NC STATE UNIVERSITY

Probabilistic Risk Assessment based Model Validation Methods using a Bayesian Network

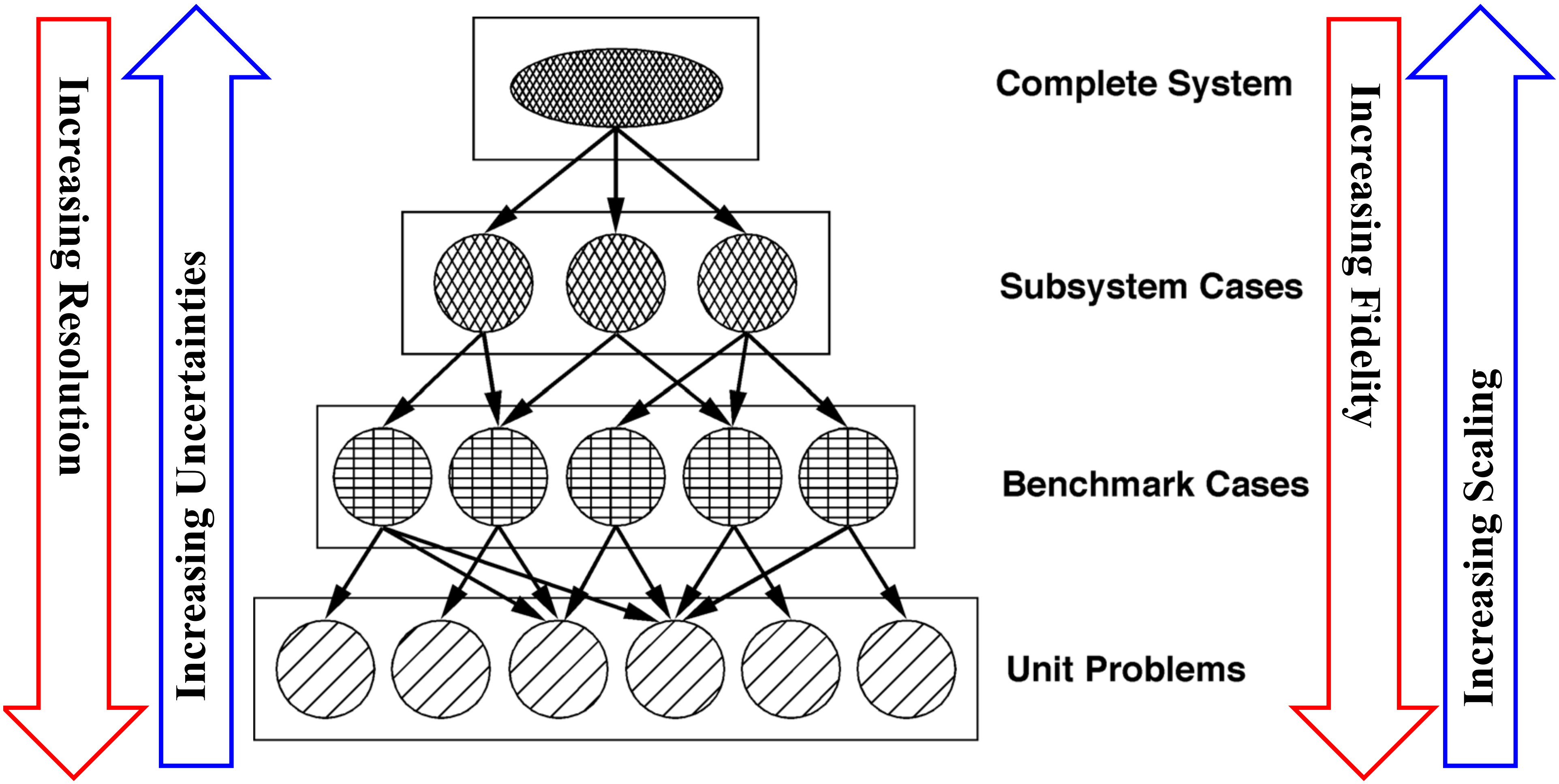
January, 2017

Abhinav Gupta, *Director*

Center for Nuclear Energy Facilities & Structures

North Carolina State University
Raleigh, NC 27695-7908

Motivation



(Sources: AIAA Guide, 1998; Oberkampf and Smith, 2014)

NC STATE UNIVERSITY

Background

- Current model validation metrics:
  - **Visual graphical comparison**
  - **Mean value:** Coleman and Stern (1997); Sprague and Geers (1999); Oberkampf and Trucano (2000); Easterling (2001); Oberkampf and Barone (2006)
  - **Classical Hypothesis testing:** Hills and Trucano (2002); Dowding and Rutherford (2003); Chen et al. (2004); Hills (2006)
  - **Bayesian Hypothesis testing:** Kennedy and O'Hagan (2001); Zhang and Mahadevan (2003); O'Hagan (2006); Bayarri et al. (2007); Chen et al. (2008); Babuška et al. (2008); Rebba and Mahadevan (2008)
  - **Probability distribution, CDF:** Ferson et al. (2008); Ferson and Oberkampf (2009); Roy and Oberkampf (2011); Voyles and Roy (2015)

- Inference in system-level validation metrics:
  - **A building block or hierarchical approach:** AIAA Guide (1998); Hasselman et al. (2002); Bayarri et al. (2005); Korb et al. (2013)
  - **Bayesian network:** Mahadevan and Rebba (2005); Rebba and Mahadevan (2006); Jiang and Mahadevan (2007)

- Limitations:
  - Still needs quantitative validation metric under uncertain nature
  - Mostly limited to mathematically defined system-level model
  - Cannot identify the critical component
  - No basis for effectively improving system-level simulation model

NC STATE UNIVERSITY

Objective

- Develop a novel method to quantitatively assess a system-level simulation model based on component-level validation information and effectively improve simulation model in performance perspective.

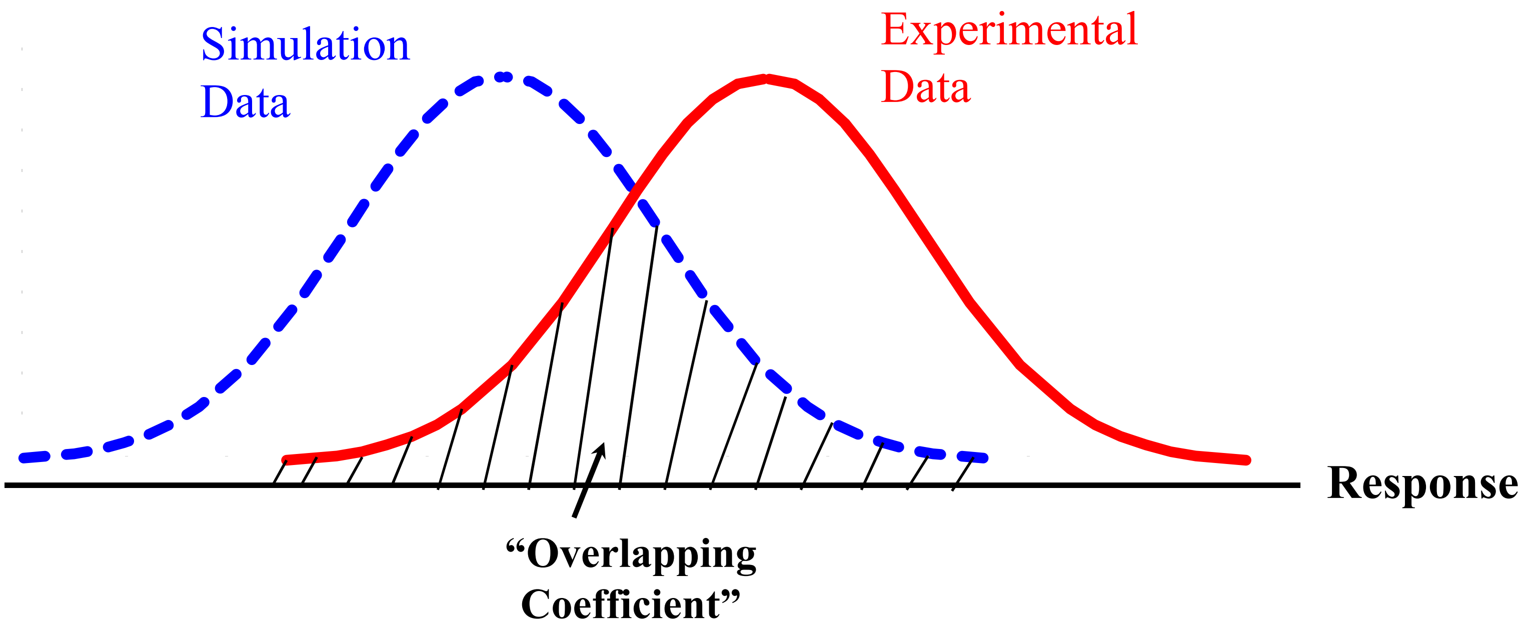
- **BN based PRA**: identify critical scenario and accordingly pick out important component.

- **Concept of Overlapping Coefficient (OC)**: specify quantitative validation metric (0 ~ 100%) under uncertain nature.

- **Bayesian updating**: incorporate newly observed data.

- **Response surface**: establish the unknown relation between low-level and upper-level data.

OC for Model Validation

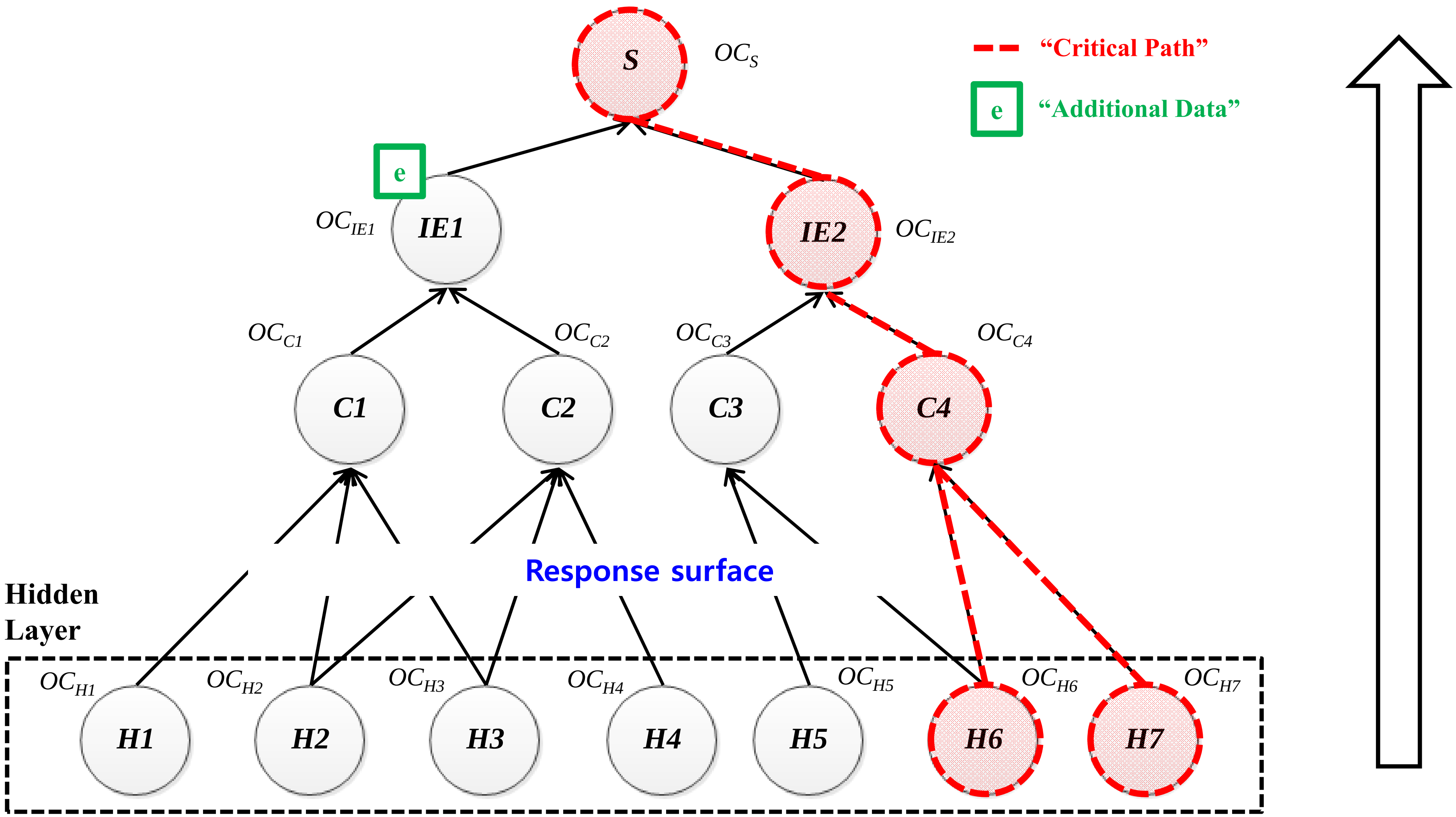


Response quantities are represented as distributions in simulation and experimental data because of uncertainties.

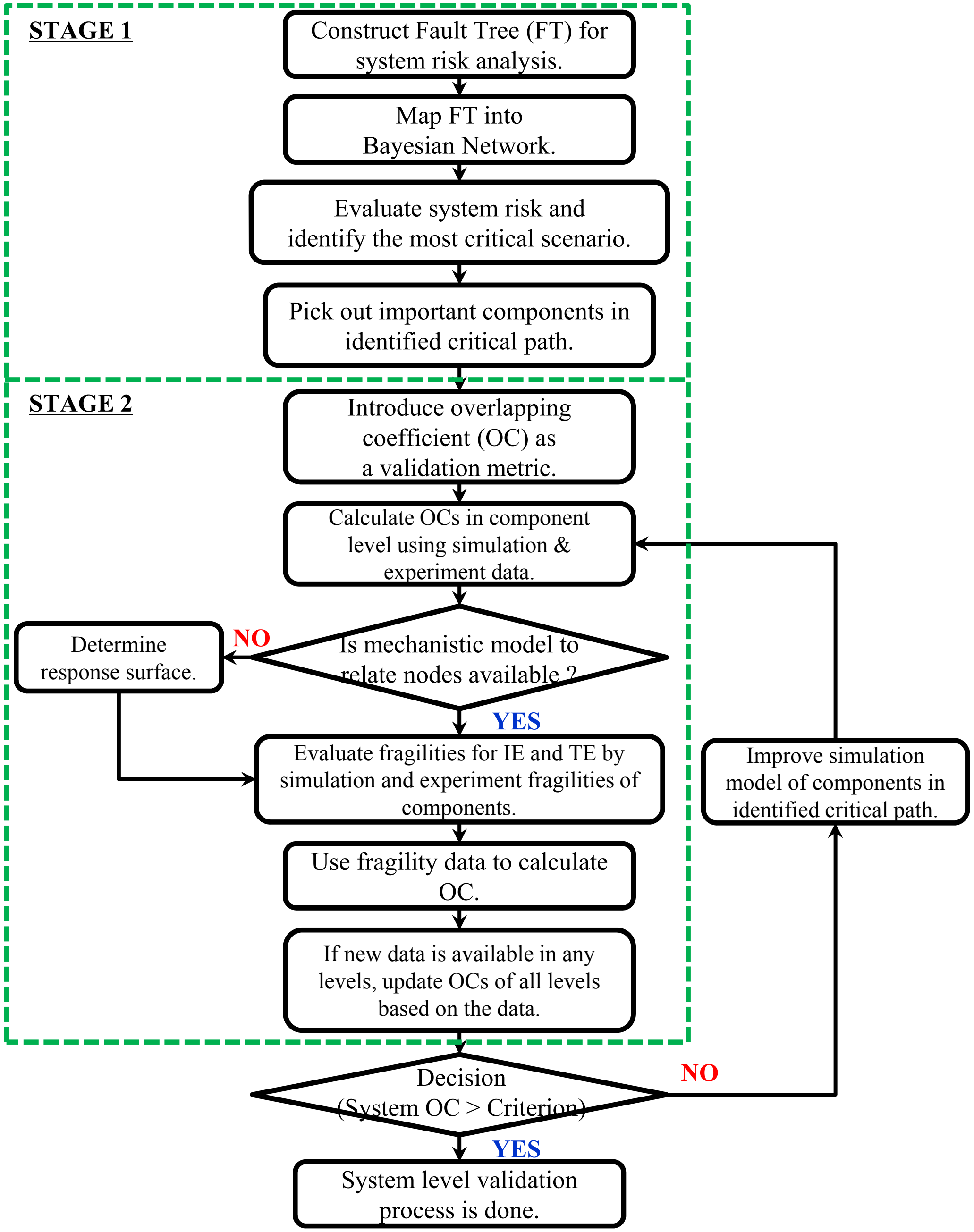
"Shaded Area" so called *Overlap Coefficient* (OC) can represent the accuracy of simulation model compared to experimental data.

OC is in 0 to1 range. "1" means perfect match.

Conceptual Process



Proposed Method: Flowchart

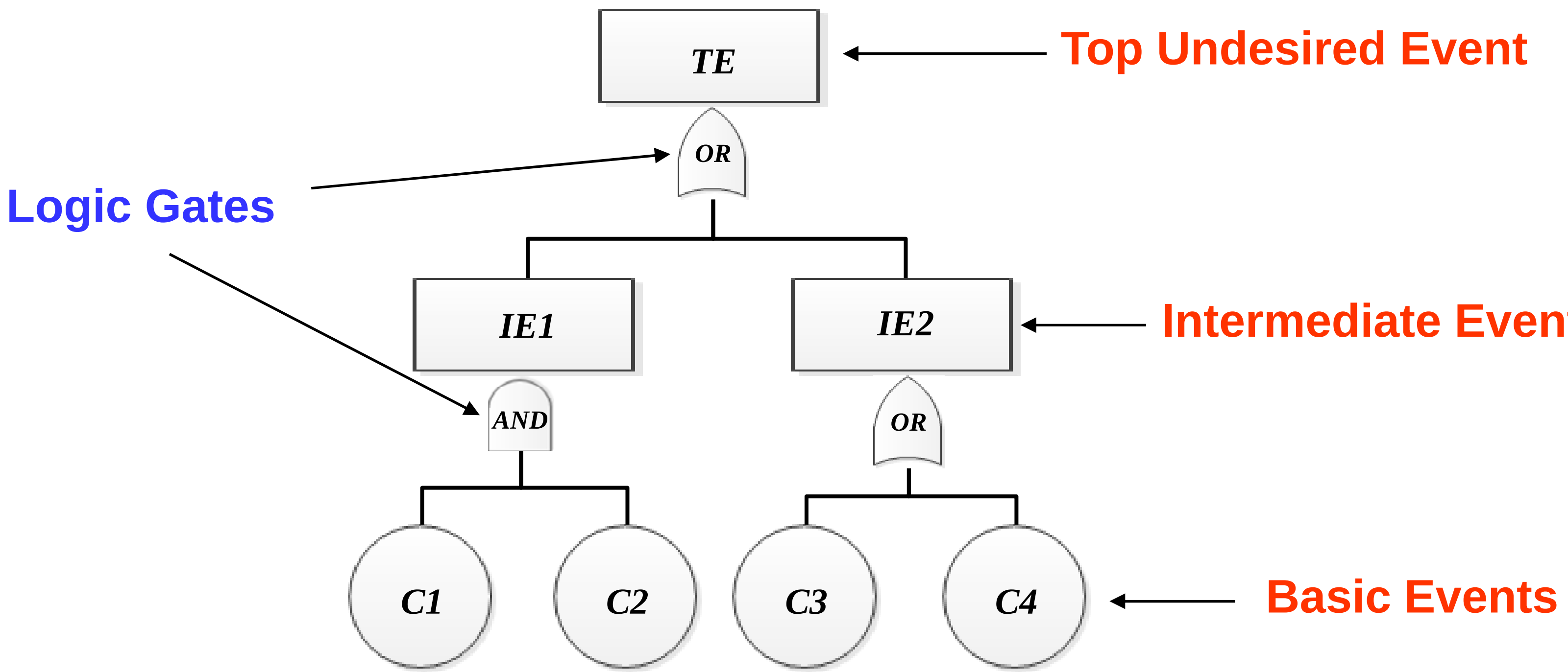


Fault Tree Analysis

- Fault tree analysis typically includes two ways:
  - Qualitative evaluation: logical expression of top event (*TE*) and minimal cut-sets
  - Quantitative evaluation: probabilities of all events and importance measures
- For example,

$$TE = (C1 \cap C2) \cup (C3 \cup C4) \quad \text{: Logical expression}$$

$$= C1 \cdot C2 + C3 + C4 \quad \text{: Minimal cut-sets}$$



Bayesian Network

- Bayesian network is the directed acyclic graph:
  - Each node represent variable and has conditional probability table (CPT).
  - Arcs signify direct relationship between the linked nodes.
  - The CPT assigned to nodes specify how strongly the linked nodes influence each other.

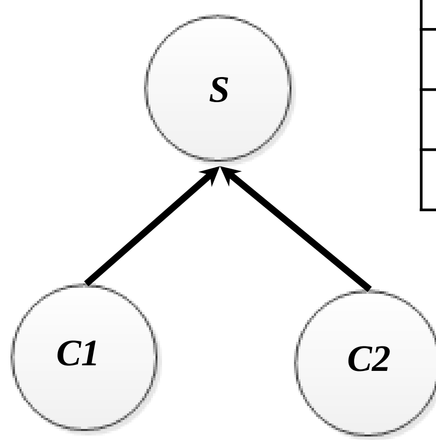
- For example,

Legend
: Nodes
: Arcs or arrows

$$P(C1,C2,S)=P(S\,|\,C1,C2)\cdot P(C1)\cdot P(C2)$$

CPT of S

C1	C2	P(S=1|C1, C2)
1	1	1
1	0	0
0	1	0
0	0	0



CPT of C1

	C1 = 1	C1 = 0
P (C1)	0.1	0.9

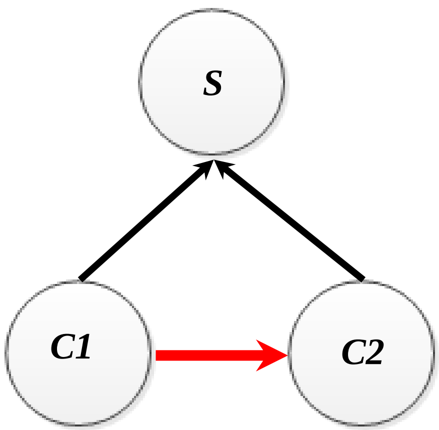
CPT of C2

	C2 = 1	C2 = 0
P (C2)	0.3	0.7

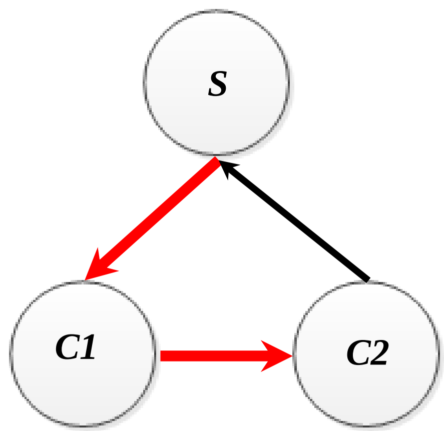
NC STATE UNIVERSITY

Bayesian Network

- The Bayesian network (BN) can manage various kinds of statistical dependencies of components more than fault tree (FT) analysis can do:



(a)



(b)

<BNs which FT cannot specify >

Bayesian Inference

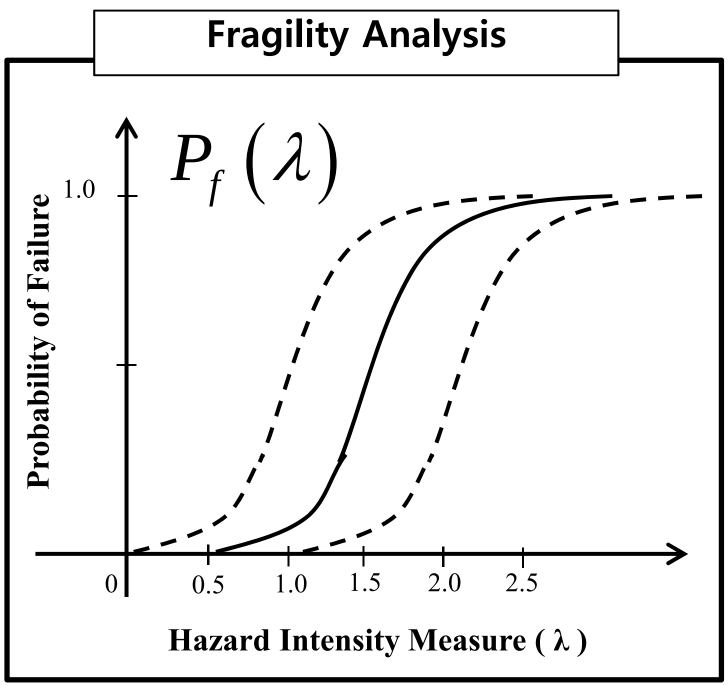
- Bayesian inference can deal with uncertainties of event occurrence probabilities (**p**), and incorporate newly obtained data (**D**) in any levels.
- Markov Chain Monte Carlo (MCMC) is utilized to obtain the posterior PDF

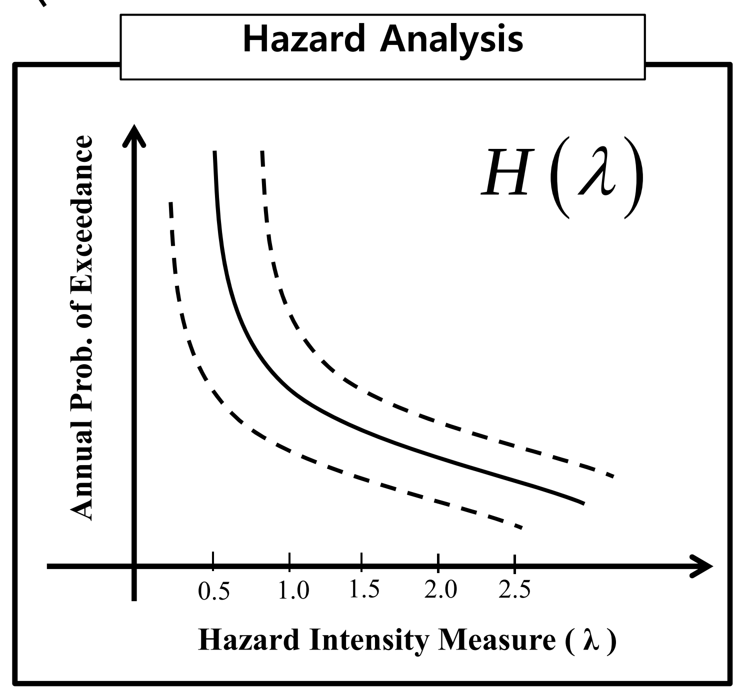
$$\underbrace{f\left(\mathbf{p}\,|\,\mathbf{D}\right)}_{\text{POSTERIOR}} = \frac{\overbrace{f\left(\mathbf{D}\,|\,\mathbf{p}\right)}^{\text{LIKELIHOOD FUNCTION}}\;\overbrace{f\left(\mathbf{p}\right)}^{\text{PRIOR INFORMATION}}}{\int f\left(\underset{\text{DATA}}{\mathbf{D}}\,|\,\mathbf{p}\right) f\left(\mathbf{p}\right) d\mathbf{p}}$$

Risk Calculation in PRA

- Risk (annual probability of occurrence) for an event is evaluated by convolution of hazard curve and the corresponding fragility curve as following:

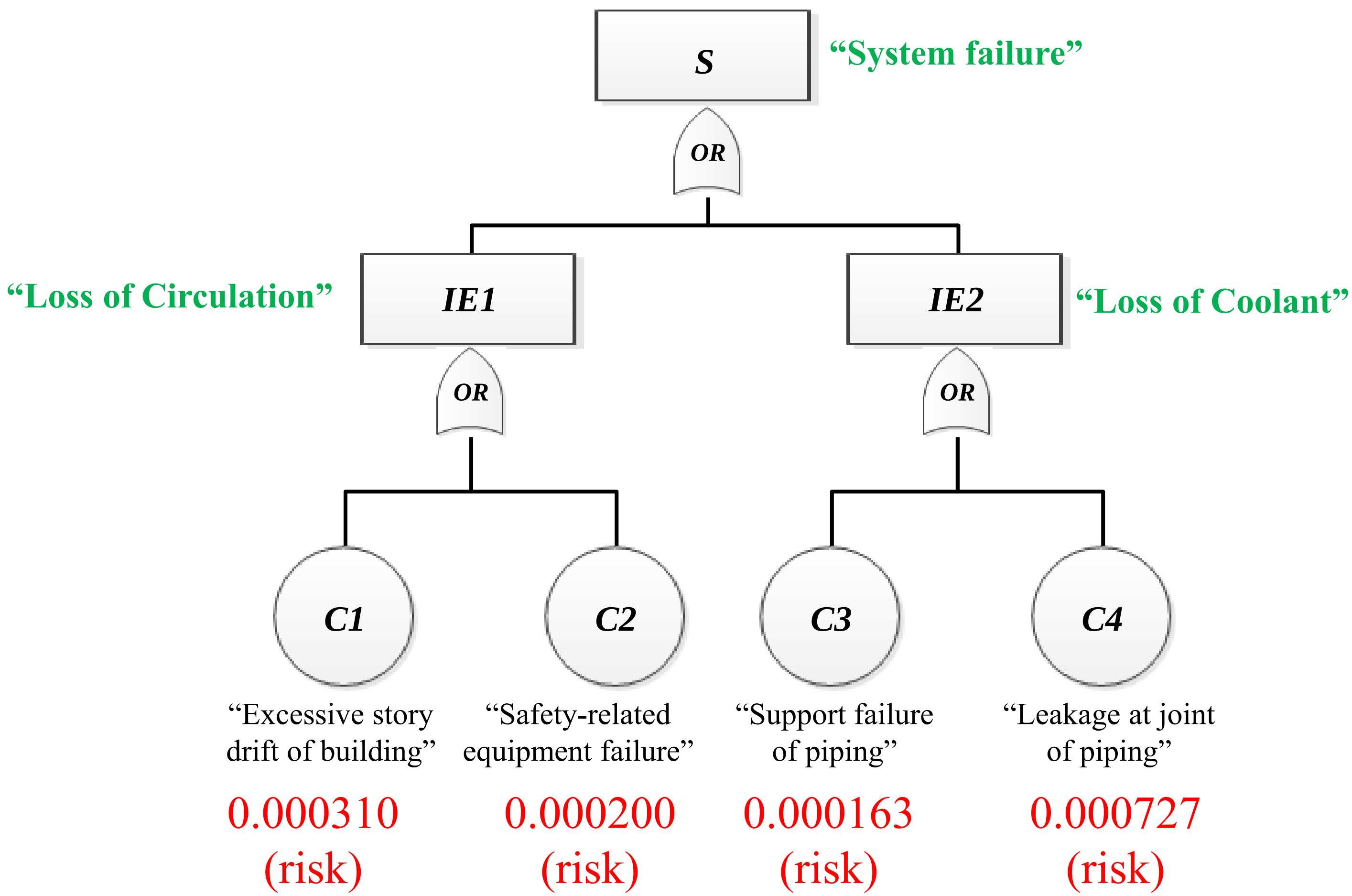
$$Risk = \int P_f(\lambda) \cdot \left| \frac{dH(\lambda)}{d\lambda} \right| d\lambda$$





(*PRA: probabilistic risk assessment)

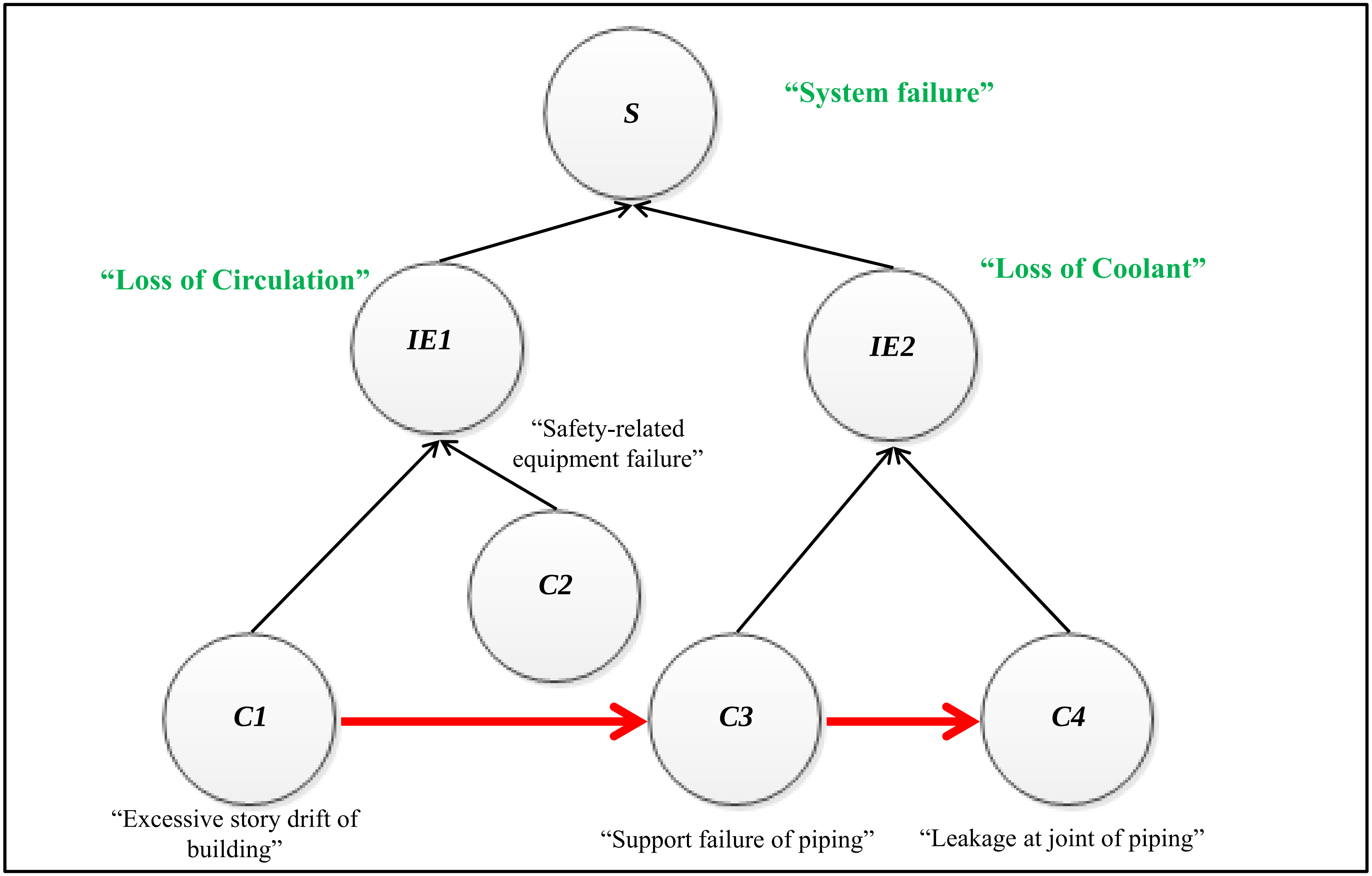
Ex.1: Fault tree for B-P System Failure



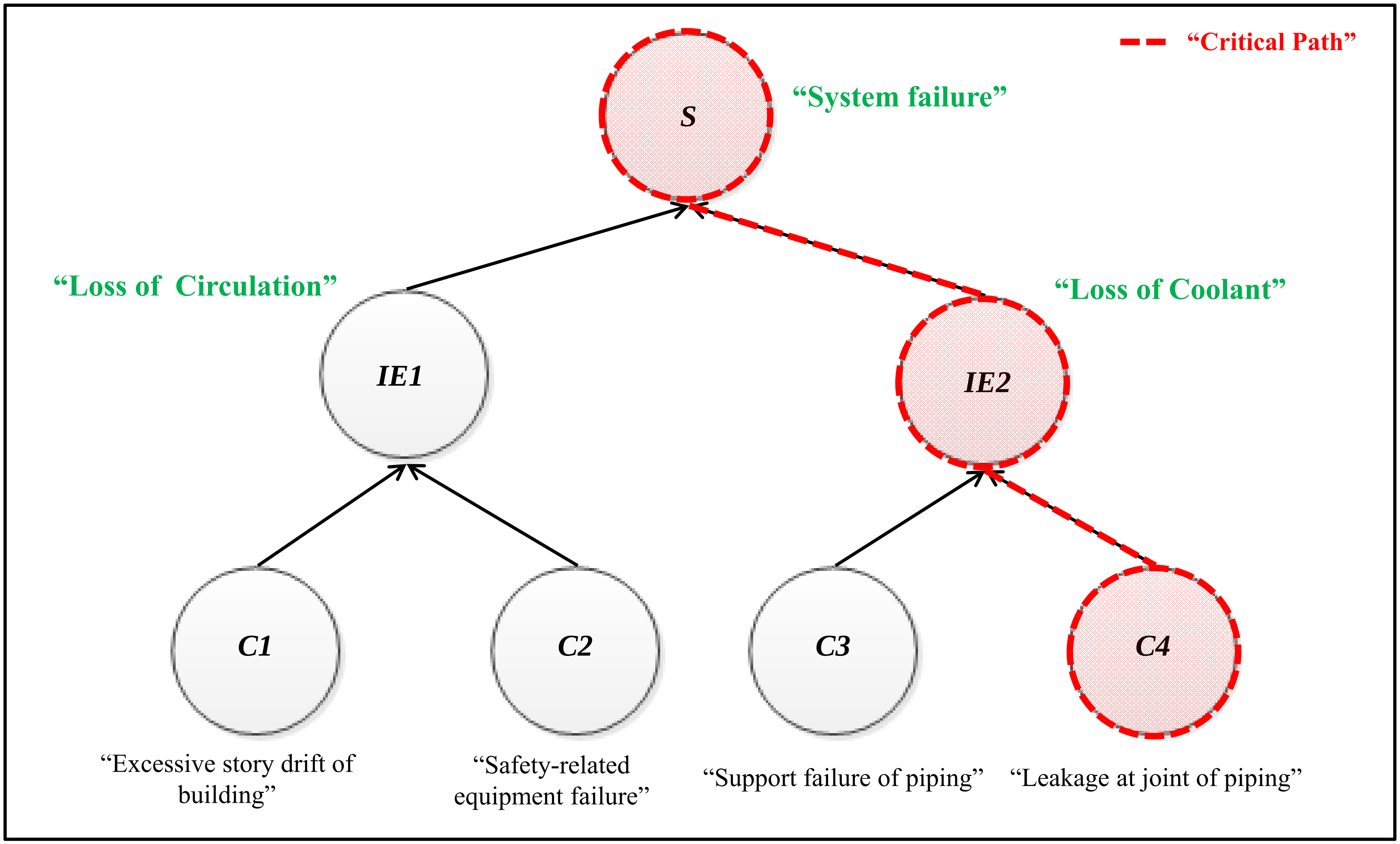
Risks of C1, C2, C3 and C4 are evaluated by convolution of given seismic hazard curve and the seismic fragility curves.

Ex.1: Additional Dependences in BN

$*Ci \sim \text{Normal}\left(\mu, \text{c.o.v} \cdot \mu\right)$

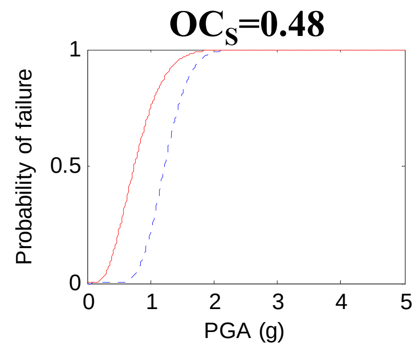


Bayesian Network & Critical Scenario (Ex)

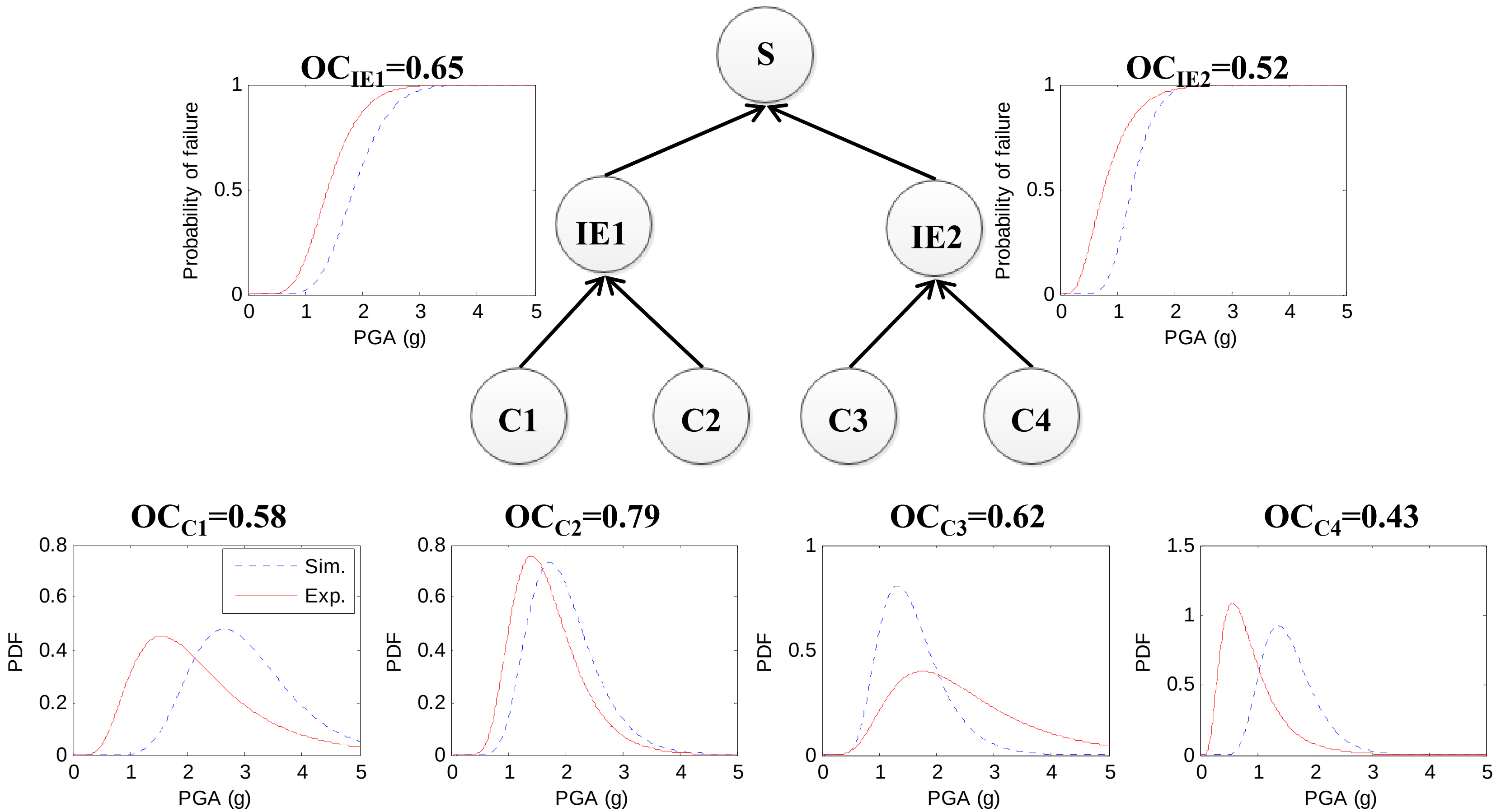


NC STATE UNIVERSITY

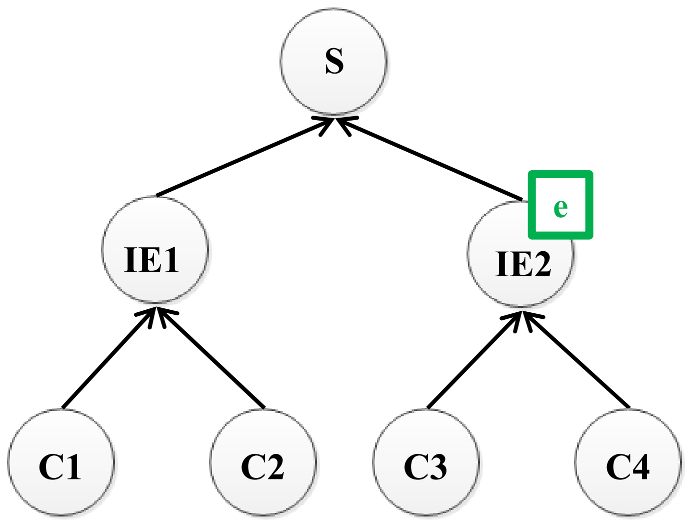
OC in All Levels (Ex)



Model Acceptance Decision:
System-level OC < 0.8
→ Improve Sim. Model Accuracy



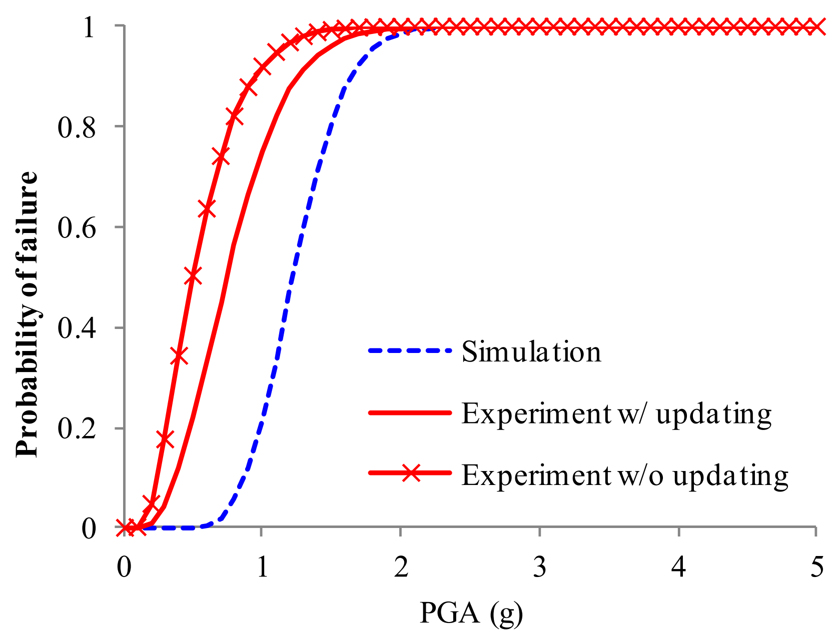
Additional Data (Ex)



Additional Discrete data in IE2 node

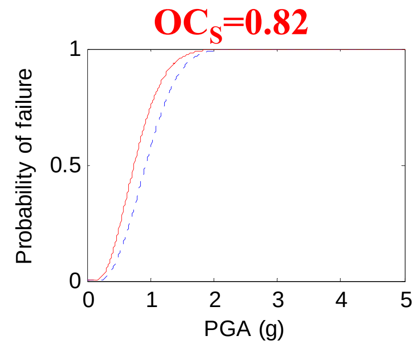
PGA	$*N_f$ / $*To$
1.0g	6/30
1.5g	15/30

$*N_f$: Number of failures, *To*: Total occurrence

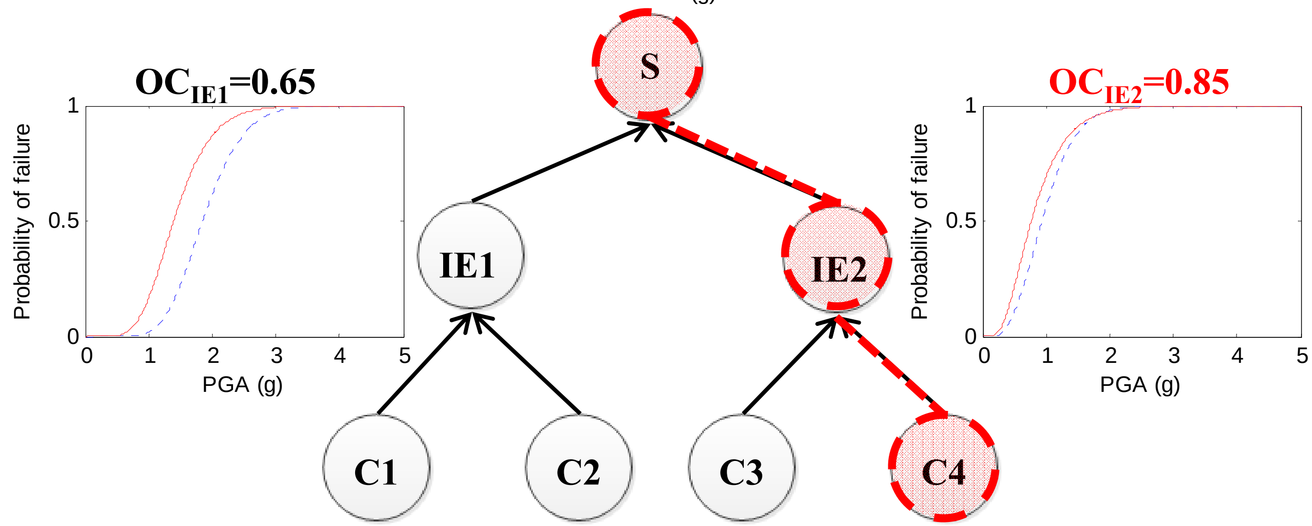


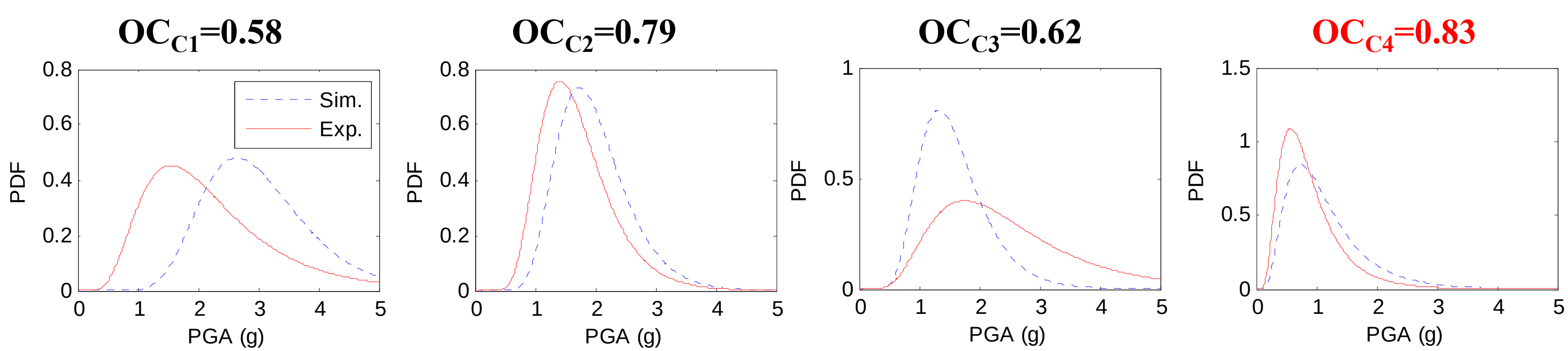
Fragility curves of system failure with/without discrete data of IE2

Improve Simulation Model (Ex)



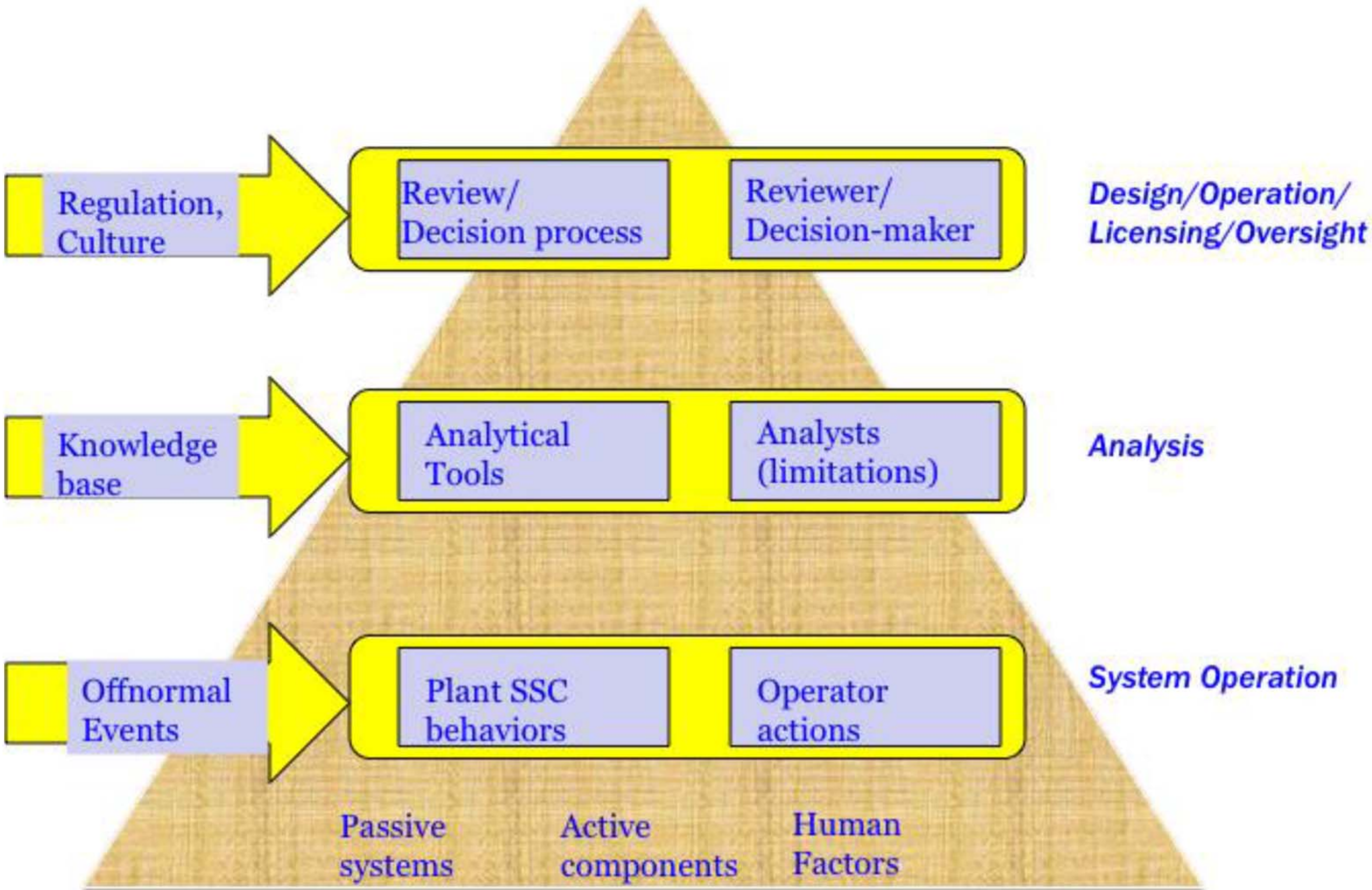
Model Acceptance Decision:
System-level OC > 0.8
"OK"





NC STATE UNIVERSITY

Challenges: Data Driven Risk-Based Decision Making



"Humans in the Loop" at All Levels

Wrap Up: Conclusions

- ❑ Address uncertainty in decision making by interfacing technology expertise with advances in system, data and computational sciences.
  - ✓ data-analytics for enabling decisions
  - ✓ knowledge management through machine learning
  - ✓ computational intelligence for safety, security, and risk-mitigation.

- ❑ Innovative environment is needed for:
  - ✓ Scalability and narrative processing of high volume data to generate, evolve, predict, interact, and continually update decision tools.
  - ✓ A data management framework to drive decision-making, knowledge base, and accident management and at the same time also involve human expertise at all levels.
  - ✓ Identifying existing tools and addressing gaps by developing new tools consistent with the framework for a data-driven approach.