From Cold Start to Critical: Formal Synthesis of Hybrid Controllers

PI: Dane A. Sabo
dane.sabo@pitt.edu

Advisor: Dr. Daniel G. Cole
dgcole@pitt.edu

Monday 20th October, 2025

Project Summary

Overview

This research will develop a methodology for creating autonomous hybrid control systems with mathematical guarantees of safe and correct behavior. Nuclear power plants require the highest levels of control system reliability, where failures can result in significant economic losses or radiological release. Currently, nuclear operations rely on extensively trained human operators who follow detailed written procedures to manage reactor control. However, reliance on human operators prevents introduction of autonomous control capabilities and creates fundamental economic challenges for next-generation reactor designs. Without introducing automation, emerging technologies like small modular reactors face significantly higher per-megawatt staffing costs than conventional plants, threatening their economic viability.

To address this need, we will combine formal methods from computer science with control theory to build hybrid control systems that are correct by construction. Hybrid systems use discrete logic to switch between continuous control modes, similar to how operators change control strategies. Existing formal methods can generate provably correct switching logic from written requirements, but they cannot handle the continuous dynamics that occur during transitions between modes. Meanwhile, traditional control theory can verify continuous behavior but lacks tools for proving correctness of discrete switching decisions. By synthesizing discrete mode transitions directly from written operating procedures and verifying continuous behavior between transitions, we can create hybrid control systems with end-to-end correctness guarantees.

Intellectual Merit

The intellectual merit lies in unifying discrete synthesis and continuous verification to enable end-to-end correctness guarantees for hybrid systems. This research will advance knowledge by developing a systematic, tool-supported methodology for translating written procedures into temporal logic, synthesizing provably correct discrete switching logic, and developing verified continuous controllers. The approach addresses a fundamental gap in hybrid system design by bridging formal methods from computer science and control theory.

Broader Impacts

This research directly addresses the multi-billion dollar operations and maintenance cost challenge facing nuclear power deployment. By synthesizing provably correct hybrid controllers, we can automate routine operational sequences that currently require constant human oversight, enabling a shift from direct operator control to supervisory monitoring. Beyond nuclear applications, this research will establish a generalizable framework for autonomous control of safety-critical systems including chemical process control, aerospace systems, and autonomous transportation.

Research Description

1 Objectives

The goal of this research is to develop a methodology for creating autonomous control systems with event-driven control laws that have guarantees of safe and correct behavior.

Nuclear power relies on extensively trained operators who follow detailed written procedures to manage reactor control. Based on these procedures and operators' interpretation of plant conditions, operators make critical decisions about when to switch between control objectives. While human operators have maintained the nuclear industry's exceptional safety record, reliance on human operators has created an economic challenge for next-generation nuclear power plants. Small modular reactors face significantly higher per-megawatt staffing costs than conventional plants, threatening their economic viability. Autonomous control systems are needed that can safely manage complex operational sequences with the same assurance as human-operated systems, but without constant supervision.

To address this need, we will combine formal methods from computer science with control theory to build hybrid control systems that are correct by construction. Hybrid systems use discrete logic to switch between continuous control modes, similar to how operators change control strategies. Existing formal methods generate provably correct switching logic but cannot handle continuous dynamics during transitions, while traditional control theory verifies continuous behavior but lacks tools for proving discrete switching correctness. We will bridge this gap through a three-stage methodology. First, we will translate written operating procedures into temporal logic specifications using NASA's Formal Requirements Elicitation Tool (FRET), which structures requirements into scope, condition, component, timing, and response elements. This structured approach enables realizability checking to identify conflicts and ambiguities in procedures before implementation. Second, we will synthesize discrete mode switching logic from these specifications using reactive synthesis tools such as Strix, which generates deterministic automata that are provably correct by construction. Third, we will develop and verify continuous controllers for each discrete mode using standard control theory and reachability analysis. We will classify continuous modes based on their transition objectives, and then employ assume-guarantee contracts and barrier certificates to prove that mod

From Cold Start to Critical: Formal Synthesis of Hybrid Controllers for Nuclear Reactor Startup

White Paper

NSF Program: Cyber-Physical Systems (CPS)

Principal Investigator: Dane A. Sabo
Email: dane.sabo@pitt.edu

Faculty Advisor: Dr. Daniel G. Cole

Email: dgcole@pitt.edu

University of Pittsburgh
Department of Mechanical Engineering and Materials Science
Swanson School of Engineering

Project Summary

Overview

This research will develop a methodology for creating autonomous hybrid control systems with mathematical guarantees of safe and correct behavior. Nuclear power plants require the highest levels of control system reliability, where failures can result in significant economic losses or radiological release. Currently, nuclear operations rely on extensively trained human operators who follow detailed written procedures to manage reactor control. However, reliance on human operators prevents introduction of autonomous control capabilities and creates fundamental economic challenges for next-generation reactor designs. Emerging technologies like small modular reactors face significantly higher per-megawatt staffing costs than conventional plants, threatening their economic viability.

To address this need, we will combine formal methods from computer science with control theory to build hybrid control systems that are correct by construction. Hybrid systems use discrete logic to switch between continuous control modes, similar to how operators change control strategies. Existing formal methods can generate provably correct switching logic from written requirements, but they cannot handle the continuous dynamics that occur during transitions between modes. Meanwhile, traditional control theory can verify continuous behavior but lacks tools for proving correctness of discrete switching decisions. By synthesizing discrete mode transitions directly from written operating procedures and verifying continuous behavior between transitions, we can create hybrid control systems with end-to-end correctness guarantees.

Intellectual Merit

The intellectual merit lies in unifying discrete synthesis and continuous verification to enable end-to-end correctness guarantees for hybrid systems. This research will advance knowledge by developing a systematic, tool-supported methodology for translating written procedures into temporal logic, synthesizing provably correct discrete switching logic, and developing verified continuous controllers. The approach addresses a fundamental gap in hybrid system design by bridging formal methods from computer science and control theory.

Broader Impacts

This research directly addresses the multi-billion dollar operations and maintenance cost challenge facing nuclear power deployment. By synthesizing provably correct hybrid controllers, we can automate routine operational sequences that currently require constant human oversight, enabling a shift from direct operator control to supervisory monitoring. Beyond nuclear applications, this research will establish a generalizable framework for autonomous control of safety-critical systems including chemical process control, aerospace systems, and autonomous transportation.

Research Description

2 Objectives

The goal of this research is to develop a methodology for creating autonomous hybrid control systems with mathematical guarantees of safe and correct behavior for nuclear reactor operations. Nuclear reactors are quintessential hybrid cyber-physical systems where continuous neutron kinetics and thermal-hydraulics interact with discrete control mode decisions and trip logic. Hybrid systems combine continuous dynamics with discrete mode transitions, formally expressed as $\dot{x}(t) = f(x(t), q(t), u(t))$ for continuous states and $q(k+1) = \nu(x(k), q(k), u(k))$ for discrete transitions.

If this research is successful, we will be able to do the following:

1. *Synthesize written procedures into verified control logic.* We will develop a methodology for converting written operating procedures into formal specifications. These specifications will be synthesized into discrete control logic using reactive synthesis tools. This process uses structured intermediate representations to bridge natural language and mathematical logic. Control engineers will be able to generate mode-switching controllers from regulatory procedures with little formal methods expertise, reducing barriers to high-assurance control systems.
2. *Verify continuous control behavior across mode transitions.* We will develop methods using reachability analysis to ensure continuous control modes satisfy discrete transition requirements. Engineers will be able to design continuous controllers using standard practices while ensuring system correctness and proving mode transitions occur safely at the right times.
3. *Demonstrate autonomous reactor startup control with safety guarantees.* We will implement this methodology on a small modular reactor simulation using industry-standard control hardware. This trial will include multiple coordinated control modes from cold shutdown through criticality to power operation on a SmAHTR reactor simulation in a hardware-in-the-loop experiment. Control engineers will be able to implement high-assurance autonomous controls on industrial platforms they already use, enabling users to achieve autonomy without retraining costs or developing new equipment.

3 Limits of Current Practice

Nuclear reactor control reveals fundamental verification gaps that motivate formal hybrid control synthesis. These gaps span current operational practices, human reliability, and even the most advanced formal methods attempts to date.

Current generation nuclear power plants employ over 3,600 active NRC-licensed reactor operators who hold legal authority to make critical decisions including departing from regulations during emergencies. This authority is both necessary and problematic. The Three Mile Island accident demonstrated how personnel error led to partial meltdown when operators misread confusing readings and shut off emergency systems. The President's Commission identified a fundamental ambiguity: operators hold full responsibility without formal verification they can fulfill this under all conditions.

Nuclear plant procedures exist in a hierarchy from normal operating procedures through Emergency Operating Procedures to Severe Accident Management Guidelines. Despite rigorous development processes including technical evaluation and simulator validation testing, these procedures

fundamentally lack formal verification. No mathematical proof exists that procedures cover all possible plant states or that required actions can be completed within available timeframes. This gap between procedural rigor and mathematical certainty creates the first major limitation of current practice.

The division between automated and human-controlled functions reveals the fundamental hybrid control challenge. Highly automated systems handle reactor protection and emergency core cooling, while human operators retain strategic decision-making. Current practice treats continuous plant dynamics and discrete control logic as separate concerns without formal integration. No application of hybrid control theory exists that could provide mathematical guarantees across mode transitions.

Human factors compound these verification gaps. Multiple independent analyses show that 70–80% of all nuclear power plant events are attributed to human error rather than equipment failures. More significantly, the IAEA concluded that human error was the root cause of all severe accidents at nuclear power plants. The persistence of this ratio despite four decades of improvements suggests fundamental cognitive limitations rather than remediable deficiencies.

The Three Mile Island accident provides the definitive case study. When a pressure-operated relief valve stuck open, instrumentation showed only commanded position, not actual position. This information gap proved decisive. When Emergency Core Cooling pumps automatically activated, operators shut them down based on incorrect assessment. Operators faced over 100 simultaneous alarms, overwhelming their cognitive capacity. The core suffered 44% fuel meltdown before stabilization.

Human Reliability Analysis methods quantify these limitations. Nominal Human Error Probabilities stand at 0.01 (1%) for diagnosis tasks under optimal conditions. These rates degrade dramatically under accident conditions: inadequate time increases error probability 10-fold, extreme stress by 5-fold, high complexity by 5-fold. Under combined adverse conditions, human error probabilities approach 0.1 to 1.0—essentially guaranteed failure.

The underlying causes are fundamental and cannot be overcome through training alone. Response time limitations constrain effectiveness. Visual perception requires 100-200 milliseconds, decisions 200-400 milliseconds. Reactor transients evolve in seconds. Protection systems must respond in milliseconds, 100-1000 times faster than humans. Working memory capacity is limited to 7 ± 2 chunks, explaining why TMI's 100+ alarms exceeded operators' processing capacity. These are not training failures—they are fundamental properties of human cognition.

Recent efforts to apply formal methods to nuclear control show both promise and remaining gaps. The High Assurance Rigorous Digital Engineering for Nuclear Safety (HARDENS) project represents the most advanced application to date. Completed in nine months at a fraction of typical costs, HARDENS produced a complete Reactor Trip System with full traceability from NRC requirements through formal specifications to verified binaries.

The project employed impressive formal methods. FRET handled requirements elicitation. Cryptol provided executable specifications. SAW performed verification. Automatic code synthesis generated formally verifiable implementations. This comprehensive approach demonstrated that formal methods are technically feasible and economically viable for nuclear protection systems.

Despite these accomplishments, HARDENS has a fundamental limitation directly relevant to our work. The project addressed only discrete digital control logic without modeling or verifying continuous reactor dynamics. Real reactor safety depends on interaction between continuous

processes—temperature, pressure, neutron flux—and discrete control decisions. HARDENS verified the discrete controller in isolation but not the closed-loop hybrid system behavior.

Experimental validation presents the second major limitation. HARDENS produced a demonstrator at Technology Readiness Level 3–4 rather than a deployment-ready system. The gap between formal verification and actual deployment involves integration with legacy systems, long-term reliability under harsh environments, and regulatory acceptance of formal methods as primary assurance evidence.

These three converging lines reveal the research imperative. Current practice lacks formal verification of procedures and mode transitions. Human operators contribute to 70–80% of incidents despite continuous improvements, suggesting fundamental rather than remediable limitations. HARDENS demonstrated feasibility of formal methods but addressed only discrete logic, leaving hybrid dynamics unverified. The continuous dynamics of reactor physics combined with discrete control logic demand hybrid automata or differential dynamic logic that can verify properties spanning both domains. This gap defines the research opportunity.

4 Research Approach

This research will overcome the identified limitations by combining formal methods from computer science with control theory to build hybrid control systems that are correct by construction. We accomplish this through three main thrusts that progress from written procedures to verified implementations.

Commercial nuclear power operations remain manually controlled despite significant advances in control systems. The key insight is that procedures performed by human operators are highly prescriptive and well-documented. Written procedures in nuclear power are sufficiently detailed that we may translate them into logical formulae with minimal effort. This translation forms the foundation of our approach.

To formalize these procedures, we will use temporal logic, which captures system behaviors through temporal relations. Linear Temporal Logic provides four fundamental operators: next (X), eventually (F), globally (G), and until (U). These operators enable precise specification of time-dependent requirements.

Consider a nuclear reactor SCRAM requirement: "If a high temperature alarm triggers, control rods must immediately insert and remain inserted until operator reset." This natural language statement translates into temporal logic as: $G(HighTemp \rightarrow X(RodsInserted \wedge (\neg RodsWithdrawn\ U\ OperatorReset))$ The specification precisely captures the temporal relationship between alarm, response, and persistence requirement.

The most efficient path to accomplish this translation is through NASA's Formal Requirements Elicitation Tool (FRET). FRET employs FRETish, a specialized requirements language that restricts requirements to easily understood components while eliminating ambiguity. FRET enforces structure by requiring all requirements to contain six components: Scope, Condition, Component, Shall, Timing, and Response.

FRET provides functionality to check realizability of a system. Realizability analysis determines whether written requirements are complete by examining the six structural components. Complete requirements neither conflict with one another nor leave any behavior undefined. Systems that are not realizable contain behavioral inconsistencies that represent the physical equivalent of software bugs. Using FRET during autonomous controller development allows us to identify and resolve these errors systematically. FRET exports requirements in temporal logic format com-

patible with reactive synthesis tools, enabling the second thrust of our approach.

Reactive synthesis is an active research field focused on generating discrete controllers from temporal logic specifications. The term reactive indicates that the system responds to environmental inputs to produce control outputs. These synthesized systems are finite in size, where each node represents a unique discrete state. The connections between nodes, called state transitions, specify the conditions under which the discrete controller moves from state to state. This complete mapping constitutes a discrete automaton.

We will employ state-of-the-art reactive synthesis tools, particularly Strix, which has demonstrated superior performance in the Reactive Synthesis Competition through efficient parity game solving algorithms. Strix translates linear temporal logic specifications into deterministic automata automatically while maximizing generated automata quality. Once constructed, the automaton can be straightforwardly implemented using standard programming control flow constructs.

The discrete automata representation yields an important theoretical guarantee. Because the discrete automaton is synthesized entirely through automated tools from design requirements and operating procedures, we can prove that the automaton—and therefore our hybrid switching behavior—is correct by construction. This correctness guarantee is paramount. Mode switching represents the primary responsibility of human operators in control rooms today. Human operators possess the advantage of real-time judgment—when mistakes occur, they can correct them dynamically. Autonomous control lacks this adaptive advantage. By synthesizing controllers from logical specifications with guaranteed correctness, we eliminate the possibility of switching errors.

While discrete system components will be synthesized with correctness guarantees, they represent only half of the complete system. The continuous modes will be developed after discrete automaton construction, leveraging the automaton structure and transitions to design multiple smaller, specialized continuous controllers. This progression from discrete to continuous design addresses a key challenge in hybrid system verification.

The discrete automaton transitions mark decision points for switching between continuous control modes and define their strategic objectives. We will classify three types of high-level continuous controller objectives: Stabilizing modes maintain the hybrid system within its current discrete mode, corresponding to steady-state normal operating modes like full-power load-following control. Transitory modes have the primary goal of transitioning the hybrid system from one discrete state to another, such as controlled warm-up procedures. Expulsory modes are specialized transitory modes with additional safety constraints that ensure the system is directed to a safe stabilizing mode during failure conditions, such as reactor SCRAM.

Building continuous modes after constructing discrete automata enables local controller design focused on satisfying discrete transitions. The primary challenge in hybrid system verification is ensuring global stability across transitions. Current techniques struggle with this problem because dynamic discontinuities complicate verification. This work alleviates these problems by designing continuous controllers specifically with transitions in mind. By decomposing continuous modes according to their required behavior at transition points, we avoid solving trajectories through the entire hybrid system.

To ensure continuous modes satisfy their requirements, we will employ three complementary techniques. Reachability analysis computes the reachable set of states for a given input set. We will use reachability to define continuous state ranges at discrete transition boundaries and verify that requirements are satisfied within continuous modes. Assume-guarantee contracts will be employed when continuous state boundaries are not explicitly defined. For any given mode, the input range

for reachability analysis is defined by the output ranges of discrete modes that transition to it. This compositional approach ensures each continuous controller is prepared for its possible input range. Barrier certificates will prove that mode transitions are satisfied. Control barrier functions provide a method to certify safety by establishing differential inequality conditions that guarantee forward invariance of safe sets.

Combining these three techniques will enable us to prove that continuous components satisfy discrete requirements and thus complete system behavior. To demonstrate this methodology, we will develop an autonomous startup controller for a Small Modular Advanced High Temperature Reactor (SmAHTR). SmAHTR represents an ideal test case with well-documented startup procedures that must transition through multiple distinct operational modes: initial cold conditions, controlled heating to operating temperature, approach to criticality, low-power physics testing, and power ascension to full operating capacity.

We have already developed a high-fidelity SmAHTR model in Simulink that captures the thermal-hydraulic and neutron kinetics behavior. The synthesized hybrid controller will be implemented on an Emerson Ovation control system platform, which is representative of industry-standard control hardware. The Advanced Reactor Cyber Analysis and Development Environment (ARCADE) suite will serve as the integration layer, managing real-time communication between the Simulink simulation and the Ovation controller. This hardware-in-the-loop configuration enables validation of the controller implementation on actual industrial control equipment.

5 Metrics of Success

This research will be measured by advancement through Technology Readiness Levels, progressing from fundamental concepts to validated prototype demonstration. The work begins at TRL 2–3 and aims to reach TRL 5, where system components operate successfully in a relevant laboratory environment.

TRLs provide the ideal success metric because they explicitly measure the gap between academic proof-of-concept and practical deployment. This gap is precisely what our work aims to bridge. TRLs capture both theoretical rigor and practical feasibility simultaneously. The nuclear industry already uses TRLs for technology assessment, making this metric directly relevant to potential adopters.

Moving from current state (TRL 2–3) to target (TRL 5) requires achieving three intermediate levels. TRL 3 demonstrates that each component works in isolation. Startup procedures must be translated into temporal logic specifications that pass realizability analysis. A discrete automaton must be synthesized with interpretable structure. At least one continuous controller must be designed with verified transition requirements. This level proves the fundamental approach on simplified sequences.

TRL 4 demonstrates a complete integrated hybrid controller in simulation. All startup procedures must be formalized with continuous controllers existing for all discrete modes. Verification must be complete for all mode transitions. The integrated controller must execute complete startup sequences with zero safety violations. This level proves that formal correctness guarantees can be maintained throughout system integration.

TRL 5 demonstrates the verified controller on industrial control hardware through hardware-in-the-loop testing. The discrete automaton must be implemented on Emerson Ovation hardware and verified to match synthesized specifications exactly. The ARCADE interface must establish stable real-time communication between Ovation hardware and SmAHTR simulation. Complete

autonomous startup sequences must execute across the full operational envelope. This level proves that the methodology produces verified controllers implementable with current technology.

This research succeeds if it achieves TRL 5, establishing both theoretical validity and practical feasibility. Success provides a clear pathway for nuclear industry adoption and broader application to safety-critical autonomous systems.

6 Broader Impacts

Nuclear power presents both a compelling application domain and an urgent economic challenge. Recent interest in powering artificial intelligence infrastructure has renewed focus on small modular reactors for hyperscale datacenters. According to the U.S. Energy Information Administration, advanced nuclear power entering service in 2027 is projected to cost $88.24 per megawatt-hour. With datacenter electricity demand projected to reach 1,050 terawatt-hours annually by 2030, operations and maintenance costs represent approximately 23–30% of total levelized cost, translating to $21–28 billion annually for projected datacenter demand.

This research directly addresses the multi-billion dollar O&M cost challenge. Current nuclear operations require full control room staffing for each reactor. These staffing requirements drive high O&M costs, particularly for smaller reactor designs where the same overhead must be spread across lower power output. The economic burden threatens the viability of next-generation nuclear technologies.

By synthesizing provably correct hybrid controllers, we can automate routine operational sequences that currently require constant human oversight. This enables a fundamental shift from direct operator control to supervisory monitoring where operators oversee multiple autonomous reactors rather than manually controlling individual units. The transition fundamentally changes the economics of nuclear operations.

The correct-by-construction methodology is critical for this transition. Traditional automation approaches cannot provide sufficient safety guarantees for nuclear applications where regulatory requirements and public safety concerns demand the highest levels of assurance. By formally verifying both discrete mode-switching logic and continuous control behavior, this research will produce controllers with mathematical proofs of correctness. These guarantees enable automation to safely handle routine operations that currently require human operators to follow written procedures.

Small modular reactors represent an ideal deployment target. Nuclear Regulatory Commission certification requires extensive documentation of control procedures, operational requirements, and safety analyses written in structured natural language. These regulatory documents can be translated into temporal logic specifications using FRET, synthesized into discrete switching logic using reactive synthesis tools, and verified using reachability analysis and barrier certificates. The infrastructure of requirements is already complete as part of licensing. This creates a direct pathway from existing regulatory documentation to formally verified autonomous controllers.

Beyond nuclear applications, this research will establish a generalizable framework for autonomous control of safety-critical systems. The methodology of translating operational procedures into formal specifications, synthesizing discrete switching logic, and verifying continuous mode behavior applies to any hybrid system with documented operational requirements. Potential applications include chemical process control, aerospace systems, and autonomous transportation. These domains share similar economic and safety considerations that favor increased autonomy with provable correctness guarantees. By demonstrating this approach in nuclear power—one of

the most regulated and safety-critical domains—this research will establish both technical feasibility and regulatory pathways for broader adoption across critical infrastructure.

References Cited

References

[1] U.S. Nuclear Regulatory Commission. “10 CFR Part 55 - Operators’ Licenses.” *Code of Federal Regulations*, 2024.

[2] Princeton University. “Nuclear Reactor Operator Training.” *Princeton Plasma Physics Laboratory*, 2023.

[3] J. G. Kemeny et al. “Report of the President’s Commission on the Accident at Three Mile Island.” U.S. Government Printing Office, October 1979.

[4] U.S. Nuclear Regulatory Commission. “Guidelines for the Preparation of Emergency Operating Procedures.” NUREG-0899, August 1982.

[5] U.S. Department of Energy. “Human Performance Improvement Handbook.” DOE-HDBK-1028-2009, June 2009.

[6] World Nuclear Association. “Safety of Nuclear Power Reactors.” *World Nuclear Association Information Library*, 2020.

[7] International Atomic Energy Agency. “Lessons Learned from Severe Nuclear Accidents.” IAEA Technical Report, 2016.

[8] Y. Wang et al. “Human Error Analysis of 190 Events at Chinese Nuclear Power Plants from 2007-2020.” *Nuclear Engineering and Design*, vol. 395, 2025.

[9] J. Kiniry et al. “High Assurance Rigorous Digital Engineering for Nuclear Safety (HARDENS).” NRC Final Technical Report ML22326A307, September 2022.

[10] U.S. Energy Information Administration. “Levelized Costs of New Generation Resources in the Annual Energy Outlook 2022.” Report, March 2022.

[11] Environmental and Energy Study Institute. “Data Center Energy Needs are Upending Power Grids and Threatening the Climate.” Web article, 2024.

e transitions occur safely and as defined by the