U.S. DOE Light Water Reactor Sustainability Program
Risk-Informed Systems Analysis (RISA) Pathway
Stakeholder Engagement Virtual Meeting
November 13, 2024

LWRS Program Research on Risk Assessment of Safety-related DI&C Systems

Congjian Wang, Tate Shorthill,
Edward Chen, Jisuk Kim,
Jooyoung Park, Svetlana Lawrence

INL/MIS-24-81813

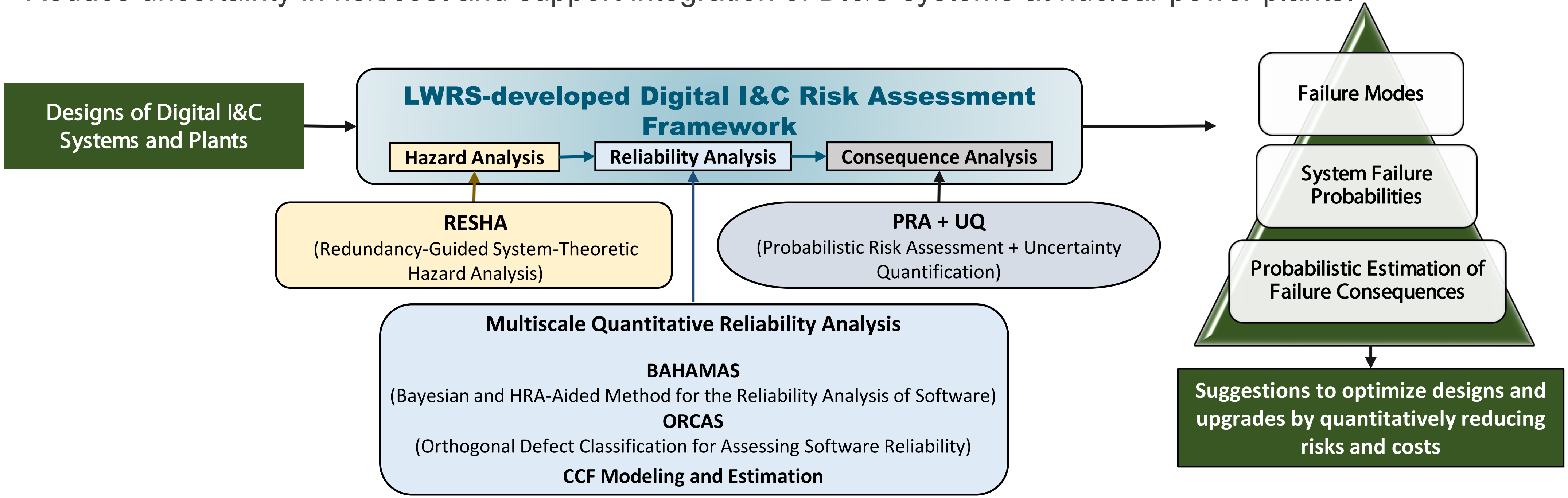
November 13, 2024

Project "Digital I&C Risk Assessment"

U.S. DOE Light Water Reactor Sustainability (LWRS) Program, Risk-Informed Systems Analysis (RISA) Pathway

Overview of Digital I&C Risk Assessment within LWRS RISA Pathway

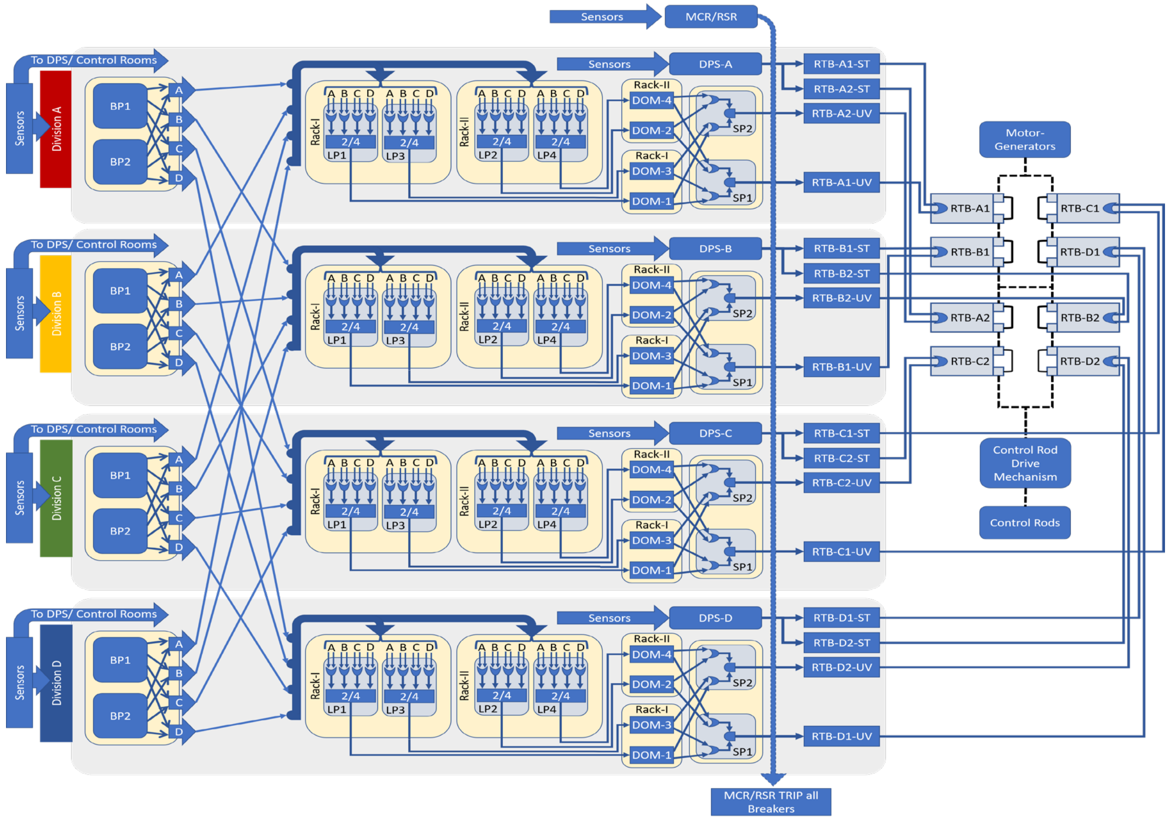
- Offer a capability of design architecture evaluation of various DI&C systems to support system design decisions on diversity and redundancy applications
- Develop systematic and risk-informed tools to address CCFs and quantify corresponding failure probabilities for DI&C technologies
- Support and supplement existing risk-informed DI&C design guides by providing quantitative risk-informed and performance-based evidence
- Reduce uncertainty in risk/cost and support integration of DI&C systems at nuclear power plants.



Value Proposition

- **The framework** is envisioned and developed as **an integrated risk-informed tool** to support vendors and utilities with optimization of design solutions from economical perspectives GIVEN the constrain of meeting risk-informed safety requirements.

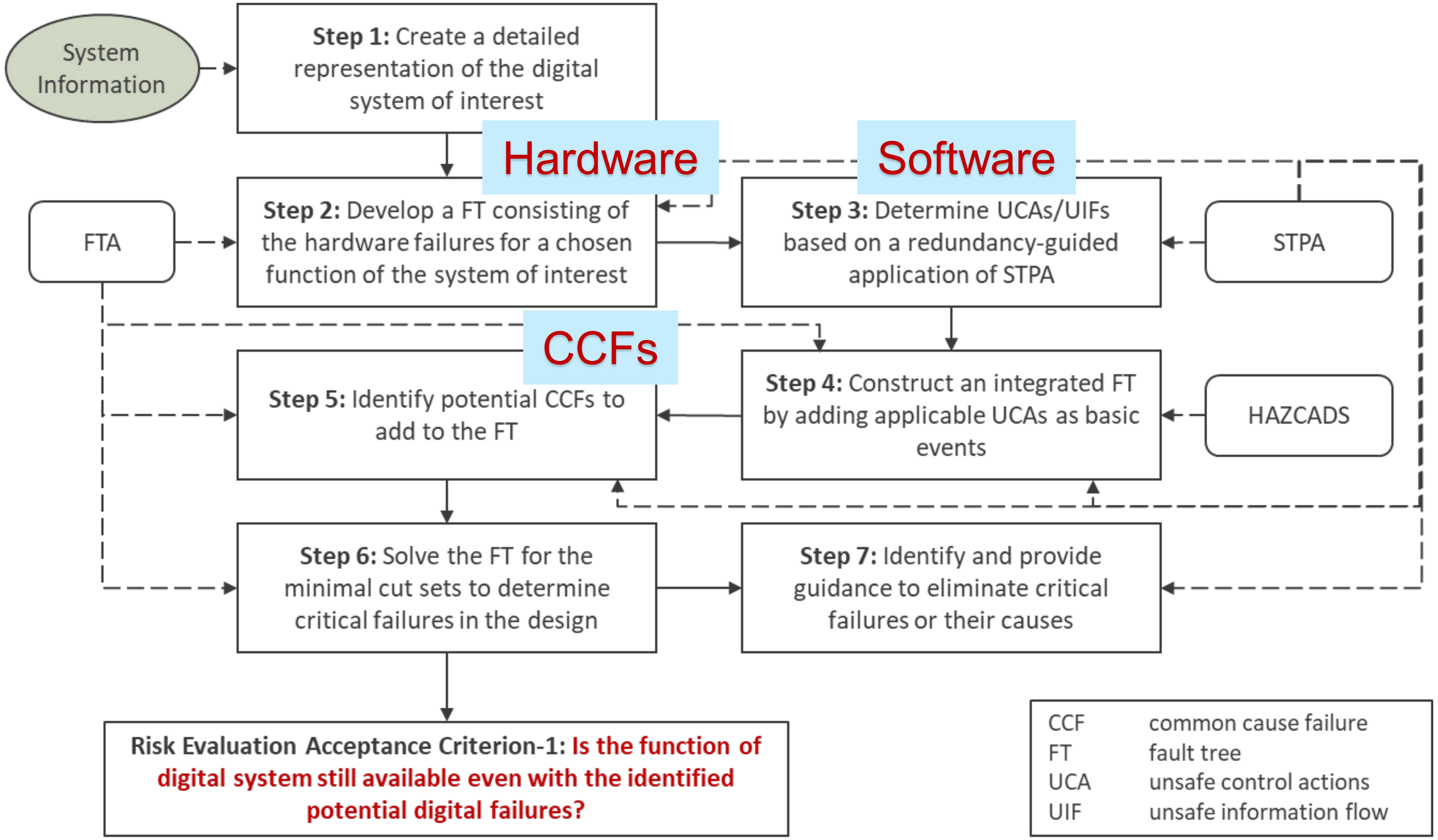
- **Quantitative Risk Analysis**
  - Software reliability metrics → DI&C system reliability → Plant safety analysis
- **Risk-informed Design**
  - Management strategy of CCFs
    - Identification and elimination
  - Level of redundancy
    - 4 divisions vs. 2 divisions
    - 4 vs. 2 local logic processors per division
  - Level of diversity
    - Design: Analog? Digital? A combination of both?
    - Software: Design requirements, programming language, etc.
    - Hardware Equipment: Manufacturers, designs, architectures, etc.



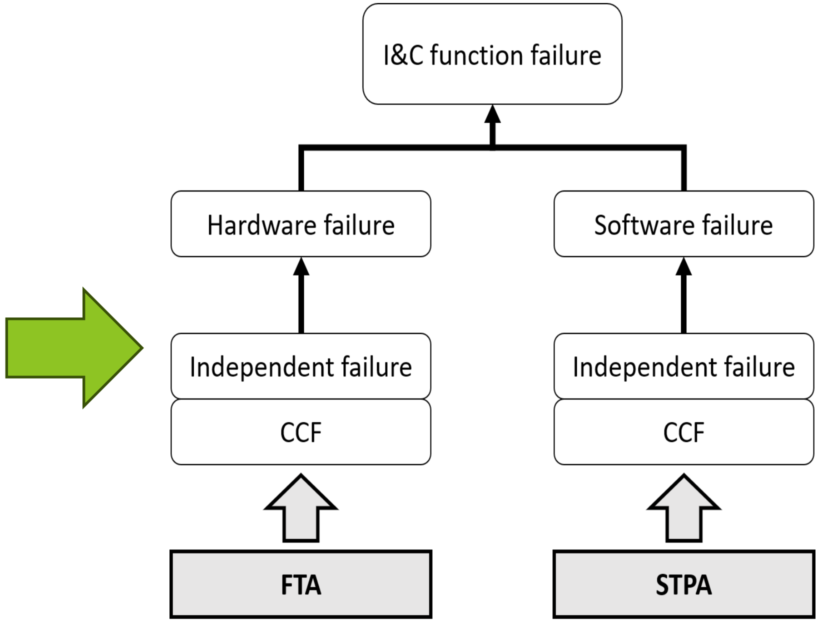
A Four-Division Digital Reactor Trip System

Hazard Analysis via Redundancy-guided System-theoretic Hazard Analysis (RESHA)

- Incorporates **Fault Tree Analysis (FTA)** and **System-Theoretic Process Analysis (STPA)**.
- Reframes STPA in a **redundancy-guided** way to **identify various CCFs**.
- Identifies and traces failures in both **Unsafe Control Actions** (UCAs) and **Unsafe Information Flows** (UIFs).



CCF	common cause failure
FT	fault tree
UCA	unsafe control actions
UIF	unsafe information flow



Integrated Fault Tree Schema via RESHA

Quantitative Software Reliability Analysis

- **Methods developed within this project:**
  - **BAHAMAS (**Bayesian and HRA-Aided Method for the Reliability Analysis of Software)
    - Developed for the conditions with limited testing/operational data or for reliability estimations of software in early development stage.
    - Provide an estimation of failure probabilities to support the design of software and target DI&C systems.
  - **ORCAS** (Orthogonal Defect Classification for Assessing Software Reliability)
    - Developed for the conditions with sufficient testing/operational data.
    - A more refined estimation of software failure probabilities can be provided.

	BAHAMAS	ORCAS
Applicable conditions	• Limited testing/operational data • For reliability estimations of software in early development stage	• Sufficient testing/operational data • For reliability estimations of software in development or testing stage
Key assumption	Software failures can be traced to human errors in the software development life cycle	Sufficient data is available through testing (e.g., T-Way testing)
Ways to identify root causes	STPA + ODC + HRA in SDLC	STPA + ODC
Ways to quantify failure rates of root causes	HRA in SDLC, i.e., Technique for Human Error Rate Prediction	Software reliability growth modeling

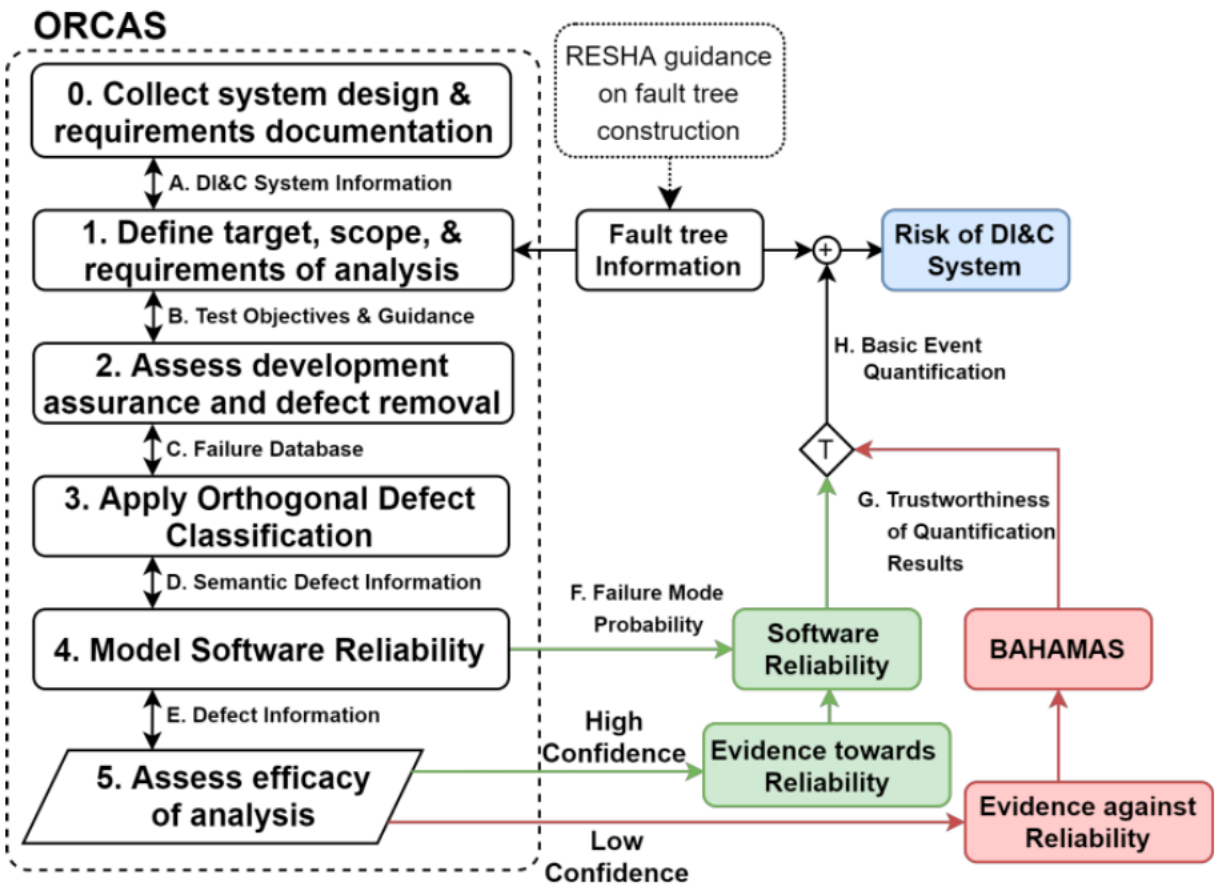
BNN	Bayesian Belief Network
ODC	Orthogonal Defect Classification
HRA	Human Reliability Analysis
SDLC	software development life cycle

Bayesian and HRA-Aided Method for the Reliability Analysis of Software

- **BAHAMAS** tracks human errors in the software development and their influence on the existence of specific types of defects which ultimately influence the probability of software failure

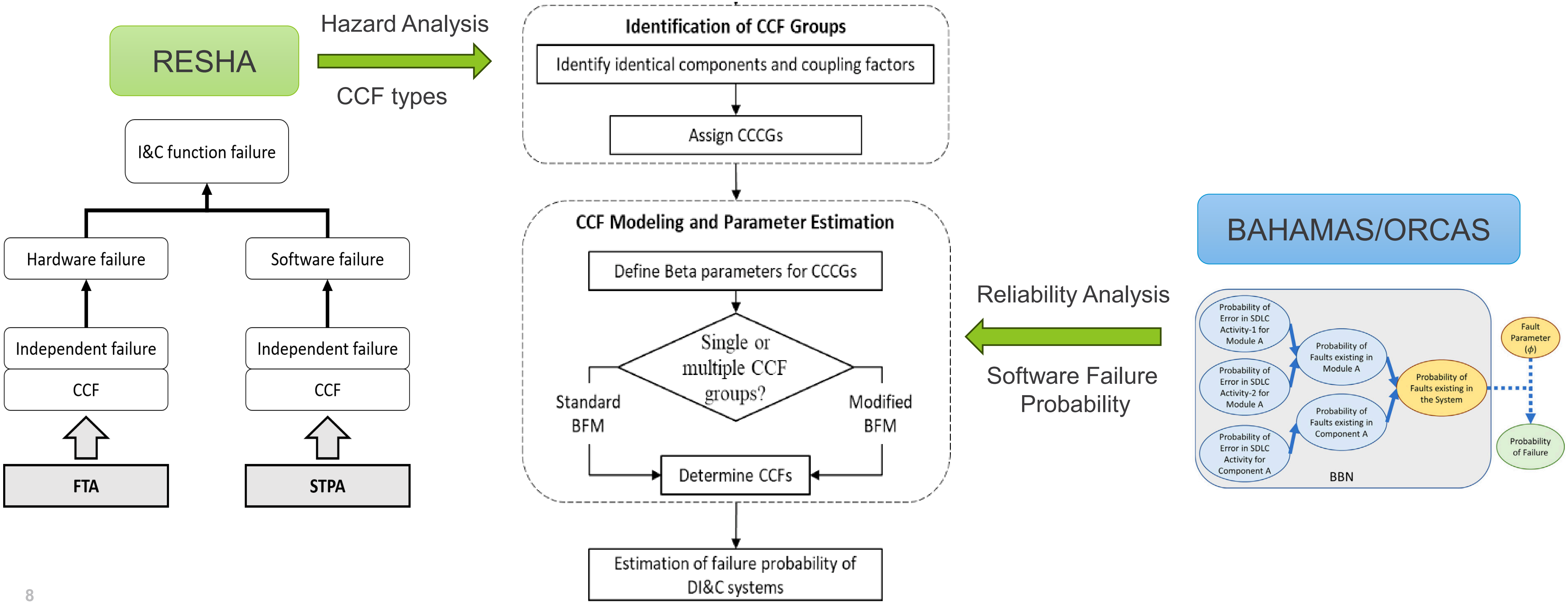
Orthogonal Defect Classification for Assessing Software Reliability

- **ORCAS** leverage software comprehensive testing, ODC and software reliability growth models to quantify the software failure probability of specific UCAs/UIFs

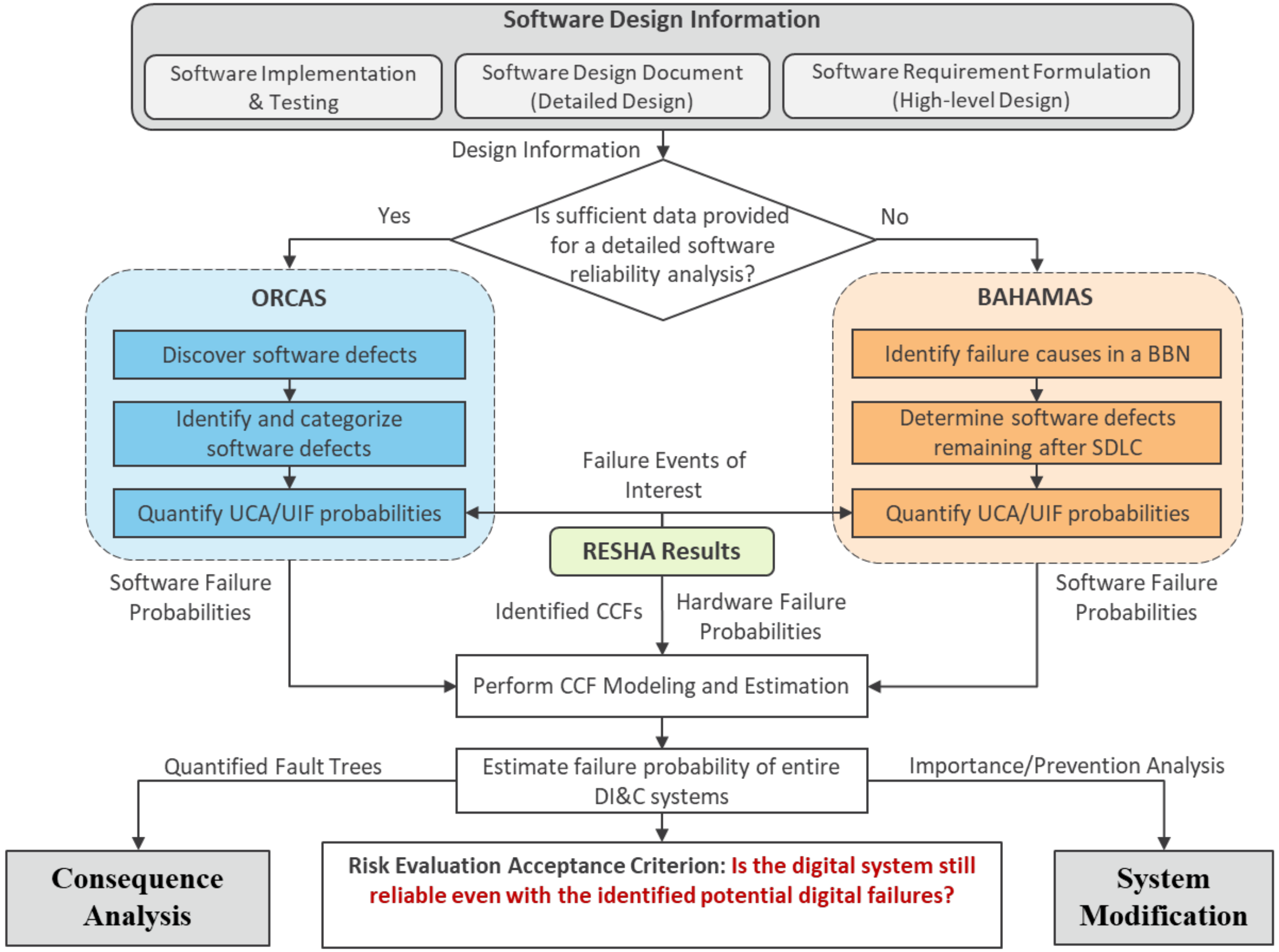


CCF Modeling and Estimation

- A CCF modeling flowgraph is developed for software CCF modeling and estimation based on modified **Beta Factor Model (BFM)** and **Partial Beta Factor (PBF)** Model.



Multiscale Quantitative Reliability Analysis



BAHAMAS	Bayesian and human-reliability-analysis-aided method for the reliability analysis of software	ORCAS	orthogonal-defect classification for assessing software reliability	SDLC	software development life cycle
				UCA	unsafe control action
		RESHA	redundancy-guided systems-theoretic method for hazard analysis	UIF	unsafe information flow
CCF	common cause failure				

LWRS
LIGHT WATER REACTOR SUSTAINABILITY

Major Accomplishments in FY-24 (I)

- Developed a novel approach to evaluate the reliability of ML-integrated control systems.
  - A journal article published.
  - Results were included in the June M4 technical report.

Reliability Engineering and System Safety 250 (2024) 110266

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

ELSEVIER

Check for updates

Evaluating the reliability of machine-learning-based predictions used in nuclear power plant instrumentation and control systems

Edward Chen [a,*], Han Bao [a], Nam Dinh [b]

[a] Idaho National Laboratory, Idaho Falls, ID, USA
[b] North Carolina State University, Raleigh, NC, USA

A R T I C L E I N F O

Keywords:
Machine learning
Reliability
Trustworthiness
Out-of-distribution detection

A B S T R A C T

The field of data-driven, neural-network-based machine learning (ML) has seen significant growth, with applications in various information and control systems. Despite promising real-world uses, the reliability of models remains questionable. Conventionally, reliability is assessed based on predictive fidelity, accuracy, and training effectiveness; however, quality developmental procedures and excellent training performance metrics do not guarantee operational reliability. Instead, an ML model's predictive performance depends on the training set's representativeness to the intended operational space. It is known that ML algorithms excel at interpolation but struggle with extrapolation tasks. Anomalies and feature drift can also reduce operational performance. Determining whether a new sample is an interpolation or extrapolation task involves out-of-distribution (OOD) detection for assessing its proximity to the existing training data. Thus, we present a real-time, model-agnostic individual prediction reliability evaluation method called Data Auditing for Reliability Evaluation (DARE) for applying OOD detection to the training dataset. We demonstrate on a feedforward neural network ML-integrated digital twin for predicting fuel centerline temperatures during loss-of-flow transients. DARE acts as a "data supervisor" in determining the model's applicability under different operating conditions. In this manner, we demonstrate how training data can serve as inductive evidence to support the reliability of ML predictions.

1. Introduction

Significant research has been conducted on the integration of machine learning (ML) methods into various information and control systems. ML has been applied to enhance plant diagnostics [1–3], automate the scheduling of maintenance tasks [4–6], enable autonomous control [7], develop digital twins [8,9], etc. In such data-driven models, a training dataset typically defines a model's function by learning the latent correlation between the input and target values. The function realized through training is governed by the multiplication of nondescript weights and biases that is generally difficult to interpret. The model is also intended to generalize over a range of cases extending beyond the discrete points in the training data. Thus, ML model validation and verification are typically conducted to assess generalizability using holdout testing sets, which are samples not previously seen by model but have similar attributes to the training data (e.g., distribution, features, skewness, range). K-fold cross-validation [10] is an example of holdout data training and validation. From K-fold, the predictive accuracy on the holdout set is the assumed accuracy in post-training operation. This is based on a closed-world assumption [11] in which new input data are assumed to be drawn independent and identically distributed (i.i.d) relative to the training data.

Although ML model methods for training, validation, and verification have advanced significantly, data-driven models can experience major performance reductions when applied to real-world operational environments [12]. In addition to reduced predictive accuracy, poor-performing ML models can also present safety risks. For example, misclassifications in self-driving vehicle algorithms [13] have led to public losses. The root cause of ML model failures may originate from regressional inconsistencies [14], inherent distributional rigidness [15], metric optimization failures [16], and unintended adversarial examples [17]. Thus, if these models are to be adopted in real world systems, it is paramount that their reliability and trustworthiness be guaranteed.

Reports by the National Institute for Standards and Technology [18] and the U.S. Nuclear Regulatory Commission [19] have also indicated that trustworthiness in ML presents a critical barrier to its adoption, and will play a vital role in the safe, accountable, and secure operation of data-driven ML systems.

One possible way to develop trust in ML predictions is to assess how

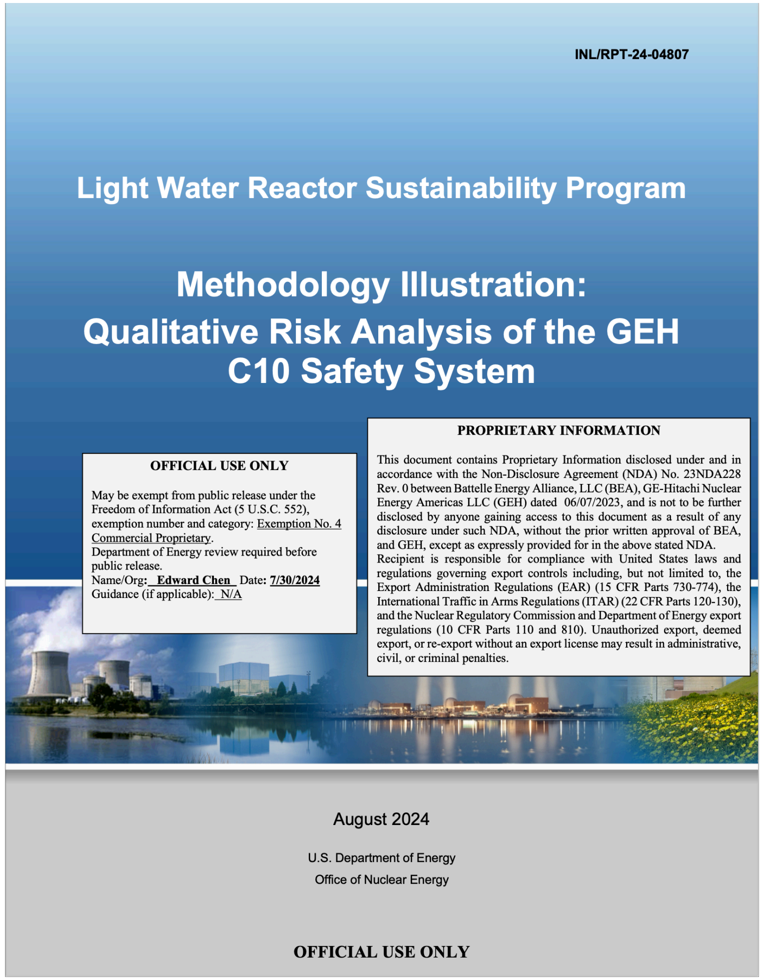
* Corresponding author.
E-mail address: Edward.chen@inl.gov (E. Chen).

https://doi.org/10.1016/j.ress.2024.110266
Received 15 December 2023; Received in revised form 29 May 2024; Accepted 7 June 2024
Available online 8 June 2024

Major Accomplishments in FY-24 (II)

- Collaboration with PWROG for CCFs quantification for DI&C system
  - Results were included in a proprietary technical white paper.

- Initiated collaboration with GE Hitachi for function-based risk assessment of multi-function DI&C systems.
  - Results were included in August M3 technical report.


INL/RPT-24-04807

Light Water Reactor Sustainability Program

Methodology Illustration:
Qualitative Risk Analysis of the GEH C10 Safety System

OFFICIAL USE ONLY

May be exempt from public release under the Freedom of Information Act (5 U.S.C. 552), exemption number and category: Exemption No. 4 Commercial Proprietary.
Department of Energy review required before public release.
Name/Org: Edward Chen Date: 7/30/2024
Guidance (if applicable): N/A

PROPRIETARY INFORMATION

This document contains Proprietary Information disclosed under and in accordance with the Non-Disclosure Agreement (NDA) No. 23NDA228 Rev. 0 between Battelle Energy Alliance, LLC (BEA), GE-Hitachi Nuclear Energy Americas LLC (GEH) dated 06/07/2023, and is not to be further disclosed by anyone gaining access to this document as a result of any disclosure under such NDA, without the prior written approval of BEA, and GEH, except as expressly provided for in the above stated NDA. Recipient is responsible for compliance with United States laws and regulations governing export controls including, but not limited to, the Export Administration Regulations (EAR) (15 CFR Parts 730-774), the International Traffic in Arms Regulations (ITAR) (22 CFR Parts 120-130), and the Nuclear Regulatory Commission and Department of Energy export regulations (10 CFR Parts 110 and 810). Unauthorized export, deemed export, or re-export without an export license may result in administrative, civil, or criminal penalties.

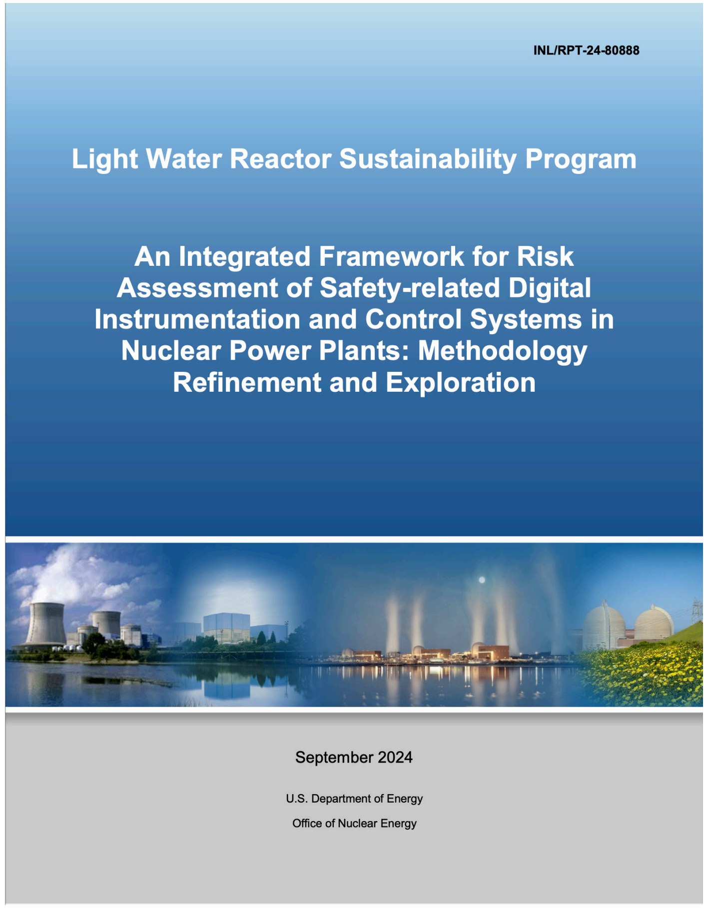
August 2024

U.S. Department of Energy
Office of Nuclear Energy

OFFICIAL USE ONLY

Major Accomplishments in FY-24 (III)

- Refined the reliability analysis methods of a safety-critical DI&C system
  - Developed a novel approach to evaluate inter-system CCFs in highly redundant and diverse DI&C systems
  - Root cause correlation analysis via ORCAS and natural language processing (NLP)
  - Leverage LLM to perform hazard analysis and reliability analysis
  - Methodology improvement and a user guidance for industry use was included in the September M2 technical report.


INL/RPT-24-80888

Light Water Reactor Sustainability Program

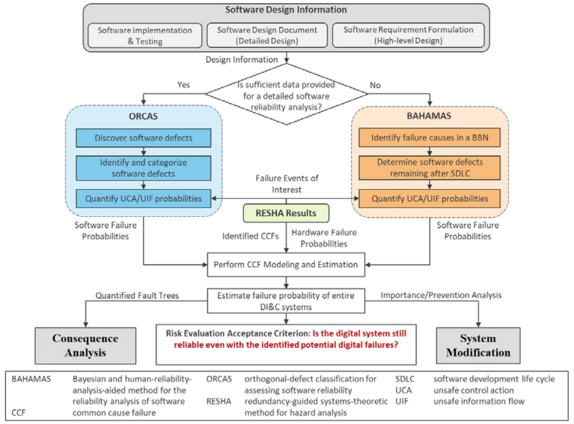
An Integrated Framework for Risk Assessment of Safety-related Digital Instrumentation and Control Systems in Nuclear Power Plants: Methodology Refinement and Exploration

September 2024

U.S. Department of Energy
Office of Nuclear Energy

Roadmap: From Risk Assessment to Design Optimization and Licensing

LWRS-developed DI&C Risk Assessment Framework



Industry

Risk-informed Design Optimization

NRC

DI&C PRA or Licensing Standards

Safety Assurance Case (SAC)

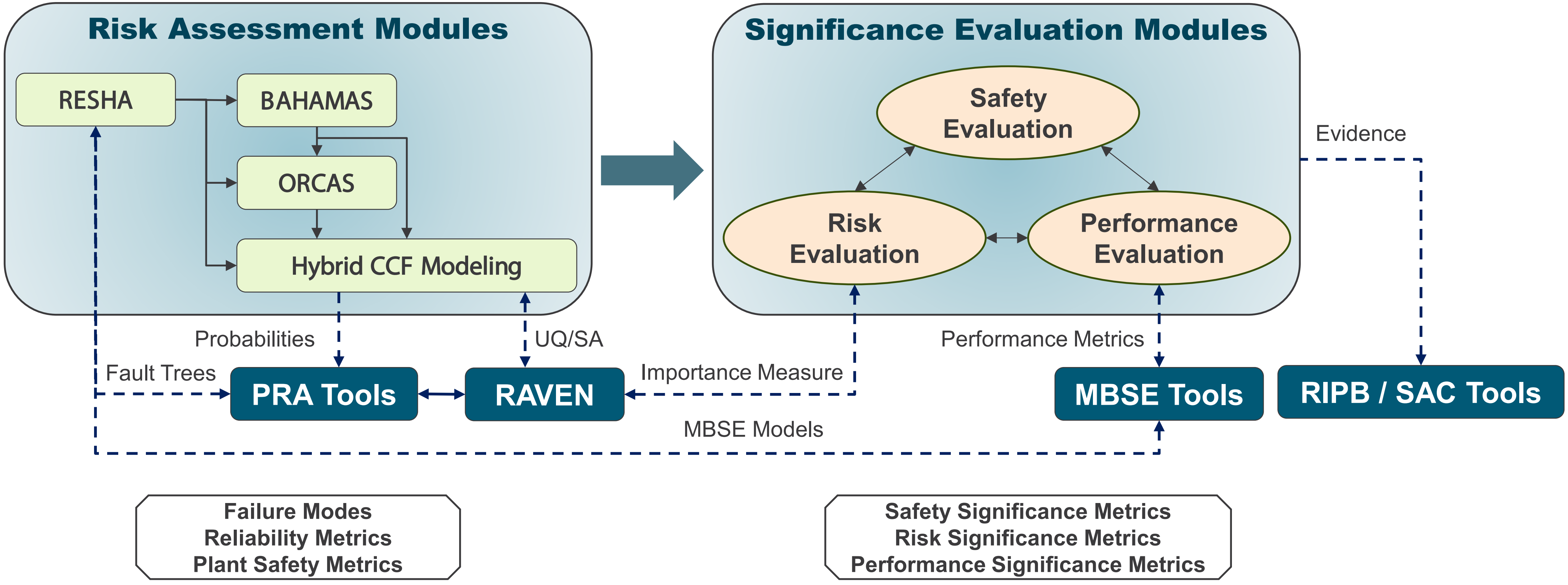
Risk-informed Performance-based (RIPB) evaluation

Diversity and Defense-in-Depth (D3) Application

*PRA: probabilistic risk assessment

Roadmap: Risk Assessment Framework Development

- Software for Hazard Identification and Evaluation of Digital Systems (SHIELDS)



Research Activities in FY-25

- Improve and further develop the current methods for risk assessment of multi-function DI&C systems
  - Keep supporting the need of DI&C reliability analysis from the industry.
- Refine the current methods:
  - Intra- and inter-system CCF modeling
  - Align better with international standards and existing risk-informed approaches and guides.
- SHIELDS framework development
- Develop capabilities on risk-informed evidence generation and evaluation to support DI&C safety assurance and design optimization with the industry and other research institutions.
- Develop novel approaches to inform risk management and design optimization of advanced (semi-) autonomous DI&C systems designed for existing LWR fleets.

Collaborators

- **PWROG engagement**: Digital I&C reliability analysis and CCF evaluation
- **GE Hitachi**: function-based risk assessment of multi-function DI&C platforms
- **KAERI:** safety analysis of advanced DI&C technologies
- **NEI/Halden**: DI&C D3 application and safety assurance
- **NRC**: Risk evaluation and design optimization of AI-aided DI&C systems
- **North Carolina State University**:
  - DI&C hazard analysis using large language model
  - CCF analysis and parameter estimation using model-based approaches.

LWRS
LIGHT WATER REACTOR SUSTAINABILITY

Sustaining National Nuclear Assets

lwrs.inl.gov