Are simple approximations for safety function calculations too good to be true?

Summary

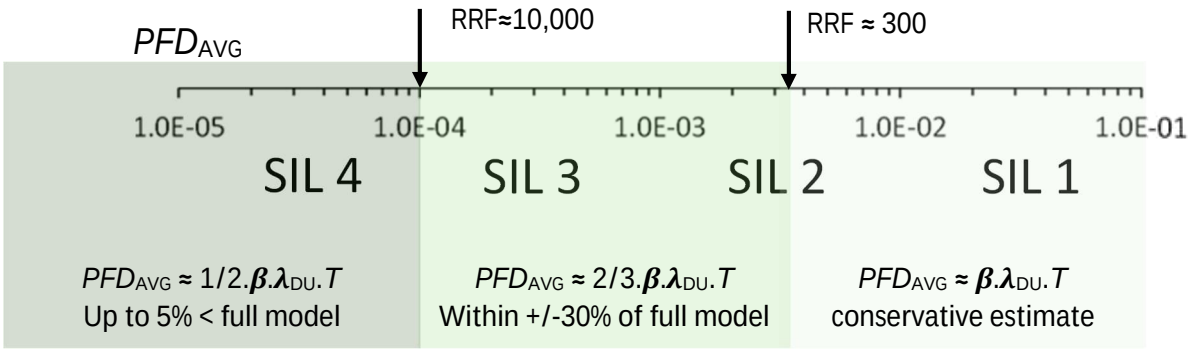
The standards IEC 61508, ANSI/ISA 84 and ISO/TR 12489 provide very detailed calculation models for estimating probability of failure. Most users rely on commercial software packages because the calculations seem to be so complicated.

It may seem incredible that simple approximations can be used to estimate the integrity level achieved by any safety function, and those simple approximations are just as accurate as the detailed models.

This paper compares the accuracy of simple approximations with fully detailed 'MooN' calculation models based on the standards. It shows that these approximations remain valid across the entire range of safety function applications.

The error between the approximations and the full models is < 30%. The error is not significant when compared with the range of uncertainty in the models. The models can only predict failure probability to within half an order of magnitude at best, due to the inevitable variability in equipment performance.

The probability of failure of most fault tolerant MooN safety functions acting in low-demand mode can be estimated as: $PFD_{AVG} \approx 2/3.\beta.\lambda_{DU}.T$



$PFD_{AVG} \approx 1/2.\beta.\lambda_{DU}.T$ may be used as a closer though less conservative estimate if RRF > 10,000

$PFD_{AVG} \approx \beta.\lambda_{DU}.T$ may be used as a closer and more conservative estimate if RRF < 300

Frequency of dangerous failure of any fault tolerant MooN safety function acting in high-demand mode or continuous mode can be estimated as $FDF \approx \beta_D.\lambda_{DD} + \beta.\lambda_{DU}$

This approximation is within 10% of the fully detailed model for all practical applications. The error is less than 1% for most applications. The common cause detected failure term $\beta_D.\lambda_{DD}$ can be omitted if there is a dependable fault reaction that will put the equipment into a safe state in response to coincident failure being detected in all N channels.

Background

The functional safety standards IEC 61508 and ANSI/ISA 84 describe complicated calculations that may be applied to estimate the likelihood of failure for safety functions. ISO/TR 12489 provides a more detailed analysis of several different approaches to estimating failure likelihood. The calculation models all assume that device failure rates remain constant over the useful operating life of the devices.

The second edition of IEC 61511 introduced the new requirement that the '*reliability data uncertainties shall be assessed and taken into account when calculating the failure measure*'. This requirement forces us to question the basic assumption that failure rates remain constant. In practice we find that failure rates vary by more than a factor of 10 between different applications. Failure rates can also vary by a factor of 10 over the operational life of the equipment in any individual application.

Single channel architectures (1oo1 or NooN) have no redundant elements and zero fault tolerance. Single channel architecture is generally limited to $PFD_{AVG} > 0.003$ and $RRF < 300$ unless unusually high diagnostic coverage can be achieved. Single channel architectures are usually only applied for SIL 1 and low-range SIL 2.

Fault tolerant architectures (MooN) are usually applied only for target $RRF > 100$ (i.e. $PFD_{AVG} < 0.01$). MooN safety functions with annual inspection and testing typically achieve at least mid-range SIL 2 ($RRF > 300$).

References

The fully detailed MooN models used in this study were developed by the 61508 Association in 2023.

The models are described in the paper 'A simplified MooN model for safety functions (2024)'. The models were validated against the equations given in IEC 61508-6.

A spreadsheet based on these models is available for public use under a Creative Commons licence: The MooN SIF model (2023).

Refer to Abbreviations below for explanation of the symbols and abbreviations used in the equations.

MooN models

Detailed MooN model for low-demand mode, synchronised testing

Synchronised testing in this context means that periodic inspection and testing is carried out on all N channels within a day or two. The average probability of failure on demand for functions operating in low-demand mode with synchronised testing can be estimated as:

$$PFD_{AVG} \approx \frac{2^{N-M+1}}{N-M+2}.\binom{N}{N-M+1}.\left((1-\beta_D).\lambda_{DD}.MTTR + \frac{(1-\beta).\lambda_{DU}.T}{2}\right)^{N-M+1}$$

$$+ \beta_D.\lambda_{DD}.MTTR + \beta.\frac{\lambda_{DU}.T}{2}$$

The average test interval T takes proof test coverage PTC into account: $T = PTC.T_1 + (1-PTC).T_2$

Detailed MooN model for low-demand mode, staggered testing

The average probability of failure on demand for functions operating in low-demand mode can be improved by a factor of at least 2 simply by staggering intervals for periodic inspection and testing.

The N channels are still each inspected and tested at an average interval T, but the inspection and testing of individual channels is offset at intervals of T/N.

The primary effect of staggered inspection and testing is to reduce the weighted average test interval that applies to common cause failures. This is based on the assumption that all of the other channels would be inspected and tested promptly if a fault or failure were detected in any one channel.

Staggering routine service, overhaul and renewal intervals should also reduce equipment failure rates.

$$PFD_{AVG} \approx St_{M,N}.\binom{N}{N-M+1}.\left((1-\beta_D).\lambda_{DD}.MTTR + \frac{(1-\beta).\lambda_{DU}.T}{2}\right)^{N-M+1}$$

$$+ \beta_D.\lambda_{DD}.MTTR + \beta.\frac{\lambda_{DU}.T/N}{2}$$

Simplified approximation for MooN in low-demand mode

This simplified approximation may be used for any low-demand mode application of MooN architecture: $PFD_{AVG} \approx 2/3.\beta.\lambda_{DU}.T$

A reciprocal version of the approximation can be used to relate the risk reduction target to the mean time between dangerous undetected failures of each channel: $RFF \approx 3/2.MTBF_{DU}/(\beta.T)$

The value of β depends on the chosen MooN voting architecture. Refer to Table D.5 in IEC 61508-6.

β is typically ≈ 0.1 for 1oo2 architecture and ≈ 0.15 for 2oo3 architecture. The value increases with higher M but decreases with higher levels of fault tolerance (N-M).

The approximation is valid regardless of whether the channels are inspected and tested at synchronised or staggered intervals.

Staggered testing and/or renewal can be modelled simply by using the weighted average test interval $T = PTC.T_1/N + (1-PTC).T_2/N$.

Explanation

The PFD_{AVG} estimated by a fully detailed MooN model includes two main terms: a term for common cause failures and a term for coincident failure of independent channels.

The largest term in the model represents the contribution from common cause failures, modelled as $1/2.\beta.\lambda_{DU}.T$.

The term representing the combination of independent failures is modelled using the failure probability of a single channel raised to the power N-M+1, i.e. $\approx ((1-\beta).\lambda_{DU}.T/2)^{N-M+1}$.

The exponent N-M+1 is the number of coincident channel failures that would cause the function to fail. A safety function with MooN architecture requires M out of N channels to work correctly for the safety function to act successfully. The hardware fault tolerance level is defined as N-M.

The contribution from independent failures becomes negligible for architectures that can tolerate more than one single faulted channel. The exponent N-M+1 > 2 when the fault tolerance N-M > 1. The probability of failure of a single channel raised to the power of 3 or more becomes too small to have any significance.

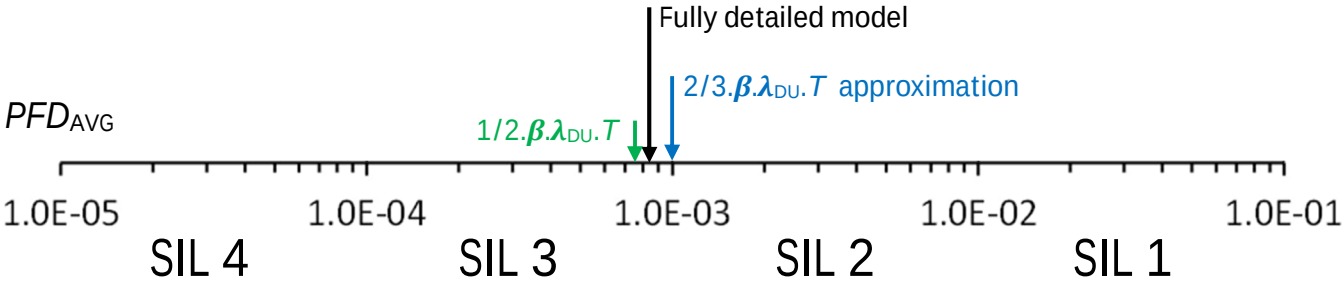
For architectures with N-M = 1 the contribution from independent failures increases with M. The increase is in proportion to the number of different ways that N-M+1 failures can occur out of a choice of N channels. For example, there is only 1 way of having 2 coincident failures in a 1oo2 architecture, but there are 6 ways of having 2 failures in a 3oo4 architecture. A 3oo4 architecture is 6 times more likely to fail due to independent failures than a 1oo2 architecture.

The common cause failure term is always larger than the independent channel failure term in all practical safety function applications unless common cause failures can be reduced to < 2%.

For example, consider a safety function designed to achieve *RRF* = 1,000 with 1oo2 architecture:

According to the simple approximation $2/3.\beta.\lambda_{DU}.T$, a target of $\lambda_{DU} \approx 0.015$ pa for the overall channel failure rate is low enough to achieve $PFD_{AVG} = 10^{-3}$ with $\beta = 0.1$ and $T = 1$ y.

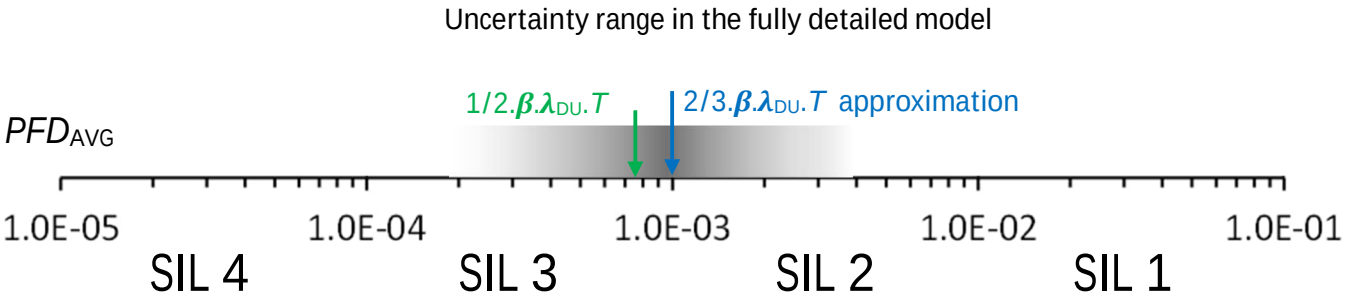
The fully detailed 1oo2 model results in an estimate of $PFD_{AVG} \approx 8.3 \times 10^{-4}$. The contribution to PFD_{AVG} from common cause failures alone is $1/2.\beta.\lambda_{DU}.T \approx 7.5 \times 10^{-4}$, representing more than 90% of the total.



The difference between the full model and $2/3.\beta.\lambda_{DU}.T$ is 17%. The difference is not significant given that all failure rates vary over at least one order of magnitude in operation.

Failure performance varies because it depends on environmental factors and on how effectively the condition of the equipment is maintained throughout its operational life.

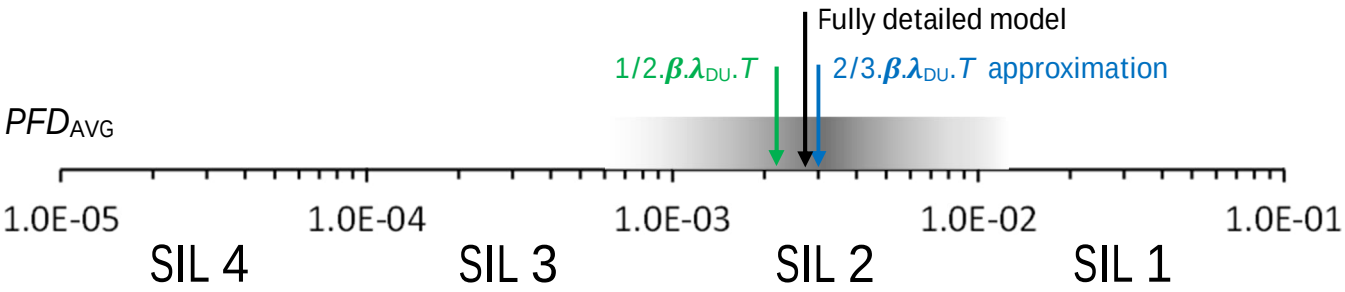
Variation in performance leads to uncertainty in the fully detailed model. The following diagram illustrates how the approximations are well within the uncertainty range of the fully detailed model.



Similarly, consider a safety function designed to achieve *RRF* = 300 with 1oo2 architecture:

According to the simple approximation $2/3.\beta.\lambda_{DU}.T$, a target of $\lambda_{DU} \approx 0.045$ pa for the overall channel failure rate is low enough to achieve $PFD_{AVG} = 3 \times 10^{-3}$ with $\beta = 0.1$ and $T = 1$ y.

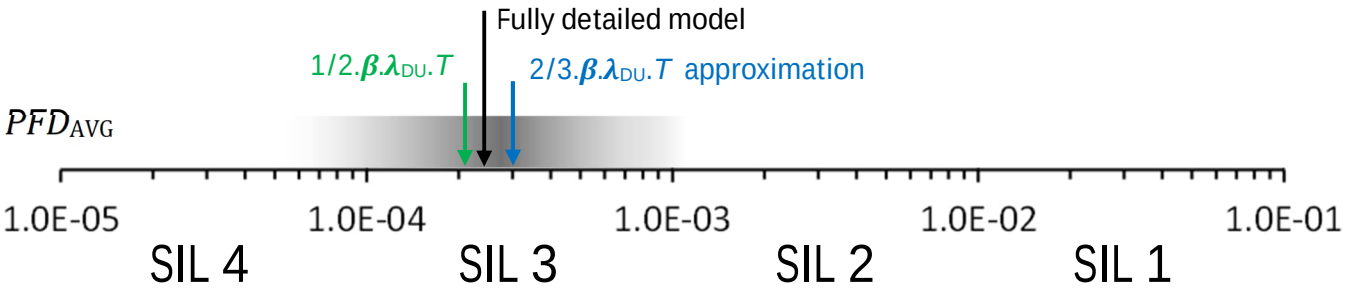
The fully detailed 1oo2 model results in an estimate of $PFD_{AVG} \approx 2.8 \times 10^{-3}$. The contribution to PFD_{AVG} from common cause failures is 80% of the total, $1/2.\beta.\lambda_{DU}.T \approx 2.25 \times 10^{-3}$



Finally, consider a safety function designed to achieve *RRF* = 3,000 with 1oo2 architecture:

According to the simple approximation $2/3.\beta.\lambda_{DU}.T$, a target of $\lambda_{DU} \approx 0.0045$ pa for the overall channel failure rate is low enough to achieve $PFD_{AVG} = 3 \times 10^{-4}$ with $\beta = 0.1$ and $T = 1$ y.

The fully detailed 1oo2 model results in an estimate of $PFD_{AVG} \approx 2.32 \times 10^{-4}$. Common cause failures contribute 96% of the total: $1/2.\beta.\lambda_{DU}.T \approx 2.25 \times 10^{-4}$



Detailed MooN model for high-demand or continuous mode

The frequency of dangerous failure (*FDF*) for safety functions with MooN architecture can be estimated as:

$$FDF = \beta_D.\lambda_{DD} + \beta.\lambda_{DU} + \frac{\mathrm{N!}}{(\mathrm{M}-1)!}.\lambda_{DU}.\left((1-\beta_D).\lambda_{DD}.MTTR + (1-\beta).\lambda_{DU}.(T/2+MRT)\right)^{\mathrm{N-M}}$$

The common cause detected failure term $\beta_D.\lambda_{DD}$ can be omitted if there is a dependable fault reaction that will put the equipment into a safe state when coincident failure of all N channels is detected.

Simplified approximation for MooN in high-demand or continuous mode

The approximation $FDF \approx \beta_D.\lambda_{DD} + \beta.\lambda_{DU}$ is within 10% of the fully detailed model for any practical applications. The error is usually < 1%.

For example, consider a 1oo2 architecture with $\lambda_D \approx 0.1$ pa, 50% diagnostic coverage and $\beta = \beta_D = 0.1$.

$\lambda_{DU} \approx 0.05$ pa, $\lambda_{DD} \approx 0.05$ pa, $T = 1$ y, $MTTR = 0.01$ y:

$$FDF = \beta.\lambda_D + \lambda_{DU}.\left((1-\beta_D).\lambda_{DD}.MTTR + (1-\beta).\lambda_{DU}.(T/2+MRT)\right)$$

FDF ≈ 0.1 x 0.1 pa + 0.05 pa x (0.9 x 0.05 pa x 0.01 y + 0.9 x 0.05 pa x 0.5 y)

FDF ≈ 10^{-2} pa + 1.1 x 10^{-3} pa ≈ 1.1 x 10^{-2} pa

An error of 10% is not significant given the wide variation in λ_D that should be expected during operation.

Diagnostic coverage of 90% can typically be achieved in high-demand and continuous mode functions. The error is then typically < 1% because λ_{DU} would usually be < 0.01 pa:

FDF ≈ 0.1 x 0.1 pa + 0.01 pa x (0.9 x 0.09 pa x 0.01 y + 0.9 x 0.01 pa x 0.5 y)

So then *FDF* ≈ 10^{-2} pa + 5.9 x 10^{-5} pa ≈ 1.006 x 10^{-2} pa ≈ 1 x 10^{-2} pa

Detailed analysis for MooN low-demand

The error between the simple approximations and the fully detailed models varies with β, λ and T.

Error range expected for final elements in process sector applications

Process sector applications generally use pneumatic, hydraulic or electrically actuated valves as final elements. Electrical contactors or circuit breakers might also be used.

Frequencies of undetected dangerous failure that can be achieved for these final elements are typically in the range 0.03 pa to 0.003 pa. That corresponds to about 300 FITS to 3000 FITS, or 3 x 10^{-7} failures per hour to or 3 x 10^{-6} failures per hour.

This may be expressed as mean time between undetected dangerous failure in the range 30 y to 300 y.

The common cause failure fraction β_{INT} for 1oo2 architecture is typically ≈ 0.1.

It is difficult to achieve β_{INT} < 0.05 unless complete diversity is achieved between the channels.

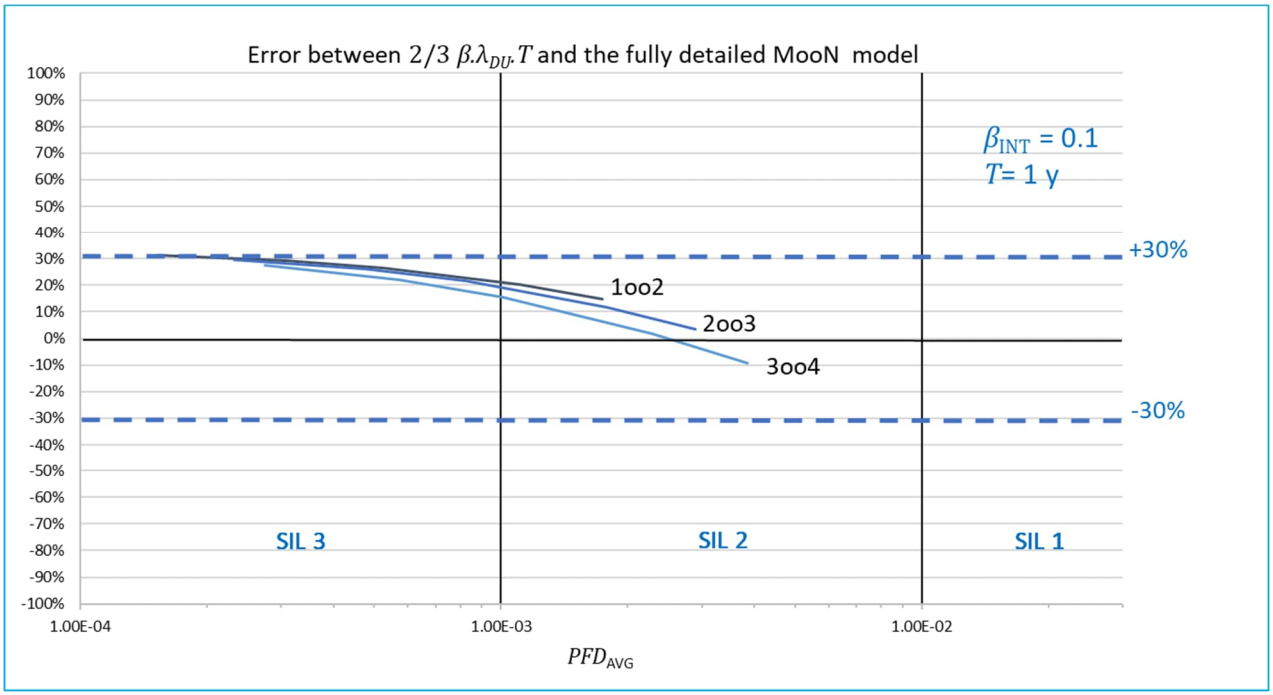
The average test interval T is usually in the range between 1 year and 4 years.

The architecture of final element subsystems is usually limited to 1oo1 or 1oo2.

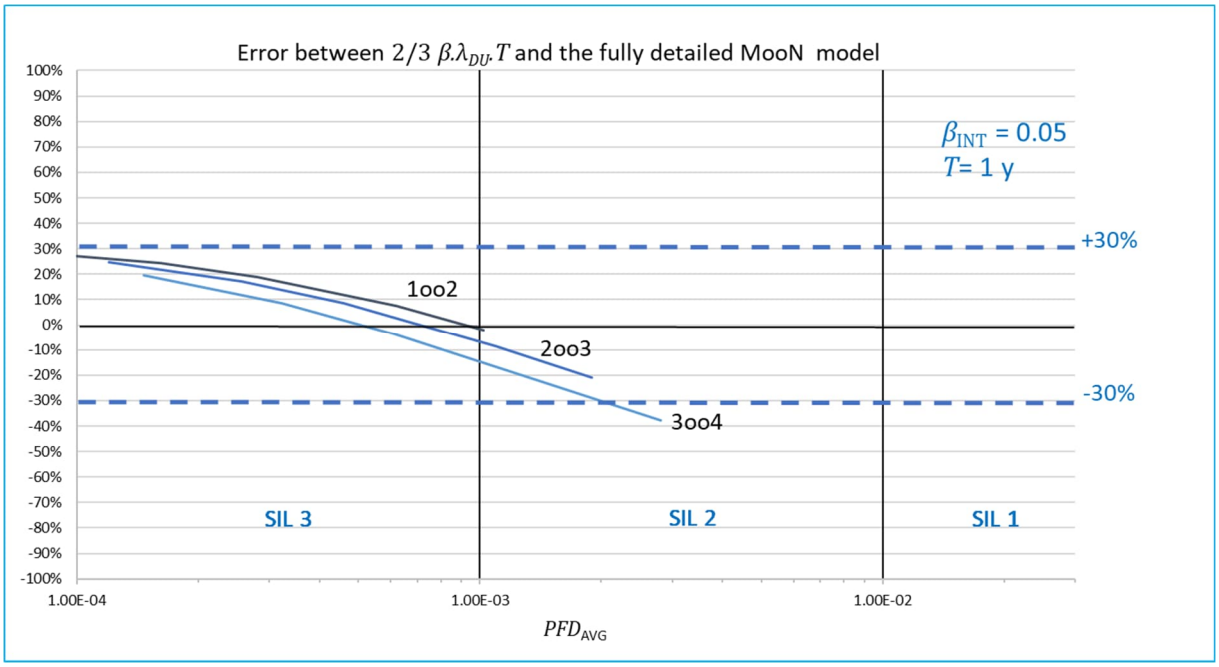
2oo3 or 3oo4 may be used for final element architecture in some applications. For example, a safety function on a chemical reactor may depend on closing at least 3 out of 4 reactant feed valves.

The following chart shows the error between the simple approximation and a fully detailed MooN model for these architectures with λ_{DU} in the range 0.03 pa to 0.003 pa. Performance in high SIL 2 to mid SIL 3 range is achieved with β_{INT} ≈ 0.1, and T = 1 y.

The simple approximation of $2/3.\beta.\lambda_{DU}.T$ produces an estimate of PFD_{AVG} that is about 10 to 20% higher than the detailed model. The maximum error is about +30% for PFD_{AVG} < 3 x 10^{-4} (RRF > 3,000).



Reducing common cause failure factor to $\beta_{INT} \approx 0.05$ reduces PFD_{AVG} by a factor of 2.



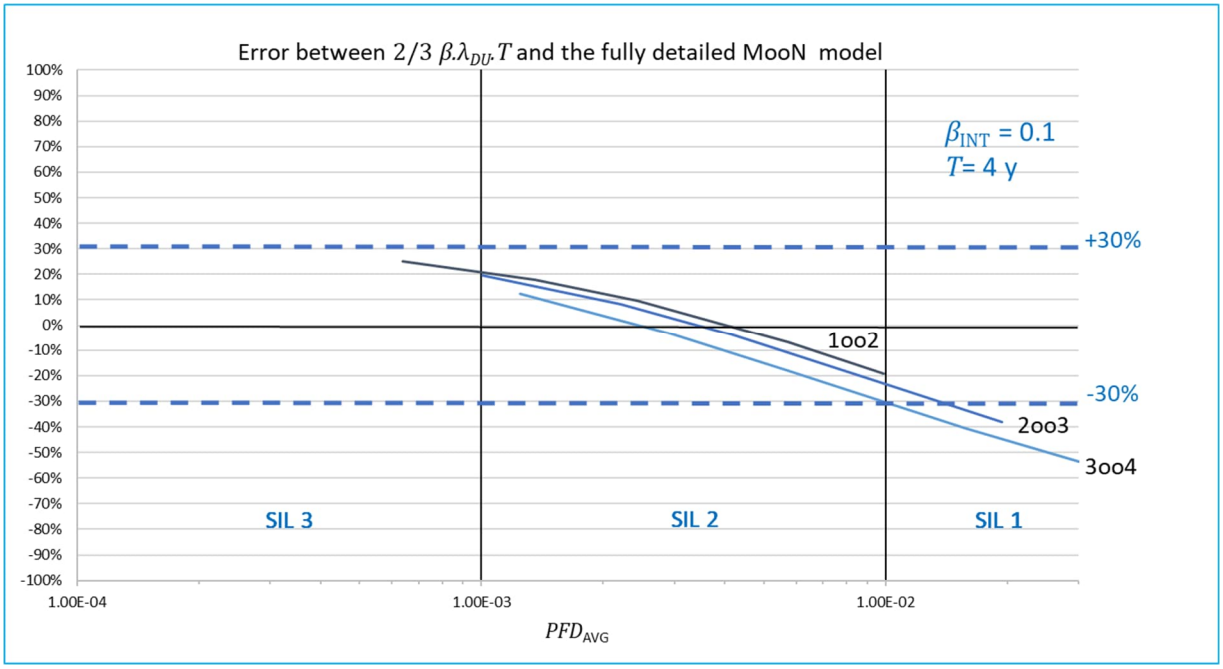
The simple approximation of $2/3.\beta.\lambda_{DU}.T$ produces an estimate of PFD_{AVG} that is typically about 15% higher than the detailed model with $\beta_{INT} \approx 0.05$ and T = 1 y..

The maximum error is about +20% for $PFD_{AVG} < 2 \times 10^{-4}$ ($RRF > 5{,}000$).

The error introduced by the approximation is increased if the average test interval is extended to 4 y.

The approximation is still within +/-20% of the detailed model in the SIL 2 range for T = 4 y.

The error is > 30% in the SIL 1 range for 3oo4 architecture with T = 4 y. It would be unusual to design 3oo4 final element systems with T > 2 y. $PFD_{AVG} \approx \boldsymbol{\beta.\lambda_{DU}.T}$ could be used for a more conservative estimate of MooN functions if the performance target is in the SIL 1 range.



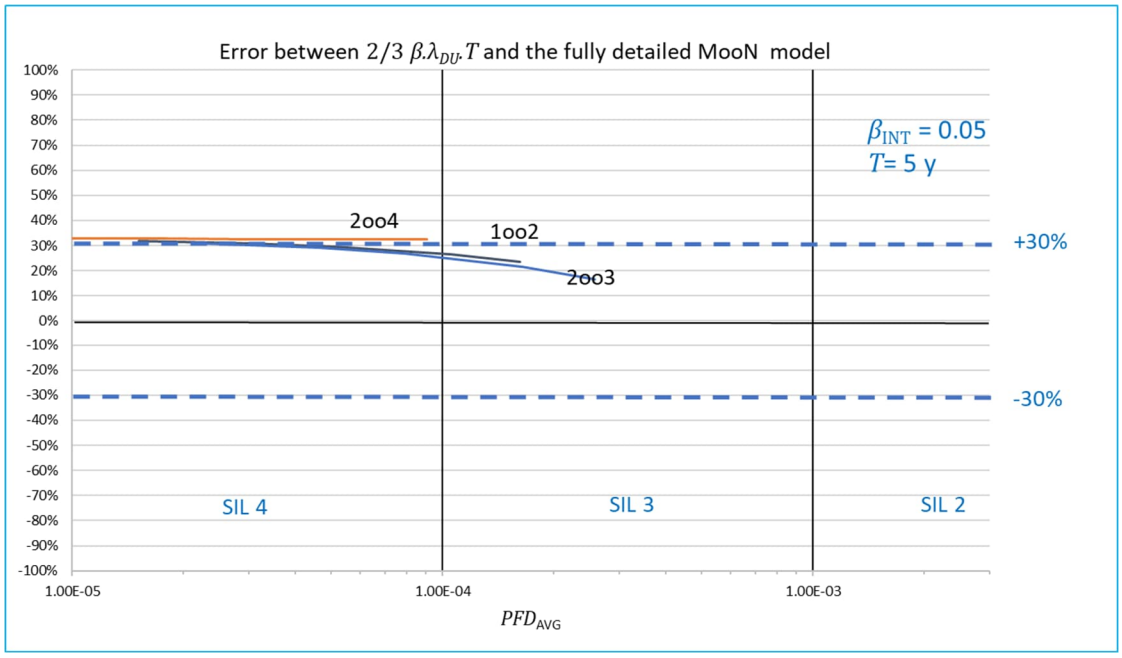
Error range expected for electronic sub-systems

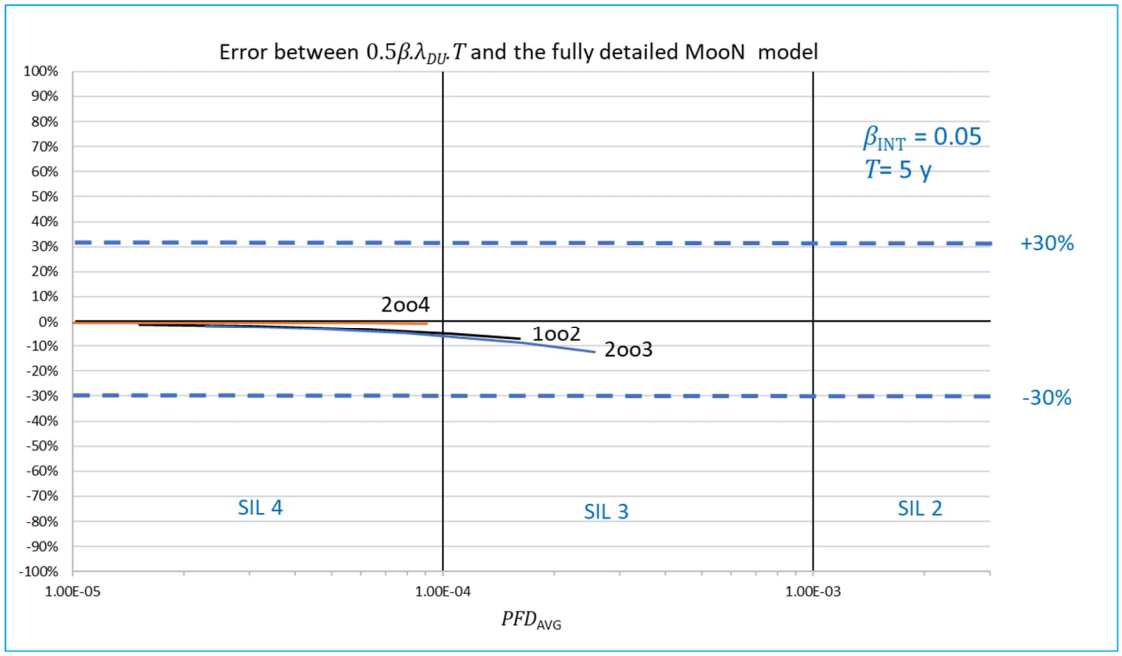
Frequencies of dangerous failure that can be achieved for electronic sub-system elements are typically in the range 0.003 pa to 0.0003 pa. That corresponds to about 30 FITS to 300 FITS, or 3×10^{-8} failures per hour to or 3×10^{-7} failures per hour.

Diagnostic coverage (*DC*) of at least 60% is typically achievable in electronic subsystems. The most common architectures used are 1oo2, 2oo3 and 2oo4. Note that the performance of the adaptive architecture 1oo2D is similar to the performance of 2oo3.

The following charts show the error between the simple approximations and a fully detailed MooN model for 1oo2, 2oo3 and 2oo4 architectures with λ_D in the range 0.003 pa to 0.0003 pa with *DC* = 60%.

Performance in high SIL 3 to high SIL 4 range is achieved with $\beta_{INT} \approx 0.05$, and $T = 5$ y.





The charts show that $1/2.\beta.\lambda_{DU}.T$ is a useful approximation for sub-systems that achieve $PFD_{AVG} < 1 \times 10^{-4}$.

A fully detailed analysis would be required to justify any claim that a whole safety function could achieve $PFD_{AVG} < 1 \times 10^{-4}$ from end to end.

Diagnostic coverage ≥ 60% enables safety integrity performance to be achieved with intervals as long as 5 or 10 years for periodic testing. Annual inspection is still recommended so that deterioration can be detected before it leads to failure.

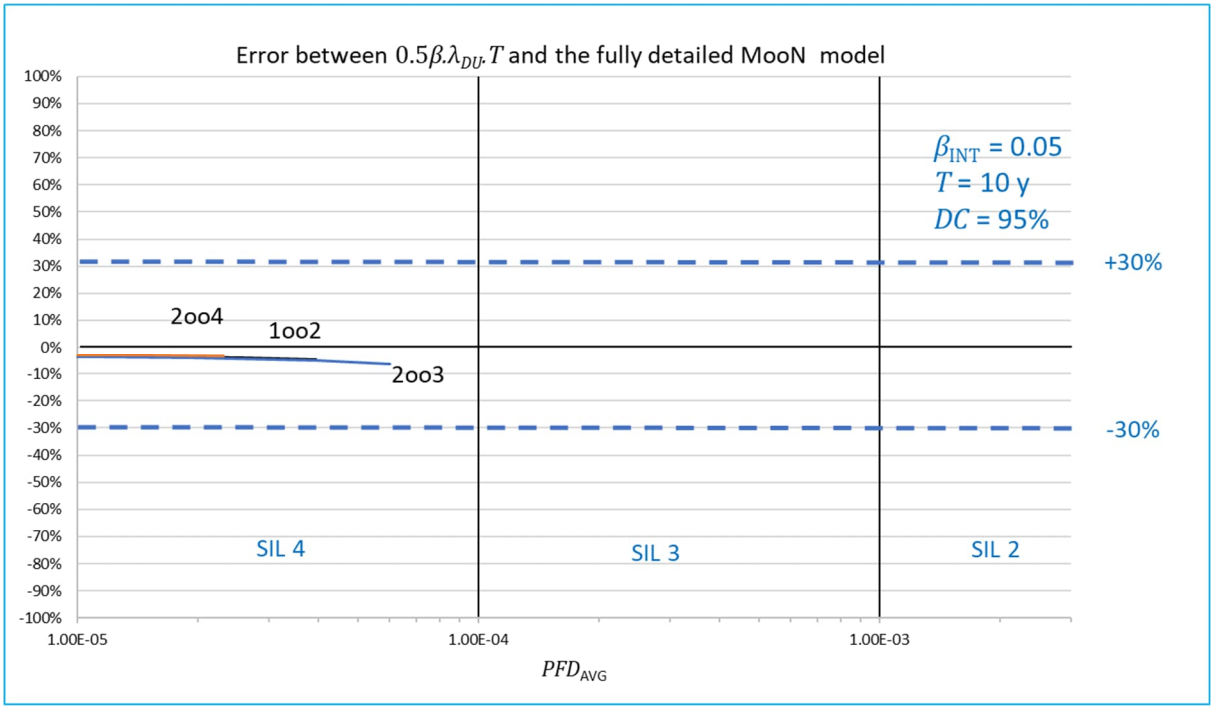
Electronic sub-systems applied for sensors would typically have $T < 5$ y.

Logic solver systems may have $T > 10$ y but would generally have $DC \geq 90\%$.

Final element subsystems depending on variable speed motor drives may also have $T > 10$ y and would also generally have $DC \geq 90\%$.

A claim of $DC \geq 90\%$ would need to be justified in detail. The claimed diagnostic coverage depends on successful automatic or manual fault reaction within the assumed *MTTR*, as well as depending on the technical integrity of the diagnostic function itself.

The chart below shows the error between the simple approximation $1/2.\beta.\lambda_{DU}.T$ and a fully detailed MooN model for 1oo2, 2oo3 and 2oo4 architectures with λ_D in the range 0.003 pa to 0.0003 pa with $DC = 95\%$.



NooN

NooN architectures have no fault tolerance. All N out of N channels need to work correctly for the safety function to achieve its specified performance.

NooN architectures are usually limited to SIL 1 applications, though low-range SIL 2 may be achievable with 1oo1.

The fully detailed model for NooN (including 1oo1) is:

$$PFD_{AVG} \approx N.\lambda_{DU}.T/2 + N.\lambda_{DD}.MTTR$$

where $T = PTC.T_1 + (1 - PTC).T_2$

The IEC 61508-6 version of the equation adds mean repair time *MRT* to $T/2$, but the difference is negligible because $MRT << T$.

The model can be simplified by neglecting the term $N.\lambda_{DD}.MTTR$ if $DC < 90\%$:

$$PFD_{AVG} \approx N.\lambda_{DU}.T/2$$

The resulting error is typically < 10%.

The term $N.\lambda_{DD}.MTTR$ may also be omitted with $DC \geq 90\%$ provided that a dependable automatic fault reaction puts the equipment into a safe state in response to detected faults. The term can always be omitted if the protected equipment is kept in a safe state (e.g. shut down) while the safety function is out of service during the time to restoration.

The reciprocal version of the approximation relates the risk reduction target to the mean time between dangerous undetected failures: $RRF \approx 2.MTBF_{DU}/(N.T)$

This approximation makes it clear that borderline SIL 2 performance ($RRF > 100$) can be achieved with 1oo1 architecture if $MTBF_{DU} > 50$ y and $T = 1$ y.

Detailed analysis for NooN

The fractional error introduced by neglecting $N.\lambda_{DD}.MTTR$ can be expressed in terms of the diagnostic coverage *DC*, the average test interval *T*, and the mean time to restoration *MTTR*:

$$\text{error} = (N.\lambda_{DD}.MTTR)/\,(N.\lambda_{DU}.T/2 + N.\lambda_{DD}.MTTR)$$

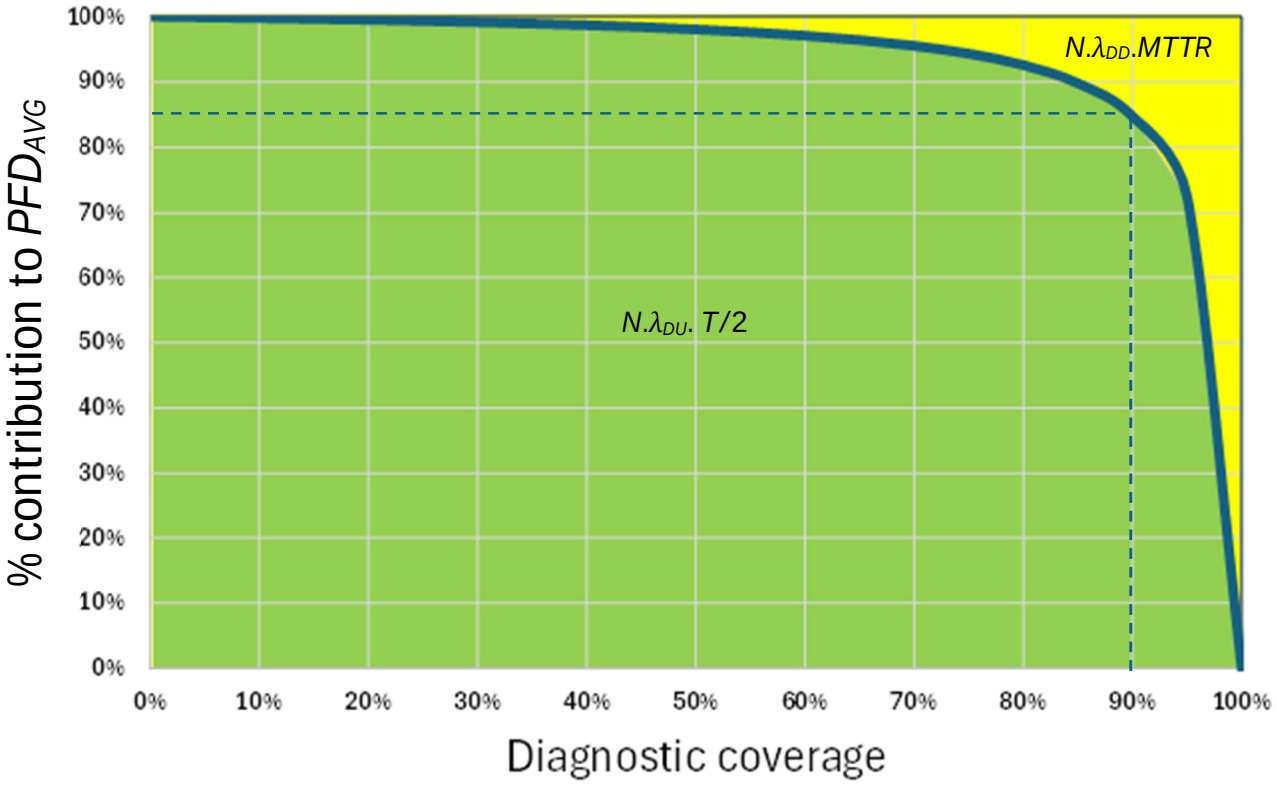
$$= (DC.\lambda_D.MTTR)/\,\big((1-DC).\lambda_D.T/2 + DC.\lambda_D.MTTR\big)$$

$$= MTTR/\,\big((1/DC - 1)\,.T/2 + MTTR\big)$$

For example, with *MTTR* = 0.01 y, *T* = 1 y, and *DC* = 0.9, the error is about 15%

$$\text{error} = 0.01/\,(0.055 + 0.01) \approx 0.15$$

The following chart illustrates the relative contribution of $N.\lambda_{DD}.MTTR$ and $N.\lambda_{DU}.T/2$ to the estimated PFD_{AVG} of a NooN architecture:



Again, a claim of *DC* ≥ 90% would need to be justified in detail. The claimed diagnostic coverage depends on successful automatic or manual fault reaction within the assumed *MTTR*, as well as depending on the technical integrity of the diagnostic function itself.

Abbreviations

TABLE 1. ABBREVIATIONS

Abbrev.	Description
β	The fraction of undetected failures that have a common cause
β_D	Of those failures that are detected by the diagnostic tests, the fraction that have a common cause
CCF	Common Cause Failure
DC	Diagnostic Coverage
FDF	Frequency of Dangerous Failure
λ	Failure Rate Subscripts: S – Safe, SD – Safe Detected, SU– Safe Undetected D – Dangerous, DD – Dangerous Detected, DU– Dangerous Undetected, DN – Dangerous Never Detected NE- No Effect Note that ISA S84 uses superscripts instead of subscripts
MooN	'M' out of 'N' voting: at least M out of N channels are required for successful operation
MRT	Mean Repair Time (= time to organise the repair after a failure has been found and then repair and restore the device to service)
MTBF	Mean Time Between Failures
MTTF	Mean Time To Failure (= MTBF + MTTR)
MTTR	Mean Time To Restoration (= time to diagnose a failure plus the MRT)
PFD_{AVG}	Average Probability of Failure on Demand
PFH_{AVG}	Average Probability of Failure per Hour (equivalent to failure rate per hour) PFH is used in place of FDF in IEC 61508 Ed.2 but was deprecated in ISO/TR 12489
PTC	Proof Test Coverage
RRF	Risk Reduction Factor
SIF	Safety Instrumented Function
SIL	Safety Integrity Level
$St_{M,N}$	Correction factor for evenly staggered testing intervals (refer to the 61508 Association publication T6A044 – Staggered Proof Testing Coefficients, available for download at https://61508.org/downloads/). $St_{M,N}$ is typically in the range 0.75 to 0.9 for commonly used architectures.
T	Weighted average proof test interval
T_1	Proof test interval
T_2	Full proof test interval (if inspection and testing at T_1 has limited coverage) Some standards use T_2 for the interval between demands, or for the planned interval between full overhaul and/or renewal of safety function equipment.

Creative Commons Licence

The document was prepared by Mirek Generowicz and Blake Merritt of I&E Systems Pty Ltd.