 'single event' or as multiple strikes over an extended period of time. The effects due to multiple events, Total Ionisation Dose (TID) and displacement damage manifest gradually in electronic components as damage is accumulated over time. These total dose effects and hard SEEs whilst relevant to electronic systems operating in the harsher space environment have a negligible effect on current semiconductor devices used in the terrestrial environment.

Whilst each form of SEE is considered in this paper the main focus will be on SEUs which are the dominant device negative functionability event mechanism affecting electronic devices in the avionics environment.

The second most prevalent SEE is the Multiple Bit Upset (MBU) that occurs when a single particle causes the upset of two or more memory cells. Fortunately MBUs only form a fraction of the total number of SEUs, hence they have little significance except for memory architectures employing Error Detection and

Correction, (EDAC) techniques. In these circumstances, dependent on the type of error correction technique employed, multiple bit errors could have significant consequences if the protected memory is used for flight or mission critical applications. MBUs are generally assumed to attribute 3% of the total upset rate [4] although rates as high as 5% have been reported.

Following MBU, Single Event Functional Interrupt (SEFI) and Single Event Latch ups (SEL) account for the majority of the remaining proportion of SEEs affecting avionic devices. SEFIs occur when an upset initiates an IC test mode or reset mode that causes the device to temporary lose functionality. SELs arise when an incident particle creates a charge disruption sufficient enough to effectively short circuit the device resulting in its permanent change of state or in some circumstances permanent damage if excessive current flows as a result of the latch-up.

The last SEE of avionics relevance that can generate soft errors in the core logic of microprocessors and microcontrollers is the Single Event Transient (SET). They are transient and non-destructive in nature and are capable of producing a soft error, (i.e. the storage of an erroneous data value in registers, memories or latches) only if it is propagated through the logic pathways of the device. This is dependent on the dynamic state of the logic at the time of the particle induced nodal voltage transition and the configuration of the logic pathways within the device. If a soft error occurs normal system behaviour can be restored by resetting or rewriting the incorrect data.

Of all the forms of SEE, SEUs are the most prevalent in avionic electronic devices; Table 2 illustrates the approximate distribution percentage values, between each type of SEE except SETs, for which no reliable data exists. This problem of limited SEE statistical data is an enduring problem in the field as the capturing and recording of SEEs during flight is impeded by:

a) Fault tolerant designs and error correction techniques.
b) SEEs incorrectly diagnosed as electrical interference or random component negative functionability events.
c) Reluctance of semi-conductor manufactures to disclose proprietary information regarding the root cause negative functionability event mechanisms and the historic negative functionability event statistics gathered from devices returned from in-flight usage.

Table 2 – Main SEE Apportionments – Avionics Environment

Single Effect Event Type	Percentage
Single Event Upset	90 %
Multiple BIT Upset	5 %
Single Event Functional Interrupt	3%
Single Event Latchup	2%

The current convention is to discuss the rate of SEU occurrence in terms of soft error rates (SER), which are measured in failures in time, (FIT). One failure in 1 billion device operating hours is defined as 1 FIT. This term is also widely used in the semiconductor industry to state the expected occurrence rate of hard negative functionability event mechanisms.

The first efforts to calculate SEU rates were presented in two papers in 1984 paper by Tsao et al. [5] and a companion paper by Silberberg et al. [6]. The Tsao paper detailed methods of calculating SEU rates from primary & secondary cosmic rays reaching down to 40,000ft and the Silberberg paper introduced methods for calculating SEU rates resulting from secondary neutrons in the atmosphere.

The reason that semiconductors have become susceptible to SEEs in the terrestrial environment rather than existing solely in space can be partially attributed to the commercial demands for increased functionality and performance, whilst lowering power consumption and cost. To fulfil these requirements component manufactures have continued to reduce the geometry size of integrated circuits with each new generation resulting in higher gate speeds, increased feature density and reduced power consumption.

Another contributing factor also exacerbating the situation is the movement by component manufactures away from the production of specialist components designed specifically for space, military and aerospace applications, towards less robust Commercial Off The Shelf (COTS) devices.

Prior to 1980, semiconductors with features sizes greater than 1μm were in general immune to the effects of radiation but as features sizes have continued to decrease into the deep submicron range (<0.25μm) SEEs have become a very real threat to the reliability of avionics systems.

Figure 1 illustrates this progression of semiconductor scaling from 10μm devices in the 1970s to recent technology with feature sizes as low as 45nm.

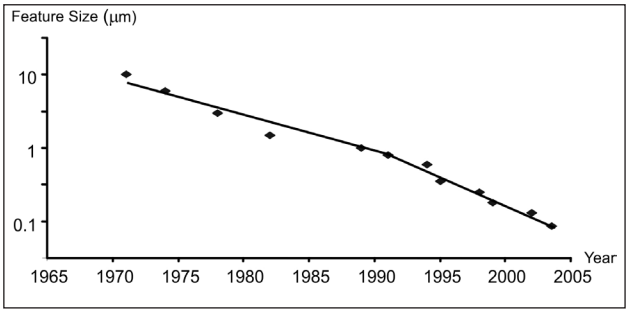

Figure 1 – Semiconductor Device Scaling[1]

To provide a sense a scale, 2000 transistors at 45nm each could fit across the width of a human hair or 30 million onto the head of a pin. This trend is set to continue well into the future, with 22nm SRAM and Microprocessors, now in production in by Intel and by 2017, devices with feature sizes as small as 8nm are planned which considering the width of a silicon atom is 0.24nm the limits of what is achievable using current silicon technology will soon be reached.

Whilst technology scaling enables the demands of system designers to be met the downside of this is an increased sensitivity to radiation. Within a memory device this is caused by a reduction in the capacitance inside a cell and a significant increase in the number of cells that could potentially be upset within each device. Less capacitance in a device due to the shrinking of process technology and reduced supply voltage means that the minimum amount of charge necessary to hold data in a device, either a logic 1 or 0 is also reduced. This quantity of charge, known as the critical charge, is therefore more susceptible to a charge disturbance caused by an incident radiation particle, thus eroding the components resistance to SEUs. The approximate critical charge of a node can be calculated using the expression:

$$Q_c = C_{node} \times VDD \quad (1)$$

where: Q_c is Critical Charge, C_{node} is Node Capacitance and VDD is Operating Voltage.

A lower nodal critical charge is therefore more likely to be 'upset' by incident particles with a lower energy, because the flux of energetic particles increases at lower energy levels.

The components most susceptible to SEU are therefore devices that contain the largest number and density of potentially volatile bits namely memories and microprocessors. Table 3 contains a list of the devices that are currently considered to be the most susceptible to SEU in aircraft avionics systems and includes the specific regions within the architecture that are most at risk.

Table 3 – SEU Sensitive Devices

Devise Type	Sensitive Areas
SRAMS and DRAMS	Memory cells and control logic.
Microprocessors and Microcontrollers	Registers, cache, sequential and combinational control logic.
FPGAs and ASICs	Combinatorial logic and sequential logic

Opto-electronics and power switching components are also susceptible, to various forms of hard and soft SEE but are not considered in this paper due to their very low probability of failure in the avionics radiation environment.

Each of the factors discussed in this section, increased functionality and performance, lack of specialist devices, lower critical charge and higher cell density all impact upon the SEU tolerance of advancing semiconductor designs.

The net effect is an increase in the overall device SER that if not adequately mitigated against using appropriate methods such as error detection and correction (EDAC) and architectural redundancy, will result in an increased system SER [1] plus potentially an increase in the number of mission or flight safety critical negative functionability events.

3. SEU Negative Functionability Event Mechanisms

3.1 The avionics radiation environment

To fully understand the negative functionability event mechanism that causes a sudden bit flip or a transient pulse in a semiconductor device installed within the avionics of a commercial aircraft, according to the principles of Mirce Mechanics it is essential to first identify and comprehend the physical nature and origins of the mechanism that lead to these events.

In essence soft errors can be attributed to two main sources, externally produced cosmic ray radiation and

[1]"Semiconductor Device Scaling" Image retrieved from David Harris, Harvey Mudd College, Introduction to CMOS VLSI Design Scaling and Economic, http://users.ece.utexas.edu/~adnan/vlsi-05-backup/lec21Scaling.ppt and Intel Corporation, http://www.intel.com/technology/itj/2008/v12i1/7-evaluation/figures/Figure_1_lg.gif.

locally produced radiation emitted from contaminates within the semi-conductor material itself or the device packaging.

As a result of this interaction between the natural radiation environment and electronic devices a number of different negative functionability mechanisms are initiated. An overview of these mechanisms is presented in section 3.2.

3.2. The mechanics of an SEU

At the simplest level soft errors are created when an incident particle strikes a sensitive region of a memory cell, register, latch or flip flop and produces sufficient excess charge within the device to reverse or flip its current data state. When a particle travels through a semiconductor, excess electron-hole pairs are generated in the path behind the traversing particle, and if the electric field in the vicinity of the PN junction has sufficient strength the holes and electrons will be swept away and collected by the oppositely charged device contacts. A soft error will then occur if the amount of collected charge exceeds the critical charge threshold level of the device. This threshold level is based on the sensitivity of each device to excess charge and is dependent on many factors. The basic mechanics of a soft error negative functionability event can be explained in more detail by considering the mechanism as a three phase process consisting of the following three phases:

- Charge deposition,
- Charge collection
- Device response.

Each of these phases are described below in sections 3.2.1 to 3.2.4.

3.2.1 Charge Deposition

The first phase in the generation of an SEU is the deposition of charge within the device as a result of atomic and nuclear interactions between the incident particle and the atomic lattice of the semiconductor material. As a particle traverses through a material two types of interaction can take place dependent not only on the particles charge but also on its trajectory. A key distinction to make at this point is that neutrons by their very nature carry no charge and hence can only interact by kinetically striking the atomic nuclei of the semiconductor material. Charged particles can also interact in this way but are also capable of transferring energy electromagnetically to the valence electrons orbiting the nuclei of the semiconductor material. This diversity between particles results in charge being deposited in a device by two fundamental methods of interaction, known as:

a) Direct ionisation: caused directly from incident charge particles.
b) Indirect ionisation: caused by secondary particles created from nuclear reactions between incident charged or uncharged particles and the atoms of the semiconductor material.

Both of these forms of ionisation may result in a soft error if sufficient excess charge is deposited. Each type of ionisation mechanism is described briefly in the following sections 3.2.2 and 3.2.3.

3.2.2 Direct Ionisation

When a charged particle transits through a material it will lose energy along its path primarily through interactions with the materials electrons, leaving a trail of atoms with ' kicked out ' orbital valence electrons. The particle will then come to rest in the semiconductor material only when it has lost all of its energy, after travelling a total distance through the material known as the particles range.

The size of the disturbance and the subsequent probability that the incident particle will cause an effect is dependent on the amount of energy deposited. This can be described in terms of the incident particles linear energy transfer (LET), which is the ability of a particle to release its energy into a material. LET can be defined as the energy loss per unit path length as a particle travels through a material. In Silicon for every 3.6 eV transferred to the material, one electron hole pair is created. As LET is a function of a particles energy level, mass and the materials density, the highest LETs will therefore typically occur when very energetic particles with greater mass transit through denser materials.

3.2.3 Indirect Ionisation

Indirect ionisation can occur as a result of charged and uncharged particles, but in the avionics environment it is the very penetrating and uncharged neutron that has the highest LET potential.

The probability of a neutron striking the nuclei of a semiconductor material is very remote but when it does occur an elastic or inelastic nuclear collision will take place. These two forms of nuclear collision can be differentiated by the amount of energy exchanged. In an inelastic collision large energies are exchanged in the creation of reaction fragments whereas in

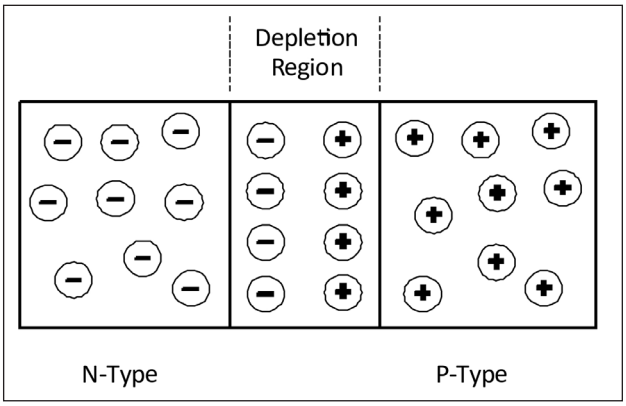

Figure 2 – Depletion Region

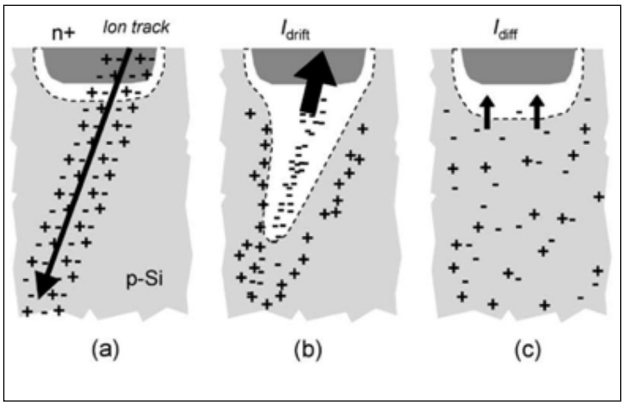

Figure 3 – Charge Generation and Collection[2]

elastic collisions involving the slight deflection of a neutrons trajectory, much less energy is transferred. The reaction particles produced by inelastic collisions are generally much heavier than the incident neutron and are consequently much less likely to produce an SEU than elastic reaction products which are normally less energetic and tend to remain close to the original impact location.

After a collision the charged reaction products of both inelastic and elastic collisions will then interact with the semiconductor material through direct ionisation. If the quantity of collected charge, deposited by secondary fragments exceeds the critical charge threshold of the device an SEU will occur.

3.2.4 Charge Collection and Device Response

Within electronic devices the reverse biased PN semiconductor junction is the most sensitive area to the deposition of additional charge caused by an incident ionising particle. PN junctions are the point at which P-type and N-type semiconductor materials come into contact. P-type and N-type semiconductor materials are created by a process known as doping which adds an impurity to the silicon to create either a net loss or gain of valence electrons throughout the materials atomic lattice.

Silicon has four valence electrons, which it uses to form co-valent bonds with four adjacent silicon atoms. If an impurity e.g. phosphorus is added that has five valence electrons to make an N-type material, four valence electrons of the phosphorus atom will bond with the silicon leaving a free electron to carry negative charge. Conversely, to form a P-type material an impurity with three valence electrons e.g. boron is added resulting in a missing electron or hole from one of the four covalent bonds. The created hole then acts as a free positive charge carrier. Within an N-type material electrons are defined as the majority charge carriers and holes as minority charge carriers. In a P-type material holes are the majority charge carriers and electrons the minority charge carriers.

At the PN junction of an isolated semi-conductor the excess electrons from the N-type side combine with the holes in the P-type material, at the same time, holes from the P-type material diffuse over to the N-type side to combine with free electrons. This process as shown in Figure 2 creates a layer devoid of majority charge carriers known as the depletion region. The presence of positive and negative charges on each side generates an electric field that creates a voltage potential across the junction.

In isolation the junction remains in a state of equilibrium. However if a forward biased external voltage is applied the width of the depletion region is reduced enabling a diffusion current to flow. In a reverse biased condition the width of the depletion layer is increased creating a barrier to current flow by majority charge carriers but allowing a small reverse current to flow via minority carriers.

If an incident particle strikes a reverse biased junction as shown in Figure 3 the presence of a strong electric field provides an efficient method of charge collection for the deposited charge, sweeping the electrons and holes to the device contacts using a combination of drift, diffusion and field funnelling charge transport mechanisms before they can

[2]"Charge Generation and Collection," Figure from [1], original label "Fig. 2. Charge generation and collection phases in a reverse-biased junction and the resultant current pulse caused by the passage of a high-energy ion".

recombine. In a forward biased junction deposited charge is more likely to recombine and hence less likely to result in an SEU.

Figure 3 (a) shows the transient disruption to the electrostatic potential of the node caused by the generation of electron hole pairs as a particle transits through the PN junction. This well documented 'funnelling effect' (b) extends into the semiconductor substrate increasing the quantity of charge collected and results in a rapid current rise. Following this phase is a longer period of electron diffusion (c) into the depletion region until all the excess charge carriers are removed by a mix of transport charge mechanisms.

At the device node the resultant quantity of collected charge is dependent on a wealth of factors relating to the device characteristics, node size, doping, etc., in addition to the specific properties of the incident particle such as strike location and LET.

If the total charge collected exceeds the critical threshold level of the node, which is also a function of the device characteristics, primarily operating voltage and nodal capacitance, a change or "Single Event Upset" of the devices data state will occur.

3.3 Radiation sources of negative functionability event mechanisms

To clarify the relationships between the different types of radiation present in the avionics environment and the physics of the SEU mechanism described in this section, Table 4 below has been compiled to provide a summary of the characteristics of each type of radiation particle and the resultant negative functionability event mechanisms that can be induced.

3.4 SEU in memories

Complementary Metal Oxide Semiconductor (CMOS), SRAM and DRAM are used extensively throughout avionic electronic circuits for the storage of data. DRAM devices are typically used as main system memory whereas the lower power consuming and faster SRAM is usually embedded within processors, FPGAs and ASICs. Due to the their distinct circuit topologies and operating characteristics SRAM and DRAM devices behave differently to the deposition of charge caused by an incident particle strike.

DRAM devices generally consist of a single transistor and capacitor for each bit of data while SRAM devices are constructed in most cases from six transistors although more can be added to act as redundant elements or to provide additional functionality. As a result of these architectural differences between DRAM and SRAM, the number of bits stored per unit of volume, also known as the bit density, is greater in DRAM devices.

Another significant difference between the two types of memory is that SRAM cells have an active feedback mechanism provided by a cross coupled "restoring" transistor. DRAMs in contrast have to be periodically refreshed; hence any charge disruption will remain unless corrected by dedicated circuitry. A soft error in a DRAM cell typically consists of the relaxation of the stored charge package resulting in a 1 to 0 transition, [7].

In a SRAM cell as shown in Figure 4, if a particle hits the reverse biased junction (formed between the drain and substrate of the transistor) in the "off " state a voltage transient is induced at the drain of the struck transistor as current flows through the

Table 4 – Negative Functionability Event, Mechanism Summary

Radiation Type	Radiation Source	Method of Charge Deposition	Negative Functionability Event Mechanism
Thermal neutrons	Secondary cosmic ray neutrons	Indirect Ionisation	Interaction between thermal neutrons and materials containing the Boron-10 isotope creates secondary ionising particles
Low energy alpha particles	Radioactive decay of uranium and thorium impurities located within the device materials	Direct Ionisation	4 to 9 MeV alpha particle, creating an electron hole funnel
High energy neutrons (10 MeV - 1 GeV)	Secondary cosmic ray neutrons	Indirect Ionisation	High energy neutron collisions with silicon nuclei

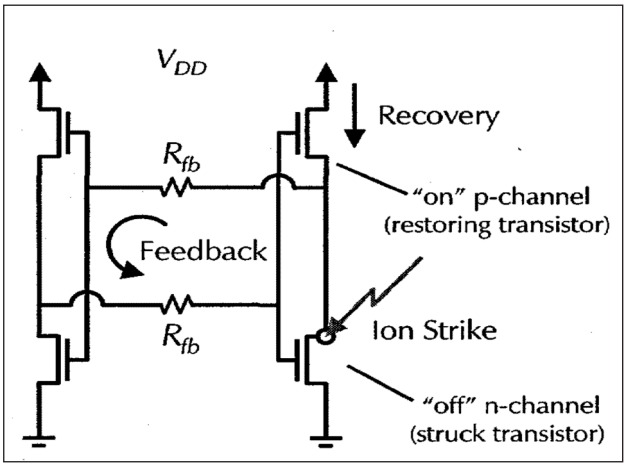

Figure 4 – SRAM Cell Negative Functionability Event Mechanism[3]

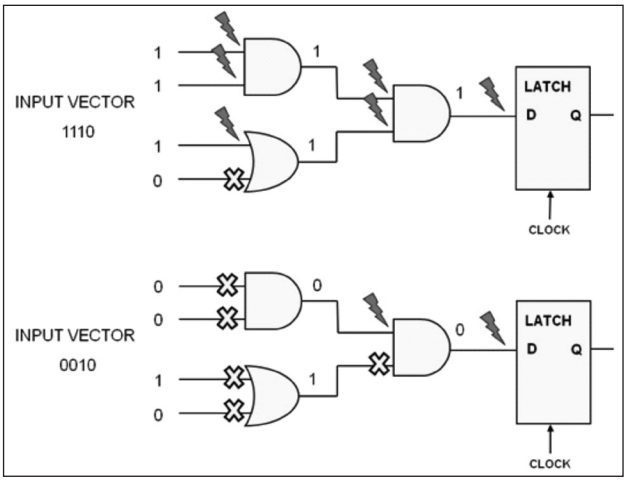

Figure 5 – Logical Masking

restoring transistor as it attempts to restore charge to the disrupted transistor. Charge stored at the drain of the struck n-channel transistor representing a digital 1 is flipped to a 0. This voltage transient acts like a write pulse and if the node does not recover in time the incorrect data value will be stored within the cell. Alternatively a 0 to 1 transition occurs if the drain of a p-channel transistor is struck in the "off" state. In this scenario the n-channel transistor then acts as the restoring transistor. If the transistors are hit in the same location but are in the "on" state the logic level will be reinforced and thus no bit flip will occur.

The probability of an incident particle reaching the critical charge threshold level of the circuit and causing a soft error within an SRAM cell is thus dependent on the dynamics of the cell feedback mechanism, inherent circuit design features and also the shape and magnitude of the collected charge pulse.

3.5 SEE in logic

Electronic circuits contain a mix of combinational and sequential logic elements that are used to interface between the major electronic circuit components such as microprocessors and microcontrollers, although both these devices also contain a highly integrated synthesis of combinational and sequential circuitry.

The fundamental difference between sequential and combinational circuits elements are that sequential circuits have memory and combinational do not. Combinational logic is dependent solely on the state of the inputs at the same instant in time whereas in sequential logic the outputs are derived from the current inputs plus the sequencing history of previous inputs. Examples of combinational and sequential logic circuits are as follows:

- Combinational Circuits: Multiplexers, Adders, Comparator, Arithmetic Logic Unit.
- Sequential Circuits: Flip-Flops, Latches, Registers, Counters.

Within a logic element when a sensitive node is struck by an ionising particle resulting in an SEU the possibility exists that due to the various inherent masking effects within the circuit the upset is not propagated to the observable outputs. In these circumstances no erroneous value is captured by the sequential circuit elements therefore the possibility of a soft error occurring is zero. For a soft error to manifest, the SEU generated pulse must survive electrical, logical and timing (or temporal) masking effects. This section will now describe each of these masking effects in more detail.

Electrical masking occurs if the magnitude of the erroneous particle induced signal is insufficient to latch into sequential elements due to the electrical attenuation of the signal on its transitory path through the logic gates of the circuit.

Logical masking is the blockage of a signal's propagation through a circuit as a result of the logical architecture and the status of other circuit inputs. This effect in illustrated in Figure 5 which shows two

[3] "SRAM Cell Negative functionability events Mechanism," Figure from "Mechanisms and Modeling of single-event upset" by P. Dodd, Sept 1998, http://www.osti.gov/bridge/servlets/purl/1264-Amehmk/webviewable/1264.PDF.

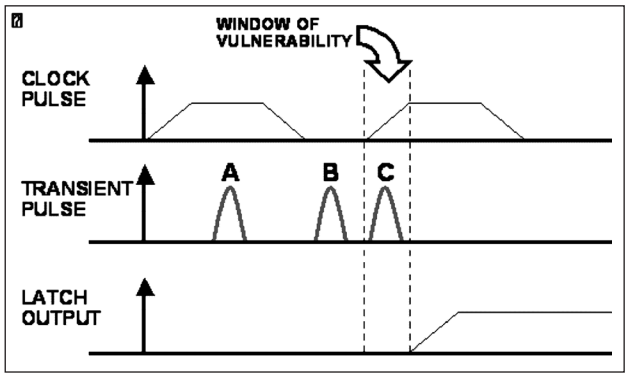

Figure 6 – Timing Masking

different input vectors into basic combinational logic consisting of two AND gates, one OR gate circuit and a latch. If a node is struck it will generate a voltage transition from either high to low or low to high.

Dependent on the design of the circuit's logic gates the particle induced voltage transition will be prevented from progressing through the circuit or it will simply be masked due to another logical value taking precedence at the gates inputs. In Figure 5 the disparity between the sensitive nodes marked with an incident particle strike and the masked modes, marked with crosses, is shown for two different input vectors. A voltage transition from either high to low or low to high will have no effect on the logical outcome of the circuit at any of the masked nodes.

It is clear that logical masking as a function of the circuits input values and design is a significant factor that must be taken into account when modelling the SEU response of a circuit. The final masking effect is contingent on the arrival time of the particle generated pulse at the latch input. A latch will capture and output its input state only at a rising edge of a clock pulse therefore any spurious SEU induced transients occurring outside of this time period termed as the 'Window of Vulnerability' will be masked. Figure 6 illustrates the timing masking effect on three transient pulses. Pulses A and B are outside the window of vulnerability and hence are masked by the circuit, but pulse C that occurs within the susceptible time window is captured by the latch and output as a valid data state.

In the event of a particle induced voltage transient surviving all three masking effects, becoming latched and hence interpreted by the system as a valid data state the possibility still exists that the soft error will not manifest as a loss in function at the system level. The soft error could be prevented from escalating to a negative functionability event due to any one of the following reasons:

a) Masked by subsequent logic in the circuit,
b) Erroneous value overwritten in a later clock cycle,
c) Soft error occurs within a redundant or unused circuit element,
d) Erroneous value irrelevant to software functionality.

As a result the outcome of an SEU is not only dependent on the characteristics of an individual particle strike, but also on many aspects of the circuit design and functionality.

4. Conclusion

The main objective of this paper was to demonstrate the necessity of addressing all physical causes that lead to the transition of a system from positive to negative functionability state during its life. Addressing the reliability and safety characteristics of a system in isolation from the investigation of the impact of the natural environment is not sufficient. Hence, results of the research performed in [3], presented here; have shown the significant impact of cosmic radiation on the occurrence of negative functionability events, and consequently, the necessity for addressing them when considering the reliability and safety characteristics of avionics, at the design stages of a systems development.

This paper presents the research results obtained by applying the principles of Mirce Mechanics to the scientific understanding of the physical mechanisms that lead to the occurrence of the Single Event Upset (SEU), which is the principal SEE affecting avionic devices. It is caused when a sole incident particle creates a charge disturbance of sufficient magnitude in a memory cell, flip-flop, latch or register to reverse or flip its currently stored data state. Alternatively, in logic or support circuitry a transient voltage pulse can be generated that dependent on the right conditions can propagate through the logic of the device and become latched into a memory cell. Voltage spikes on power supply lines and noise can also cause transient errors.

In summary this paper advocates that any system reliability and safety considerations must include the full understanding of the complex interactions between functionability significant processes and to determine the influence of each discrete factor, on the

functionability trajectory through life of a complete avionics system. Then and only then, can accurate and meaningful reliability and safety predictions become possible, enabling the ultimate goal of reducing the probability of the occurrence of negative functionability events during the life of manmade, managed and maintained systems.

Acknowledgement

Authors wish that acknowledge the financial support obtained from the Research Fund of the MIRCE Akademy that enabled this research to be performed.

References

1. R. Baumann, " Radiation-induced soft errors in advanced semiconductor technologies," IEEE Transactions on Device and Materials Reliability, vol 5, No 3, pp. 305–316, Sept. 2005.
2. J. Knezevic, Scientific Scale of Reliability, Proceedings of International Conference on Reliability, Safety and Hazard, Bhabha Atomic Research Centre, 2010, Mumbai, India.
3. Zaczyk, I., Analysis of the Influence of Atmospheric Radiation Induced Single Event Effects on Avionics Failures, Master Diploma Dissertation, MIRCE Akademy, 2008.
4. International Electro technical Commission, Technical Specification – IEC TS 62396-1. Process management for avionics – Atmospheric radiation effects – Part 1: Accommodation of atmospheric radiation effects via single event effects within avionics electronic equipment, 1st Edition 2006-03.
5. 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.
6. 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.
7. P. Dodd and L. Massengill, "Basic mechanisms and modelling of single-event upset in digital microelectronics," IEEE Trans. Nucl. Sci., vol. 50, No 3, pp. 583–602, June 2003.

Life Cycle Reliability and Safety Engineering
Vol.2 Issue 2 (2013) 11-33

Modelling Air Traffic Control System Complexity and its Impact on Human Reliability in Relation to Organisational Context

***Saleh H. Al-Ghamdi, #Oliver Straeter, #Prof. Dr.Habil.**
**General Authority of Civil Aviation (GACA), Saudi Arabia*
 # University Kassel, Germany
dc97sha@hotmail.com

Abstract

This paper proposes a systems view for understanding an Air Traffic Control system (ATC) from the perspective of human reliability and safety of ATC system. In particular it uses two models which are based on systems theory. These models are the Viable System Model (VSM) and the Connectionism Assessment of Human Reliability (CAHR) model. Managing complexity in an ATC environment is fundamental to improving safety and reducing human error. While there have been several studies focusing on human error in ATC systems, this paper starts an investigation of ATC system failure caused by either human error or the organizational structure/context in which people work. It also proposes a new framework for assessing human error. A qualitative and quantitative analysis is performed by using the VSM and the CAHR respectively. Both interact with each other in order to model complex systems: while the Viable System Model (VSM) models the structure of a system from its technological and organizational perspective, and the Connectionism Assessment of Human Reliability (CAHR) models the human reliability influences. Together they propose a resilient approach towards complex system assessment, which was applied to the ATC environment.

Key words*: Air Traffic Control; Stafford Beer; VSM; CAHR, Safety; Human Error, Human Performance, HRA.*

1. Introduction

Human Reliability Analyses (HRA) models are used to estimate the probabilities of human errors that can potentially fail the defenses. However, this estimation needs to take into account the work environment and task conditions under which the work is done, since these can provide an important influence on the likelihood of error. For example, bad weather, long shift times, and high workload all can increase significantly the likelihood of human errors. In turn, work environment and task conditions are often influenced by organizational factors like work rules, duty times, and so on. Therefore, the error estimation process needs to account for these contributing factors. Human reliability analysis employs a set of tools to estimate the likelihood of required human actions being performed when needed. These likelihoods can then be incorporated into the overall risk assessment, so they can be combined with other probabilities, such as those of equipment faults and other hazardous states, to estimate the overall likelihood of hazardous events.

2. Using System Theory in the Analysis of Human Reliability

The purpose of Human Reliability Analysis (HRA) is to properly represent the human element in a safety case and provide evidence that a system is able to perform safely from the human point of view. Therefore HRA needs to meet two key objectives. The first objective is to be in alignment with the modeling of the technical system. Technical systems are usually represented in the approach of system theory. Secondly, HRA need to be in alignment with the human characteristics. Modeling these goes beyond system theory as the human has a much broader variety in behaviour and is much less determined than technical systems. However human modeling needs to be compatible with system theory in order to fulfill the objective representing the human element in a safety case. 1st generation HRA models did favour the system theoretical view while tolerating less accuracy in the human modeling. 2nd generation HRA models did focus on accuracy in the human modeling, but caused problems in fitting this to the classical way of

safety case modeling. The approach here is solving this trade-off by using the VSM model for linkage of the human element to system theory and using the CAHR approach to model the richness of human behaviour. The VSM model allows for recursive and complex modeling of the human element while preserving the system theoretical view. In addition the CAHR approach allows for accuracy in the human modeling. Overall the approach provides a path from system theory to human modeling that is needed to improve safety cases according to the human as well as organizational elements.

Moreover accidents/incidents that are caused by human error are better seen as a complex interaction of technical, social, organizational and managerial factors as well as the environment. Therefore, a system approach is necessary to capture the broader view of human error by considering the total system view.

2.1 Overview of the VSM

The Viable System Model (VSM) is perhaps one of the most insightful and powerful tools available today for studying the structure of organizations. It focuses on the resources and relationships necessary to support an organization's viability rather than on the organization's formal structure, thus offering a way to overcome the traditional over-emphasis on hierarchical relationships. Its basic assumption is that viable organizations emerge when people find successful strategies for working together, to the extent that they are able to develop and maintain a group identity in spite of environmental disturbances. These strategies entail creating, in one form or another, organizational mechanisms for the invention, re-invention, development and maintenance of the organization over time. People, supported by all kinds of other resources, constitute these mechanisms. These resources create policies, and provide intelligence, cohesion, co-ordination and implementation capacity for the organization, i.e. they provide its functional capacity. The structural problem is in creating the conditions for people to relate to each other in such a way that they enhance the organization's chances for viability beyond survival. This requires respect for their autonomy in a cohesive and creative structural context. For instance, it is not good enough for an enterprise to have a well-designed business process relating it with its customers if it is not well supported by organizational processes. These are the processes both maintaining its autonomy and cohesion with other business processes, and ensuring that its meaning is aligned to the meaning of the organization as a whole. These processes, underpinned by structural mechanisms, support the effective implementation and adaptation of the organization's policies (Espejo *et al*, 1999).

The VSM has been developed by Stafford Beer in the 1950's. As a manager and a management consultant, Beer was searching for a completely new way of organizing and managing complex social systems. Drawing on his extensive knowledge in neurophysiology, he discovered that every viable system has exactly the same structure. This means, that companies are being designed in analogy to the human body which is able to quickly and effectively adapt to a constantly changing and highly complex environment. Beer used his findings to stipulate "the rules whereby an organization is survival-worthy: it is self-regulated, it learns, adapts, evolves", (Beer, 1979). The generic VSM (see figure 2.1), is made up of six inter-related components and the communication channels between them:

- System 1: Operations or implementation units where the company produces what the customers demand; e.g. a production area;
- System 2: Co-ordination, to co-ordinate the activities of the various Systems 1;
- System 3: Management and Control, to inform each System 1 what is required from them and to monitor performance; e.g. setting and monitoring performance targets;
- System 3* (three star): an additional Auditing function which cross checks that

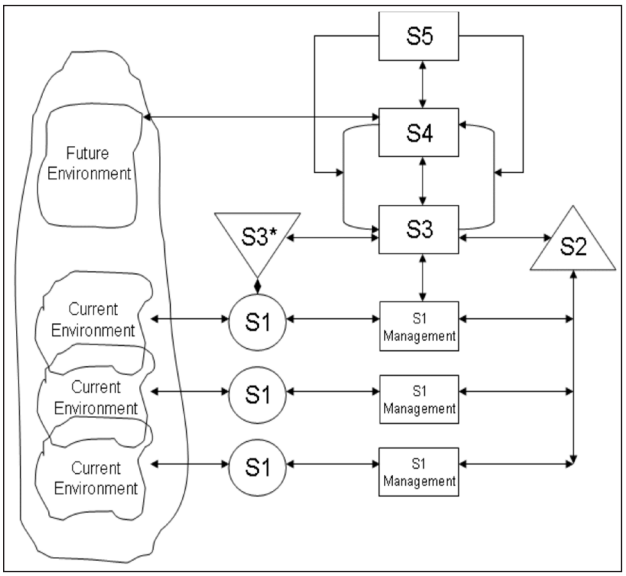

Figure 2.1 Viable System Model

the Systems 1 and System 3 are mutually effective;

- System 4: Intelligence looks at the external environment to help determine future opportunities and threats; e.g. strategic market research;
- System 5: Policy making which includes maintaining a balance between the current needs of the company (managed by System 3, the "inside and now") and the future needs (identified by System 4, "the outside and then").

Beer argued that each System 1 should be given maximum autonomy compatible with the need for System 1 in order to operate properly within the strategic framework determined by System 5. Where necessary System 3 has to ensure that each System 1 learns how to fulfill this role effectively. This autonomy is sometimes called 'empowerment' in management-systems. The System 2 comprises all functions that coordinate the relation and the balance of interests between Systems 1, e.g. any kind of rules and regulations, controlling, Human Resources and IT-support. System 3 (operative management) needs to cope with all the variety of information being produced by each of the System 1s (operations). However, it can hardly deal with all variety of all Systems 1. Referring to Ashby's Law of Requisite Variety, only variety (of the corporation) can absorb variety (of the environment). Where there is inequality in the level of variety then systems must be put in place to amplify or attenuate this variety. It is at the interface between the six VSM Systems that variety control has to be built in to ensure the correct balance of variety (Ashby, 1956).

The VSM allows for the analysis, redesign and control of even highly complex and diversified organizations, as this basic structure containing six subsystems repeats on each level of recursion. When the "System in focus" is the globally diversified corporation, then the Systems 1 could be the subsidiaries in each country. In each country then the different product lines form the Systems 1 and so on. When defining the Systems 1, the only prerequisite is that the Systems 1 are viable Systems themselves, i.e. "able to maintain a separate existence", (Beer, 1985). The VSM can be used to audit the effectiveness of the Organisation as a whole and each of its components, as well as the information channels between them and the overall variety balance. Whenever there is any shortfall then the organization is at an increased risk of being unable to adapt to changes in the external environment and is at risk of failing.

Viable systems depend on other viable systems at a minimum of three levels: (1) systems at the next level down, or those systems that comprise or produce the system; (2) systems at the same level that have direct input and output linkage; and (3) larger embracing systems. This observation leads to the concept of recursion. Systems are built up from other, usually simpler, systems. Beer states this as the Recursive Systems Theorem:

"In a recursive organizational structure, any viable system contains and is contained in, a viable system." (Beer, 1979).

In a previous paper titled *"The Air Traffic Control System as a Viable System: The Case of the Saudi System"* (Al-Ghamdi, et al,(2010), the ATC system and its complexity was modeled as a viable system showing how various parts of this system interact to make the whole.

2.2 Overview and application of the CAHR Model in ATC

The previous section described the Viable System Model (VSM) which can be used in the analysis of the main factors that affect the ATC Controller performance in relation to organizational context. In order to evaluate the safety performance of a system, it is necessary to use a complementing model for assessing ATC Controller performance that can be used in conjunction with the VSM and applied at any recursion level. For the objective to assessing human reliability in Air Traffic Control System, the most appropriate model that can complement to the VSM is the CAHR model. It allows to build upon the VSM structure and allows to evaluate the VSM model performance dynamically. CAHR enables this by using a connectionism approach to depect the interrelations of PSFs, performance and cognitive activities and allows herewith to assess virtually all combinations between these, as long as they were observed in events.

Since our main focus is on the human side of the ATC System, the CAHR model will be applied at the lowest recursion level of the VSM, i.e. the sector level because it is the lowest operational unit which contains the human part (the ATC Controller). However, before this model is applied, a brief description of the CAHR Model is required to gain good understanding especially when applying the model in ATC System.

This model will provide the objectivity required to link and translates the subjective outcomes of the VSM analysis into a meaningful set of reliability figures for the controllers. These figures are represented in the form of probabilities of human error (HEP's).

2.2.1 Introduction

CAHR stands for Connectionism Assessment of Human Reliability and combines event analysis and assessment in order to use past experience as the basis for human reliability assessment. The method uses the connectionism algorithm which is a term coined by modeling human cognition on the basis of artificial intelligence models. It refers to the idea that human performance is affected by the interrelation of multiple conditions and factors (internal and external) rather than by singular ones (Everdij and Blom, 2008).

A connectionism approach that is representing the complexity of human cognition process is used here to cope with the problem of uncertainty, i.e. that no cognitive model can be certain. It represents the relationship of cognitive processes to human performance, its interdependencies and the relationships to contextual and situational conditions in some kind of knowledge management system. The connectionism approach assumes that the brain is a complex net of cells (i.e. more realistic than the classical model of input, process, output), (Everdij and Blom, 2008).

The approach is implemented as a database used for analyzing operational disturbances which are caused by inadequate human actions or organizational factors. CAHR has a generic underlying model that is applicable to all observable events and to allow the collection of all information on human errors from events. The information is stored in a data base that contains generic structure for the event analysis that is extendable by the description of further events. The knowledge base contains information about the system state and the task as well as for error opportunities and Performance Shaping Factors (PSFs). The generic structure represents that Human performance is depending on multiple relationships between PSFs and errors and context. Dependencies between failures, PSFs and situational characteristics have to be considered in HRA (Human Reliability Assessment). Therefore, Human failures always have to be seen in relationship to the performances of humans in technical systems (Straeter, 2001).

In detail, the CAHR technique is based on the evaluation of the operator's task from the incident description and identification of interactions between

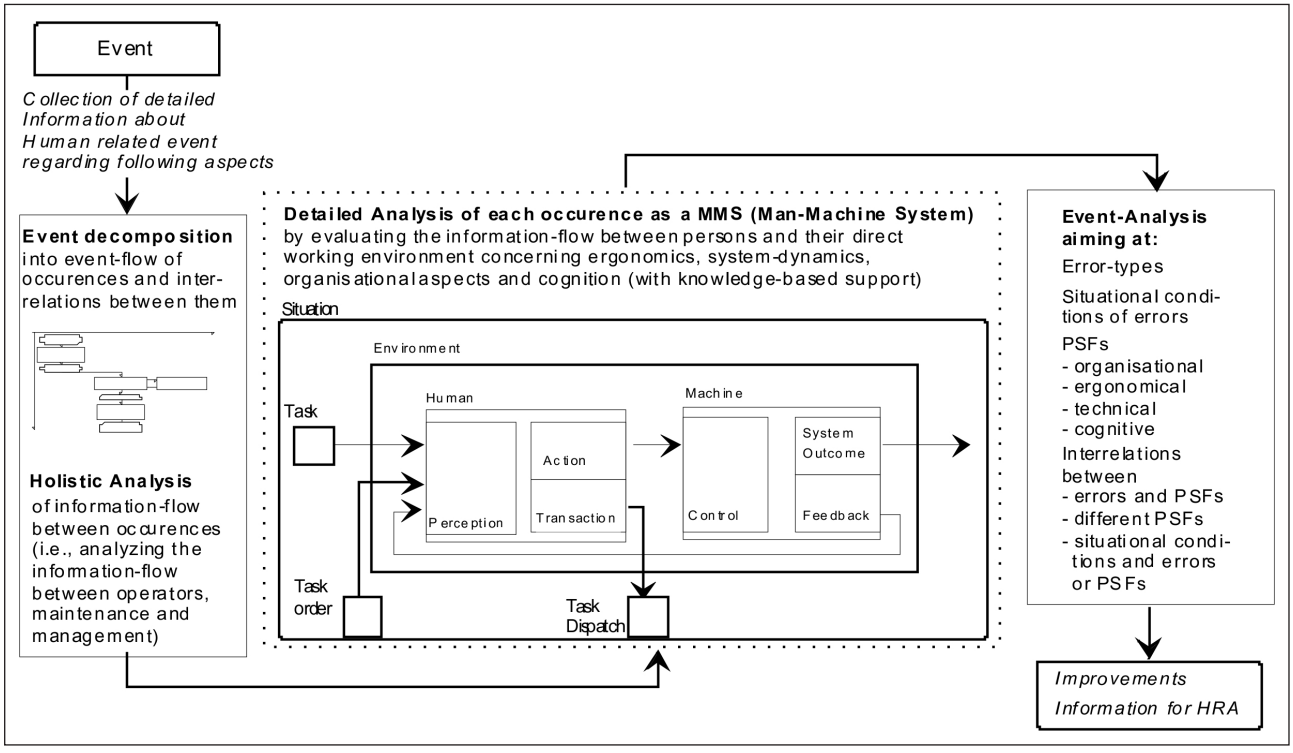

Figure 2.2 Overview of the event analysis procedure (Straeter, 2001)

various PSFs. In general, Performance Shaping Factors (PSFs) are defined here as causes or conditions necessary for the occurrence of an error. Straeter (2000) considered a weighting scheme for each PSF. Since the available data sources (i.e. databases) offered a high-level event description, it was possible to move away from a judgment based categorisation of PSFs towards a more analytical method. Straeter (2000) determined the frequencies with which a shaping factor was observed in connection to a human error of a certain type. Therefore, it enables to represent and evaluate dependencies and context on the qualitative side and suggests considering the Human Error Probability (HEP) as driven by human abilities and the difficulty of situation. Several validation studies have been performed on the approach in different industries like nuclear and aviation (Straeter, 2001).

Also there have been several ATM applications where CAHR is also known under the heading 'ATM virtual advisor' for Human Reliability (Trucco *et al.*, 2006).

Since CAHR is a data-driven HRA technique based on highly detailed databases of incident reports in the nuclear industry, it can also be used in aviation industry. Therefore using the available ATC incident reports, it is possible to define the categorisation of PSFs. However, ATC still lacks a high-level database that captures human performance in the event of an ATC related incident/accident.

2.2.2 Overview about the CAHR Method

There are three key elements to the tool (Straeter, 2000):

1) A framework for structured data collection.
2) A method for qualitative analysis.
3) A method for quantitative analysis.

2.2.3 Framework for Event Evaluation and Structured Data Collection

Figure 2.2 provides a general overview about the event analysis procedure with the MMS (Man-Machine System) as a major part of the framework. The main id ea is creating a detailed analysis of the important information flows in the event.

The event evaluation consists of four steps as follows;

1) Event decomposition

   At this stage, a complex event is broken into smaller piece called Man-Machine System (MMS) units. A key point to mention here is that it is important to have the relevant MMS units that enable us to understand the whole event, (Straeter, 2001).

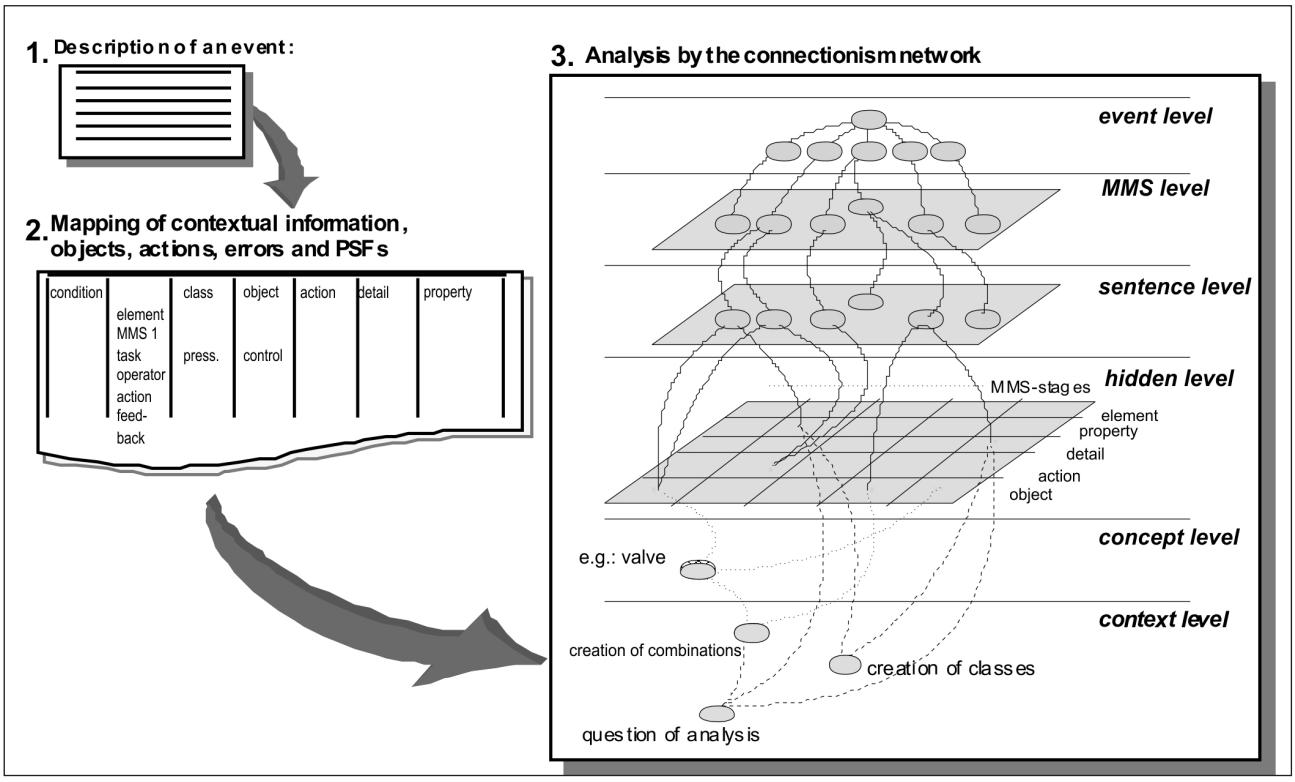

Figure 2.3 Overview of the event evaluation procedure

2) Detailed analysis of human failures

In the second stage a detailed analysis is performed for each MMS unit which is based on the MMS approach. The system outcome of an MMS defines whether an erroneous occurrence took place. The error could be the result of one or more weak points within the MMS. Here, the definition of an error is the consequence of any deficiency in a stage or in information flow of the MMS. Therefore it is important to find out the necessary information about the errors and the influencing factors for the complete analysis of the event.

3) Analysis of cognitive demands

The third stage of the event evaluation is the analysis of cognitive demands. The cognitive issues are obviously related to the human part in the MMS. It deals with the description of cognitive processes, cognitive errors and internal PSFs.

4) Description of improvement measures

The fourth stage is the description of improvement measures. This is an important step to specify clearly the measures for improvement in order to avoid the errors types that were found and to evaluate their benefit in the future.

2.2.4. Qualitative analysis

In the previous section, the description of event evaluation and date collection was explained. The evaluation method of the event description has to identify qualitative information related to the errors, the PSFs, the organization or operators. Therefore an advance evaluation algorithm based on the connectionism theory was developed. Figure 2.3 shows the overview about the event evaluation procedure.

The connectionism network represents the events data as nodes and relations within a network. It also shows the relationships among objects, actions, errors and PSFs. Since the connectionism is related to neural networks which have the ability to learn, this model is also able to learn and therefore can generate similarities between events and to also organize itself. This is important to be able to produce the required information for HRA Analysis.

The level and complexity of the detailed data requires also a highly sophisticated database model that is able to capture all of the necessary data in order to record, evaluate and assess the occurrences related to human errors or human performance. The whole system can be comprised into two major parts (Straeter, 2001);

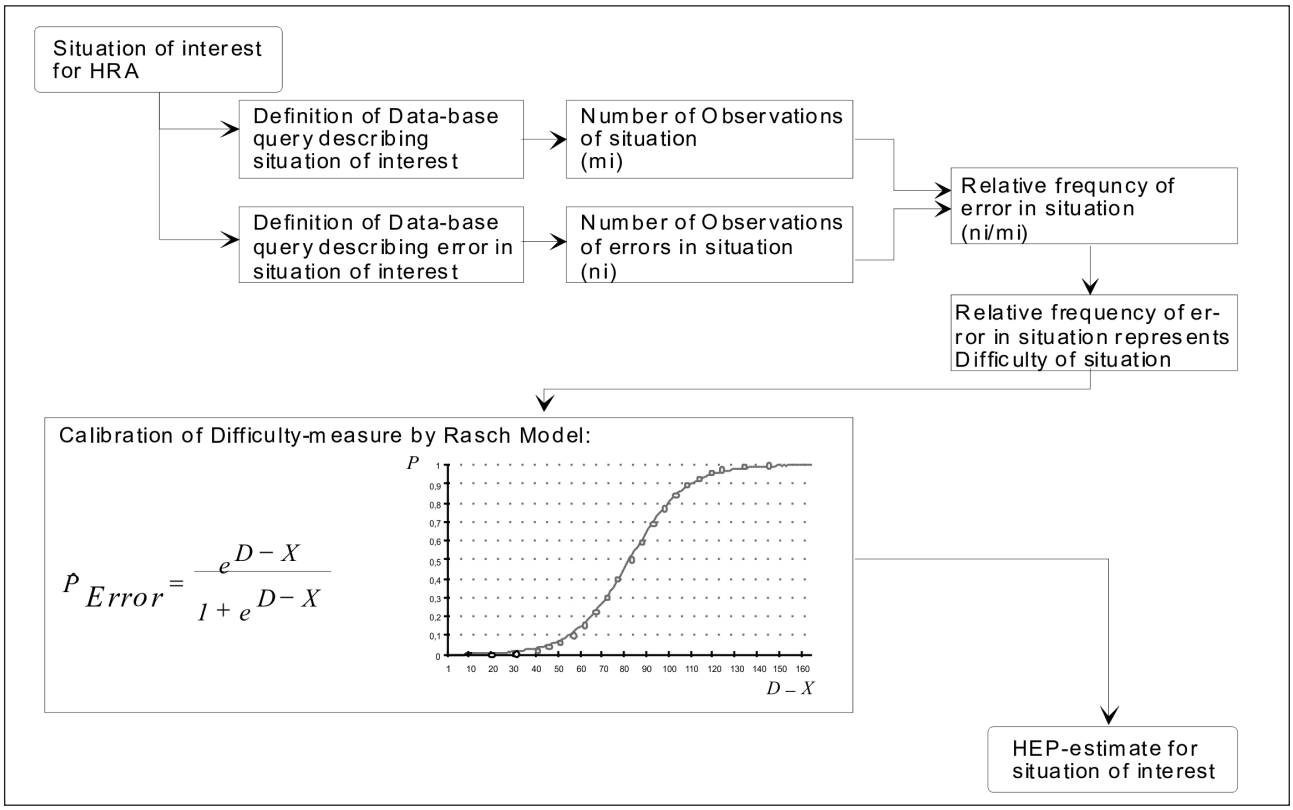

Figure 2.4 The quantification process of CAHR

1) The description of the event that has occurred.
2) The qualitative and quantitative evaluation of human reliability information.

2.2.5 Quantitative Analysis

As previously mentioned, that the CAHR technique is based on the evaluation of the operators (i.e. the ATC controllers) task from incident description and identification of interactions between various PSFs.

Therefore the collected occurrences or events will be evaluated for quantitative analysis of these similar error prone PSFs which will lead us tocalculating the frequencies of occurrence of these situations. Consequently these probabilities can be translated into human error probabilities (HEPs).

These frequencies are not directly able to estimate probabilities but they represent the difficulty an operator may have in a certain situation. Consequently, probabilities were estimated from these relative frequencies by a psychological measurement method, as will be discussed later. Figure 2.4 provides an overview of the quantification process of CAHR.

Several validations of this calibration method were conducted, in particular for nuclear industry as well as for aviation and ATC (overview in Straeter, 2004)

The calibration circumvents the issue of event data that a figure for the denominator is required. Since the incidents reports are used as the source of information, it is not possible to derive the actual number of opportunities from the events. Therefore for estimating HEPs a calibration approach is needed.

As the CAHR approach exploits also the positive performance that can be observed in any event as well (see Hollnagel *et al*., 2005), it has to possibilitiy to generate a prior estimate for the reliability by.

$$\frac{n_i}{m_i} = \frac{n_i}{n_i + o_i} \tag{1}$$

Of course, this relative number does not represent a HEP but in psychological terms this number means: if the relative number of errors is high, it seems to be a difficult situation for the operator; if the number is low, the situation seems to be easier. Hence, event analyses are able to represent the difficulty that operators or controller have with some error prone situations. In psychological measurement, Rasch(1980) assumes a simple functional relationship between difficulty and probability, which is expressed in the following equation with

(S_n) as proportionality factor (Straeter, 2000).

$$P_{\text{Failure of Type i}} = \frac{e^{(D-X)}}{1+e^{(D-X)}} \tag{2}$$

and

$$D - X = \left(\frac{n_i}{m_i} - 0{,}5 \right) * s_n \quad and \quad s_n \approx 12{,}8$$

The Rasch model performs a calibration by considering situational conditions (D) and cognitive abilities (X). For instance, if the relative frequency is lower than 0,5, the conditions (D) are such demanding that abilities (X) are not able to cope with them effectively (D-X<0). The functional relationship between conditions, abilities and probability is a probabilistic relationship (also called sometimes ogivian-relation). It states that a probability of failure of Type i can be inferred from the relative frequency of errors of Type i in the observed sample. If, in the easiest case, the relative frequency in the sample is 0,5 the probability is also P=0,5 and in this case DX equals 0. For relative frequencies smaller than 0,5 the formula postulates an overrepresentation of errors in the sample with a certain probability. For relative frequencies greater than 0,5 the formula postulates an under-representation of errors in the sample with a certain probability. (Sn) describes the amount of over- / under-representation in the sample of events. The factor was derived by calibration with THERP data.

The calibration using the Rasch model represents a simplified Baysian approach, which is also used to estimate technical reliabilities when the number of opportunities is lacking (e.g., in case of common causes). It considers that event experience is always incomplete information for generating a HEP=n/N due to the event-threshold and considers the proportion of errors in relation to positive performance (abilities) of a Human in events and uses this proportion as a prior information to correct the observed relative number. It is making a hypothesis about how the proportion will continue below the event-threshold. This hypothesis is manifested in the parameter (Sn) that is determining the ogivian-shaped calibration function as presented in Figure 2.4.

The quantification approach of CAHR has obtained a first validation in Straeter & Reer (1999). The validation showed that one big advantage of the estimation procedure is that it is based on a minimal square analysis of as much as possible anchor points

one can get. Several other validation studies have been performed in Maritime, Aviation, occupational health, automobile and ATM (Apostolakis *et al*., 2004; Trucco *et al*. 2006). This ensures that the calibration model robust (Straeter, 2001).Consequently, the whole system of CAHR will be used in order to evaluate and assess the ATCO HEP in the Air Traffic Control System.

The following section will illustrate how the two models (VSM & CAHR) can be used to assess the Air Traffic Controller reliability in ATC System.

3. General Approach of linking VSM and CAHR

The use of the CAHR and the VSM can be a very useful and strong tool in the analysis of reliability and safety issues in air traffic management systems. Figure 3.1 shows how the two models can be integrated as a framework for safety/reliability assessment in ATM/ ATC systems. It shows the general logic of how the CAHR approach could be used in the validation of the VSM model. Firstly, it shows the qualitative validation of the causal factors and hence of the suggested best mitigations. Secondly it shows that the quantitative path of the CAHR model can be used to find the most important contributions to incidents and hence the most suitable mitigations.

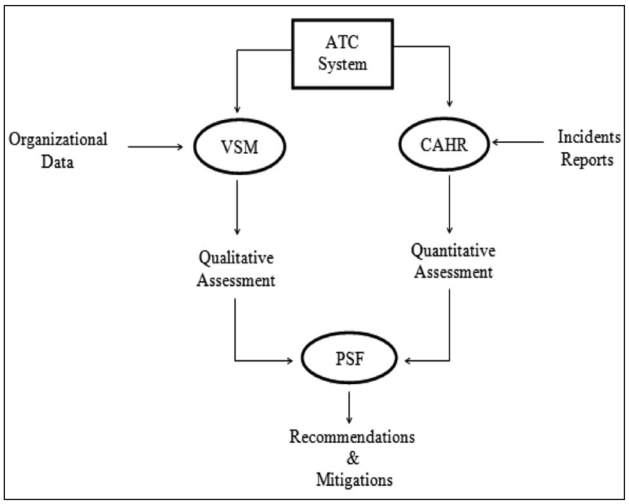

Figure 3.1 Framework of CAHR & VSM (Al-Ghamdi, 2010)

An Assessment of safety needs two branches, the engineering perspective of system modelling and the human perspective of behaviour (respectively errors). The engineering side is represented by the VSM model. The human perspective is represented by the CAHR model. Both can be dynamically linked in order to analyse the behaviour of the system either in qualitative or quantitatively terms. As Figure 3.2 outlines, the CAHR model provides data for particular links in the VSM model. Each link is represented by a

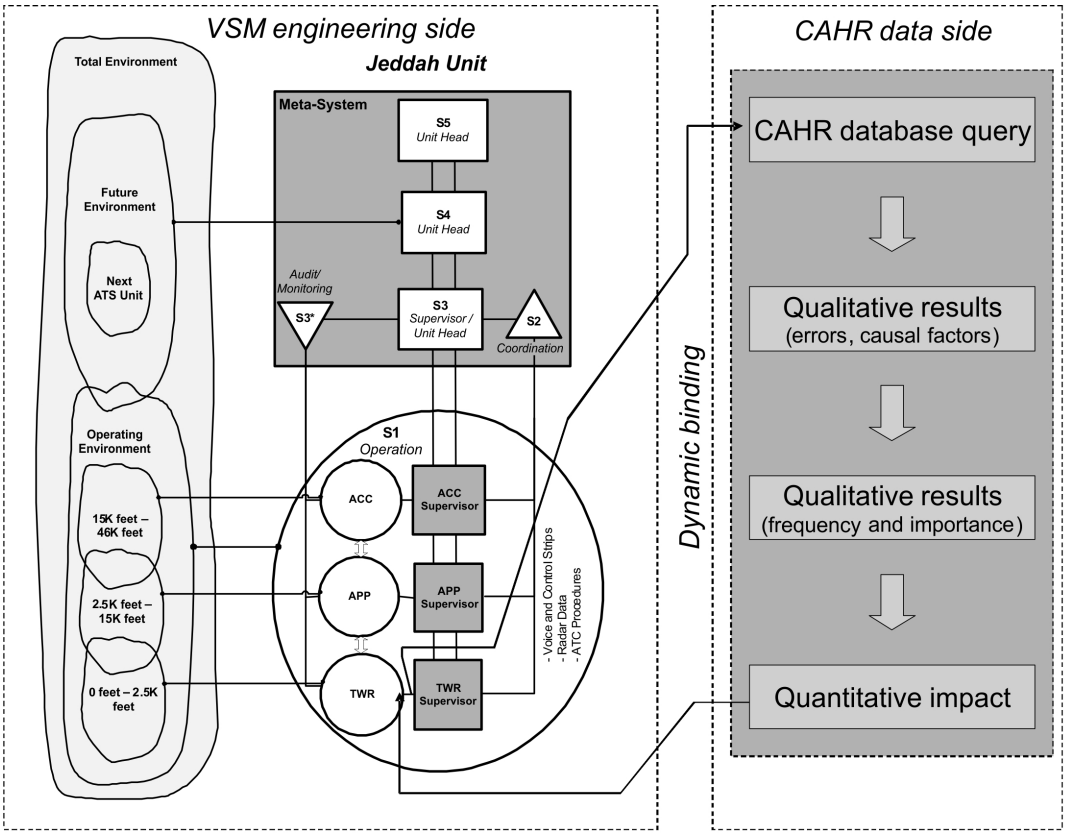

Figure 3.2Exemplified dynamic link between TWR supervisor and TWR controller

database query of CAHR and CAHR hence provides in a dynamic way different data for each branch of the VSM approach. As an example only one dynamic link is represented in the figure: the dynamic link between TWR supervisor and TWR controller. Others are e.g. the link of APP supervisor and APP controller or unit head to ACC supervisor.

In the following sections it will be demonstrated and discussed how the CAHR is applied to a set of incidents reports to reveal the relative PSFs that caused the air traffic controller to make errors. Then generic and specific cases are used to show how the CAHR can be applied to the VSM to assess ATC related tasks and produce human reliability figures.

4. Application of the CAHR method to a set of incidents

With the help of the CAHR method, 42 events related to air traffic control were analysed and evaluated regarding error types and causes for human failures. Existing incident reports were used for the application of the CAHR Method and the identification of the main factors influencing ATC Controller Performance.

The events were analysed and evaluated on the basis of the Man-Machine-System approach. An event can either be the result of a failure in a Man-Machine-System of the pilot who lost control over a plane or due to a controller who has to monitor the airspace and to coordinate different flights within this airspace. The events could have been split into 152 sub-events, which consist of separate Man-Machine-System (MMS) of either pilots or controllers (Proll, 2010).

4.1 Variance / Uncertainty in the Data

At the beginning of the analytical section, it is important to realise the uncertainties contained in the data. The number of sample events is determining the variance or uncertainty inherently contained in the data-set. The general formula for the uncertainty in the average due to the incompleteness of the data-set is determined by:

s=1/sqrt (n) (Bortz, 1989) while "n" is the number of samples.

The larger the number of events analysed, the smaller is the inherent uncertainty in the data. Any investigation has preferably a small dispersion or standard variance. The opposing nature of the function causes small standard variances if the number of samples is large, i.e., uncertainty in the set of events decreases.

In this example the number of samples used is 42. Hence, from the formula: $s=1/sqrt\ (n)$, $n=42 \rightarrow s=1/sqrt\ (42) \approx 0{,}15$ follows a standard variance of approximately 0,154. A standard variance of ≤0,15 is suitable to assume that the results represent a real difference.

As an example, there were 10 errors of a type A observed and 7 errors of type B. the deviation between both is 3. According to the above formula, the inherent variation in the frequency is ≈0,15 hence the lower bound for type A errors is ≈8,5 (10 – 10*0,15) and the upper bound for Type B errors ≈7,81 (7 + 7* 0,15). As the difference between 8,5 and 7,81 is still greater 0, it can be concluded that the difference is a valid difference that will be evident also in a larger set.

In the following only those difference will be considered that meet the condition in such a way that the difference is also valid taking into account the inherent uncertainty in the data of ≈0,15

4.2 Evaluation of Errors

In the beginning of this analysis, common information on the errors that occurred was analysed on the basis of CAHR. At first, events are investigated in respect to absolute and relative terms and by comparing the errors of pilots and controllers.

Note that the term error in this investigation is understood as an error within the MMS, i.e. the context a human is working in. Hence the term error has not whatsoever connotation with the human as the one being responsible or even guilty for the error. The term is rather more reflecting the inappropriateness of the contextual conditions comprised in the MMS to perform adequately.

4.2.1 Distribution of errors

In order to make more precise statements about the relative allocation on controllers and pilots, the relative error frequency need to be determined and compared. The controllers' and pilots' relative amount of errors is represented. It is the average per person involved in the 42 events. The results in the diagram are calculated by dividing the total of the errors represented in the figure above by the number of controllers involved. The same can also be applied to the pilots.

The formula is: n(rel,c)=n(err,c) / n(c)

With:

n(rel,c) → relative amount of errors per controller

n(err,c) → total amount of errors per controller

n(c) → amount of controllers involved

The calculation for the pilot's errors is analogous. The variable "c" (controller) of the formula is then replaced by a "p" (pilot).

The calculation described above is used in each of the relative considerations during the course of this work. The results are shown in Figure 4.1.

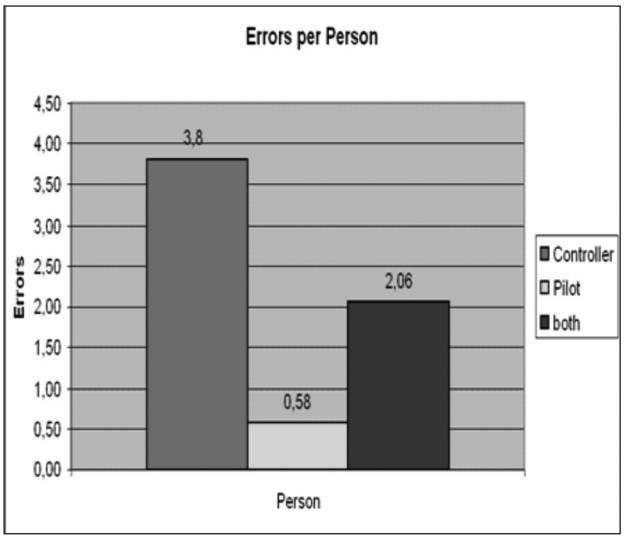

Figure 4.1 Errors per Person

The result of the calculations represented in the diagram is that a controller has an average amount of errors of 3,81 and a pilot an average amount of 0,58. Therefore the identified controller's error rate is as 6,5 times as higher than the one of the pilots. The quotient of the amount of errors found n(er,c+p) and the amount of the Man-Machine-Systems involved n(c+p) results in an average of n(rel,c+p)=2,07 errors per Man-Machine-System (pilots or controllers may be involved).

The calculation proves that Human error identification is not equally distributed on both sides. The majority of errors were identified on the controllers' side; pilots seem to be less focused in the event reporting.

4.2.2 Definition of different types of errors

A more detailed analysis of the events focuses on the different types of errors. Herewith it is possible to determine the relation between the error type and the task. Thus the connection between task, action, error and cause becomes obvious.

In this section the essential types of errors are defined to secure the comprehension of the subsequent chapters. Interpretation and informational content of the following diagrams and figures can be facilitated. The following descriptions are not related to controllers or pilots but intended to be universal.

The error type's taxonomy is a result of the 42 flight events that were investigated with CAHR. The taxonomy was found during the analysis and compilation of the event descriptions into the structure used by the CAHR methodology.

Definition of Error Types:

- Omit - *Omit* means that a task was neglected. The usual cause may be that workload is too high for the person responsible.
- Failed - The intended objective of a task is not attained. Thus, this is the description of failure through wrong execution of tasks.
- Not allowed - In this case actions were conducted against existing rules. The error may occur consciously but also unconsciously, provided that the violation of job execution was not intended.
- Not enough - While executing a task, the actions were not performed with appropriate emphasis.
- Too late - Errors are made by running a task later as it should have been executed.
- Wrong - The task performed was wrong; a different task or action should have been performed.
- Incorrect - When the error type *incorrect* ap pears, the execution of a task was wrong.
- Malfunctioned - The error type *malfunctioned* exclusively occurs in relation to computer systems.
- No response/reply - The confirmation of an order or message between controller and pilot remains absent.
- Not received - Technical failures in communication during the placing of an order from controller to pilot may occur. For instance radio messages are not transferred.
- Too much - Errors may also occur when a task is repeated too often. An error of this category may be described by a controller that gives information too often or several times, leading in turn to a consecutive pilot error for instance.

The following error types are subsumed in the figure under the category "all others"

- Not completely - An error may be described as not completely when a task is executed incompletely.
- Not noticed - The error type *not noticed* describes deficits in attention. e.g. the task of a pilot was not executed as he was not attentively.
- Too early - Errors are made by running a task earlier as it should have been executed.
- Very bad - The error *very bad* only occurs in situations depending on the environment. This may be influenced by causal factors such as *bad weather/ rain / poor visibility*.

4.2.3 Types of errors of controllers

Taking into consideration the frequency of controllers' errors as described above, this section gives an investigation of the types of errors. The taxonomy of the database (CAHR) allows to analyse the errors more precisely and to structure the errors.

Figure 4.2 provides the results. The main priorities, shown in the diagram, are of the categories *omit* with a frequency of 94 and *failed* with a frequency of 83, in other words 177 of 262 errors may be referred to these two categories.

A third priority is summarised to one group, consisting of the error types *not allowed, too late, not enough and incorrect*. Their absolute frequency spans from 14 to 18. Their total of 53 errors constitutes an overall contribution of about one-fifth of all errors identified.

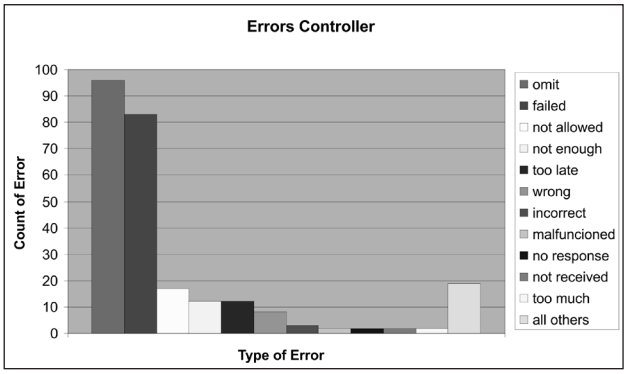

Figure 4.2 Errors of Controllers

There are further error types such as *bad, too much, not received* etc. in the range of frequency, equal or less than three. Their occurrence of 33 from 262 added together only counts about one-eighth. Due to their rareness and the uncertainty in the data-set, these errors have little or no relevance and are therefore neglected in the further analysis.

4.2.3.1 Error Investigation of Different Controller Types

The incidents allowed distinguishing Air-traffic controllers in the 'Area Control Centre' (ACC), Approach Controller (APP), Tower Controllers (TWR) and Supervisors (SUP). Controllers in the ACC (ACC Controller) have to monitor the major part of controlled airspace and the 'en route'flight-phase. Controllers of the 'Approach Control Center' (APP Controller) observe the airspace around the related airport. Tower controllers (TWR) are observing the runway as well as the surveillance zone around the airport and grant takeoff and landing clearances. The supervisor controllers (SUP) are in a monitoring position to all of the other controllers mentioned. Supervisors are also responsible for organisational issues that may include management tasks. In addition to that, they can support the other controllers in case of high workload. The human reliability of controllers is investigated in the following sections.

4.2.3.2 Absolute distribution of errors

The first step is to analyse the four types of controllers with respect to their specific errors. The Figure 4.3 shows the absolute frequency of errors per controller.

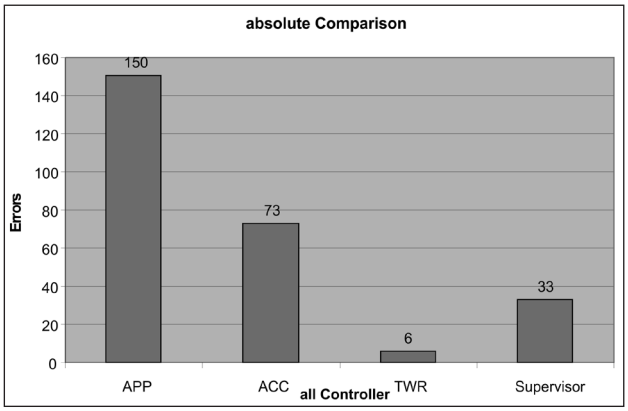

Figure 4.3 Comparison of the absolute number of errors of Controllers

It is exemplified that APP controllers, with an amount of 150 errors, have an error frequency that is nearly twice as high as the errors of Tower and ACC controllers taken together. Consequently the main issue of human reliability seems to be in the APPs' field of competence. Nevertheless, a frequency of 73 errors is in the responsibility of ACC controllers and demonstrates a lack of human reliability in

this category as well. Management errors with an abundance of 33 represent a further emphasis.

4.2.3.3 Relative distribution of errors

In order to make more detailed statements about the controllers' error distribution, the figures detected are summarized into a relative result. This result describes the averagenumber of errors per controller, shown in Figure 4.4.

The figure shows that ACC controllers have the highest error rate in relative terms in opposition to the findings in Figure 4.3. The figures are representing the relative number of errors per event of ACC, APP and TWR controllers. The relative number represents the importance of the errors in relation to the number of events. Through the relative consideration of errors, the ACC controllers now show a higher rate of errors with 5,21 per controller (73 encountered errors in 14 events related to ACC controllers), whereas the APP controllers have a lower rate with 4,55 errors (150 encountered errors in 33 events). Supervisor controllers follow with a rate of 2,36 errors (33 encountered errors in 14 events). The lowest rate is represented by the tower controllers with only 0,86 per controller (6 encountered errors in 7 events). The average error rate of controllers is 3,81 for all controller types.

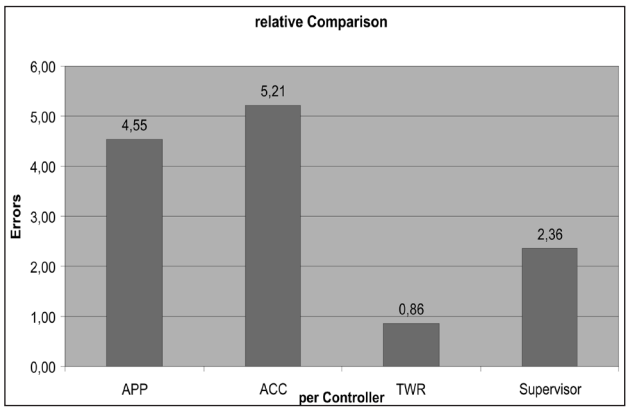

Figure 4.4 Comparison of the relative number errors of Controllers

4.3 Investigation of Causal Factors (PSF)

4.3.1 Causal factors identified

Subsequently to the evaluation of the flight events it is appropriate to regard the sources of the errors, the causal factors or Performance Shaping Factors (PSF) - and their importance

The causal factors can be derived either from the pilot's Man-Machine-System or the controller's Man-Machine-System. Causal factors are linked to the whole MMS (Man-Machine-System) and do not only focus on the pilots' or controllers' behaviour.

In total there were 126 causes identified. The result is shown in Figure 4.5. The factors are defined as follows:

- **Workload** - A significant and frequently source that leads to human error in reliability was *highworkload*. Consequently it can be stated, that the workload under multitasking demands, is too high to remain reliable, meaning one of his tasks may be neglected. Errors of omission may be the result.
- **Attention and Communication** - Further frequent sources of errors derive from the field of attention. Causal is a deficit in the *attention* of a pilot or a controller, which leads to an activity that is executed too early, too late, not exact or omitted completely. These results may also occur when communication is incorrect or inoperative. Reasons for an error in communication may be of technical origin, as defect auditory transmission systems (radio equipment), disturbance of transmitting systems or wrong phraseology used by pilots or controllers.
- **Organisation, Instruction and Job Responsibility** – *Organisation,instruction* and *job responsibility* are mainly caused by errors stemming from incoherencies between responsibility and task-control. *Job responsibility* as a source is often linked to failures in management of the supervisor controller, whereas the source o*rganisation* is mostly due to mistakes in flight planning (e.g. the flight plan in message correction due for flight data specialist). *Instruction comprises situations where instructions given by controllers misfit with aircraft capabilities or current flight operations (e.g., poor vectoring).*
- **Design** - In most cases, problems of the category *design* concern insufficient operation of visual and auditory warning systems. But they can also be due to inappropriate arrangement of control elements, which may influence the Man-Machine-System negatively.
- **Knowledg**e - An error of this category often results from a lack of expertise of the person that is responsible.
- **Experience** - As trainees or beginners do not have much experience in their field of responsibility, they tend to do more mistakes than their

colleagues, having more professional experience. Over-confidence may also have an impact on errors.

- **Procedure Adherence** - In fixed workflows there should be adherence to standard procedures that are determined. In case they are neglected, an error may occur due to the source mentioned above. An example for this is the handing over of an airplane, right in a conflict or difficulty, from one controller to another.
- **Functionality** - *Functionality* often concerns technical problems controllers and pilots have to deal with. Often the machine or system is inoperative, such as the breakdown of the flight recorder or the TCAS System, for instance.
- **Misunderstanding** - *Misunderstandings* occur when communication fails between the persons involved. A problem may be a misunderstanding concerning the level of flight, between pilot and controller. As a result the pilotexecutes a wrong flight level.
- **Training** - *Training on the job* aims to improve the working skills of a person as well as his sense of responsibility. Besides, the controller is trained how to react in critical situations and is therefore professional trained. But training should not be used as a universal solution. To resolve a certain problem its reason has to be investigated properly. Otherwise there is the risk of missing the objective.
- **Violation of Tasks** - *Violation* in this context means a deliberate or knowingly action though the person knows that the action is wrong. Violation may therefore entail suspension.
- **Judgement** - *Wrong judgement* often induces wrong behaviour. A controller may for instance fail to assess the speed of two airplanes one behind the other, which may result in wake vortex issue.
- **HMI-Human Machine Interaction** - The object of this source is the Man-Machine-System, thus the interaction between the human being and the machine he is handling with. Disturbance in Human-Machine-Interaction may occur due to bad visual warnings in a conflict or turbulence. Consequently the Man-Machine-System does not work properly.
- **Complexity** - Problems of this category often concern a certain task that may become too complicated for one person. This is, for instance, when a controller has to supervise a large sector with a lot of airplanes he has to coordinate.
- **Automation** - *Automation* concerns the centralisation of recurring workflows in the transmission process from human hands to a machine. In other words, it is difficult to transfer the complexity of human action to a machine.
- **Ambiguity** - *Ambiguity* may result from insufficient information concerning a workflow. The information given can be interpreted in different ways. Therefore the person involved has no precise instructions of how to manage a task.
- **Airspace Design** - This issue concerns the allocation of flight routes within the airspace. Problems may occur due to bidirectional routes, e.g. there is one route with air traffic in two directions.
- **Night Time** - The visibility due to darkness complicates the work of pilot and controller, particularly for tower controllers.
- **Fixation** - Pilots or controllers are executing tasks too late or neglect them completely as he/ she was focused on another task.
- **Distraction -** A problem can derive from a distraction when a controller or pilot is noticing a disturbance too late. A distraction is especially serious in a critical situation.

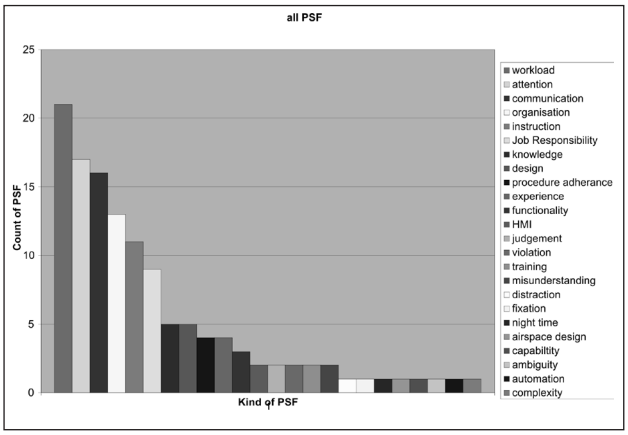

Figure 4.5 Causal factors identified

4.3.2 Importance of causal factors for specific working environments

The list of PSF does not reflect yet, which PSF is important for a specific workplace. In order to determine this, the absolute frequency of occurrences need to be contrasted against the relative frequency of how many PFS are related to the specific workplace. The specific workplace might for instance be the pilot

workplace, the controller or supervisor. This specific workplace need to be part of the database query to the CAHR database (e.g., query of the causal factors leading to errors of controllers).

The importance of causal factors depends on their absolute and relative frequency. The relationbetween these is illustrated in an X, Y – diagram. Overall such a diagram is representing the importance of the causal factors. The diagram contains four differently coloured areas that categorize the priority of the error source. The description of the four areas is as follows:

- **Red Area** - It is the most important area within the diagram as it includes those causal factors which are high in absolute and relative frequencies. This implies that the causal factors occur very often in the investigated events and have a high relative portion for the specific working environments. Causal factors in this category should be investigated intensively. The red area is deemed to be critical; the priority to mitigate these is high.
- **Blue Area** - The blue area shows causal factors where the absolute frequency is high whereas the relative frequency is low. This implies that the causal factors occur very often in the investigated events, but they have more importance in other contexts (working environments). They should be assigned having second high priority.
- **Yellow Area** - The yellow area shows the sources in which the frequency is high in relative terms but low in the absolute ones. They are mostly concerning exceptions that do not occur very often or sporadically. The sources in the yellow area are of middle priority for the specific working environment.
- **Green Area** - The green area contains error sources that do have a low frequency, relatively and absolutely. Therefore they are of low priority and may be regarded as uncritical for the specific working environment.

It can be stated that the nearer a source is to the edge of the red area, the higher the importance is. This is illustrated in Figure 4.6 exemplified by a hypothetical allocation of the causal factor "workload" for four specified working conditions.

In the diagram on the left side of the bottom "workload" is situated in the green area with a relative frequency of 20% and an absolute frequency in four of the incidents. In this context "workload" can be described as uncritically. This might be the case if there

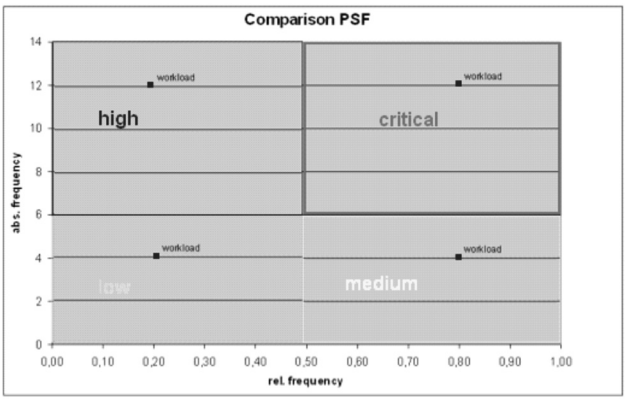

Figure 4.6 Pattern of Priority (Workload)

are 20 events with workload issues and only 4 (=20% of 20) are related to the specific working condition (e.g., the working environment of the pilot).

In case the relative frequency arises to 80% "workload" must be placed into the yellow area on the right side of the bottom. It is now of middle priority. This might be the case if there are 20 events with workload issues and 16 (=80% of 20) are related to the specific working condition (e.g., the working environment of the TWR controller).

Another possibility is that the relative frequency of workload stays the same (20%) but its absolute frequency is now 12. As a consequence "workload" shifts to the blue area on the left side of the top. The priority of the causal factor "workload" becomes high. This might be the case if there are 60 events with workload issues and 12 (=20% of 60) are related to the specific working condition (e.g., the working environment of the APP controller)

A causal factor may be described as critical if its absolute and relative frequency become high. In our example "workload" will shift to the red area which is situated on the right side of the bottom. This might be the case if there are 60 events with workload issues and 48 (=80% of 60) are related to the specific working condition (e.g., the working environment of the ACC controller).

Contrasting the relative and absolute numbers allows to gain more detailed insights into those human-factor aspects, which need urgent attention in mitigationsand hence specific suggestions for improvement may be developed.

The boundaries between the areas are not fixed but can be determined according to the specific results of the CAHR database query. Usually the boundaries are set by 50% between minimum and maximum of

the absolute frequency as well as 50% between the minimum and maximum of the relative frequency. If there are clusters of PSF, the boundaries might be shifted slightly in order to have maximum fit of the clusters and boundaries.

4.3.3 Causal Factors related to Controller

4.3.3.1 Overview of Causal Factors related to all Types of Controller

This section shows a comparison of the controller's causal factors by means of the scheme explained above. The following diagram shows the causal factors and their priorities. It is related to all the controller types "APP, ACC, TWR and Supervisor Controller". Thus, the diagram represents their common causal factors. A comparison between the single types of controllers will be provided later on.

Figure 4.7 provides a detailed overview on the main causal factors on controllers. The category "workload" is exposed in the upper right which implies that having multiple tasks simultaneously or in short intervals, inevitably leads to most errors of controllers. Often, errors of the category "omit" are related to workload. Another consequence is that tasks

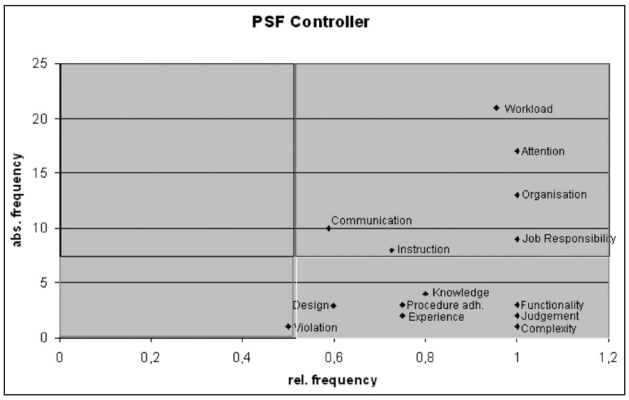

Figure 4.7 PSF Controller

are executed poorly or are even neglected because of high workload.

The source "Attention" is also of high priority as it may occur as a result of intensive working hours in which high concentration is obligatory. False assignment of tasks leading to a conflict or the confusion of airplanes´ call signs may be the consequence.

4.3.3.2 Importance Profile of different Controllers

As it was mentioned before, a comparison of the four types is also of interest for a more specific

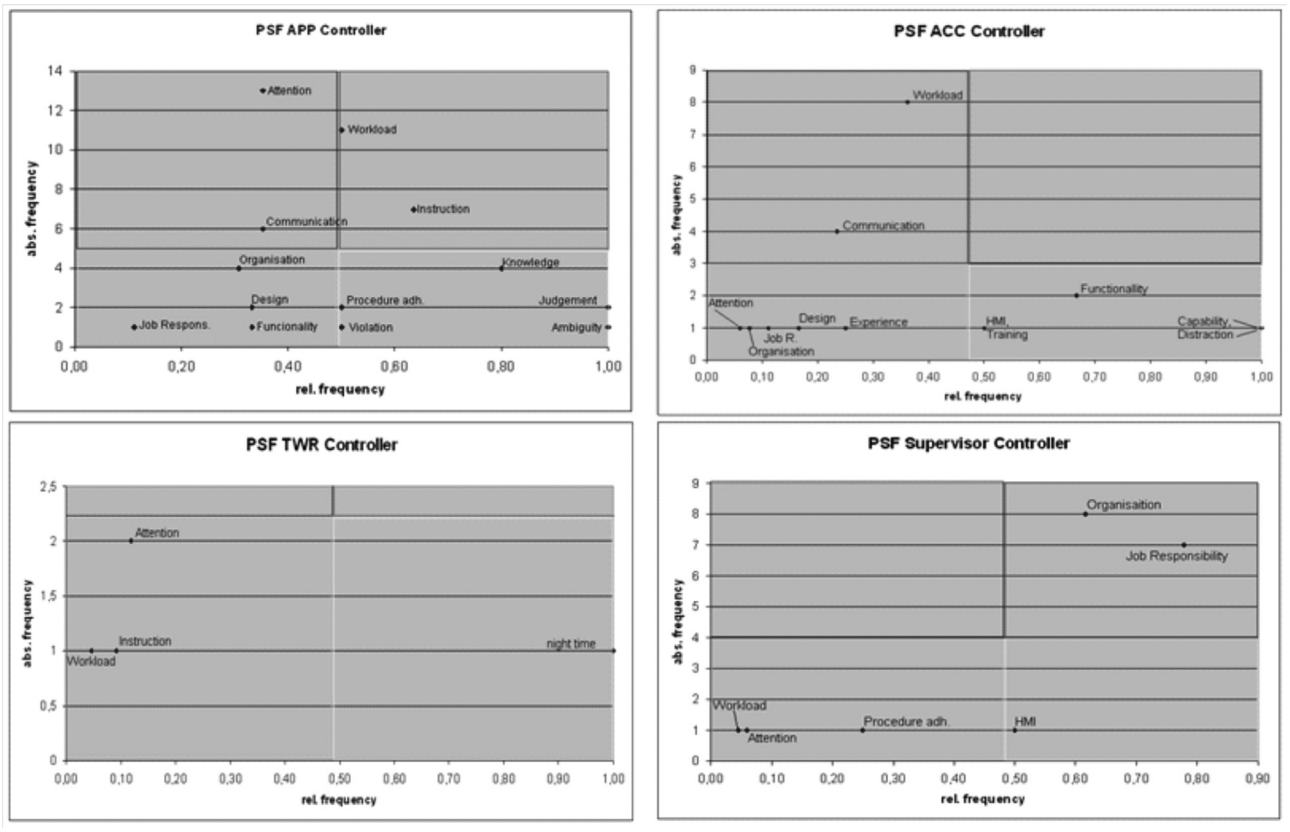

Figure 4.8 Comparison of Controllers Causal Factors

investigation of the best mitigations according to the different controllers' types. Thus, a more specific diagram is needed in which the controller types APP, ACC, TWR and SUP can be compared. The procedure is similar to the one above, just the database query is now made more specific for the different controller types (e.g., query of the causal factors leading to errors of ACC controllers). The diagram separates the causal factors per type of controller of the four mentioned types. The priorities of the causal factors can be derived for each of the types. This enables a differentiated view on the causal factors and in some cases leads to surprising results. The presentation in Figure 4.8 shows how different frequencies between the causal factors and the controller type are related.

The APP controllers diagram shows an emphasis on the causal factors "attention, workload, instruction and knowledge", whereas the one of the ACC controllers shows main errors in "communication" and "workload". The tower controller's main errors were caused by "attention" and the determining factor "night time". The supervisor controllers' main errors were caused by "job responsibilities" and "organisation".

Concerning the APP controller's priority of causal factors it can be said, that "workload" and "attention" have a higher absolute frequency than "instruction" and "knowledge", but as absolute and relative frequency correlate in the representation, there is something like a compensation, which leads to similar priorities of the four causal factors.

In contrast to that, the frequencies of "communication" and "organisation" are relatively low. They may therefore be considered to be of middle priority in the course of the problem-solution-process. Against all expectations, the priorities of "violation, procedure adherence and design" were comparatively low. Regarding the ACC controller's diagram, it is striking that none of the causal factors appears within the critical red area. Though, problems in the field of "workload" and "communication" should not be neglected as their absolute frequency is high, just as the importance of "functionality". The TWR controller's diagram shows very few causal factors. The problems that occur can be assigned to "night time" and "attention". The frequency of "workload" and "instruction" is rather low as the airspace tower controllers have to supervise is smaller and less complicated than the ones of ACC and APP controllers. As it was repeatedly mentioned the causal factors of the supervisor controller originate from the field of process organisation. "Job responsibility" and "organisation" are almost synonymous concerning their frequency and solely responsible for their errors. "Workload, attention, procedure adherence and HMI" have an absolute frequency of 1 which is not representative.

4.3.3.3Conclusion on the Importance, Assessment and Mitigation

In conclusion, each type of controller has specific conditions and causal factors, depending on the respective workspace and responsibility. Solutions to the problems that were evaluated need to consider the fact that these causal factors and conditions are different and therefore need different mitigation strategies. Often also these differences are not considered in Human Reliability assessments and for instance the issue of workload is assumed as relevant for all types of controllers. The evaluation revealed that the PSFs need to get different weights according to the type of controller. Overall the importance measures show that an investigation of causal factors and a resolution or mitigation can be made on a rational basis by analyzing the problems, providing an accord evaluation of importance and by then suggesting a specific assessment and reduction.

4.3.4 Causal Factors related the Error Types "Omission" and "Commission"

In the course of this work so far there was either an analysis of error types or their causal factors. CAHR also provides the opportunity to investigate the interdependence of errors and causal factors. In this section it is aimed to regard the dependence of errors on their causal factors. An important distinction in safety is the one between "errors of omission" and "errors of commission". While the errors of omission relate to a so called task-oriented assessment of human behaviour, errors of commission relate to a so called goal-oriented assessment of human behaviour. While the error of omission affects negligence of a task, errors of commission are related to errors in "organisation". As there are more types of errors it is necessary to separate the errors into two groups. The CAHR database allows determining groups by allocating specific information to classes that are then usable in database queries. Table 4.1 outlines the classes defined to distinguish between the primary groups of "errors of omission" and "errors of commission".

Table 4.1Errors of Omission/Commission

Omission	Commission
omit	incomplete
failed	too early
forgotten	too late
garbled transmission	too many
impossible	too much
not	too old
not heared	incorrect
not available	limited
wrong	not all time
not noticed	not allowed
not possible	not at all
without success	not completely
	not enough
	partial
	not understood
	very bad
	very weak

Figure 4.9a and 4.9b show the diagram reflecting errors of omission and commission in absolute and relative terms. The scheme is derived with the same approach as in the previous sections. The database

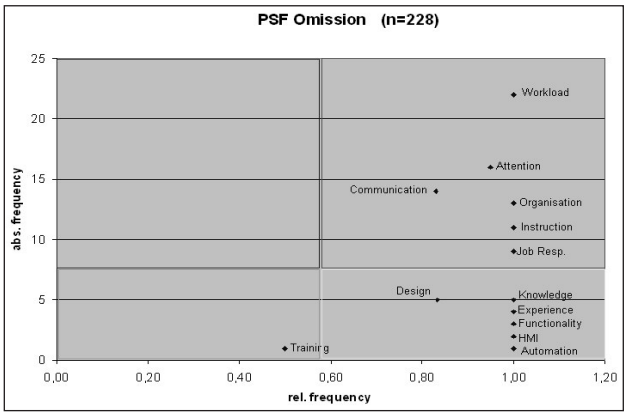

Figure 4.9a Errors of Omission

calculated the importance diagrams using a query of all events where an error of omission (respectively commission) was leading to any controller error. As in the previous importance diagrams, the red area shows the causal factors which have a high relative frequency and are of high importance.

It is obvious that the most important causal factors for omissions are *workload*, *attention*, *communication* but also *organisation*, *instruction* and *job responsibility*".

Table 4.2 Unconditional probability of errors for the different types of controllers

Controller type	Ni	Mi	Justification for Mi	Sn	D-X	P
APP	28	152	Events Represents	12.85	-4.0564	1.70E-02
	Main Causal Factors	Organizational	Both	Personal		
	Ambiguity	X				
	Attention			X		
	Communication	X				
	Complexity	X				
	Design	X				
	Experience		X			
	Fixation			X		
	Functionality	X				
	Instruction	X				
	Job Responsibility	X				
	Judgment			X		
	Knowledge		X			
	Organization	X				
	Procedure adherence		X			
	Violation		X			
	Workload		X			
ACC	13	152	Events Represents	12.85	-5.3240	4.84E-03

	Main Causal Factors	Organizational	Both	Personal		
	Attention					
	Capability	X				
	Communication	X				
	Design	X				
	Distraction			X		
	Experience		X			
	Functionality	X				
	HMI	X				
	Job Responsibility	X				
	Misunderstanding			X		
	Organization	X				
	Training		X			
	Workload		X			
TWR	3	152	Events Represents	12.85	-6.7691	2.09E-03
	Main Causal Factors	Organizational	Both	Personal		
	Attention			X		
	Instruction	X				
	Night Time	X				
	Workload		X			

The most important causal factors for commissions are *workload*, *attention*, and *communication* but also *instruction*. Thus the difference between omissions and commissions is not that much in the commonly appearing factors *workload, attention* and *communication*. The difference is more in *organisation*, *instruction* and *job responsibility* (on the side of the omissions) and *instruction* for commissions. The differing factors *communication,instruction* and *organisation* are the main distinctive causal factors for error of commission; all these factors are related to processes in the organisation.

Other causal factors distinguishing omissions and commissions are *HMI*, *experience* and *automation* as being relatively more important for omissions and *procedure adherence*, *violation* and *distraction* as being more important for commissions. However, those causal factors are of less importance due to their low absolute frequency. The results of the CAHR Database queries are presented in the following tables.

Table 4.2 provides the results of the unconditional probability of errors for the different types of controllers. It also shows the main causal factors and whether it is due to organizational factors, personal factors or both.

It can be observed that the APP controllers have a probability of P=1.70E-02 compared to those of ACC and TWR of (P=4.84E-03vs P=2.09E-03) respectively.

This means that the APP controllers have the lowest reliability figure whereas the figures are almost equal for the ACC and TWR controllers. This indicates the criticality of addressing the causal factors of the APP controllers and the urgent need to mitigate these issues that are causing their performance degradation.

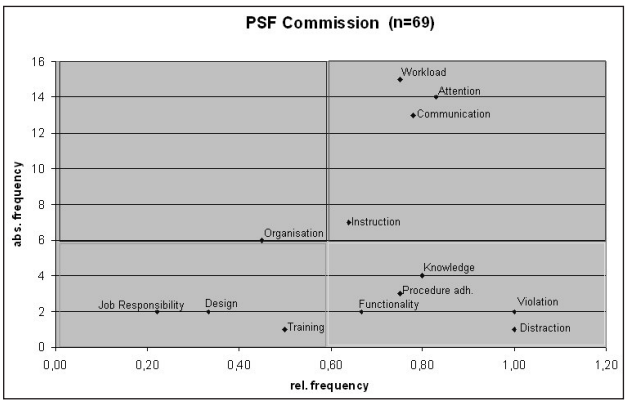

Figure 4.9b Errors of Commission

5. Assessment of specific ATM related tasks

5.1 General Approach to Assess ATM related Tasks

The database may be described as an appropriate instrument to analyse the errors of pilots and particularly of the controllers and to consider these errors in relation to their causal factors. Besides the possibility to discuss the importance of causal factors for specific problems and to find appropriate approaches to resolve them subsequently, the database CAHR also allows for a quantitative levelling of different importance of errors or causal factors. This quantitative levelling allows for using the data also in quantitative risk assessments.

The general approach to assess ATM related tasks can be accomplished by linking the CAHR results to a system engineering approach of the VSM approach. Figure 5.1 outlines this approach. In principle, the CAHR database provides the assessment for each link in the VSM approach. Figure 5.1 provides as an example the link between ATCO 5 and Sector 5, which is represented by a dynamic link using the CAHR data model. This dynamic link represents the reliability of the behaviour of ATCO 5 in Sector 5.

5.2 Generic Human Reliability Figures

In relation to figure 5.1, the following generic links will be provided with data using the CAHR quantification approach:

P (ACC supervisor -> ACC controller -> operating environment) - This quantitative line consist of the two aspects P (ACC supervisor -> ACC controller) and P (ACC controller -> operating environment)

P (APP supervisor -> APP controller -> operating environment) - This quantitative line consist of the two aspects P (APP supervisor -> APP controller) and P (APP controller -> operating environment)

P (TWR supervisor -> TWR controller -> operating environment) - This quantitative line consist of the two aspects P (TWR supervisor -> TWR controller) and P (TWR controller -> operating environment)

The results of the CAHR Database queries are presented in the following tables. Table 5.1 provides the results of the unconditional probability of errors for the different types of controllers. It also shows the main causal factors and whether it is due to organizational factors, personal factors or both. The Table provides the generic probabilities for

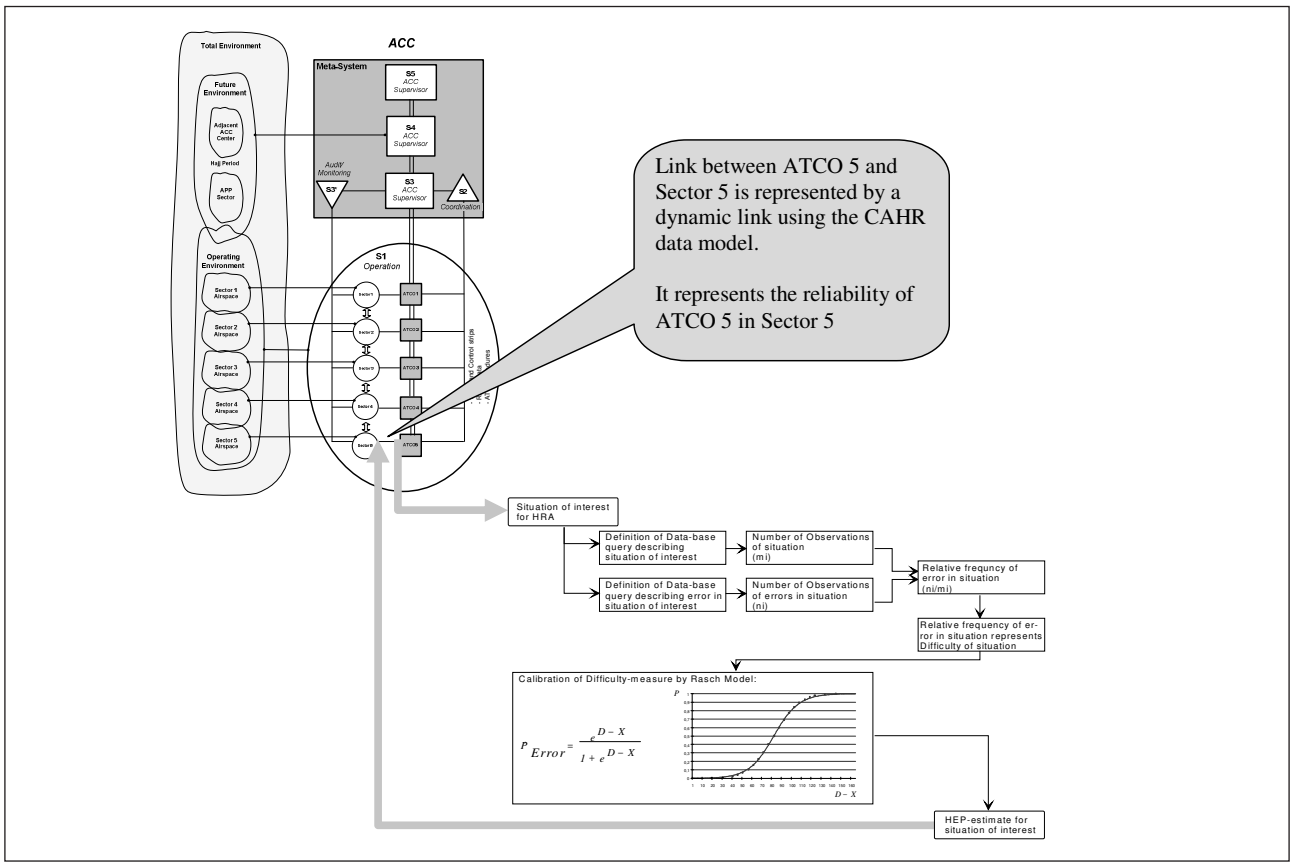

Figure 5.1 Using the CAHR results for assessing the links in the VSM model

controllers committing an error and the conditional probability that supervisor's behaviour results into controller errors. The input parameter of the estimated probabilities are given in the left side columns, where (ni) means the number of sub-events where an error occurred for the task described, For instance for the item 'supervisors' behaviour result into APP errors' 11 sub-events were identified in which an error in a supervisors' behaviour resulted into an error of an APP controller. The parameter mi then describes the number of all sub-events in total where supervisor-APP relations were observed and represented in the event description (31 sub-events). Note: (Sn) is a fixed parameter as encountered in the development of the approach (see Straeter, 2000) and D-X is the resulting value according to the formula above. Probabilities are calculated with the formula as provided in the utmost right column of the table. The parameter mi changes according to the completeness of the database with respect to the specific cases.

As an example for generic probabilities of controllers, item one in this section of the table provides the probability of P= 1,70E-02 that an error of an APP controller commit an error in his task execution. The parameter mi includes all sub-events 152 as in any events APP controllers were involved positively or negatively in all sub-events. On the other hand for the conditional probability of supervisor to controllers item one in this section of the table provides the probability of P= 1,34E-01 that an error of a supervisor results into a follow-up error of an APP controller. In other words, an APP controller recovers (corrects) erroneous behaviour of a supervisor in 86% of all cases (0,866 = 1-1,34E-01). In this example only 31 sub-events were observed with the desired combination (supervisors and APP). The estimate is only based on 31 sub-events and has therefore some uncertainties. With higher number of events this uncertainty might be reduced according to the formula *1/sqrt(mi)* as described above as well.

The uncertainty of the database need to be considered if one uses the probabilities in safety assessments. In the example given above the uncertainty amounts 17,9%.

However, the example also shows that the approach gives insights into the potential of complex conditional error probabilities such as impact of supervisors on controllers. The overall calculation as provided in the bottom of the table shows the resulting probability of a specific path in the VSM model.

As an example the first item of the table shows the probability the APP supervisor's behaviour leading to a follow-up error of a controller, which has consequences for the operating environment. The probability amounts P=1,93E-03. This means that the approach would estimate that in roughly 2 out of 1000 cases a supervisor's instruction to a controller will result into consequences on the working environment.

The most important influences of supervisors identified lie in the approach area (ni/mi = 11/31), which results into an APP overall probability of P=1,93E-03. Compared to this the probabilities of ACC and TWR are (P=1,48E-04 vs P=3.38E-06) respectively. These figures can directly be implemented into the VSM model and generally confirm the conclusions of the VSM model as repeated in Table 5.2.

The following section provides quantitative figures for the Human reliability in generic tasks and specific tasks according to the VSM approach. In principle, the CAHR database could provide such validating assessments for the more detailed conclusions of the VSM approach as well. However, given only the set of 42 events, the number of events is currently too low to generate such specific data with sufficient precision.

6. Conclusion

This paper describes the use of CAHR method to analyse ATM incidents and to then conclude on human reliability aspects, either in qualitative terms of the most important causal factors leading to incidents or in human reliability quantification. With respect to the use and findings of the CAHR approach, it was found that the method is suitable for the analysis of the 42 incidents provided.

The most important finding is that the four different types of controllers have completely different causal profiles that lead to different mitigation strategies for either APP, ACC, TWR or supervising controllers. Whereas APP controllers have to deal with considerable workload and attentional demands, supervisors show considerable impacts stemming from organisational factors or job responsibilities. ACC controllers have more HMI related issues and distractions than other controllers.

Though the absolute numbers of TWR controllers were low, it can be concluded that TWR controllers have more impacts from night shifts or environmental factors. This finding could be enhanced by analysing more events.

Table 5.1Unconditional &Conditional probabilities of Supervisors impact on ACC, APP and TWR and probabilities for controllers on the operating environment

Item	ni	mi	Justification for mi	sn	D-X	P
Unconditional probability of Controllers						
APP	28	152	No of all sub-events represents generic baseline	12.85	-4.0564	1.70E-02
ACC	13	152	No of all sub-events represents generic baseline	12.85	-5.3240	4.84E-03
TWR	3	152	No of all sub-events represents generic baseline	12.85	-6.7691	2.09E-03

Item	ni	mi	Justification for mi	sn	D-X	P
Conditional probability supervisor -> controller						
Conditional probability that supervisors' behaviour result into APP errors	11	31	Number of sub-events with APP controllers	12.85	-1.8646	1.34E-01
Conditional probability that supervisors' behaviour result into ACC errors	3	13	Number of sub-events with ACC controllers	12.85	-3.4583	3.05E-02
Conditional probability that supervisors' behaviour result into TWR errors	0	7	Number of sub-events with TWR controllers	12.85	-6.4226	1.62E-03

Probability for controller -> operating environment						
Probability that APP behaviour results into undesired state	26	152	Number of all sub-events as all working environments are addressed	12.85	-4.2254	1.44E-02
Probability that ACC behaviour results into undesired state	13	152	Number of all sub-events as all working environments are addressed	12.85	-5.3240	4.84E-03
Probability that TWR behaviour results into undesired state	3	152	Number of all sub-events as all working environments are addressed	12.85	-6.1691	2.09E-03

Overall Probability					
Item	**Conditional probability**	**and**	**Probability of behaviour**	**results into**	**Overall Probability**
P (APP supervisor -> APP controller -> operating environment)	1.34E-01	*	1.44E-02	=	1.93E-03
P (ACC supervisor -> ACC controller -> operating environment)	3.05E-02	*	4.84E-03	=	1.48E-04
P (TWR supervisor -> TWR controller -> operating environment)	1.62E-03	*	2.09E-03	=	3.38E-06

Table 5.2 Findings of the VSM model with respect to consequences of controller's errors

	Level of Altitude	Skill of Controllers	Tools	Consequences of errors	Safety indicators
ACC	16,000 feet and above	Decision Maker, Highly Strategic, Communication Skills, Cope with Stress, Team Workers.	Separation standards, Application Procedures, Inter and Intra working group agreement.	Less Severe, Less chance of errors.	No. of accidents, No. of incidents, Separation violation, e.g. TCAS*, STCA*.
APP	4000 feet Up to FL 16000 feet	Decision Maker, High response to abnormal situations, Communication and coordination skills, Cope with stress, Team Worker.	Separation standards, Application Procedures, Inter and Intra working group agreement.	More Severe, More chance of error.	No. of accidents, No. of incidents, Separation violation, e.g. MSAW*, STCA*.
TWR	Ground up to 4000 feet	Decision Maker, Communication Skills, Cope with Stress, Team Workers.	Separation standards, Application Procedures, Inter and Intra working group agreement.	More Severe, Less chance of error.	Runway incursion, Violation of Clearance.

A human reliability assessment needs to consider these distinctions. Given these findings, a human reliability approach using a "one-fits-all-controllers" will certainly lead to erroneous assessments and hence misleading mitigations.

The results also fit to the framework of the VSM safety model. Key aspects of the VSM model could have been validated like essential PSFs as impacting controllers' performance such as managerial factors on the supervisor level or workload issues on the ACC level.

Unfortunately the small number of events (42 incidents) did only allow for an incomplete validation of the VSM model. More incidents would allow completing the picture. However, key aspects like managerial influences on supervisors as well as key factors for ACC, APP or TWR controllers could verify essential elements and conclusions of the VSM approach.

In conclusion this paper demonstrated how systems theory was used to assess human reliability. The use of the VSM model in conjunction with the CAHR model presented a new framework for the analysis and assessment of any safety critical system such as the air traffic control system. This new framework could prove to be a paradigm shift in the analysis of human reliability and safety. It can be used not only as analysis but also a predictive tool that will enhance organizational performance.

Acknowledgements

The incident analysis using the CAHR method for ATC events was conducted in the Bachelor Thesis of Matthias Proll at the Institute for Human Factors and organizational Psychology of the University Kassel. The authors like to thank him for the profound support to feed this approach with data.

References

1. Al-Ghamdi, S. H. (2010), "*Human Performance in Air Traffic Control System and its Impact on Safety*". PhD thesis, Systems Research Centre, School of Engineering and Mathematical Sciences, City University, London, United Kingdom.
2. Al-Ghamdi, S. H., Panagiotakopoulos, P. D., Stupples, D. W.,(2010), "*The Air Traffic Control System as a Viable System: The Case of the Saudi System*", Journal of air traffic control, Volume 52 No. 1, Winter 2010, Publication of the Air Traffic Control Association, Inc.
3. Apostolakis, G., Soares, C.G., Kondo, S. & Sträter, O. (2004), Human Reliability Data Issues and Errors of Commission, Special Edition of the Reliability Engineering and System Safety. Elsevier.
4. Ashby, R. (1956), "An Introduction to Cybernetics". London, Chapman and Hall.
5. Beer, S. (1985), "Diagnosing the System for Organizations". Great Britain, John Wiley and Sons Ltd.
6. Beer, S. (1979), "*The Heart of the Enterprise*". New York, John Wiley and Sons.
7. Bortz, J. (1989), "*Statistics for social scientists*". Springer. Berlin, Heidelberg,New ork.
8. Eurocontrol (2002), "*Technical Review of Human Performance Models and Taxonomies of Human Error in ATM (HERA)*".
9. Espejo, R, D. Bowling & P. Hoverstadt (1999). The Viable System Model and the Viplan Software, in Kybernetes, Vol 28 Number 6/7, 661-678.

10. Everdij M.H.C and Blom H.A.P (2008), "*Safety Methods Database* ", Http://www.nlr.nl/documents/flyers/SATdb.pdf.
11. Proll, M. (2010) „*Untersuchung von Ereignissen in der Flugsicherung mit Hilfe des CAHR Verfahrens*" *[Evaluation of Air Traffic Management events using the CAHR method]*. Diplomarbeit [Diploma work]. Universität Kassel; Fachgebiet Arbeits- und Organisationspsychologie.
12. Rasch, G. (1980), "*Probabalistic Models for some Intelligence and Attainment Test*". University of Chicago, Press Chicago.
13. Straeter, O. & Reer, B. (1999), "*Comparison of the Application of the CAHR method to the evaluation of PWR and BWR events and some implications for the methodological development of HRA".* In: Modarres, M. (Ed). PSA'99 – Risk-Informed Performance-Based Regulation. American Nuclear Society. La Grange Park, Illinois, USA. (ISBN 0-89448-640-3).
14. Straeter, O. (2000), "*Evaluation of Human Reliability on the Basis of Operational Experience.* GRS-170. GRS. Koln/Germany.
15. Straeter, O. (2001), "*Overview about the CAHR method and its application in assessing errors of commission*". Proceedings of the Workshop on Errors of Commission; Washington, May 2001. OECD.
16. Sträter, O. (2004) "*Considerations on the Elements of Quantifying Human Reliability"* In Apostolakis, G., Soares, C.G., Kondo, S. & Sträter, O. (Ed.) Human Reliability Data Issues and Errors of Commission. Special Edition of the Reliability Engineering and System Safety. Elsevier.
17. Trucco, P., Leva M. & Straeter, O. (2006), "*Human Error Prediction in ATM via Cognitive Simulation: Preliminary Study*". PSAM 8.

Life Cycle Reliability and Safety Engineering
Vol.2 Issue 2 (2013) 34-40

Search for Optimal Preventive Maintenance Policy of Equipment under an Uncertainty of Detection of its Condition

Anil Rana, Prof Ajit Kumar Verma, Prof A Srividya
Indian Institute of Technology, Powaii
E-mail : ranaanil13@hotmail.com

Abstract

Preventive maintenance of equipment is generally chosen over the corrective maintenance policy in order to preclude the chances of sudden failure that incurs high opportunity and repair costs. However, the choice between a time based preventive maintenance and a condition based preventive maintenance is generally carried out under an assumption that the probability of detection of the deteriorating condition of the equipment is 1. This assumption may be far fetched, as most of the condition monitoring techniques have a probability of correct detection of equipment condition less than 1. In some other cases, even the deteriorating condition that is being measured may not have a perfect correlation with the equipment state. The uncertainties involved in use of a condition monitoring system may result in making an improper choice of the PM (Preventive Maintenance) policy resulting in sub-optimal use of resources. This paper presents a method of selection of a suitable preventive maintenance policy under the uncertainty involved in correct detection of the deteriorating condition of a equipment. A non-stationary Gamma wear process has been used to model the deteriorating condition of the equipment and the wear thresholds for alarm and time for monitoring the condition have been included as decision variables for deciding the optimal PM policy based on cost.

Key words: Gamma wear process, TBPM, CBPM

1. Introduction

For an equipment or a machinery, there are only two kinds of maintenance actions: Preventive maintenance and Corrective maintenance. Preventive maintenance can either be time based or condition based. A time based maintenance is understood to be a maintenance action where in, no condition monitoring is undertaken, instead the equipment is replaced or maintained at periodic (or fixed time or age) time intervals. Condition based maintenance on the other hand involves monitoring of the condition of the equipment. When a specified level of deterioration or wear of the subject equipment is surpassed, the equipment may be replaced or repaired. There will always be a chance of breakdown of machinery under both the above preventive maintenance policies that will give rise to a corrective maintenance incurring high cost of repair and opportunity and at times this may have some safety implications too. The optimal choice of the preventive maintenance policy will therefore be guided by the degree to which the chance of corrective maintenance is minimized.

From Barlow and Hunter [1] in 1960, till date, there have been many models and case studies on preventive maintenance policies. References from [1] to [5] are few such examples. Wang[6] provided a thorough review of time based preventive maintenance approaches in the literature. The author has discussed age dependent preventive maintenance policies, periodic preventive maintenance policies, failure rate limit policies, sequential preventive maintenance policies, repair limit polices, opportunistic maintenance polices and optimization approaches for maintenance policies. Blischke and Murthy [7] have also provided a broader view of many of the maintenance policies available in practice. Mann et al [8] provided a review of time based PM models and condition based models. Endrenyi et al [9] proposed use of RCM (Reliability Centered Maintenance) to determine the most cost effective maintenance policy for a given system. Whereas Saranga [10] proposed a structured approach method called the RCP or Relevant Condition Parameter which selects the maintenance significant items according to a risk priority number. Condition based maintenance has been explored by many researchers such as Grall et al [11], Fouadirad et al[12] and Barata et al

[13] and many others. Grall et al [11] has proposed a varying time based monitoring interval based on the extent of deterioration. The other authors have considered continuous monitoring and few others have considered joint effect of shock and deterioration as the failure process of the equipment. However, the above mentioned authors have not considered the effect of probability of detection of the condition of equipment (or wear or deterioration level) as a parameter in their models. In this paper, we consider the cost based comparison between time based PM (TBPM) and condition based PM (CBPM) based on parameters such as, monitoring time intervals, wear threshold and probability of detection of condition. The contribution of this paper therefore is a cost based PM decision model that includes the probability of detection as one of the parameters.

2. TBPM or CBPM

Time based maintenance are generally proposed as an effective strategy for less critical systems which also have a comparatively smaller degree of variability in the failure time distributions. Condition based maintenance techniques on the other hand are justified for highly critical systems that require effective maintenance planning and execution. However, before a CBPM based policy can even be applied on a equipment, availability of a particular parameter that can accurately detect the deteriorating condition of a particular failure mode of the equipment need to be analyzed. For a CBPM policy to be applied on an equipment, it is important that the wear or deterioration progress with respect to time be completely defined in terms of a continuous stochastic process. The process can then be used to define two different levels of deterioration, one the alarm level and the other the failure level. The time for the equipment deterioration to reach the failure level from the alarm level becomes an important consideration in deciding whether enough time is available for the maintainers to act before the equipment fails catastrophically.

The question, therefore, that would often arise during a condition based maintenance decision making is that what should be the interval of monitoring of the equipment condition ? or at what probability of detection of the equipment condition, should one consider the CBPM to be economically viable ? or given a probability of detection if one has to shift away from the optimal time interval of monitoring to another time schedule, would the CBPM still be advantageous over the time based maintenance? In the above arguments we have safely assumed that the corrective maintenance actions are cost and safety prohibitive and therefore need not be considered as an available option.

3. Modeling of Wear/Deterioration of Equipment Condition Using Gamma Process

The gamma process was applied in a series of papers in the fifties to model water flow into a dam, Moran [14,15,16]. However, it was proposed to model deterioration occurring random in time only in 1975. Since then it has been satisfactorily fitted to data on creep of concrete Cinlar et al[17], fatigue crack growth Lawless et al[18], corroded steel gates Frangpool et al[19], thinning due to corrosion Kallenet al [20] etc. A method for estimating a gamma process by means of expert judgment is proposed in Nicolai et al [21]. Gamma wear process which is non-stationary has been shown as the most suitable process in Pandey et al [22] that can take care of the temporal variability of the wear process. In this process the system failure behavior might be described by a damage accumulation model or shock model. The system state at any time 't' can be summarized by a random ageing variable/deterioration W_t. In the absence of repair or replacement actions, W_t is an increasing stochastic process, with W_0=0. The system will fail when the ageing variable or deterioration exceeds a predetermine threshold level W_f. The gamma process is also a reasonable extension of a deterioration process with exponential jumps. The gamma process is parameterized by α and β which can be estimated from the deterioration data . If W_t (deteriorating state) is a gamma process then for all 0 < = s < t the random variable $W_t - W_s$ (increments of deterioration between s and t) has a gamma pdf with shape parameter α(t-s) and a scale parameter β, given by :

$$f_{\alpha(t-s),\beta}(w) = \frac{\beta^{\alpha(t-s)}}{\Gamma(\alpha(t-s))}.w^{\alpha(t-s)-1}.e^{-w.\beta}I_{\{x\geq 0\}} \quad (1)$$

The gamma process has a non-negative independent increment property. The mean and variance of its degradation rate can be expressed as α(t-s)/β and α(t-s)/β^2. For such a process the deteriorating state starting from w_0, the associated failure time distribution, CDF for a given failure threshold, W_f can be expressed as

$$F_{\alpha,\beta}(w) = 1 - \frac{1}{\Gamma(\alpha.t)}.\int_0^{(W_f-w_0).\beta} e^{-u}.u^{(\alpha t-1)}.du \quad (2)$$

Consider the process of wear or deterioration of a equipment shown in figure 1 below (Grall [23]). As time progresses, the equipment condition deteriorates. The equipment is monitored at regular intervals for its deterioration or wear. There are two wear levels which are of consequence. The 'wear threshold level' shown in figure 1 is an alarm level. If on an inspection it is observed that the wear of deterioration of the equipment has crossed the threshold level, it is preventively replaced with a new one or maintained so that its condition becomes as good as new. If however, the wear crosses the 'wear limit', it is considered to have failed and the equipment needs to be correctively replaced. A gamma wear process helps include the wear levels of alarm and failure into the model calculations.

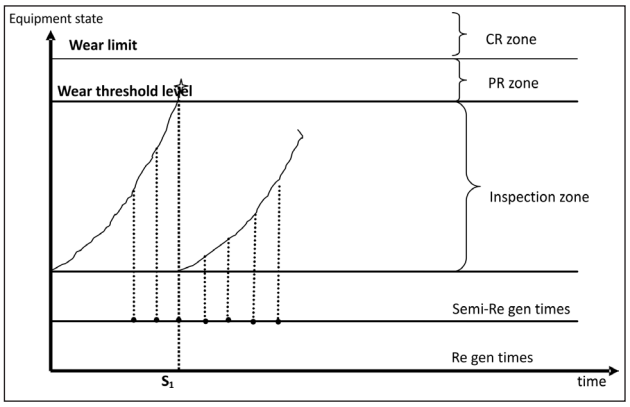

Fig. 1 Schematic evolution of the maintained system state

4. Choice between TBPM and CBPM Using an Example

We consider two main probabilities of maintenance : a probability of carrying out preventive maintenance (which can be either time based or a condition based) and a probability of corrective maintenance. When the TBPM is in force, a corrective maintenance is possible only when the equipment fails before the scheduled time 'T' is clocked. When the CBPM is in force, a corrective maintenance is possible only under the following conditions.

- If the monitoring time schedule has been clocked and the wear has already crossed the alarm level (or threshold level) 'W_{th}' but not the failure level 'W_{lim}'. Also in this case the condition of the equipment has been correctly detected.
- When the equipment wear or deterioration has reached the failure level irrespective of the time clocked by the equipment.
- If the monitoring time schedule has been clocked and the wear or deterioration is well below the alarm level (or threshold level) 'W_{th}', however due to faulty detection system the condition of the equipment has been detected as having crossed the alarm level.

To choose between a TBPM and a CBPM for an equipment, we first evaluate the optimal CBPM policy. This optimal CBPM policy based on cost is decided based on three parameters, mainly : condition monitoring interval, probability of detection of condition and a chosen wear or deterioration threshold level for alarm. A TBPM on the other hand is decided based on only one single parameter i.e. the fixed time interval for carrying out the PM. We use a renewal cycle method to calculate the probability of carrying out PM. Using a non-stationary gamma wear process to map the wear or deterioration of equipment condition, we evaluate the cost rate of its maintenance. The choice of the optimal maintenance policy, TBPM or CBPM can then be decided based on this cost rate. If we divide the time line into discrete renewal cycles with 'n' as the number of such cycles, the probability of the equipment landing up in the preventive maintenance can be explained as :

$$\begin{aligned}\text{Probabilty of PM} = \sum_{n=0}^{\infty} &\text{Probability of detection X}\\ &(\text{Probability of the detrioration}\\ &\text{crosses threshold level}\\ &\text{between}(n+1).\Delta t \text{ and } (n)\Delta t)\\ &\text{X(Probability that deterioration}\\ &\text{doesn't cross wear limit between}\\ &(n+1).\Delta t \text{ and } (n)\Delta t))\\ &+ 0.5(1\text{- Probability of detection}) \text{ X}\\ &(\text{Probability that deterioration}\\ &\text{crosses threshold level between only}\\ &\text{after } n.\Delta t))\end{aligned}$$

Expressing the probabilities in the form of gamma process (see equations 1 and 2 above) and writing the probability of PM in a mathematical equation form we get :

$$P_{PM} = \sum_{n=0}^{\infty} P_d \cdot \left[\int_0^{(W_{th})} \frac{\beta^{\alpha.(n+1)\Delta t^{\zeta}}}{\Gamma(\alpha.(n+1)\Delta t^{\zeta})} x^{\alpha.(n+1)\Delta t^{\zeta}-1} . e^{-\beta x} . dx \right.$$

$$\left. - \int_0^{(W_{th})} \frac{\beta^{\alpha.(n)\Delta t^{\zeta}}}{\Gamma(\alpha.(n)\Delta t^{\zeta})} x^{\alpha.(n)\Delta t^{\zeta}-1} . e^{-\beta x} . dx \right].$$

$$\left[1-\int_{0}^{(W_{\lim}-W_{th})}\frac{\beta^{\alpha.\Delta t^{\zeta}}}{\Gamma(\alpha.\Delta t^{\zeta})}x^{\alpha.\Delta t^{\zeta}-1}.e^{-\beta x}.dx\right]+0.5.(1-P_d).$$

$$\left[\int_{0}^{(W_{th})}\frac{\beta^{\alpha.(n+2)\Delta t^{\zeta}}}{\Gamma(\alpha.(n+2)\Delta t^{\zeta})}x^{\alpha.(n+2)\Delta t^{\zeta}-1}.e^{-\beta x}.dx-\int_{0}^{(W_{th})}\frac{\beta^{\alpha.(n+1)\Delta t^{\zeta}}}{\Gamma(\alpha.(n+1)\Delta t^{\zeta})}x^{\alpha.(n+1)\Delta t^{\zeta}-1}.e^{-\beta x}.dx\right] \quad (3)$$

The cost rate can then be given as :

$$\text{Cost Rate} = \frac{P_{PM}(\text{PM cost}) + P_{CM}*(\text{CM cost}) + \text{CI}*\text{Renewal cycle}/\Delta t}{\text{Renewal cycle}}$$

where

we have $P_{CM} = 1 - P_{PM}$

$$\text{Renewal Cycle} = \sum_{n=0}^{n=\infty}(n+1).\Delta t.P_d\left(\begin{array}{l}\text{(Prob of wear threshold being reached between } n\Delta t \text{ and } (n+1)\Delta t).\\ \text{prob of wear limit not being reached between } n\Delta t \text{ and } (n+1)\Delta t)\end{array}\right)+$$
$$\sum_{n=0}^{\infty}\left(\begin{array}{l}\text{(Prob of wear threshold being reached between } n\Delta t \text{ and } (n+1)\Delta t).\\ \text{mean time of wear limit being reached between } n\Delta t \text{ and } (n+1)\Delta t)\end{array}\right)+$$
$$\sum_{n=0}^{n=\infty}(n+1).\Delta t\left(\begin{array}{l}\text{(Prob of wear threshold being reached between } (n+1)\Delta t \text{ and } (n+2)\Delta t).0.5(1-P_d).\\ \text{(This represents time due to wrong alarm)}\end{array}\right)$$
$$\sum_{n=0}^{\infty}\left(\begin{array}{l}0.5(1-P_d).(\text{Prob of wear threshold being reached between } (n-1)\Delta t \text{ and } (n)\Delta t).\\ \text{mean time of wear limit being reached between } (n)\Delta t \text{ and } (n+1)\Delta t)\end{array}\right)_{n>1} \quad (5)$$

And the meantime to reach the wear limit between $n\Delta t$ and $(n+1)\Delta t$ can be given as $\int_{n\Delta t}^{(n+1)\Delta t} x.f(x).dx$

P_{PM} = probability of preventive maintenance

P_{CM} = probability of corrective maintenance

P_d = Probability of detection

W_{th} = wear threshold level or alarm level deterioration

W_{lim} = wear limit for failure

Δt = time interval for condition monitoring

$\alpha.t^{\zeta}$ = shape parameter of the gamma wear process

β = scale parameter of the gamma wear process

n = number of cycles

CI = cost of inspection; PM ost = cost of PM and CM cost = cost of CM

f(x) = pdf of wear (gamma process)

Assumptions

The following assumptions are being made in the model being discussed

- The equipment is preventively replaced when the condition being monitored reaches a wear threshold level. When following the time based PM, the equipment is replaced at regular time based intervals
- If the probability of detecting the correct condition of the equipment is 'P', there is 0.5*(1-P) chance of making a wrong detection on the safer side. We know that there is always a (1-P) probability of making a wrong detection. In the model we have assumed that out of this (1-P) probability there is a 50% chance of making a wrong detection that the equipment has reached the alarm level (thereby causing a corrective maintenance to take place) where in reality the wear of equipment hasn't reached the alarm level at all.
- Though 'P' is the probability of correct detection of the condition of the equipment, there is a perfect correlation between the parameter being monitored and the actual condition of the equipment
- The equipment follows a non-stationary gamma wear process with shape parameter '$\alpha.t^{\zeta}$' and scale parameter 'β' where α = 0.02278; β =1.2; ζ=1. The wear threshold for failure is a non-dimensional number = 10
- Cost of setting up a comprehensive condition monitoring system has not been included in the example

5. Results

Using the values of parameters of gamma wear process shown in the assumptions above and for a chosen value of probability of detection P_d we use equation (3) to evaluate the probability of PM (P_{PM}). We then calculate the value of renewal cycle using equation (5) and for assumed values of PM cost and CM cost (corrective maintenance) evaluate the cost rate of performing the maintenance under the chosen PM policy using equation (4). For finding out the optimal CBPM policy we evaluate the cost rates for varying values of monitoring intervals and wear alarm levels. Once the optimal CBPM policy is identified (for the existing value of P_d), we compare the optimal CBPM policy with the TBPM policy, the cost rate of which is also evaluated using the above procedure.

Using the values of a example of a equipment, we have evaluated cost rate values (in accordance with equation (4)) as shown in figure 2. The figure clearly shows that depending upon the chosen monitoring interval and the wear alarm level, the minimum cost rate values changes. In figure 2, the minimum cost rate

occurs at a wear alarm level of 8.0 with a monitoring interval of around 12 days. . It may be noted that figure 2 has been drawn up with probability of detection (P_d) equal to 1. Different sets of curves (similar to figure 2) can be obtained for different values of probability of detection of equipment condition.

In comparison to figure 2 for CBPM policies, figure 3 shows a similar plot for cost rate with a TBPM policy with renewal cycle= T; the time for maintenance interval. The plots in figure 2 display the optimal wear threshold level and the optimal time for monitoring the condition of a given equipment. The plots also display the alternatives available with the maintenance engineer in deciding the wear threshold level that he would like to choose for his equipment. Since the time between the wear threshold level and the wear limit level is crucial in making the logistics arrangement ready for the upcoming maintenance actions, a maintenance engineer may like to choose the wear threshold level that may not be an optimal solution. The time available to the maintenance engineer, once the wear threshold level has been reached can be given in accordance with the approximation formula Verma et al [24]

$$\left(\frac{\text{wearlimit}.\beta}{\lambda}+\frac{0.479}{\lambda}.e^{\left(\frac{1-\zeta}{\Gamma(\beta+\zeta)}\right)}\right)^{\frac{1}{\zeta}}-\left(\frac{\text{wearthreshold}.\beta}{\lambda}+\frac{0.479}{\lambda}.e^{\left(\frac{1-\zeta}{\Gamma(\beta+\zeta)}\right)}\right)^{\frac{1}{\zeta}} \quad (6)$$

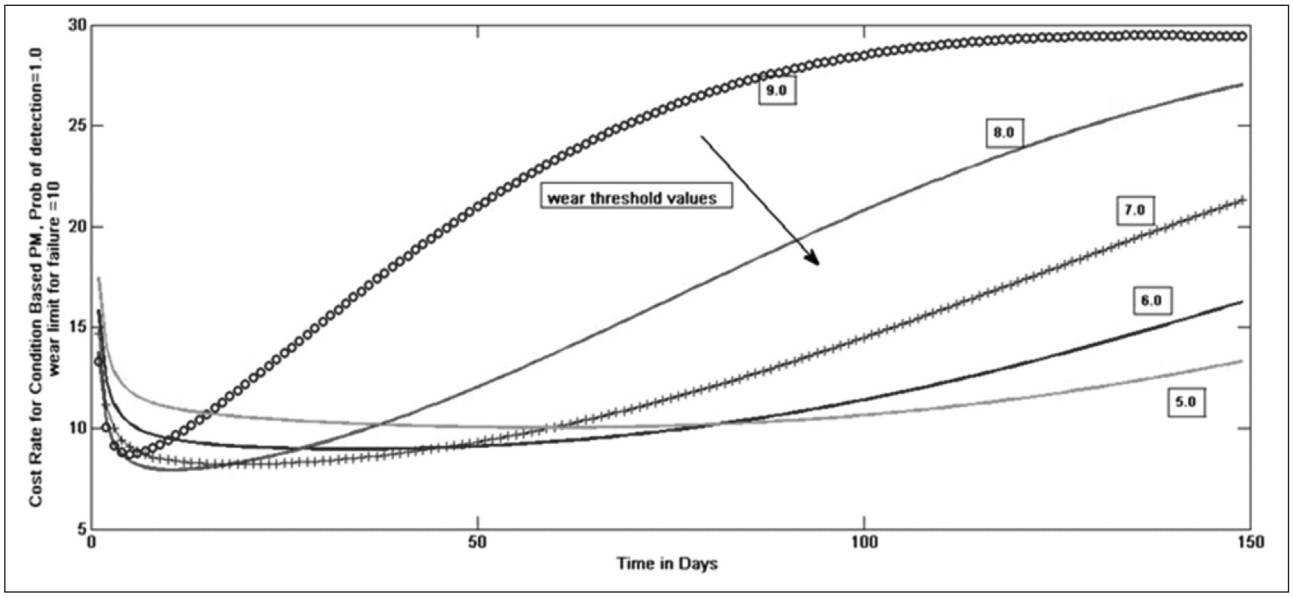

Fig.2 Cost rate for various wear threshold levels and condition monitoring intervals

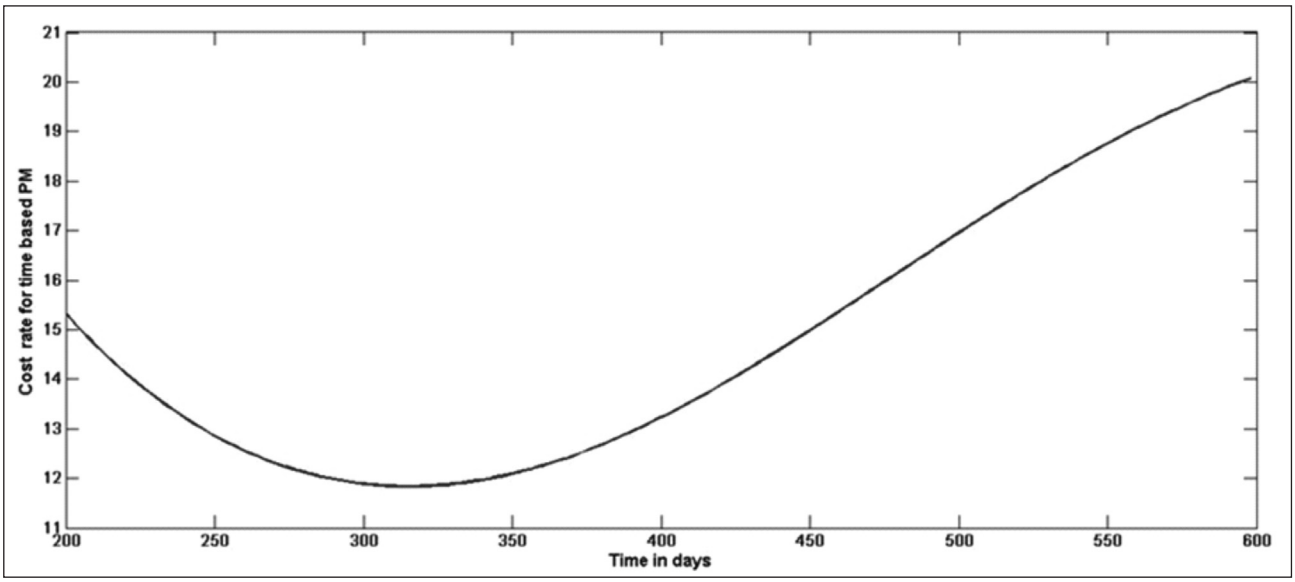

Fig. 3 Cost rate for time based PM

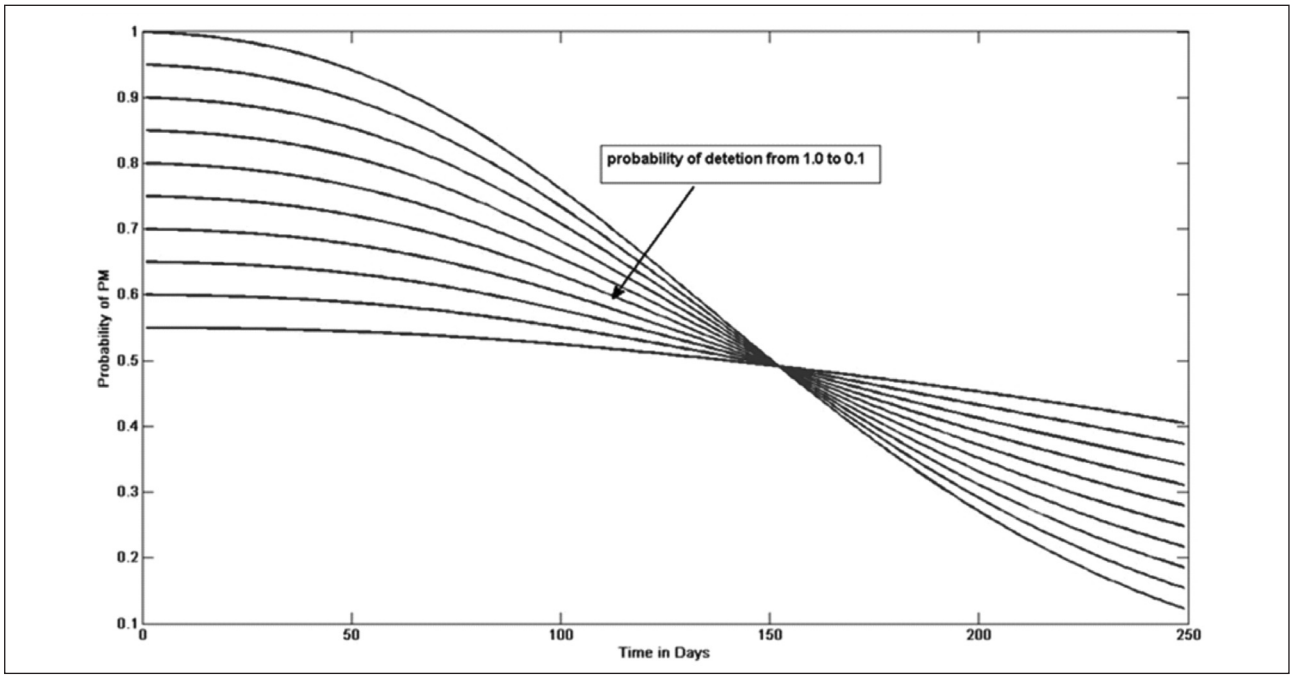

Figure 4 Probability of carrying out PM for various probabilities of detection and monitoring interval a=0.02278; β=1.2;ζ=1;wear threshold =7; wear failure limit=10

The maintenance engineer may also not be able to provide regular monitoring at every optimum time interval because of various constraints, instead he can choose the time interval suitable to him and know the consequences in terms of cost rate as per the plots displayed in figure 2. It may be noted in figure 2 that as the wear threshold level rises from 5.0 to 9.0 the optimal time interval in that particular wear threshold moves to the left. This is because the time available for the equipment to reach the wear limit for failure (assumed to be 10 in this case) becomes shorter and therefore the monitoring becomes more frequent.

Going by the solutions in figure 2 and 3 for the example being discussed, one can see that the CBPM is a better preventive maintenance policy than the TBPM, when the wear alarm threshold level is maintained at 8 and the monitoring is carried out every 12 days. However, if the logistics time delay (time in arranging resources for replacement of the equipment) does not allow for this wear alarm threshold, and if this threshold has to be maintained at 5, the TBPM seems to be a better PM policy. It may be noted that these results are drawn up for probability of detection of 1. For values less than 1 a different set of graphs will need to be drawn up.

The effect of the probability of detection on the probability of PM has been shown at figure 4 for different monitoring time intervals. The cost rate advantage of CBPM over a TBPM policy is limited by the probability of correct detection of the condition of equipment. As this probability value drops and as the monitoring time intervals are increased, the probability of PM drops too thus increasing the probability of corrective maintenance.

6. Conclusion

The paper has presented a model which shows the importance of including the probability of correct detection of condition (wear or deterioration level) as one of the key parameters for making a choice between a TBPM or a CBPM policy for a equipment. To justify the choice of a CBPM over a TBPM policy, there will always be a lower limit to the probability of correct detection of its wear or deterioration level which should not be crossed. The effect of this probability of detection of condition on the probability of carrying out a PM on a equipment (for a chosen example) has been displayed in figure 4.

References

1. R Barlow and Larry Hunter, "Optimum preventive maintenance policies", Operational Research. Vol.8, pp 90-100 feb 1960
2. H Mine and H Kawai, "Preventive replacement of a 1 unit system with a wear out state", IEEE Transactions on reliability, vol R-23, no.1, April 1974
3. T. Nakagawa, "Optimum preventive maintenance policies

for repairable systems", IEEE Transactions on Reliability, vol R-26, 1977
4. KS Park, "Optimal continuous wear limit replacement under periodic inspections", IEEE Transactions on Reliability , 37(1), 97-102. 1988
5. IB Gertsbakh, "Models of Preventive Maintenance", 1977, North Holland Publishing company
6. Wang H. A survey of maintenance policies of deteriorating systems. Eurpoean Journal of Operational Research, 139, 469-489, 2002
7. Blischke WR and Murthy DN, " Case Studies in Reliability and Maintenance", New Jersey, John Wiley and Sons. (2003)
8. Mann L, Saxena A, Knapp G, " Statistical Based or Condition Based Maintenance? Journal of Quality in Maintenance, 1, 46-59, 1995.
9. Endrenyi J, Aboresheid S, Allan N, Anders GJ, Asgarpoor S, Billinton R, The Present Status of Maintenance Strategies and the Impact of Maintenance on Reliability. IEEE Transactions on Power Systems, 16, 638-646, 2001
10. Saranga, H. Relevant Condition Parameter Strategy for an Effective Condition Based Maintenance. Journal of Quality in Maintenance Engineering, 8, 92-105, 2002.
11. Grall A, Berenguer C and Dieulle L. A condition Based Maintenance Policy for Stochastically Deteriorating Systems. Reliability Engineering and System Safety 76, pp 167-180, 2002
12. Fouladirad, Grall A, Dieulle L. On the Use of Oi-Line Detection for Maintenance of Gradually Deteriorating Systems. Reliability Engineering and System Safety. 93, pp 1814-1820, 2008
13. Barata J, Soares CG, Marseguerra M, Zio E. Reliability Engineering and System Safety, 76, pp 255-264, 2002
14. Moran PAP. A probability theory of dams and storage systems. Australian Journal of applied science 1954; 5(2):116-24
15. Moran PAP. A probability theory of dams and storage systems: modifications of the release rules. Australian Journal of applied science 1955;6(2):117-30
16. Moran PAP. The theory of storage. London:Methuen: 1959
17. Cinlar E, Bazant ZP, Osman E., Stochastic process for extrapolating concrete creep. J Engg Mech Div 1977; 103 (EM6):1069-88
18. Lawless J, CrowderM, "Covariates and random effects in a gamma process model with application to degradation and failure", Lifetime data Analysis 2004; 103(3):213-27
19. Frangpool DM, Kallen MJ, Van Noortwijk JM, "Probabilistic models for life cycle performance of deteriorating structures: review and future directions", Prog Struct Eng Mater 2004:6(4):197-212
20. Kallen MJ, Van Noortwijk JM. Optimal maintenance decisions under imperfect inspection. Reliability Engineering and System Safety 2005;90(2-3) 177-85
21. Nicolai RP, Dekker R, Van Noortwijk JM, "A comparison of models for measurable deterioration: An application to coatings on steel structures.", Reliability and Engg System Saftey 2007, in press, doi:10.1016/j.ress. 2006.09.021
22. Pandey MD, Yuan X-X, Van Noortwijk JM. A comparison of probabilistic deterioration models for life cycle management of structures. Structure Infrastructure Engineering, 2007.
23. Grall A, Laurence D, B Christophe and R Michael," Continuous time predictive maintenance scheduling for a deteriorating system", IEEE Transactions on Reliability, vol 51, no. 2, June 2002
24. Verma AK, Srividya A, Rana Anil "Approximation of MTTF calculation of a non-stationary Gamma wear process", International Journal of System Assurance Engineering and Management, DOI: 10.1007/s13198-011-0079-0.

Life Cycle Reliability and Safety Engineering
Vol.2 Issue 2 (2013) 41-52

Cosmic Phenomena in Mirce Mechanics Approach to Reliability and Safety

I. Zaczyk, J. Knezevic
MIRCE Akademy, Exeter, EX5 1JJ, UK
jk@mirceakademy.com

Abstract

The main objective of this paper is to argue that the scientific approach to reliability and safety is the only way forward for the reliability community, if accurate predictions regarding occurrences of negative functionability events are to be made and subsequently verified during the operational processes of the future man made, managed and maintained systems. For that to happen, a scientific understanding of the mechanisms that cause occurrences of functionability events of the surrounding natural environment are required. Then and only then, can accurate and meaningful reliability and safety predictions become possible, enabling the ultimate goal of reducing the probability of failure event occurrences during the life of manmade, managed and maintained systems. This paper focuses on the scientific understandings of 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 reliability and safety.

Keywords: Reliability and safety, cosmic phenomena, cosmic rays

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 reliability and safety 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 manmade and maintained systems, is the true understanding of the impact of the environment that surrounds them. The reliability and safety 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 negative functionability events a fully comprehensive understanding of the generation, behaviour and the interactions between the relevant physical phenomena must first be understood.

Consequently, the main objective of this paper is to argue that the scientific approach to reliability and safety 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. For that to happen a 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 the occurrence of functionability events to be understood. Then and only then, can accurate and meaningful reliability and safety predictions become possible, enabling the ultimate goal of reducing the probability of failure event occurrences during the life of manmade, 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 the lives of manmade, managed and maintained systems. [2]. For decades, 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 the major causes of cosmic phenomena from the physical world that can influence system functionability from a reliability and safety 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 reliability and safety.

3. 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:

a) High-energy cosmic ray neutrons.
b) Thermal or low energy cosmic ray neutrons.
c) 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.

4. 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:

a) Extra galactic cosmic rays,
b) Galactic cosmic rays,
c) Solar cosmic rays,
d) 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.

4.1 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 $cm^{-2}\,s^{-1}$.

4.2 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.

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

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.

4.3 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

[1] "Mean integral solar cosmic ray flux at solar minimum and maximum" Derived from a table from " Heliospheric Physics and Cosmic Rays ", Chapter 4 Lecture notes fall term 2003, prepared by Kalevi Mursula and Ilya Usoskin, University of Oulu.

approximately equal to the semi-major axis of Earth's orbit around the Sun) from the Earth.

5. 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, is 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.

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.

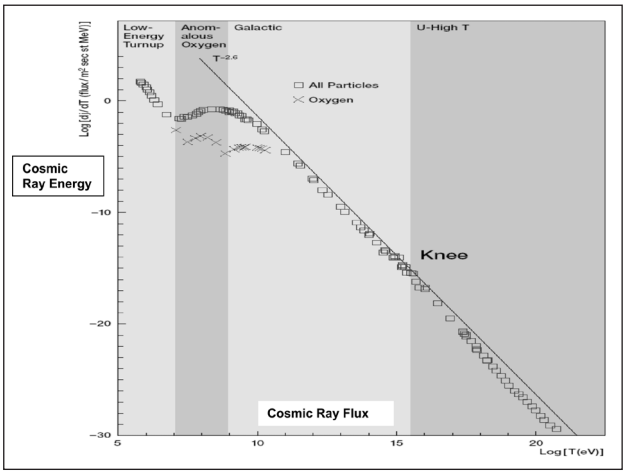

Figure 1 – Energy spectrum of cosmic rays measured at the Earth[2]

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

Table 2 – Categories of cosmic ray particles[3]

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.

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

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

Table 3 – Modulation of Cosmic Rays[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 has the affect of temporarily increasing the shielding effect of the Earth's magnetosphere. Decreases usually occur over several hours.
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.

as proposed by current astrophysics research are as follows:

a) 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.

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

c) 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.

6. Modulation of Cosmic Rays

The intensity of the secondary cosmic ray flux in the atmosphere is not constant because it is influenced

[4] "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.

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 are listed in Table 3.

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.

7. 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.

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 as in the Quebec event in the introduction 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.
- The speed of protons during SPEs is dependent on energy level and therefore typically takes 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.

Table 4 – Secondary cosmic ray particles[5]

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

8. 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

9. 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

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

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.

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 4 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.

10. High Energy Cosmic Ray Neutrons

As neutrons possess 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% [5]. 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 10^{-15} coulombs). As a result incident neutrons pose a much greater upset risk to semiconductors than alpha particles.

11. 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 paper 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 commercial passenger aircraft is increased by about an order of magnitude and varies dependent on internal location due to the different composition and distribution of materials. [6]

12. 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 an 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. 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.

13. Conclusion

This paper has demonstrated that if accurate predictions regarding the occurrences of functionability events are to be made, it is mandatory to implement the Mirce Mechanics scientific approach to understanding the competing mechanisms driving negative functionability events, as the consequence of the diverse range of interactions between manmade systems and the surrounding natural environment. Then and only then, can the reduction of the probability of the occurrence of failure events during the life of manmade, 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 the 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.

Referemces

1. 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.
2. Knezevic, J. Physical Scale of Mirce-Mechanics," Lecture Notice, Master Diploma Programme, MIRCE Akademy, Woodbury Park, Exeter, UK, 2009.
3. 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
4. Knezevic, J., Functionability in Motion, Proceedings 10th International Conference on Dependability and Quality, DQM Institute, 2010, Belgrade, Serbia.
5. Dirk. J. D., Nelson, M.E., Ziegler, F.E., Thompson, A., Zabel, T.H., "Terrestrial thermal neutrons," IEEE Trans. Nucl. Sci., vol. 50, no. 6, pp. 2060–2064, Dec. 2003
6. Zaczyk. I, "Analysis of the Influence of Atmospheric Radiation Induced Single Event Effects on Avionics Failures", Master Dissertation, MIRCE Akademy, Exeter, UK, 2010.