UCLA ENGINEERING
B. John Garrick Institute for the Risk Sciences

Assessing Safety, Reliability, and Security Risks of Autonomous Systems

Ali Mosleh
Distinguished Professor
Evelyn Knight Chair in Engineering
Director, The B. John Garrick Institute for the Risk Sciences
University of California, Los Angeles

First International Workshop on Autonomous Systems Safety
Trondheim, Norway, March 11-13, 2019

Risk

- Risk is usually associated with the uncertainty and undesirability of a potential situation or event

Risk = Uncertainty and Undesirability

- Metrics of Risk

Risk = Likelihood and Severity

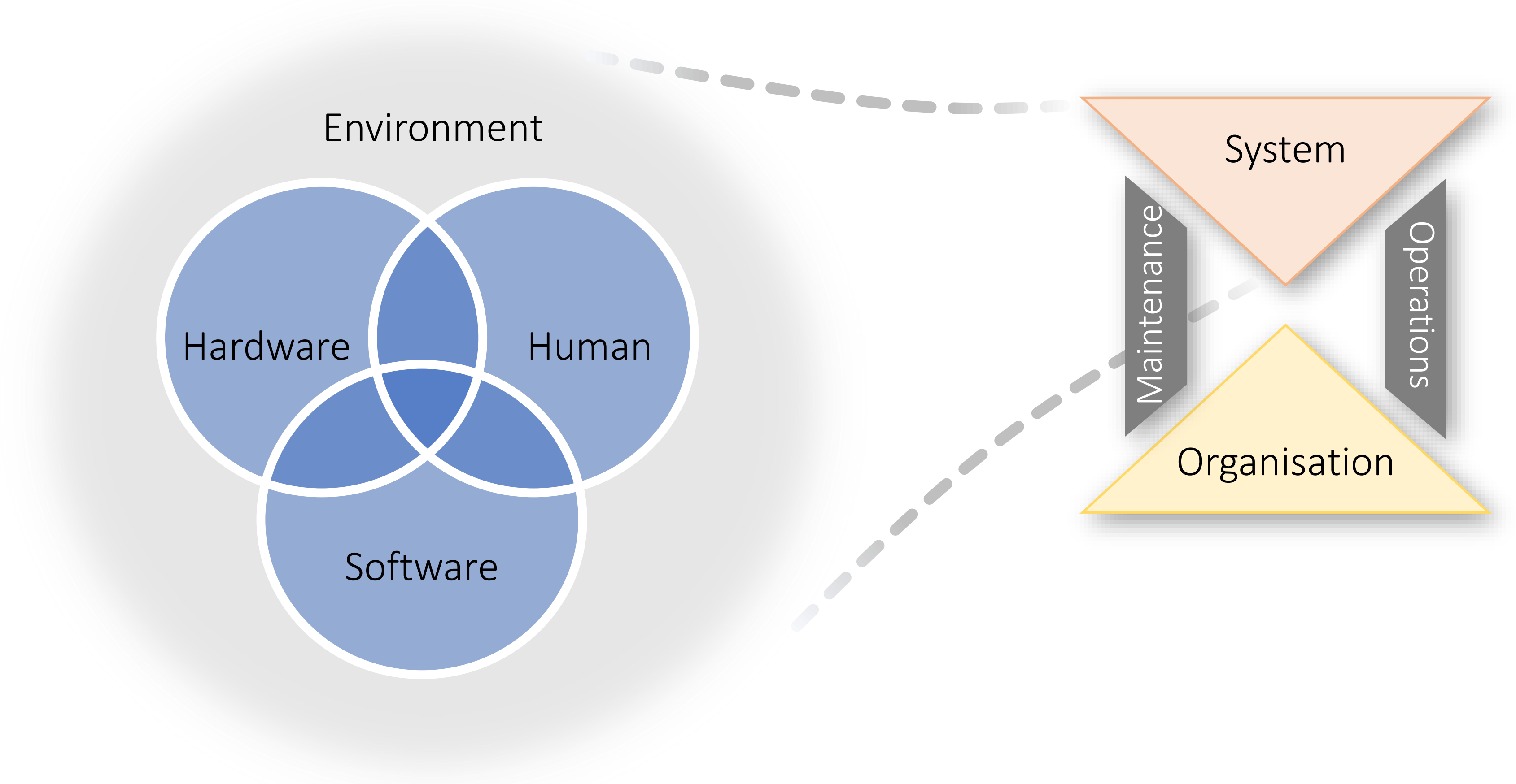
- Safety, Reliability, Security, Environmental, Economic, ….

Risk Analysis
(Safety, Reliability, Environmental, Financial, Security)

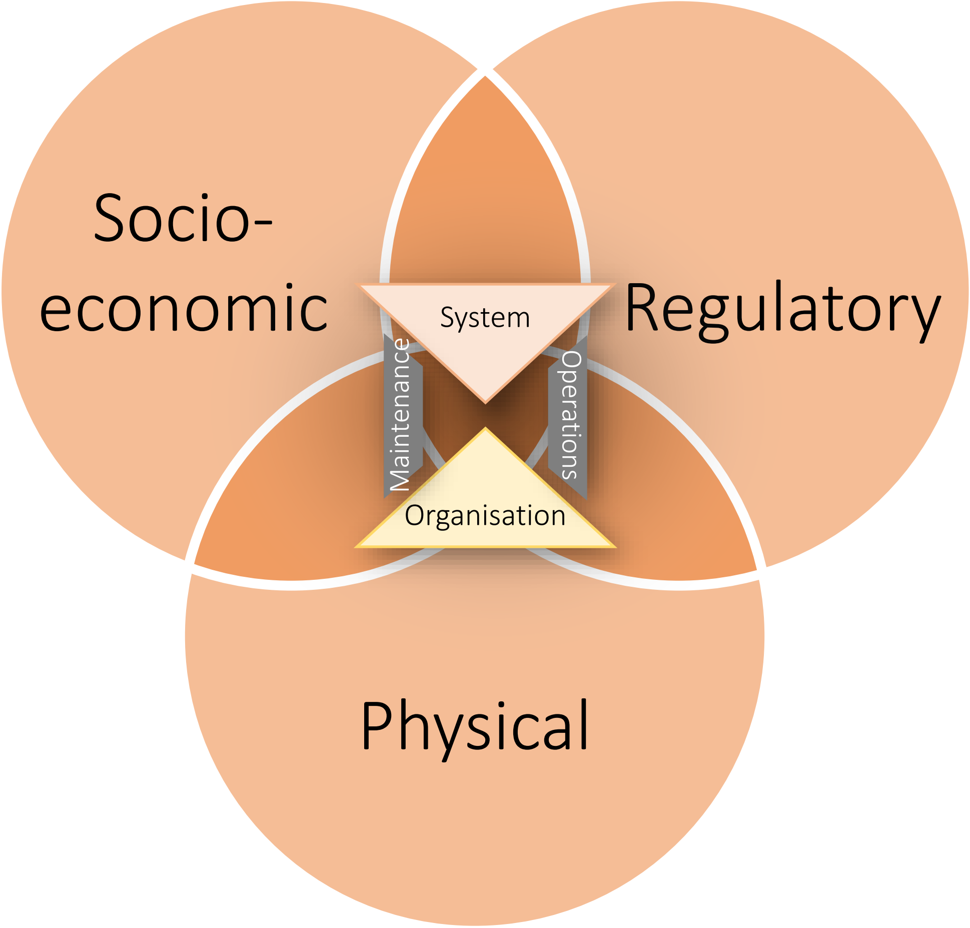
- Determine potential undesirable consequences associated with use of systems and processes
- Identify ways that such consequences could materialize
- Estimate the likelihood (e.g., probability) of such events
- Provide input to decision makers on optimal strategies to reduce the levels of risk

- ❑ What can go wrong?
- ❑ What are the consequences?
- ❑ What is the likelihood?

Engineered Systems



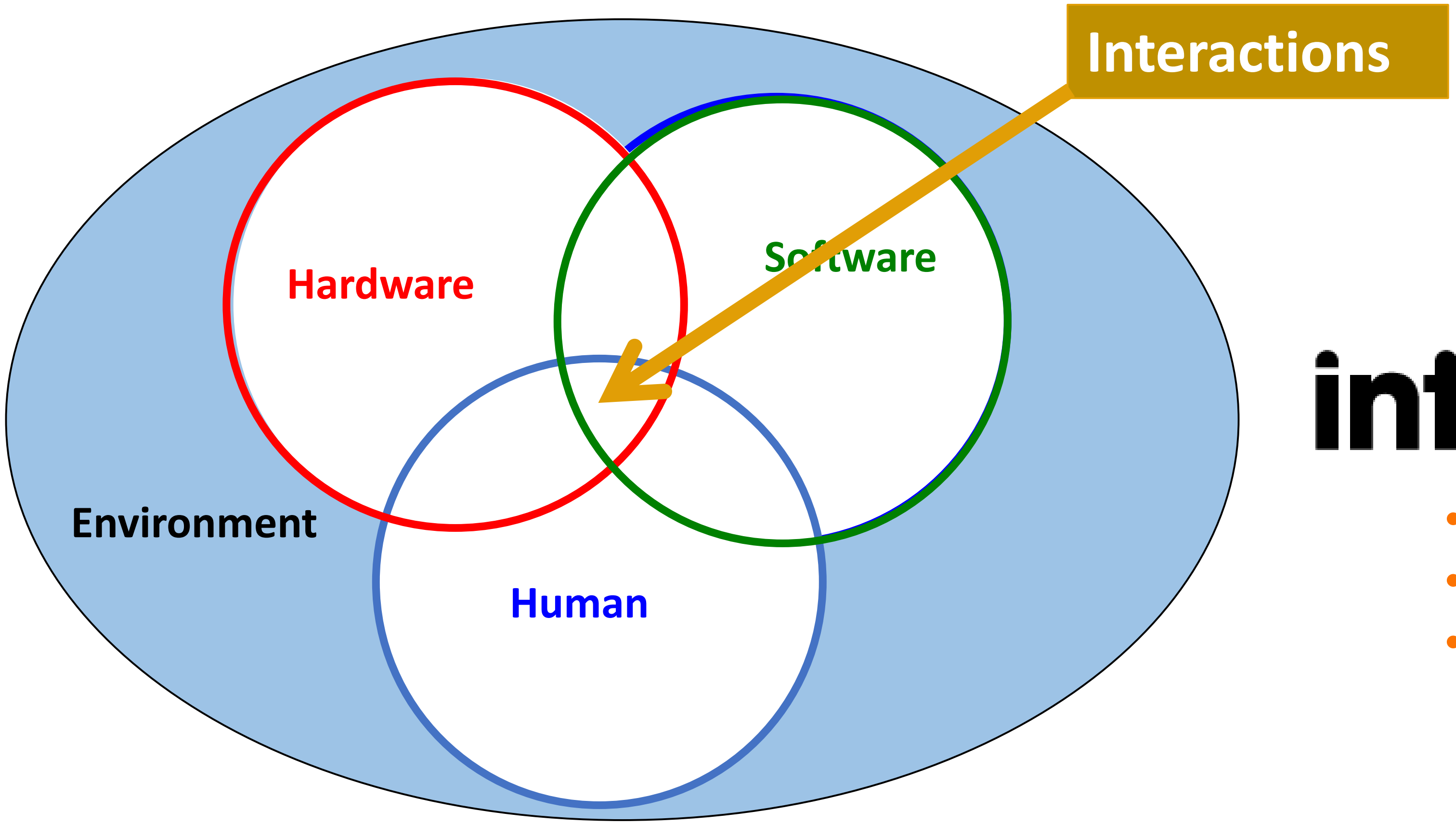
Engineered Systems

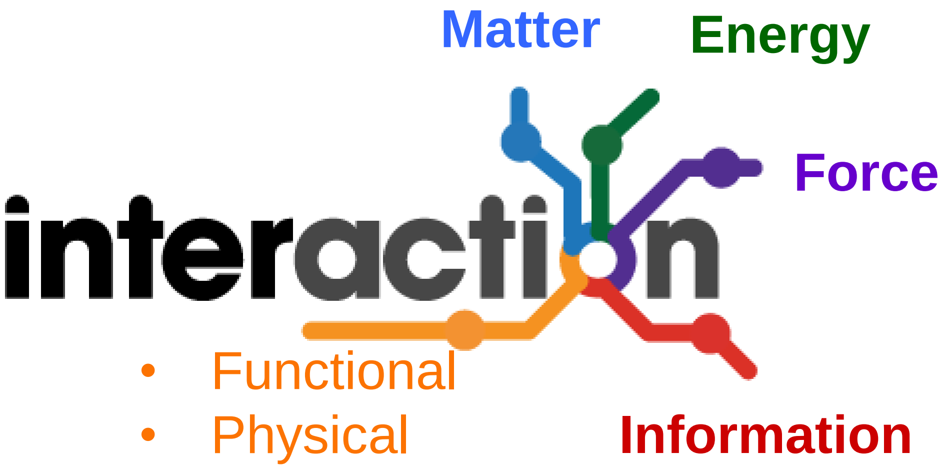


Cyber–Physical-Human (CPH) Systems

- Ultra complex, heterogeneous, distributed, open, possibly "learning systems"
- High levels of integration of the technical and social dimensions (highly interconnected socio-technical systems)
- Very high pace of development and deployment
- Higher levels of diversity of supply chain, subject to different levels of quality, reliability, and safety standards

Challenge ….





- Functional
- Physical
- Environmental

Failures of X-Ware Systems

Mars Polar Lander Crash on Mars

interaction

- Functional

Information

CRH D310 rear-ended CRH D3115 in 2011, China,
35 died, 211 injured

Matter

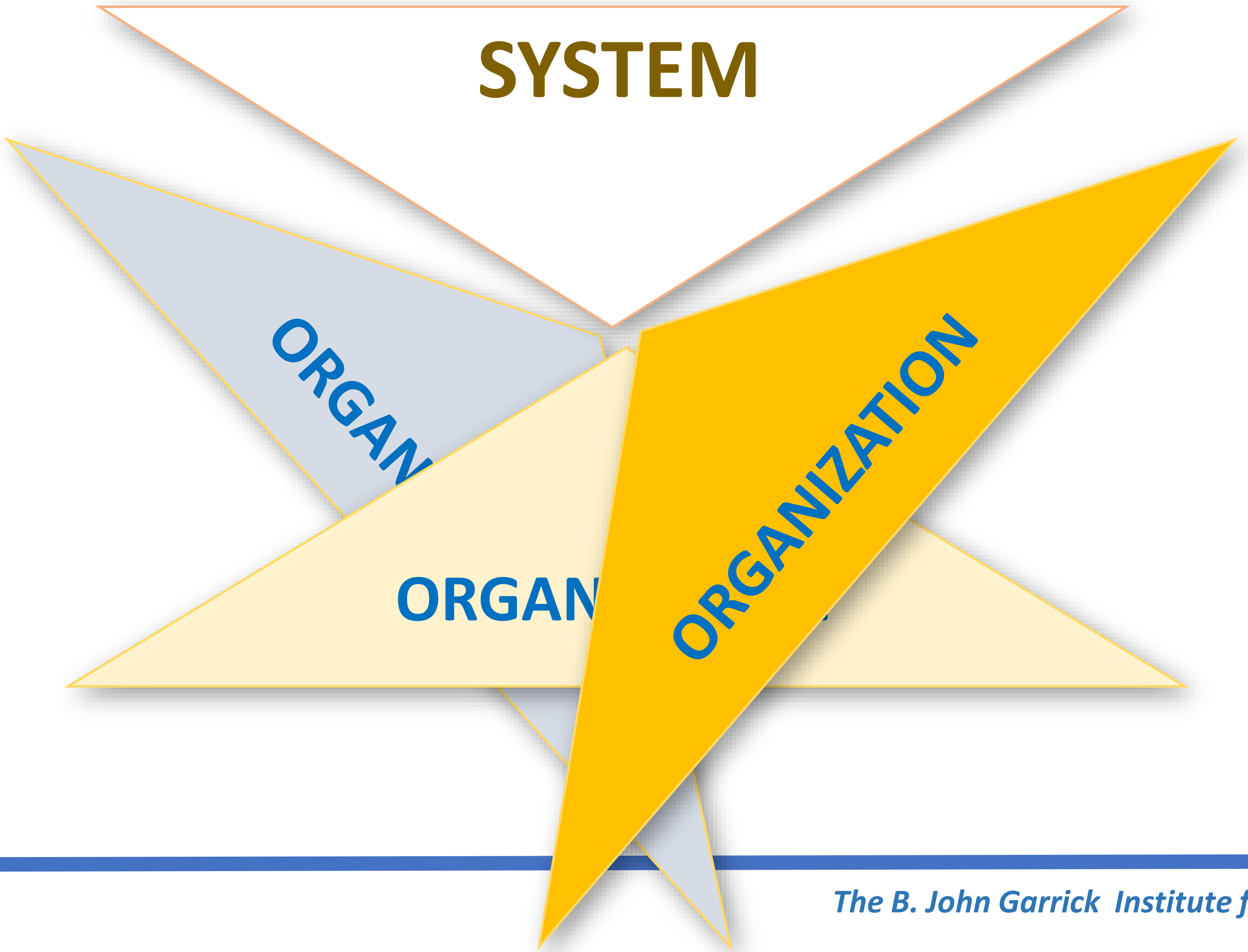
Force

interaction

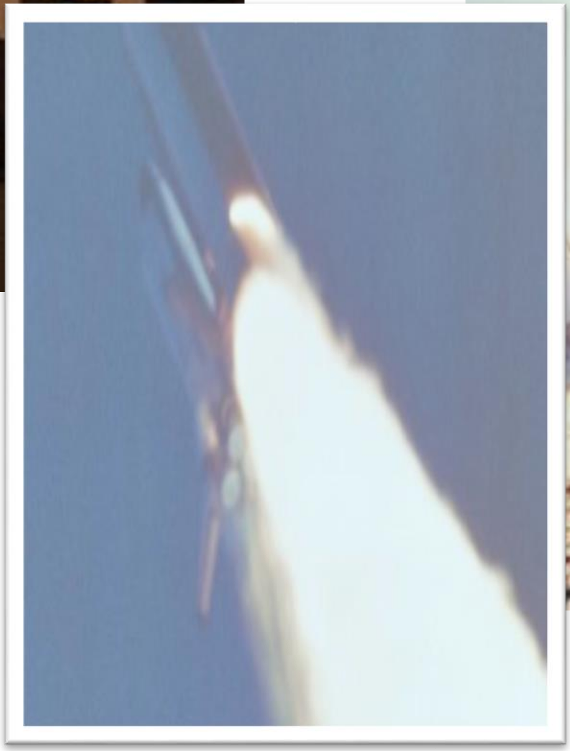
- Functional
- Environmental

Information

Organization Interface Failure



Organization Interface Failure



System Level CPH Failures

- Propagation of Failure
- Conflicts: lack of coordination of elements' behaviors
- Failure Masking: suppression of behavioral deviations

TUMBLING JUMBO

- ❑ During a flight, a China Airlines B-747 experienced a flame-out of one of the engines
- ❑ The crew failed to notice the problem, since the autopilot software was compensating for the resulting thrust imbalance
- ❑ The compensating actions kept the plane in a stable, yet abnormal state
  - The autopilot now played a critical role in the plane’s stability
- ❑ The crew finally detected the problem
- ❑ They tried to take control of the plane, by switching off the autopilot
- ❑ The plane immediately became unstable, and started to tumble

Autonomy

«A system's or sub-system's own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve its goals as assigned by its human operator(s) through designed human-machine interface (HMI)"

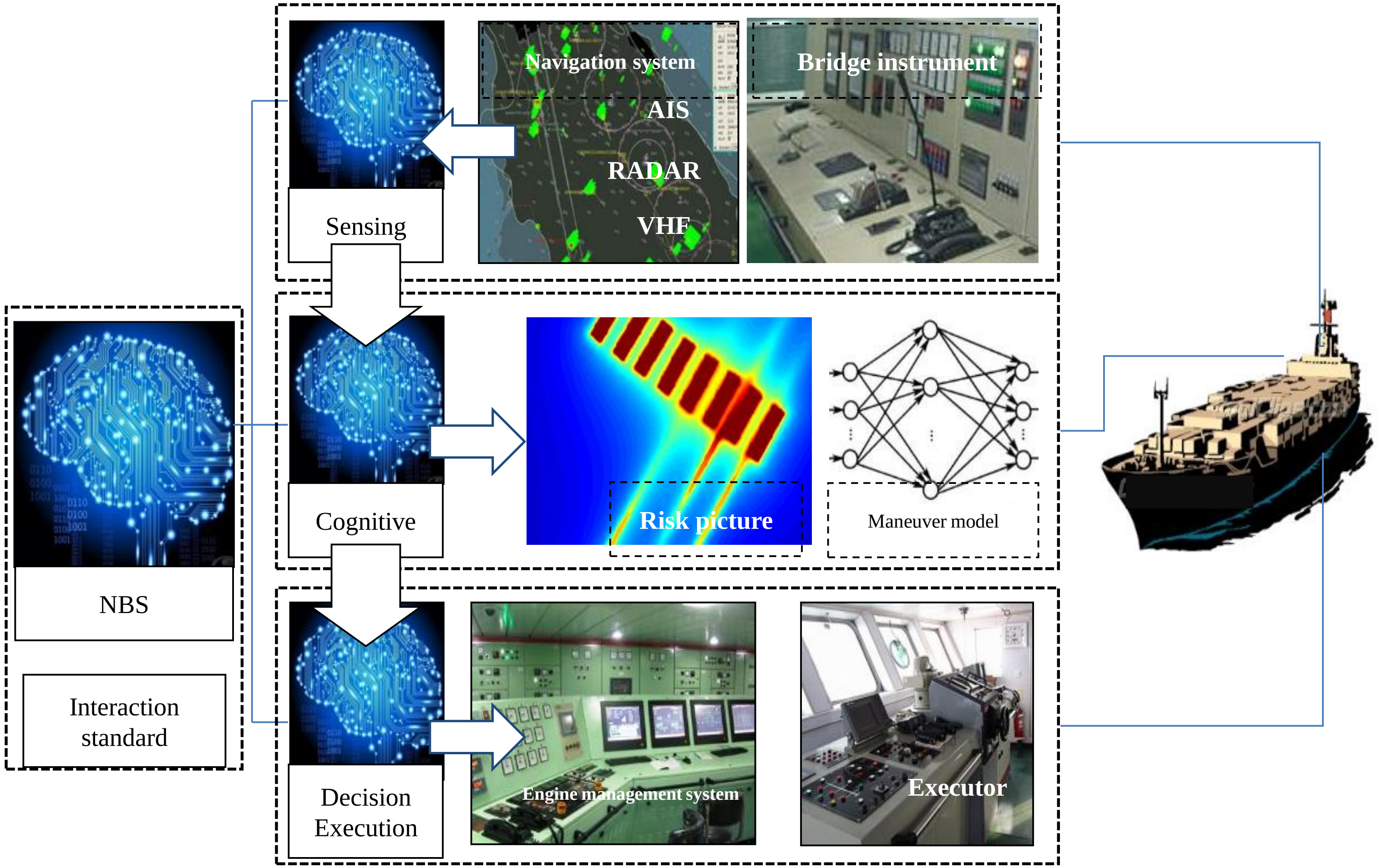
Level of Autonomy	Description
1	Fully manual control
2	The computer offers a complete set of decision/action alternatives.
3	The computer narrows alternatives down to a few
4	The computer suggests one alternative
5	The computer executes that suggestion if the human approves
6	The computer allows the human a restricted time to veto before automatic execution
7	The computer executes automatically, then necessarily informs the human
8	The computer informs the human only if asked
9	The computer informs the human only if it decides to
10	Fully autonomous Control

IBM Watson in Charge at ISS

Navigation Brain System

国家水运安全工程技术研究中心
National Engineering Research Center for Water Transportation Safety

- **"Sensing"：**data collection of ship condition and nearby navigation environment
- **"Perceiving"：** Identification of navigation related data to evaluation navigational safety situation
- **"Decision & Execution"：** make the optimum decision and execute using ship control system, and give feedback to "Sensing"



Human Still in the Loop !

- One of the inherent characteristics of all engineered system is the inevitability of interface with humans; in design, in operation, in intended use, and unintended effects.

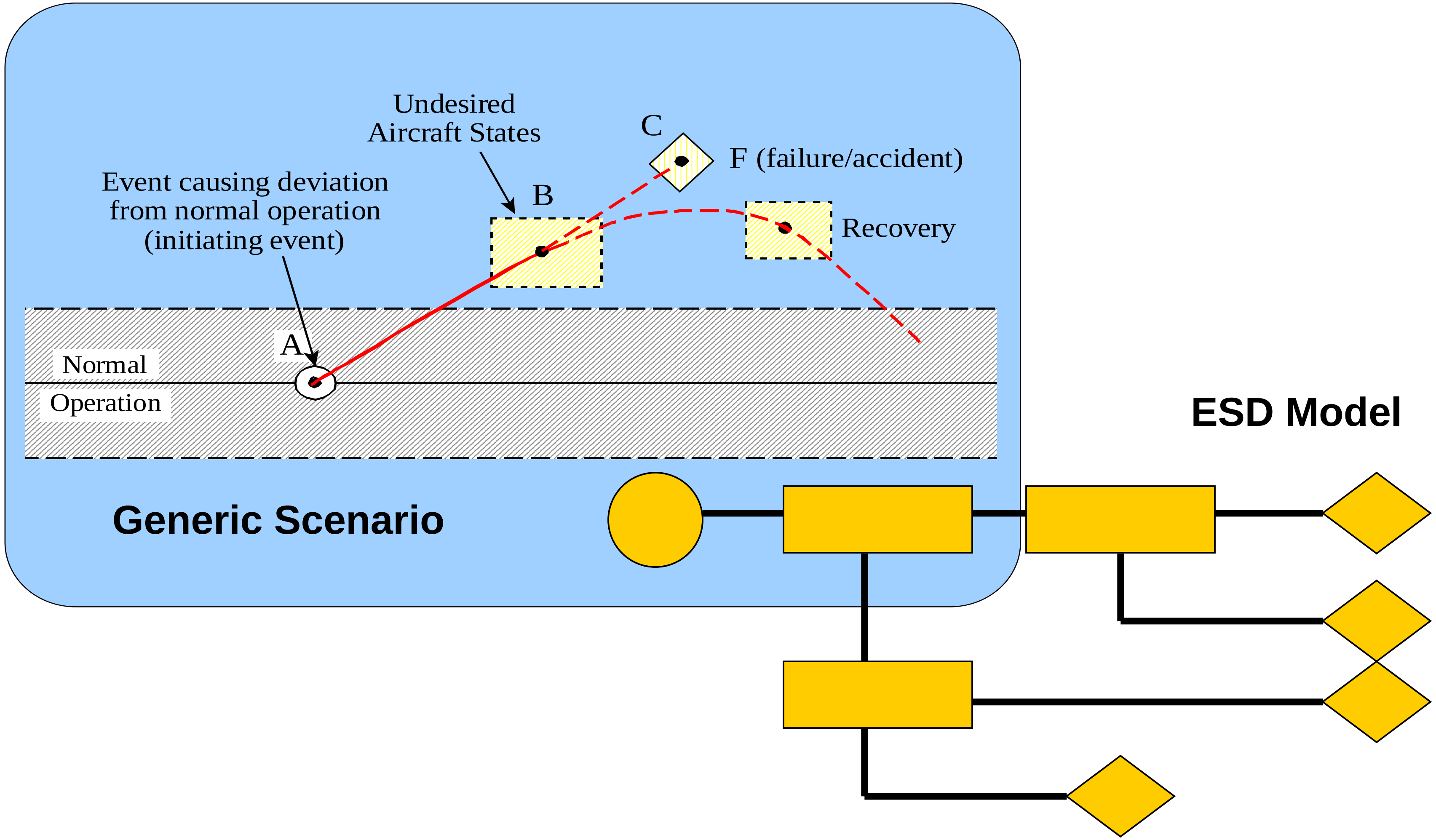
- Autonomous systems are not immune, even though one of the main motivations and the core design feature of such systems is to eliminate or reduce the need for human operators.

Questions

- How to make the case for autonomous system Safety, Reliability, and Security (SRS)
- Modeling and analysis methods for assessing autonomous systems SRS
- Human in the loop, risks and benefits
- Dealing with complexity of integrated systems of Software – Hardware – Human
- Safety standards, oversight, regulations, and liability

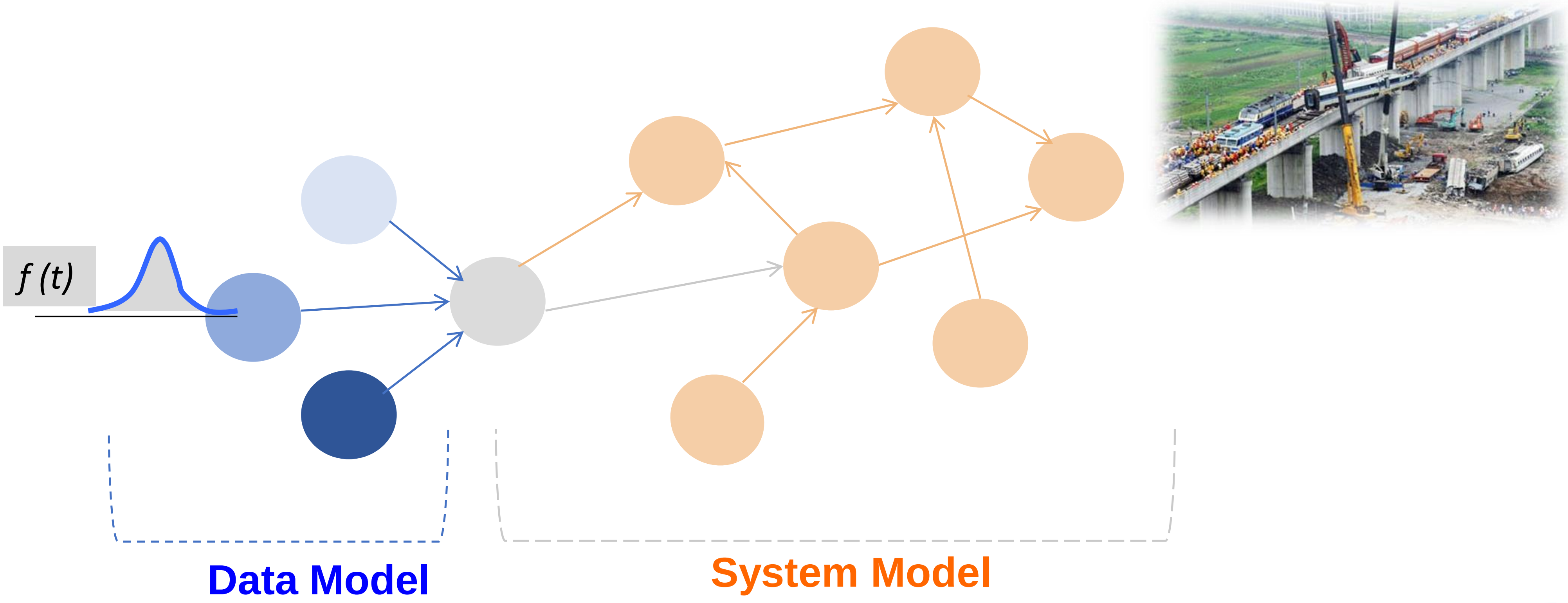
On Modeling Approaches

Modeling Scenarios: The ESD Methodology

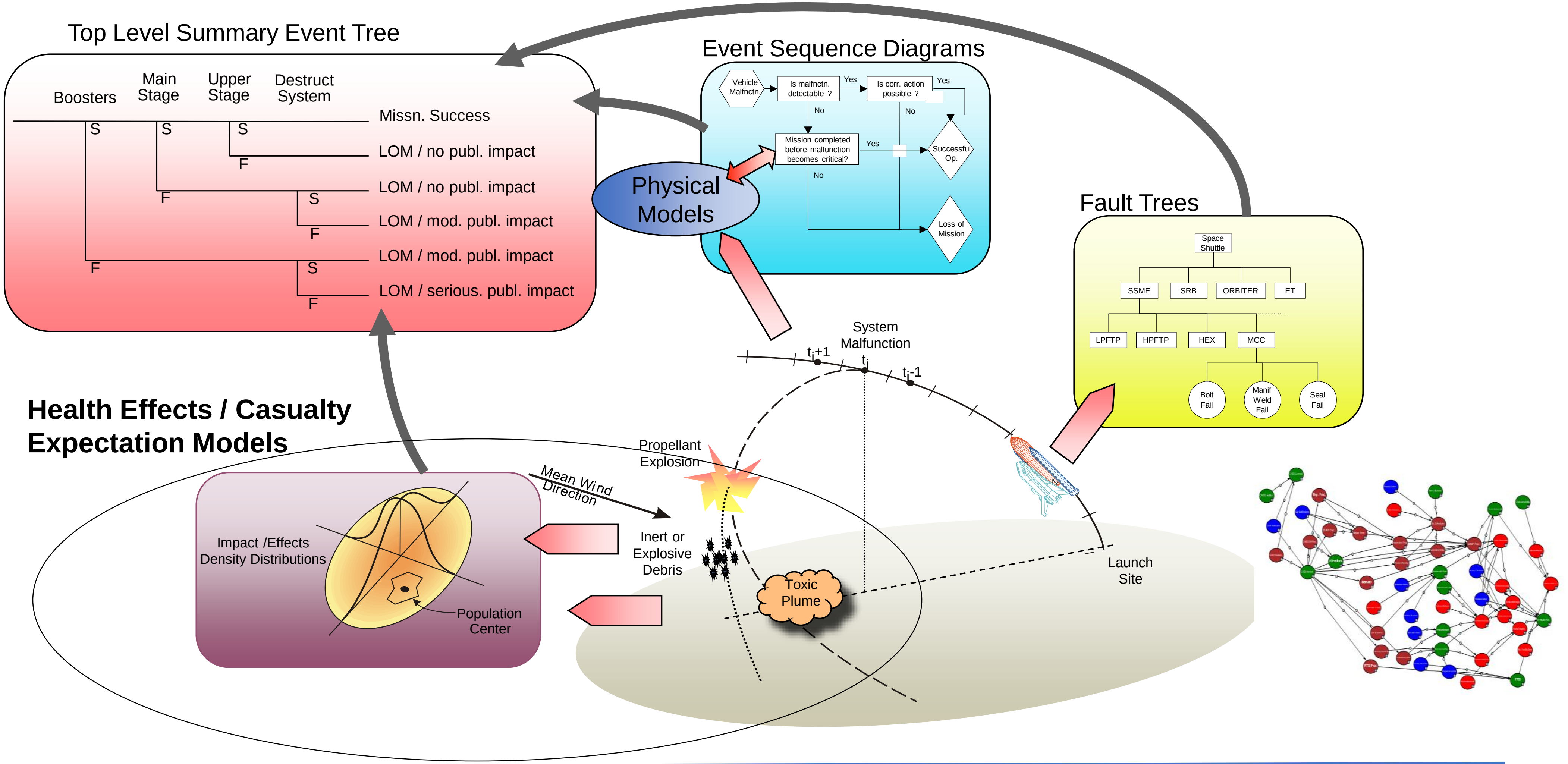


Modeling with Bayesian Network

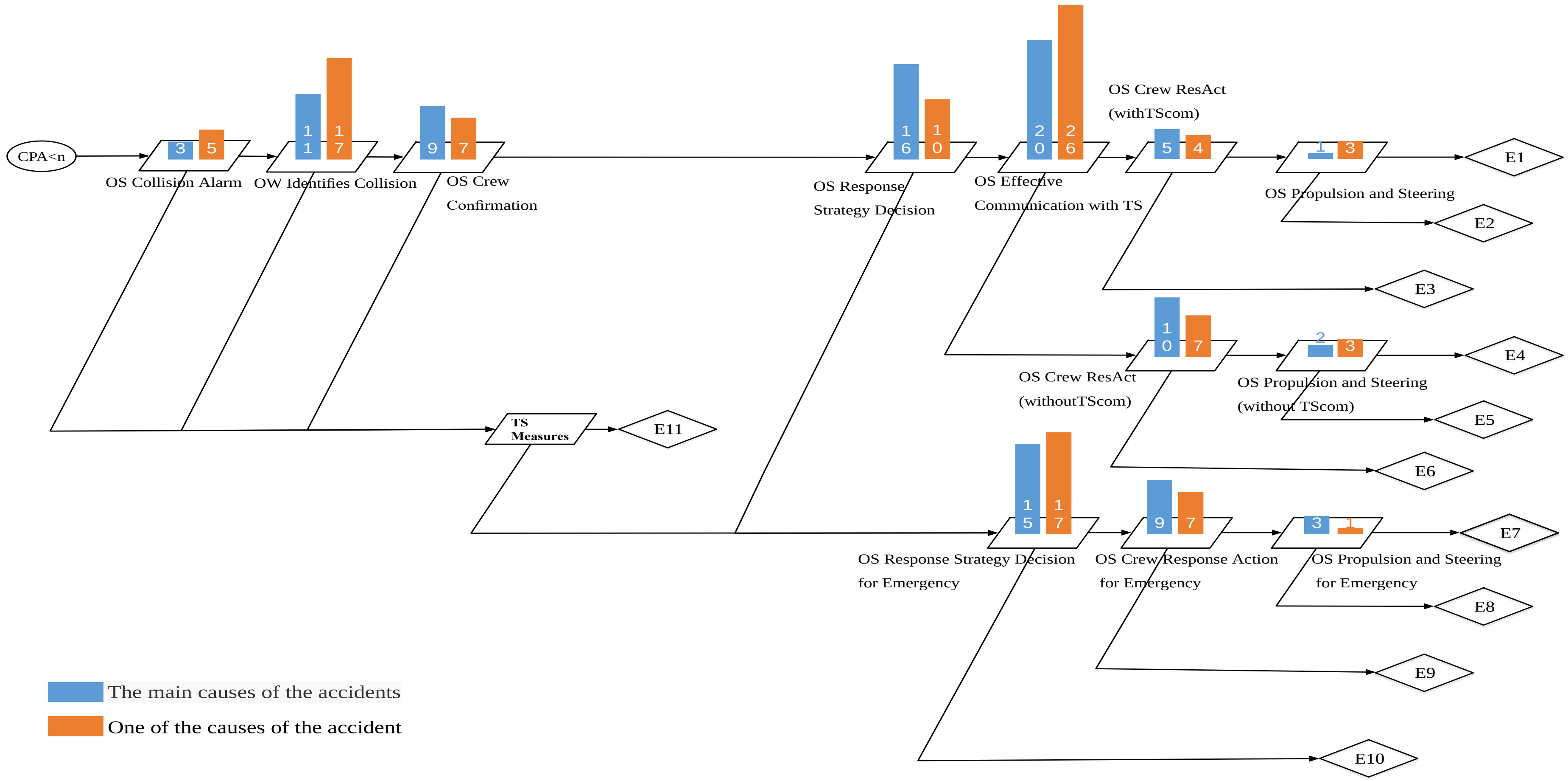
Compact and seamless integration of the ***data model*** and ***System model***



Hybrids (Mixing Phenomenological and Logic Based Models)

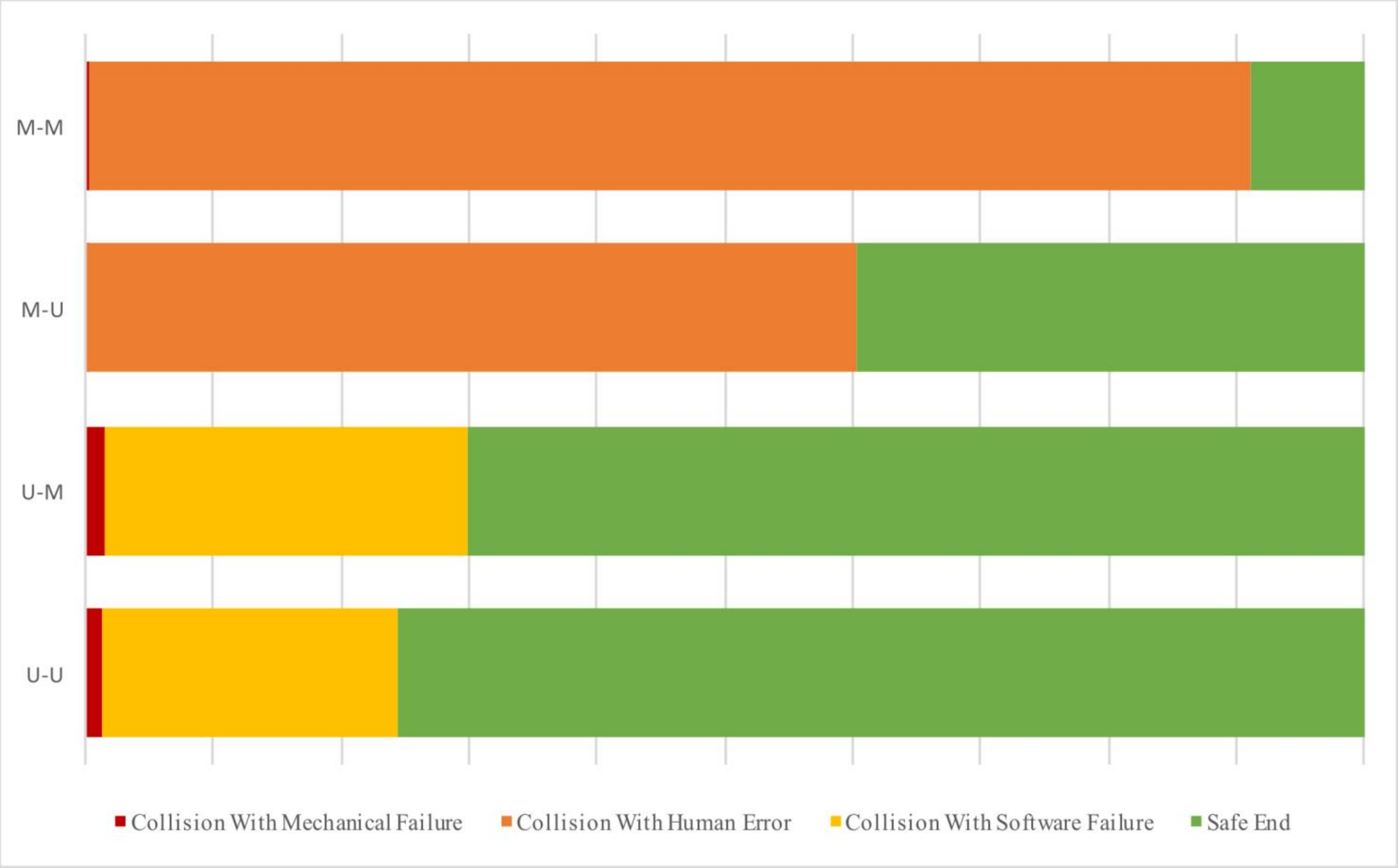


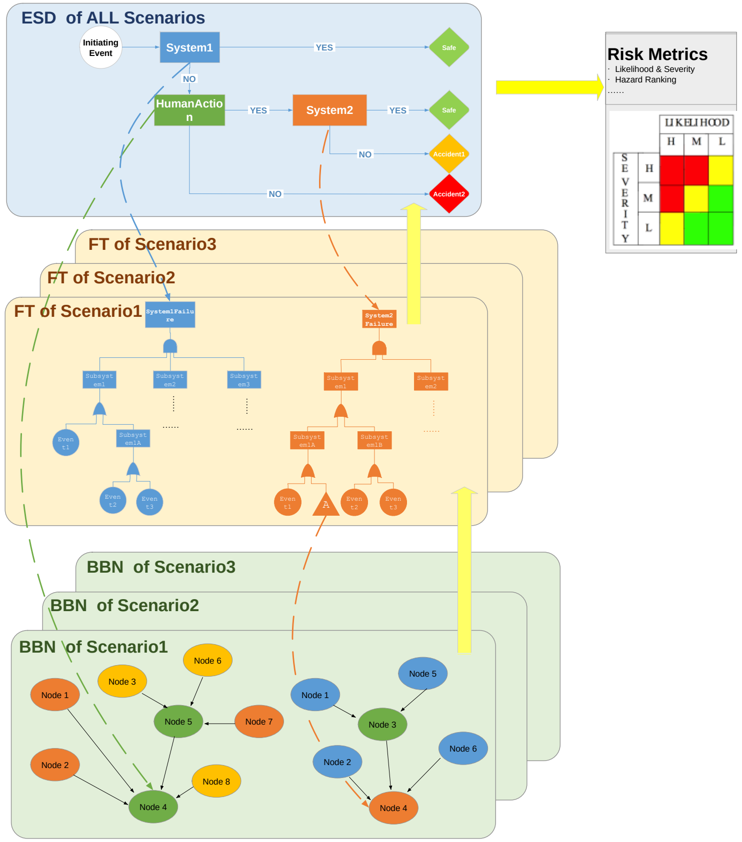
Value of Simple, Highly Abstracted, Models



Tengfei Wang et al , “**A comparative assessment of collision risk of manned and unmanned vessels” Submitted to RESS**

Comparative Assessment of Generic Collision Risk of Manned and Unmanned Vessels*





Tengfei Wang et al , "**A comparative assessment of collision risk of manned and unmanned vessels" Submitted to RESS**

Software Reliability

"Software never fails; it does exactly what it was coded to do."

Why is the number 32 768 important?

Ariane 5 rocket

first launched in 1996 by the European Space Agency (ESA)

expendable launch system (i.e. no crew)

heavy reliance on software

https://www.youtube.com/watch?v=gp_D8r-2hwk

Why is the number 32 768 important?

the Ariane 5's control software converted 64-bit floating point values to 16-bit signed integers

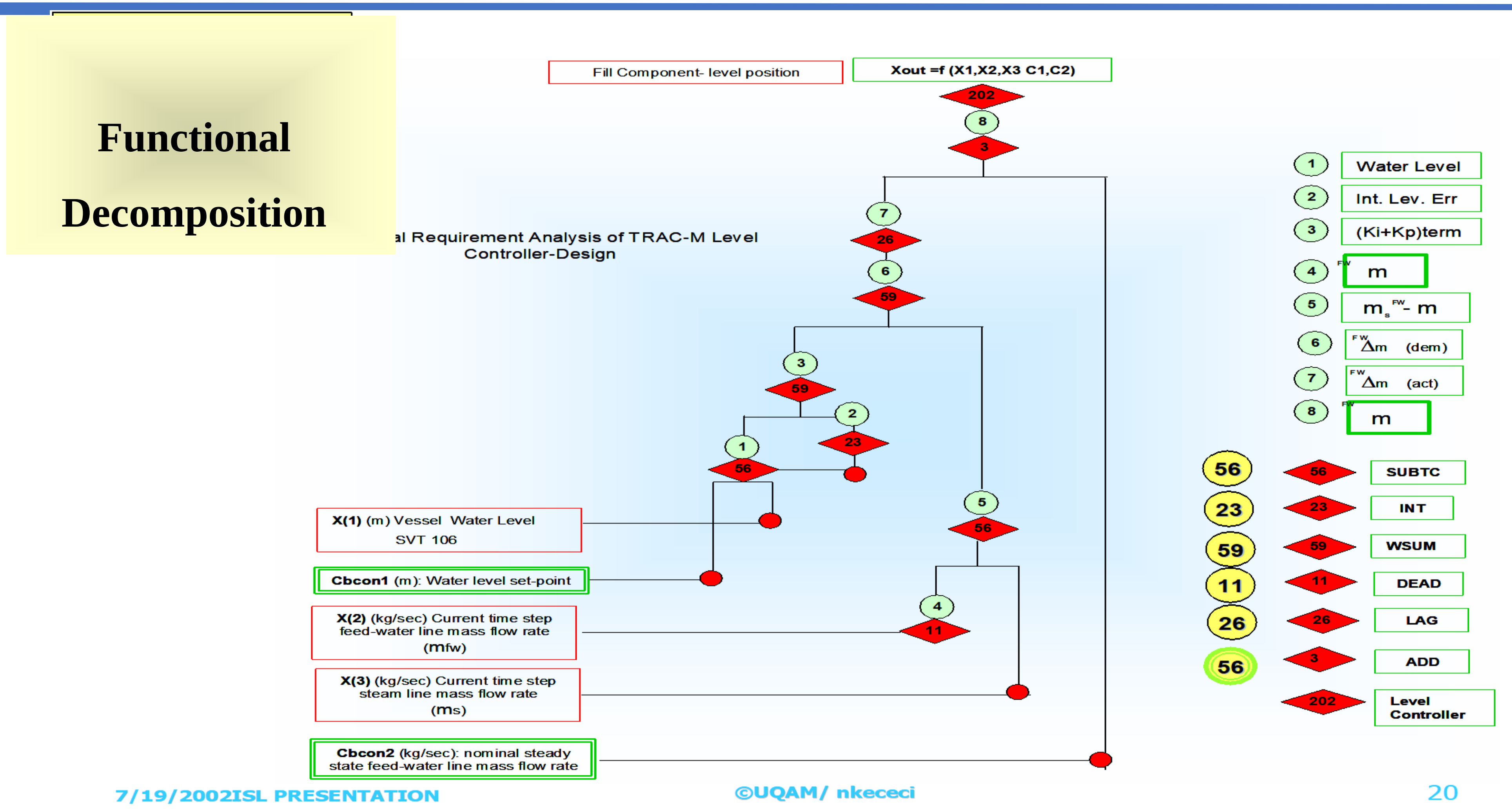
… the maximum value for a 16-bit signed integer is 32 768

What Happened ?

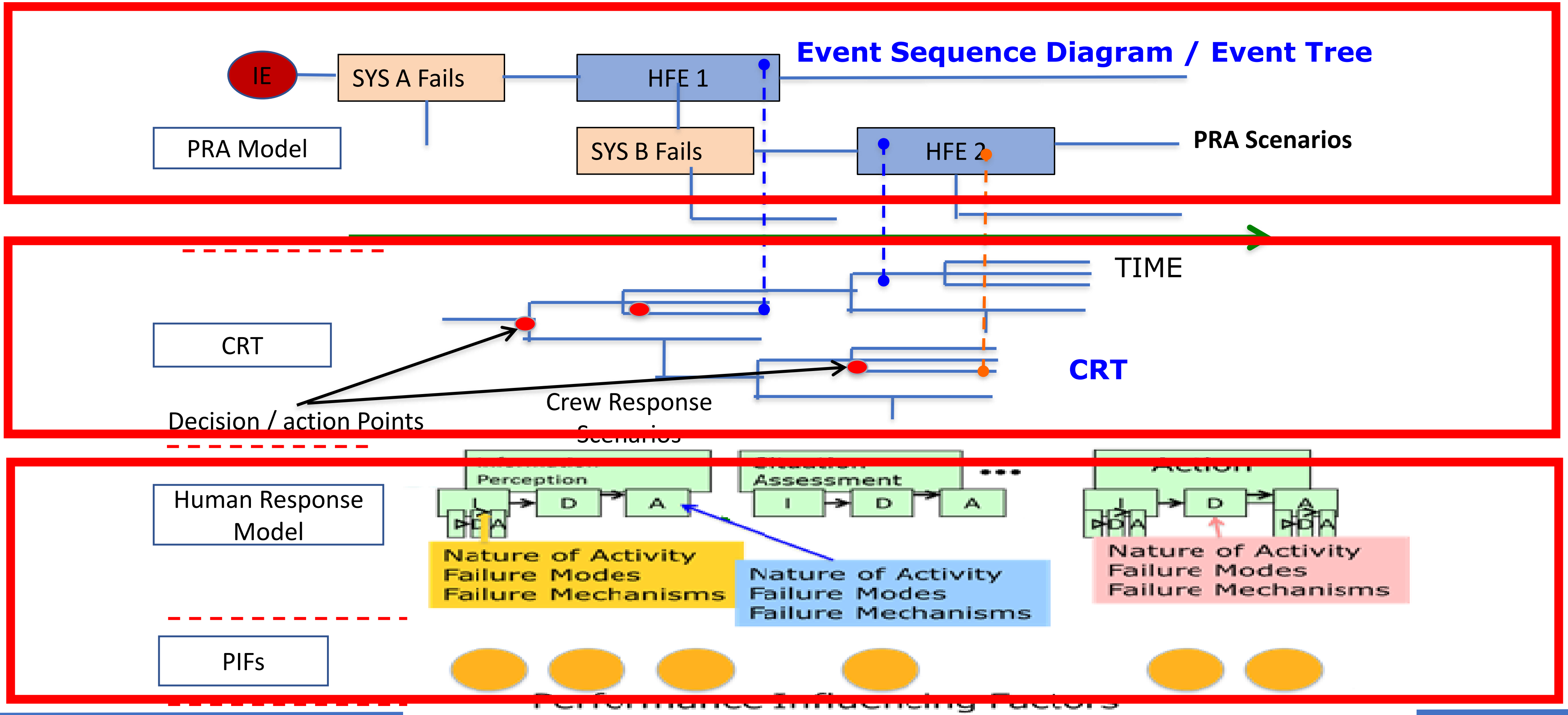
- Control software was responsible for handling the 'horizontal bias' variable …
- … **which was left unprotected by a handler because it believed the rocket physically limited the value**.
- When the number exceeded 32768, the software reset the field to 0
- The rocket self-destructed believing it to be 90 degrees misaligned

the 1996 launch was Ariane 5's first

Software Failure Modeling

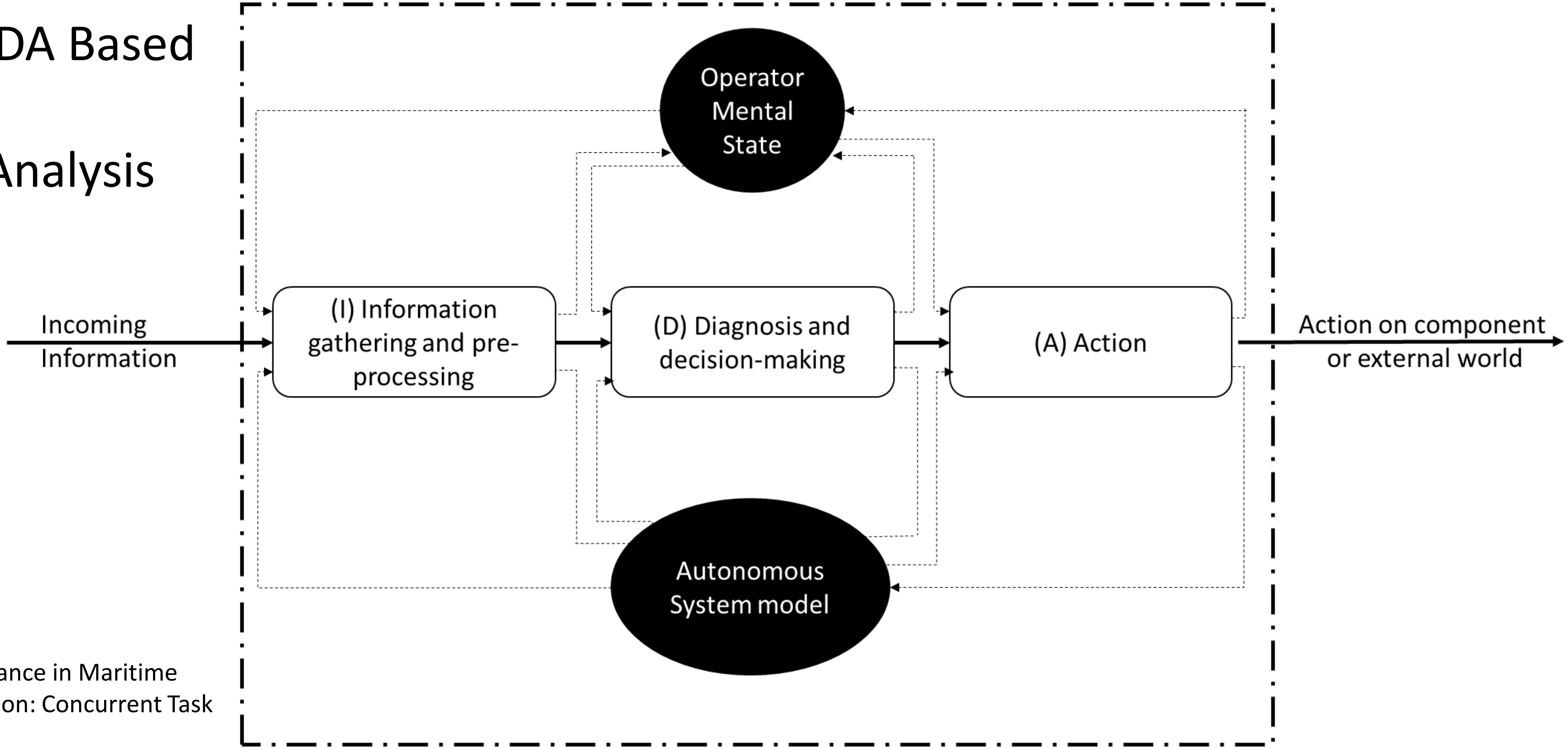


Context Analysis – Crew Response Tree



Concurrent Task Analysis*

- HRA Inspired IDA Based Model
- Parallel Tasks Analysis



* Ramos, M. et al, “Collision avoidance in Maritime Autonomous Surface Ships operation: Concurrent Task Analysis” This Confrence

Characterization of Interactions

- ❑Continuous/Discrete,
- ❑Dynamic/Static
- ❑Routine/Opportunistic (e.g., under abnormal conditions, accidents, emergencies only
- ❑Single/ Multiple (redundant/diverse) channels per interface function
- ❑Designed (or planned)/Ad-hoc interface

Characterization of Interactions

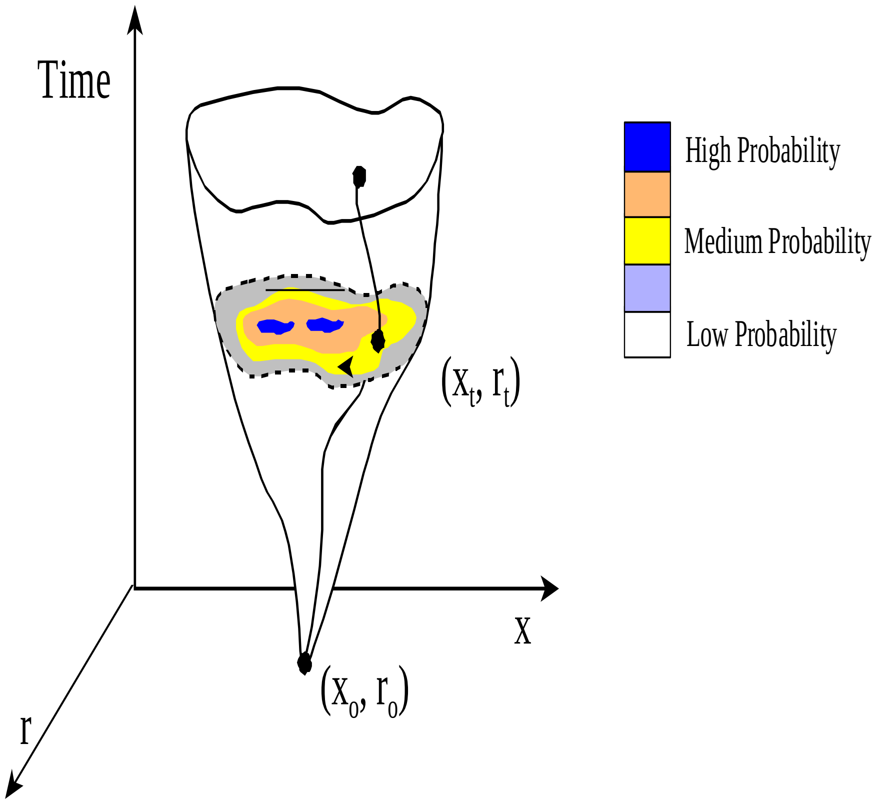
- Monitored / Unmonitored Interface
- Real time / Time-lagged Interface
- Critical/Noncritical to mission of at least one organization
- Manual / Automated
- Physical/Virtual
- Information/Mass/Energy

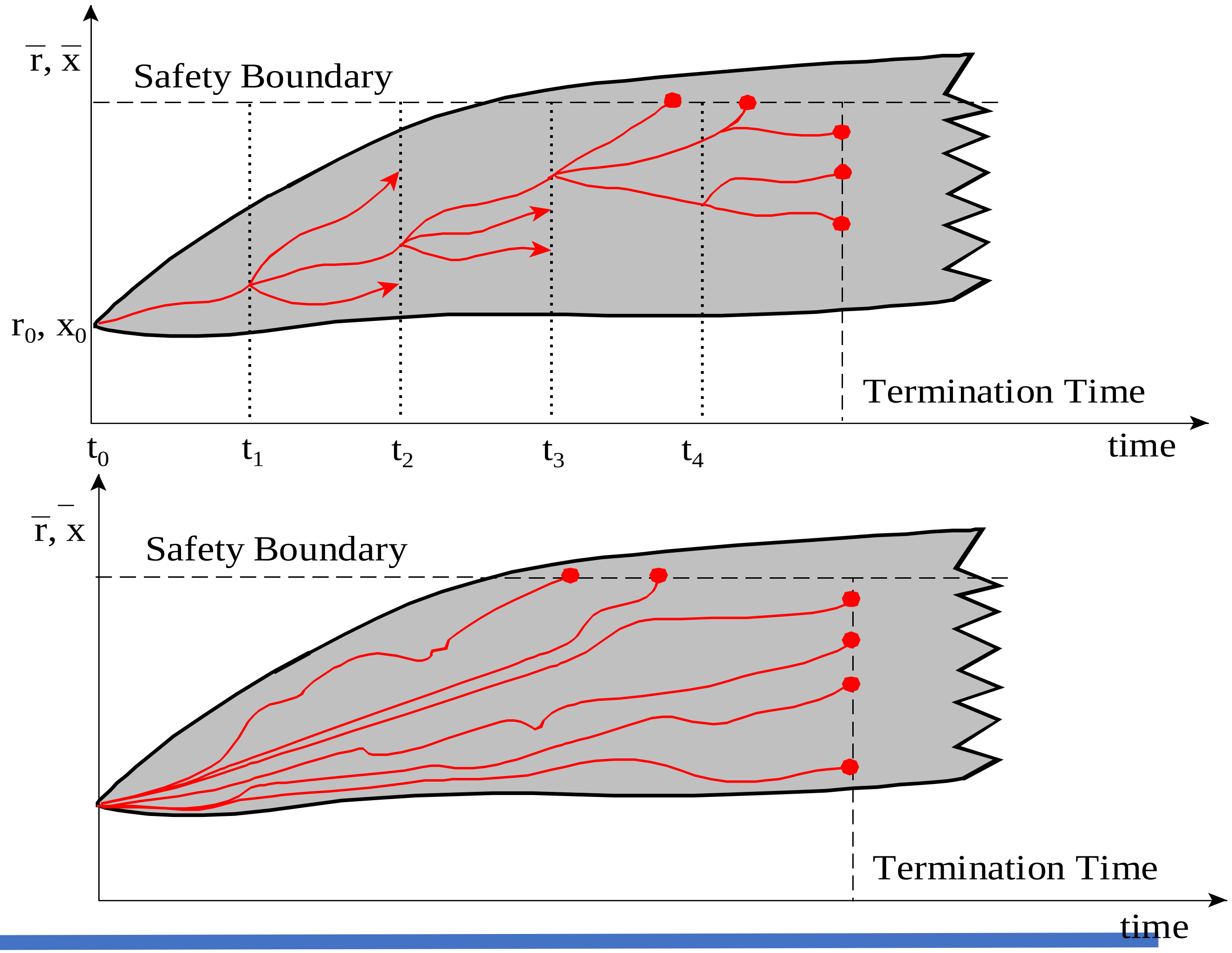
COMPLEXITIES

- Complexities due to nature of failure events
  - Systems involve hardware, software, and human, exhibiting distinct behaviors
  - Complex failure scenarios arise due to interactions of different elements
- Complexities due to the time behavior of the system
- Dimensions in which such complexities need to be addressed:
  - Representational
  - Computational

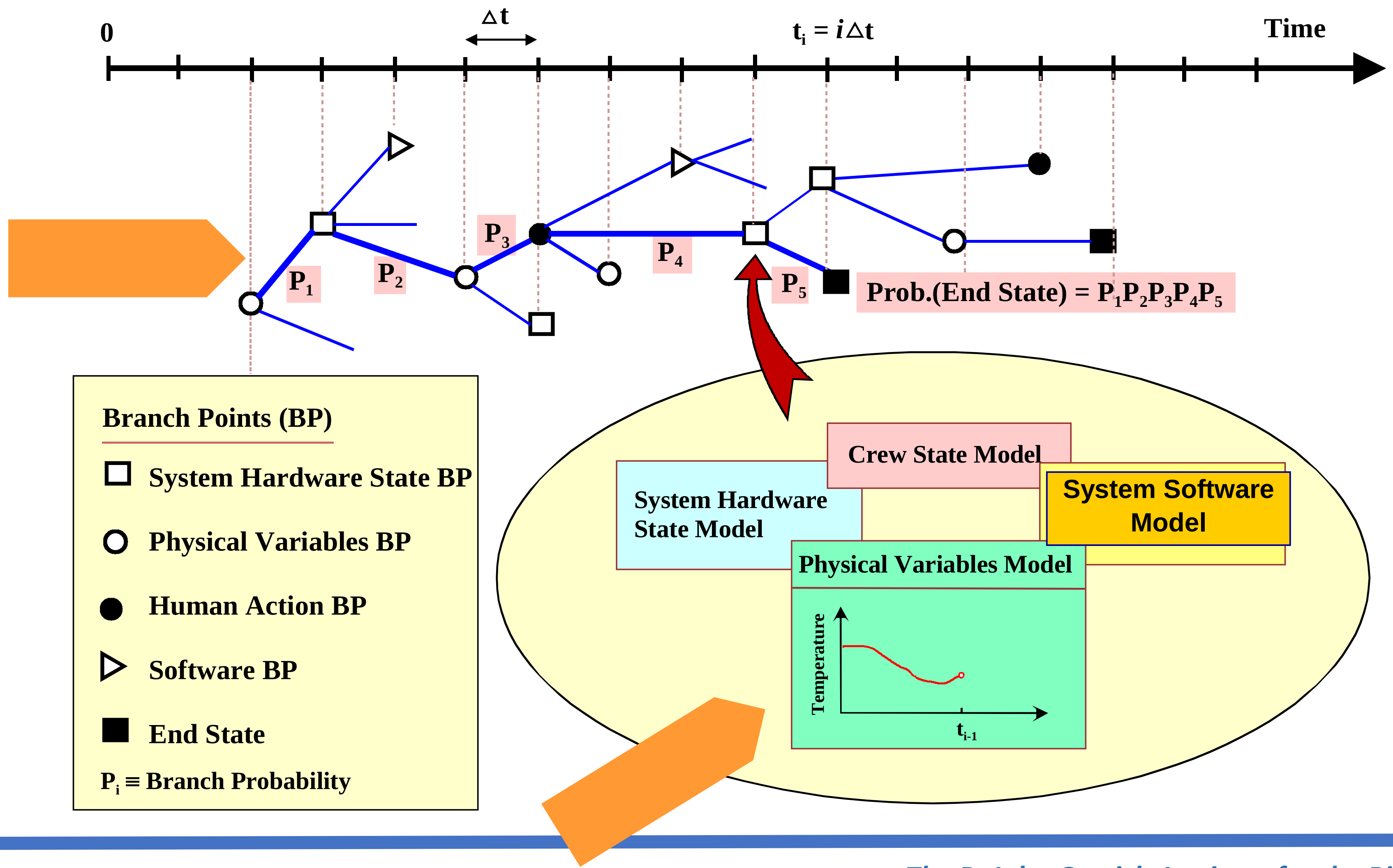
More Realism, Probabilistic Simulation

Framework and Solution Methods

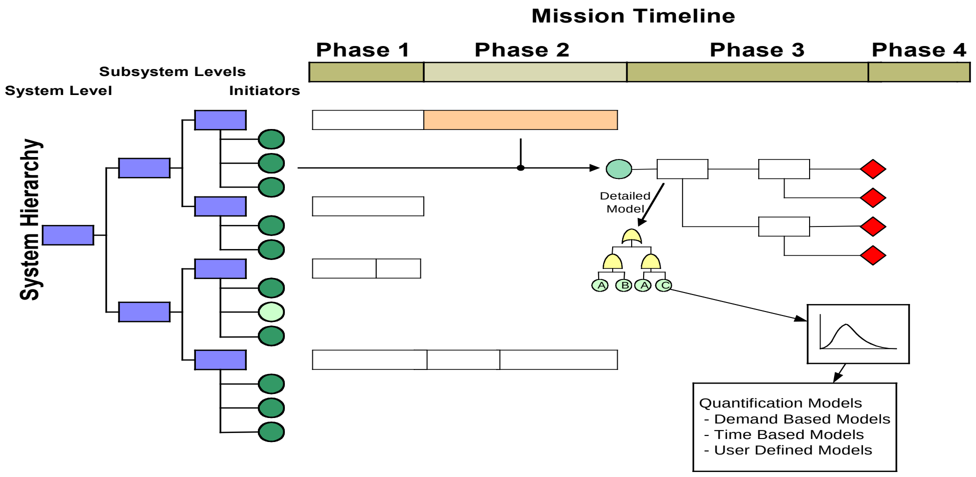




Discrete Dynamic Event Tree

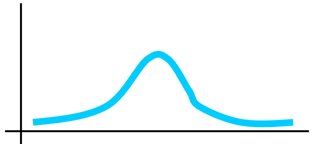


Continuous –Multi-Scale Modeling and Simulation

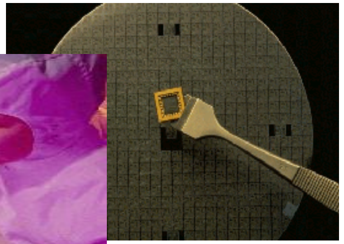


"Lambda Line"

At a given stress S



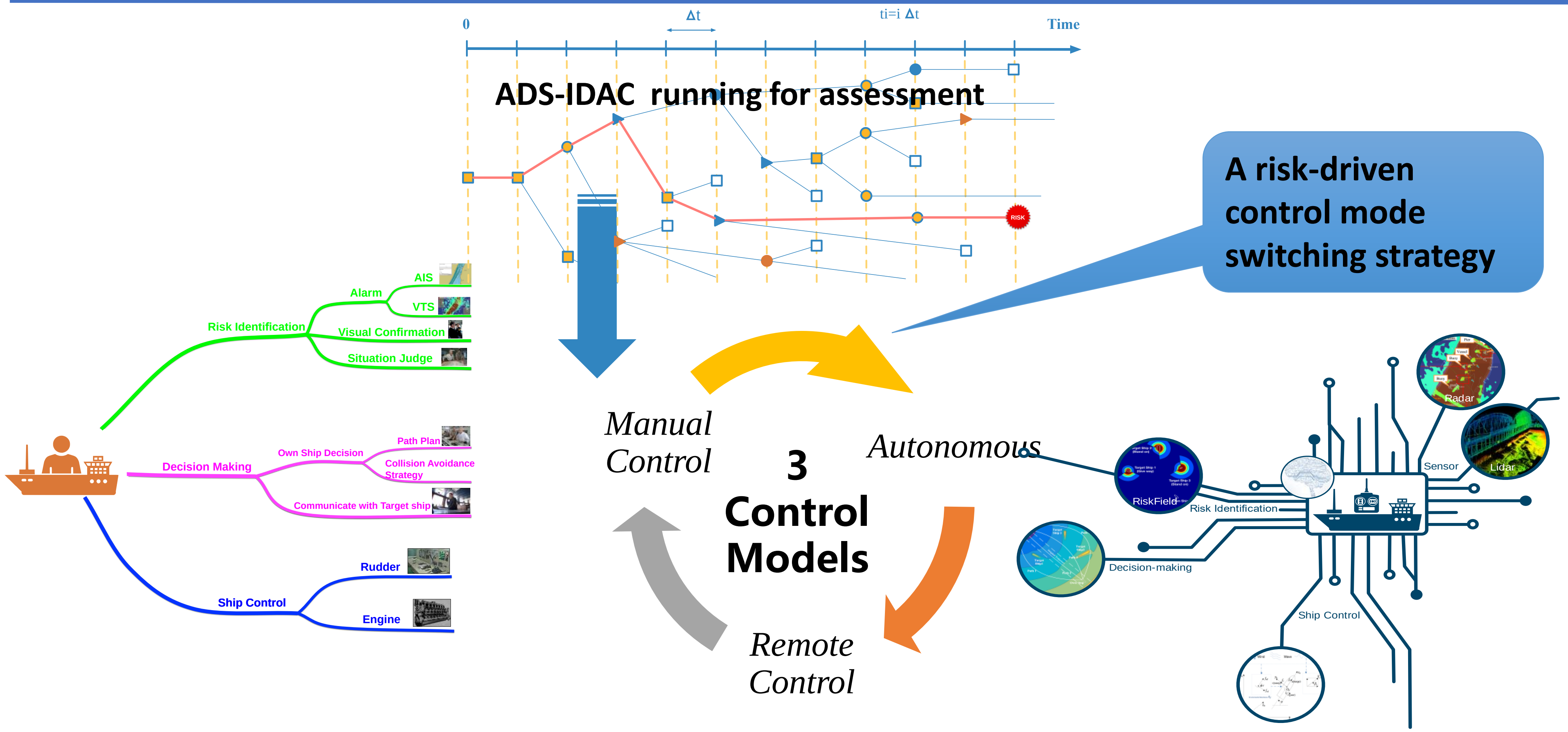
$f(t, S / K, n, b)$



Stress-Life Joint Distribution:

Where K, n and β are parameters

Real Time Risk Based Decision Support of Unmanned Ships

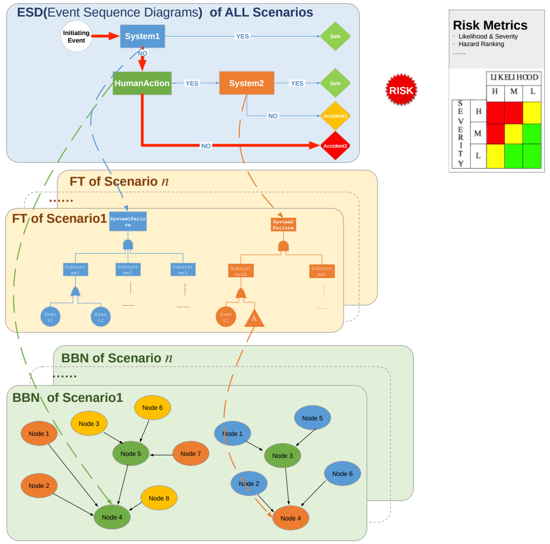


Solution Completeness and Scalability

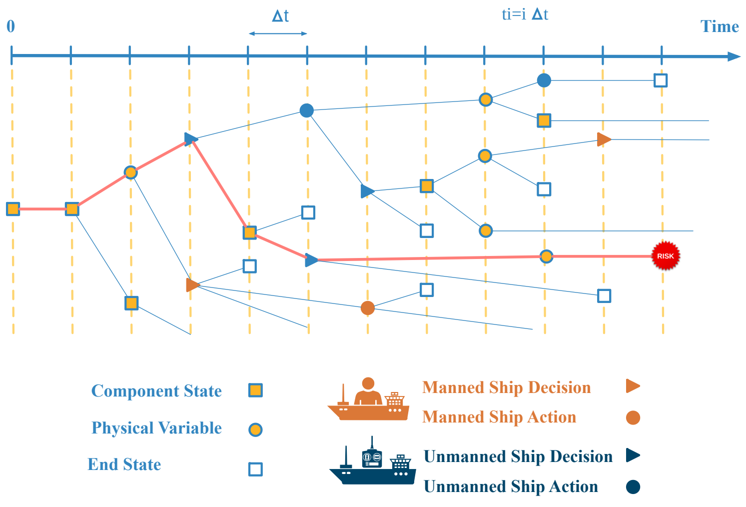
- ❑ Coverage, Scope Completeness ,
- ❑ Optimum Level of Decomposition
  - Hardware (Systems, Sub-systems, components, Parts, Failure Modes, Failure Causes …)
  - Software (Functions, Objects, Computational Routines, Line of Code,…)
  - Human (Cognitive Functions, Information Model, Task Decomposition,…)
- ❑ Interface Characterization
- ❑ Representational Effectiveness (in capturing nature of the phenomenon, inter-model compatibility, traceability, user-friendliness)
- ❑ Ability to Do a Graded Modeling and Analysis
- ❑ Scalability

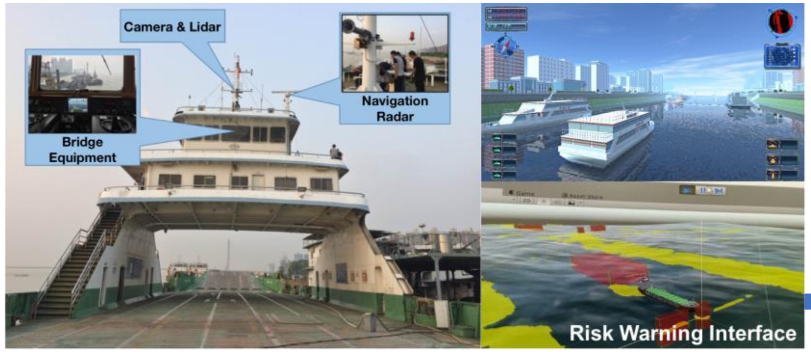
Comparative Assessment of Risks of Different Modes of Maritime Transport Using Dynamic PRA Methods

Conventional PRA



Dynamic PRA

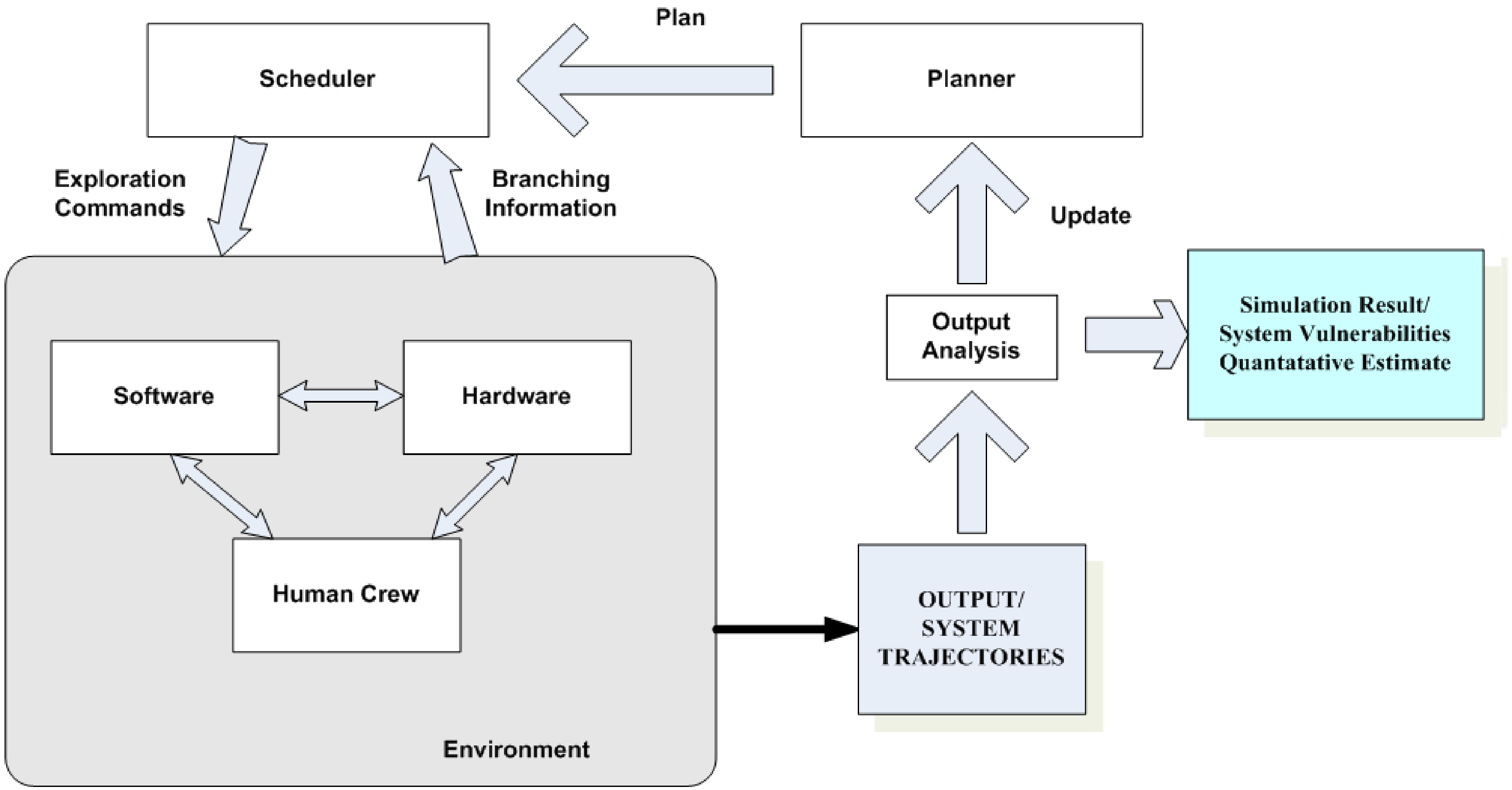




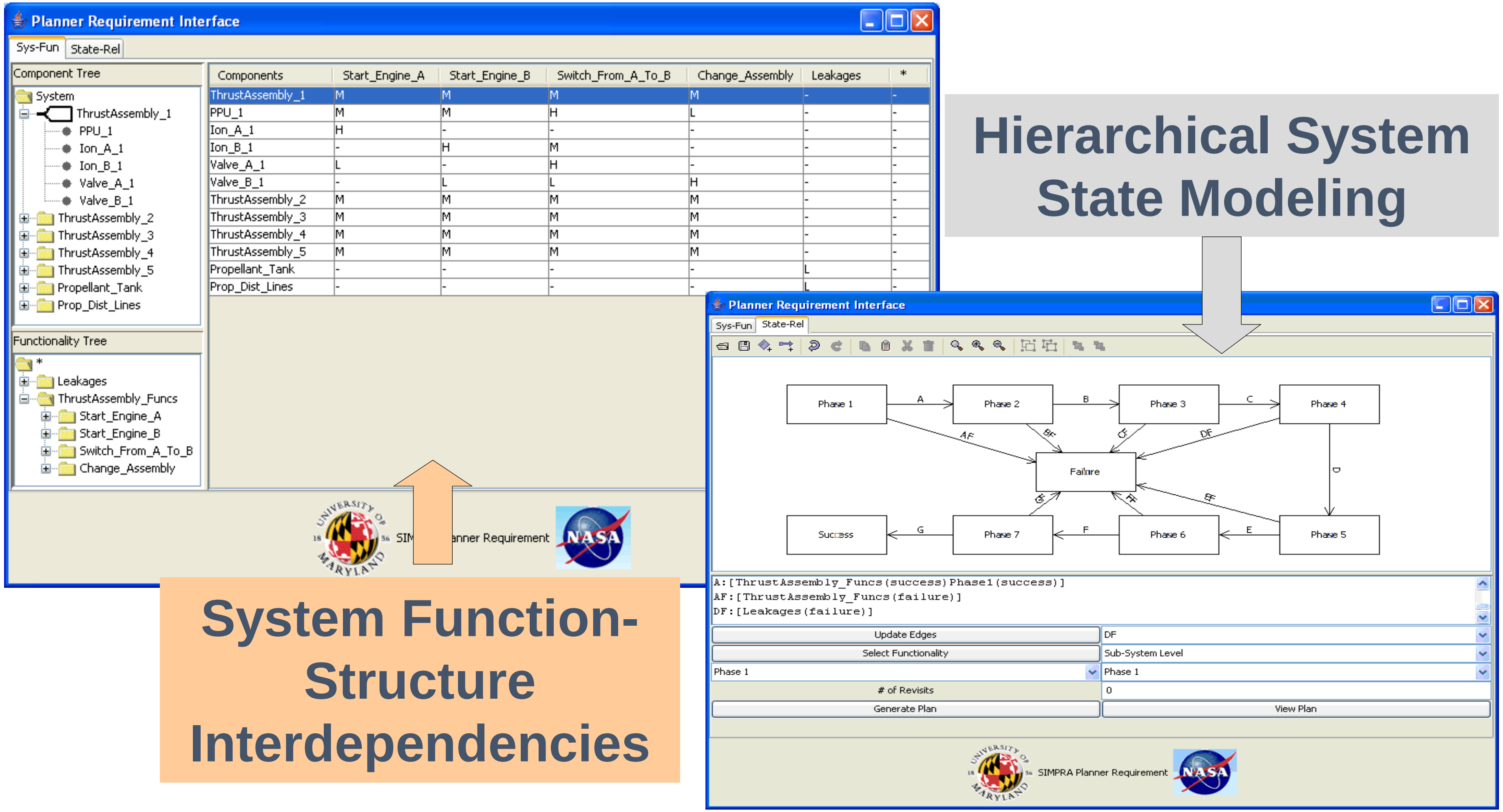
A parallel 3D simulation system has been developed on a ferry ship in the Yangtze River, China.

The Dynamic PRA can be used to greatly enhance situation awareness and help the crews make decisions in complex navigation conditions.

Overview of SimPRA Methodology



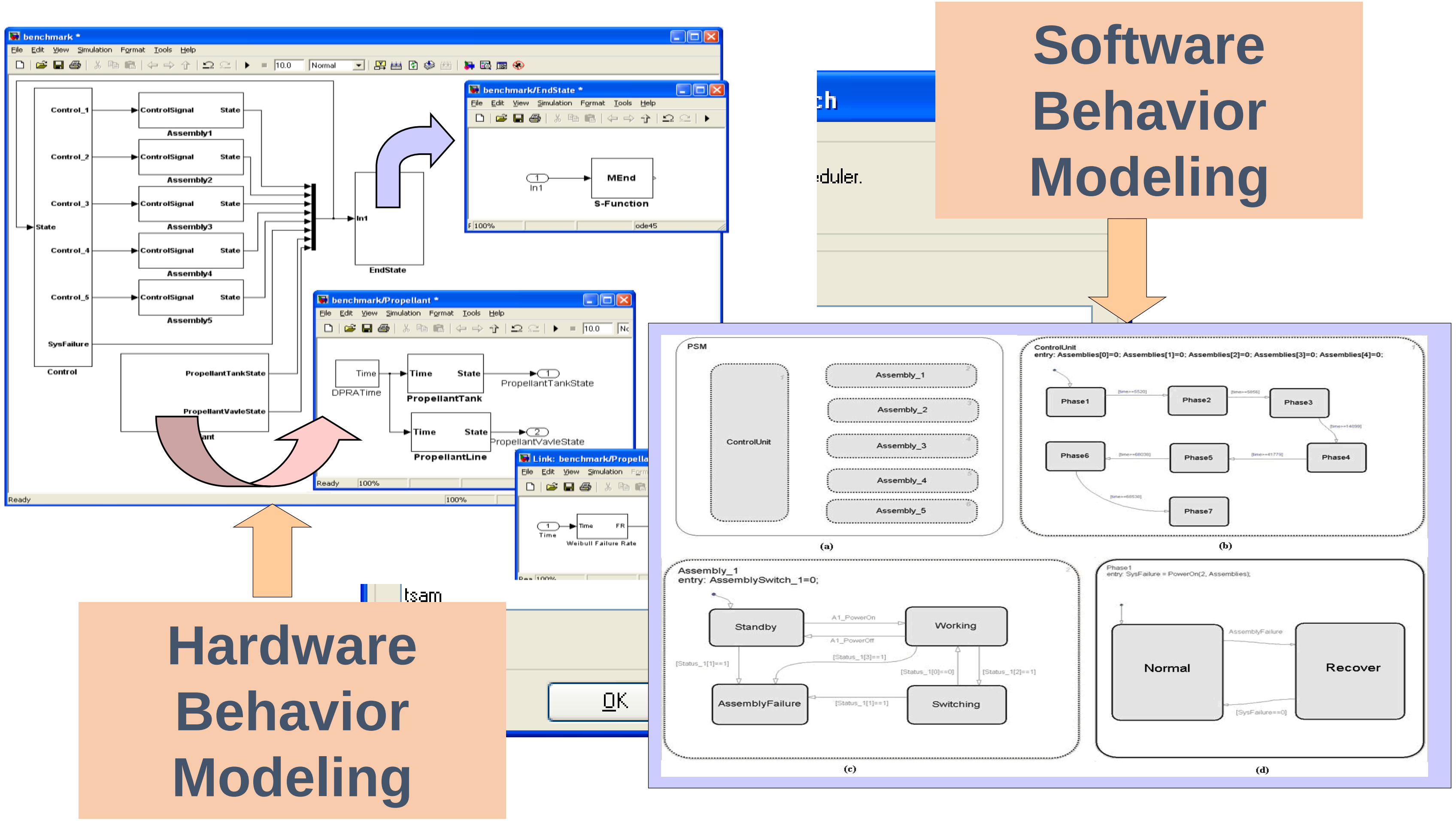
Simulation Planner Functions



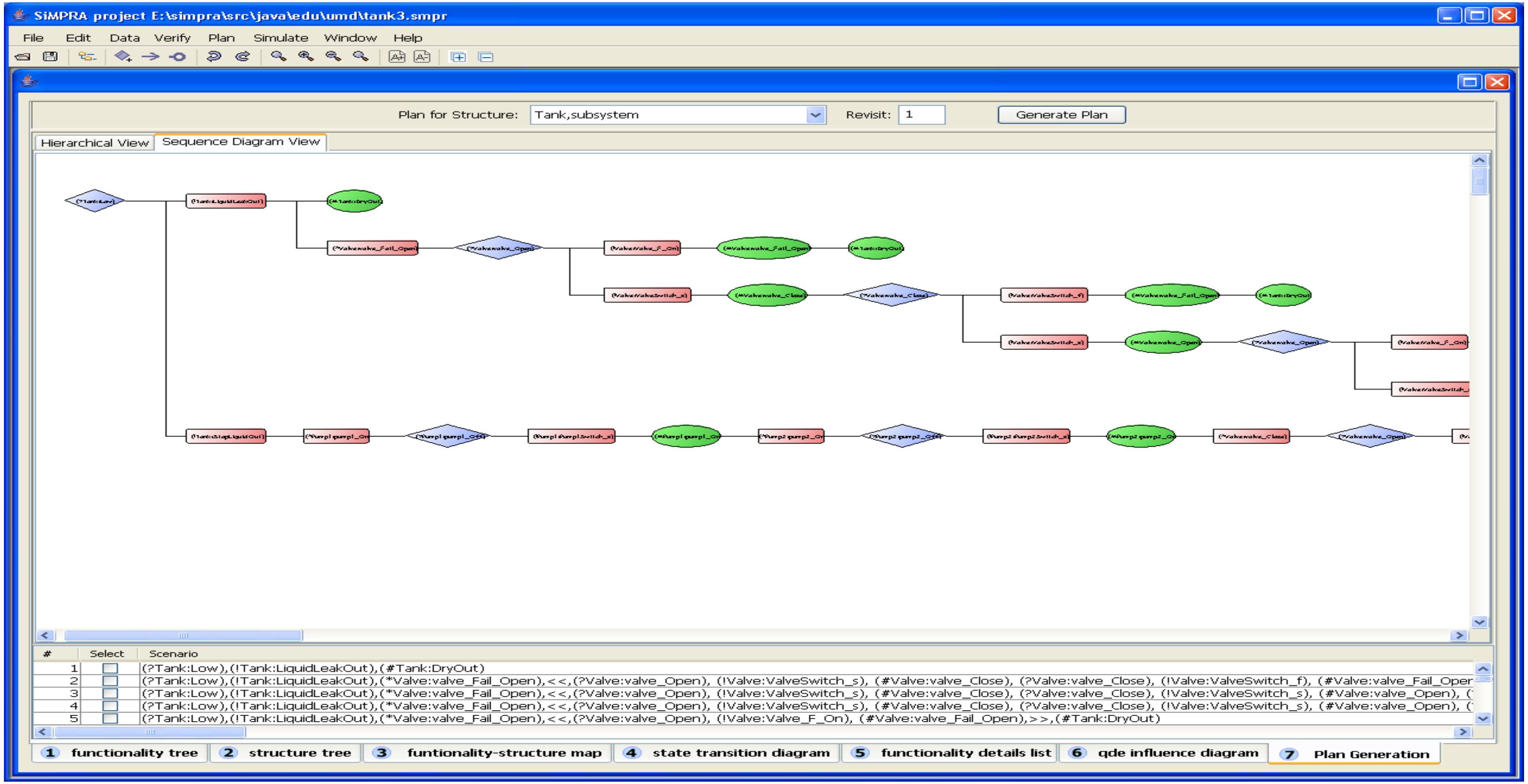
Hierarchical System State Modeling

System Function-Structure Interdependencies

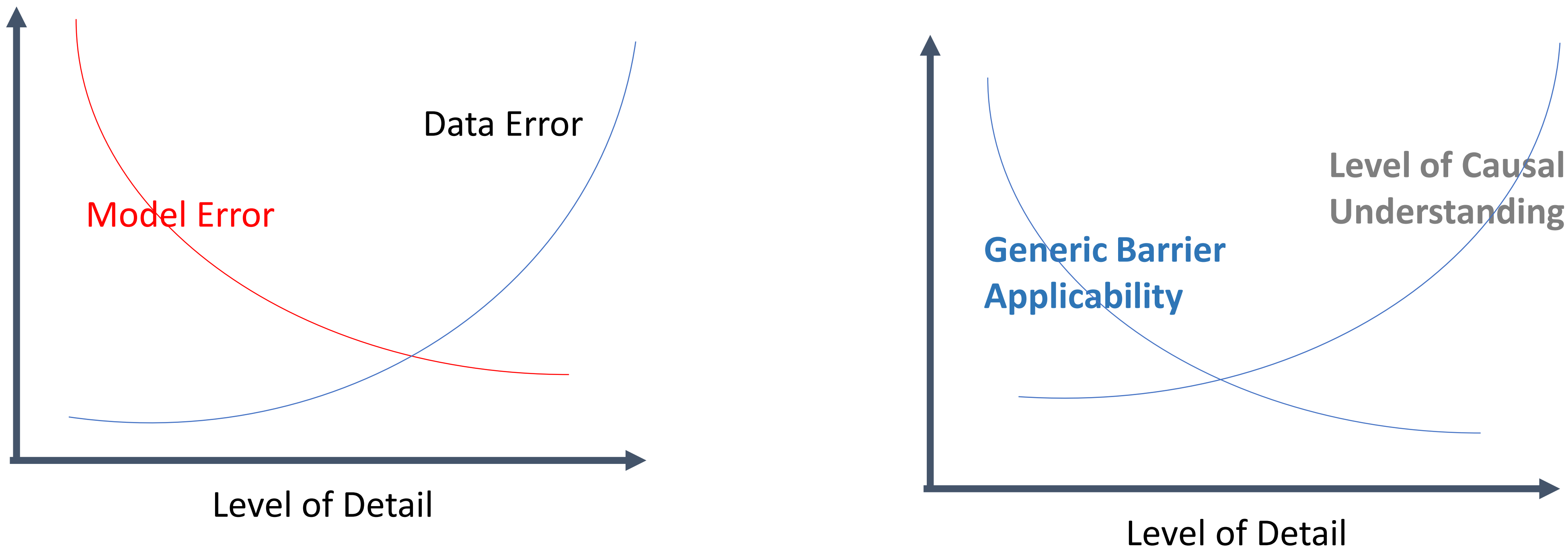
Probabilistic System Simulation Model Building



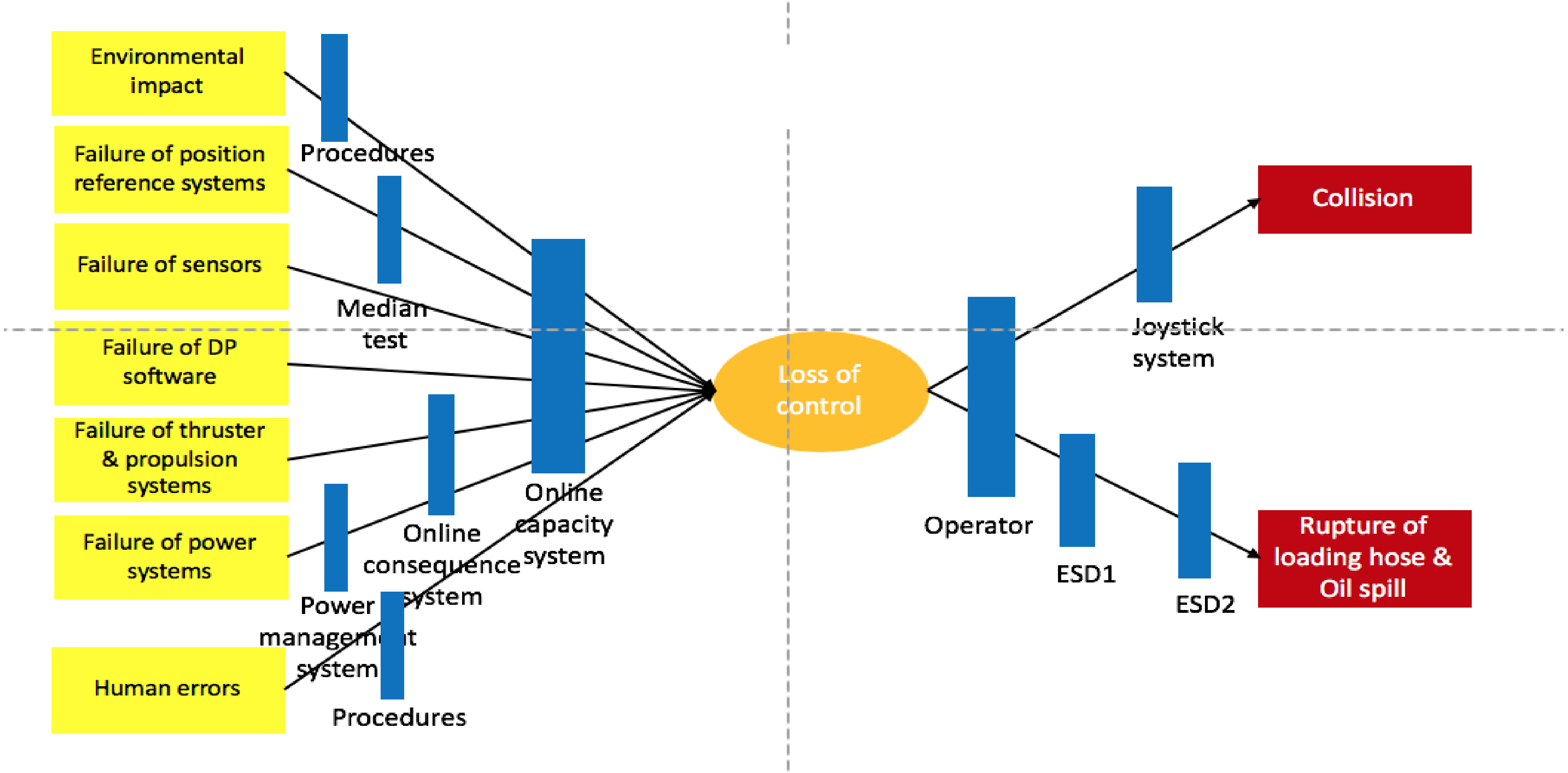
High Level Risk Scenario Generated by SimPRA Planner



Optimizing Level of Details



Generic Barrier / Defense*



* Yining Dong, Current Collaborative Work, NTNU-UCLA

Thank You !