ASPECT2019

Does SIL Live up to Expectations?

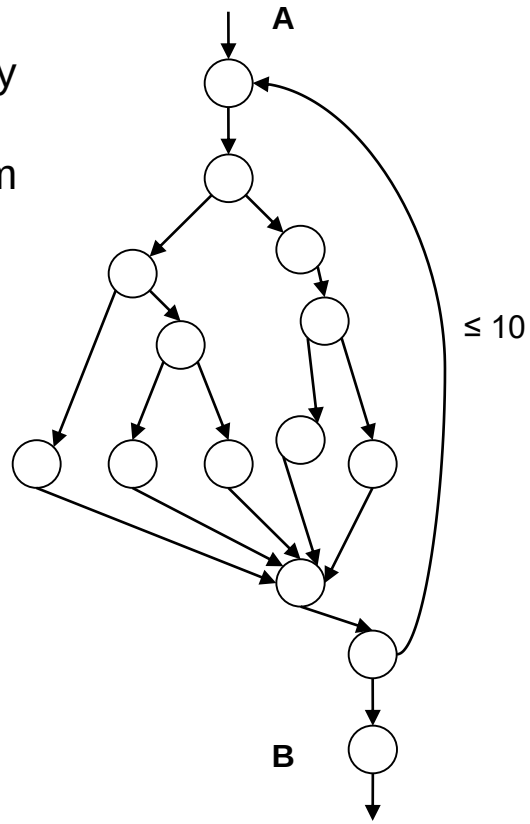
Roger Short

Parental Guidance

May contain equations and powers of 10

Why Testing Software is not Practicable

How many unique paths from A – B?



Answer

$5^{10} + 5^9 + \ldots\ldots 5^1 = 1.2 \times 10^7$

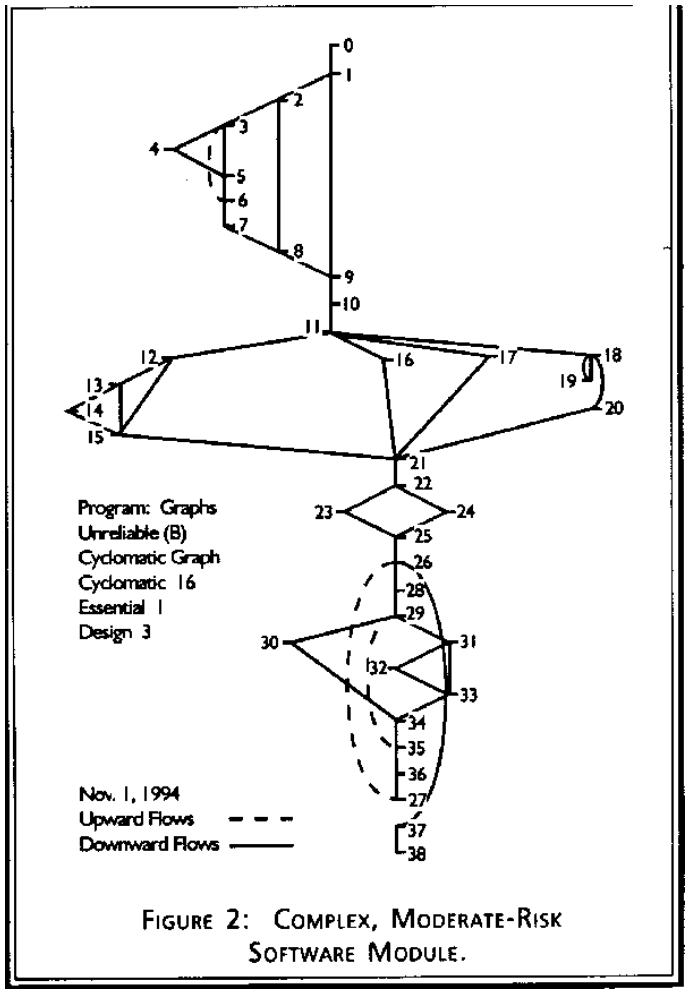


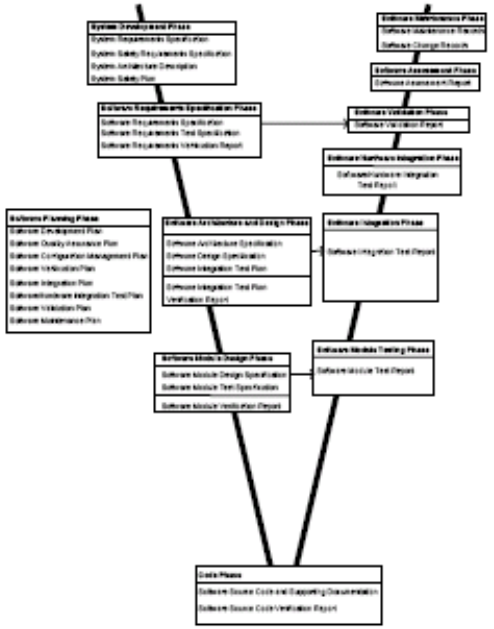
FIGURE 2: COMPLEX, MODERATE-RISK SOFTWARE MODULE.

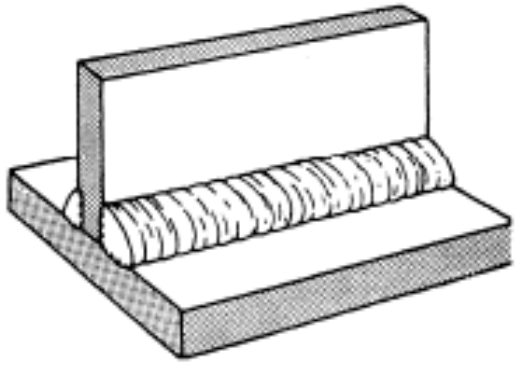
How many paths through this program?

The Solution

Safety Assurance by Process

- If you use the right process, you get the right result (like welding)





You can’t test every weld, so you rely on the welding process and the competence of the welder

The Origin of SIL

- Numerous techniques for preventing or detecting software errors had been devised by the early 1980's.
- These varied in the extent to which they could be expected to be effective in achieving their objective.
- The experts drafting what eventually became IEC 61508 ranked them in four groups in order of expected performance.
- This ranking must have been made on the basis of engineering judgement, as little or no empirical data was available about the actual effectiveness of any of the techniques.

What the Standards Say

What EN50126 says

TFFR [h^{-1}]	SIL attribution	SIL qualitative measures
$10^{-9} \leq$ TFFR $< 10^{-8}$	4	Defined in sector specific standards
$10^{-8} \leq$ TFFR $< 10^{-7}$	3	
$10^{-7} \leq$ TFFR $< 10^{-6}$	2	
$10^{-6} \leq$ TFFR $< 10^{-5}$	1	

What EN50128 implies

TFFR [h^{-1}]	SIL attribution
$10^{-9} \leq$ TFFR $< 10^{-7}$	3/4
$10^{-7} \leq$ TFFR $< 10^{-5}$	1/2

In EN50128 the qualitative measures for SIL3 and 4 are the same, as are the measures for SIL1 and 2.

What failure rate do you expect if you use SIL4 measures?

Facts or Opinions?

Through the ages people have preferred to rely on facts rather than opinions

I don't care how you made it – will it stop a **bullet?**

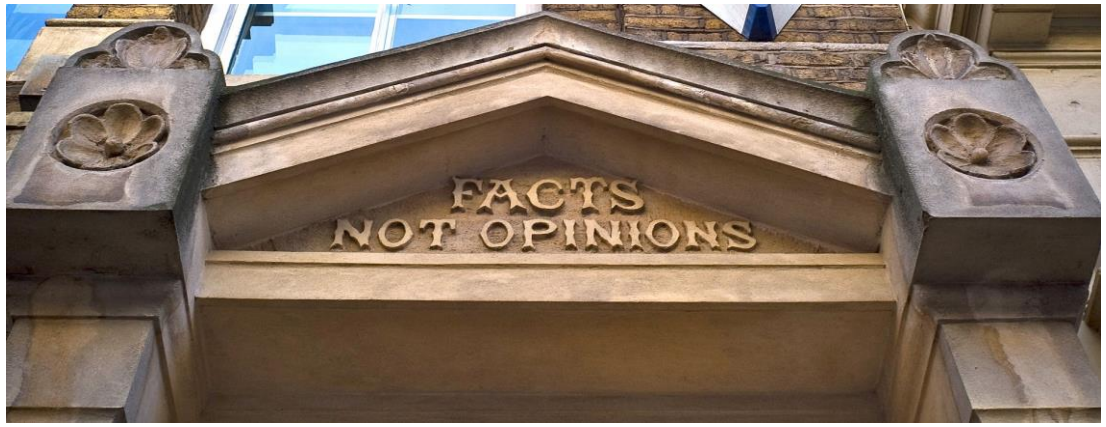


Photo by Lars Plougmann on Flickr

Entrance to Kirkaldy Testing Works in London

Is there yet enough experience of SIL4 systems to Confirm that the engineering judgement was correct?

Calculating Failure Rate

- Unsafe Functional Failure Rate

$$\lambda_{uff} = F / (NT)$$

- Where N = number of installations (only partial data available)

  T = time in service (only partial data available)

  F = number of unsafe failures (unknown)

Population Data for SIL4 Systems

- Data on the size of installed populations was sought for the following signalling systems which are generally expected to aim for SIL4:
  - Electronic interlocking systems (data found for only four types of interlocking, but limited data for remote terminal units which they might include)
  - Radio Block Centres (no data found)
  - ERTMS on-board units (comprehensive data available)
  - Axle counters (no data found)
  - TBTC trackside subsystems
  - TBTC on-board subsystems
- The available data accounts for a minority of the units believed to be in service.

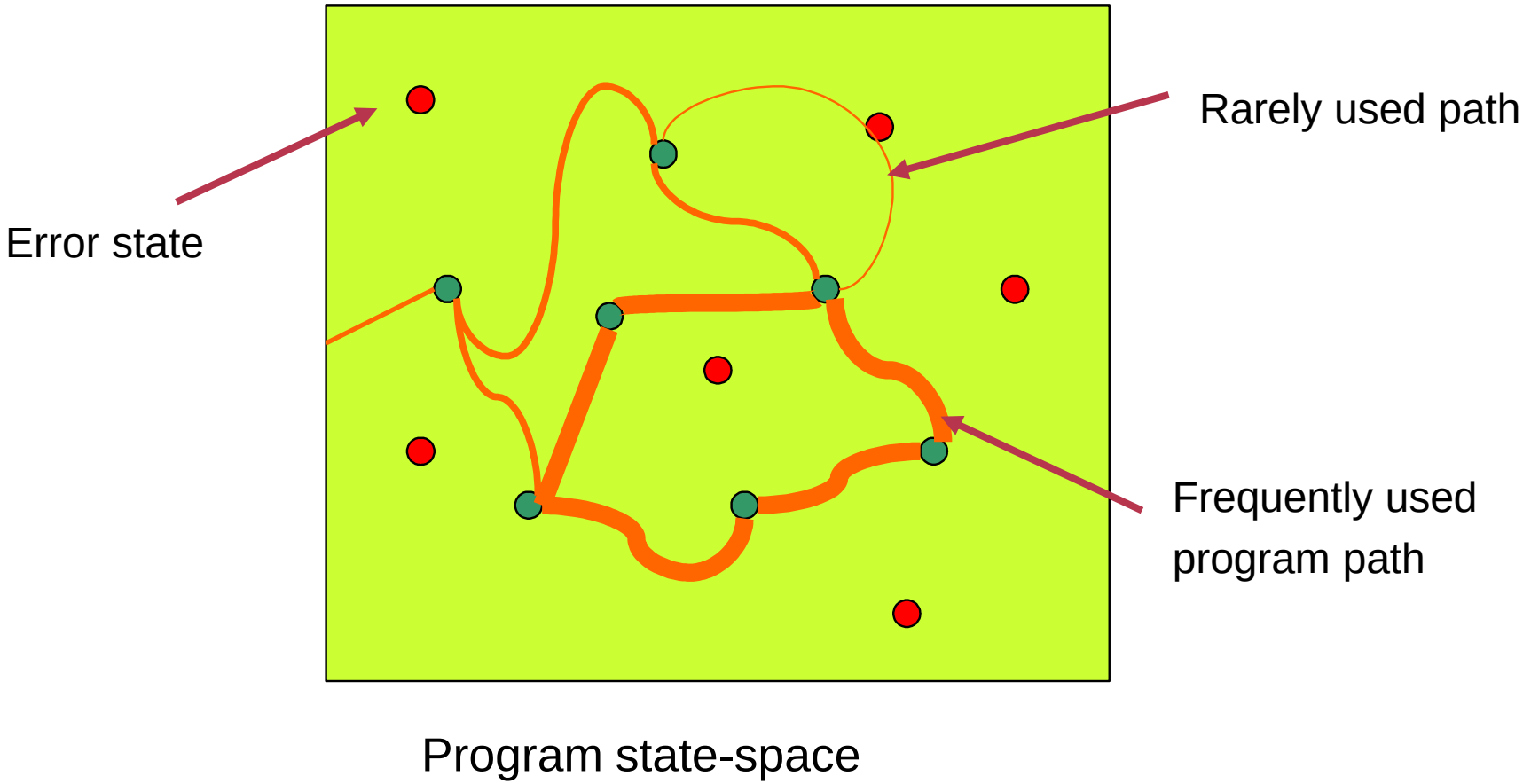
Possible Objections

- Possible objections to this exercise in evaluating the performance of the SIL concept in practice include:
  - Is it valid to combine data from various different systems? (yes, electronic component failure rates are based on data from many different sources)
  - 1000 units with the same software are only provide evidence of 1 program, not 1000 (but it is 1 program running in 1000 different situations, see ‘Activation of Defects’)
  - It is not valid to speak of rates or probabilities in connection with systematic failures (see following slides)

Random or Systematic

- “Random” failures are determined by the laws of physics and chemistry, but we don’t know when they will happen
- “Systematic” failures are determined by design factors, but we don’t know when they will happen either
- The difference is in the defences we use

Activation of Defects

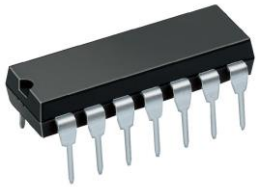


Systematic Errors Cause Random Failures

Defect

Triggering Condition

Failure



Manufacturing Defect

```
LD R1
JMPC RESET
LD PRESS_1
ST MAX_PRESS
RESET: LD 0
ST A_X43
```

Software Error

Environment
e.g. temperature,
Voltage, etc

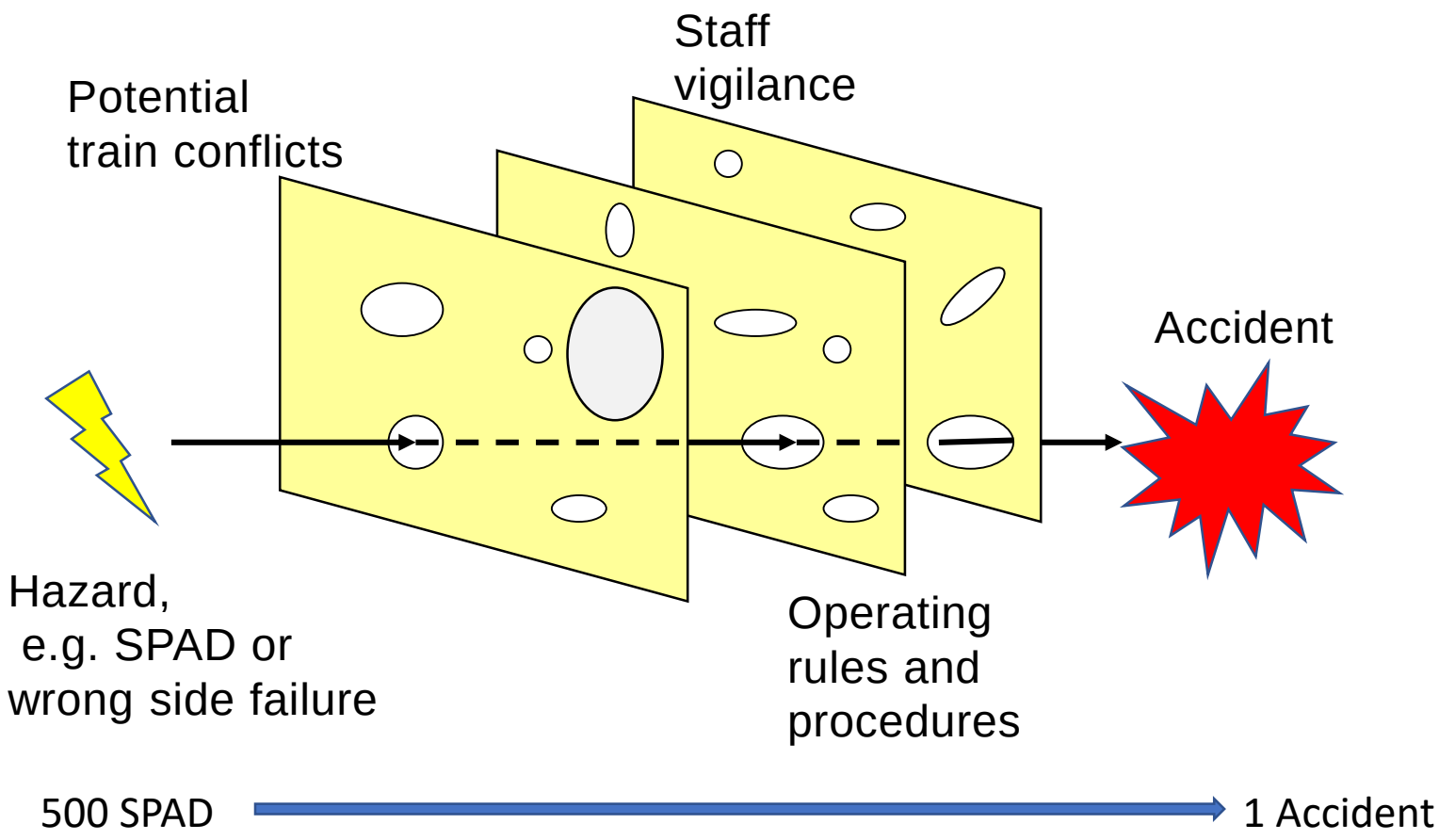
Execution of rarely used program path

Accident Data

- Failure data is not generally available
- Accidents are very public
- No accidents have been caused by failure of SIL4
- Equivalent to accident rate per equipment hour better than 1.4×10^{-9}

This result was obtained by using accidents as a surrogate for unsafe failures and making F = 1

The Swiss Cheese of Signalling Failure



Reverse Engineering the Swiss Cheese

The value of F can be deduced by multiplying the number of accidents (assumed to be 1) by the unsafe failure/accident ratio for SPADs,

$$10^3 \geq \lambda_{uff} / \lambda_{acc} \geq 10^2$$

This gives a figure for Unsafe Functional Failure Rate in the range

$$10^{-6} \geq \lambda_{uff} \geq 10^{-7}$$

Which is higher than the SIL4 equivalent figure in the Standards, but the true figure is almost certainly lower due to the systems for which no data was found.

What is needed to make SIL4 a fact

IF

- By including data for all types of interlocking in service, the number of equipment hours due to interlockings is doubled.
- There have been 25 times as many remote terminal units as interlockings in operation over the same lifespan as the interlockings.
- There have been 25 times as many axle counter evaluators as interlockings in operation over the same lifespan as the interlockings.

THEN the unsafe functional failure rate achieved to date

$$= 1 \times 10^{-8}$$

which equals the upper bound of the range for SIL4

Conclusion

Not enough data was found to convert into a **fact** the **opinion** that SIL4 measures should result in an unsafe functional failure rate better than 10^{-8}

However, there are grounds for optimism that if more data about existing systems were made available it would become a **fact**

The author hopes that the publication of this paper will provoke people into supplying data which will enable the analysis started here to be made complete.