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

Wednesday 8th October, 2025

1 Goals and Outcomes

The goal of this research is to 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, service interruptions, or radiological release. Currently, nuclear plant operations rely on extensively trained human operators who follow detailed written procedures and strict regulatory requirements to manage reactor control. These operators make critical decisions about when to switch between different control modes based on their interpretation of plant conditions and procedural guidance. However, this reliance on human operators prevents the introduction of autonomous control capabilities and creates a fundamental economic challenge 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. What is needed is a way to create autonomous control systems that can safely manage complex operational sequences with the same level of assurance as human-operated systems, but without requiring constant human 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 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. If we can formalize existing procedures into logical specifications and verify that continuous dynamics satisfy transition requirements, then we can build autonomous controllers that are provably free from design defects. This approach will enable autonomous control in nuclear power plants while maintaining the high safety standards required by the industry. This work is conducted within the University of Pittsburgh Cyber Energy Center, which provides access to industry collaboration and Emerson control hardware, ensuring that solutions developed are aligned with practical implementation requirements.

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

1. **Translate written procedures into verified control logic.** We will develop a methodology for converting existing written operating procedures into formal specifications that can be automatically synthesized into discrete control logic. This process will use structured intermediate representations to bridge natural language procedures and mathematical logic. Control system engineers will be able to generate verified mode-switching controllers directly from regulatory procedures without requiring expertise in formal methods, reducing the barrier to creating high-assurance control systems.
2. **Verify continuous control behavior across mode transitions.** We will establish methods for analyzing continuous control modes to ensure they satisfy the discrete transition requirements. Using a combination of classical control theory for linear systems and reachability analysis for nonlinear dynamics, we will verify that each continuous mode can safely reach its intended transitions. Engineers will be able to design continuous controllers using

standard practices while iterating to ensure broader system correctness, proving that mode transitions occur safely and at the right times.

3. **Demonstrate autonomous reactor startup control with safety guarantees.** We will apply this methodology to develop an autonomous controller for nuclear reactor startup procedures, implementing it on a small modular reactor simulation using industry-standard control hardware. This demonstration will prove correctness across multiple coordinated control modes from cold shutdown through criticality to power operation. We will provide evidence that autonomous hybrid control can be realized in the nuclear industry with current control equipment, establishing a path toward reducing operator staffing requirements while maintaining safety.

The innovation in this work is the unification of discrete synthesis and continuous verification to enable end-to-end correctness guarantees for hybrid systems. If successful, control engineers will be able to create autonomous controllers from existing procedures with mathematical proof of correct behavior. High-assurance autonomous control will become practical for safety-critical applications. This capability is essential for the economic viability of next-generation nuclear power. Small modular reactors represent a promising solution to growing energy demands, but their success depends on reducing per-megawatt operating costs through increased autonomy. This research will provide the tools to achieve that autonomy while maintaining the exceptional safety record required by the nuclear industry.

2 State of the Art and Limits of Current Practice

UNDER CONSTRUCTION

Basically this section is going to talk about:

1. How operating procedures are written today
2. How nuclear operators are trained and what their jobs are
3. HARDENS - an early work trying to build a reactor emergency shutdown system with formal methods, by doing a lot of this translation stuff.

Some key limits are:

1. Operating procedures are written in natural language. This makes them unavoidable ambiguous and leaves instructions up to interpretation
2. Human operators can make human errors. Discuss how most nuclear accidents are actually people driven, and not the fault of the plant itself.
3. HARDENS does not consider continuous dynamics, nor did they really test anything to validate their system works. Dan says TRL 3. I begrudgingly agree.

3 Research Approach

This research will overcome the limitations of current practice to build high-assurance hybrid control systems for critical infrastructure. Hybrid systems combine continuous dynamics (flows) with discrete transitions (jumps), which can be formally expressed as:

$$\dot{x}(t) = f(x(t), q(t), u(t)) \tag{1}$$

$$q(k+1) = \nu(x(k), q(k), u(k)) \tag{2}$$

Here, $f(\cdot)$ defines the continuous dynamics while $\nu(\cdot)$ governs discrete transitions. The continuous states x, discrete state q, and control input u interact to produce hybrid behavior. The discrete state q defines which continuous dynamics mode is currently active. Our focus centers on continuous autonomous hybrid systems, where continuous states remain unchanged during jumps—a property naturally exhibited by physical systems. For example, a nuclear reactor switching from warm-up to load-following control cannot instantaneously change its temperature or control rod position, but can instantaneously change control laws.

To build these systems with formal correctness guarantees, we must accomplish three main thrusts:

1. Translate operating procedures and requirements into temporal logic formulae
2. Create the discrete half of a hybrid controller using reactive synthesis
3. Develop continuous controllers to operate between modes, and verify their correctness using reachability analysis

The following sections discuss how these thrusts will be accomplished.

3.1 $(Procedures \wedge FRET) \rightarrow TemporalSpecifications$

The motivation behind this work stems from the fact that commercial nuclear power operations remain manually controlled by human operators, despite significant advances in control systems sophistication. The key insight is that procedures performed by human operators are highly prescriptive and well-documented. This suggests that human operators in nuclear power plants may not be entirely necessary given today's available technology.

Written procedures and requirements in nuclear power are sufficiently detailed that we may be able to translate them into logical formulae with minimal effort. If successful, this approach would enable automation of existing procedures without requiring system reengineering. To formalize these procedures, we will use temporal logic, which captures system behaviors through temporal relations. Linear Temporal Logic (LTL) 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 expressed in natural language: *"If a high temperature alarm triggers, control rods must immediately insert and remain inserted until operator reset."* This plain language requirement can be translated into a rigorous logical specification:

$$G(HighTemp \rightarrow X(RodsInserted \wedge (\neg RodsWithdrawn\ U\ OperatorReset))) \tag{3}$$

This specification precisely captures the temporal relationship between the alarm condition, the required response, and the persistence requirement. The global operator G ensures this property holds throughout system operation, while the next operator X enforces immediate response. The until operator U maintains the state constraint until the reset condition occurs.

The most efficient path to accomplish this translation is through NASA's Formal Requirements Elicitation Tool (FRET). FRET employs a specialized requirements language called FRETish that restricts requirements to easily understood components while eliminating ambiguity. FRETish

bridges natural language and mathematical specifications through a structured English-like syntax that is automatically translatable to temporal logic.

FRET enforces this structure by requiring all requirements to contain six components:

1. Scope: *What modes does this requirement apply to?*
2. Condition: *Scope plus additional specificity*
3. Component: *What system element does this requirement affect?*
4. Shall
5. Timing: *When does the response occur?*
6. Response: *What action should be taken?*

FRET provides functionality to check the *realizability* of a system. Realizability analysis determines whether written requirements are complete by examining the six structural components. Complete requirements are those that neither conflict with one another nor leave any behavior undefined. Systems that are not realizable from their procedure definitions and design requirements present problems beyond autonomous control implementation. Such systems 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.

The second category of realizability issues involves undefined behaviors that are typically left to human judgment during control operations. This ambiguity is undesirable for high-assurance systems, since even well-trained humans remain prone to errors. By addressing these specification gaps in FRET during autonomous controller development, we can deliver controllers free from these vulnerabilities.

FRET provides the capability to export requirements in temporal logic format compatible with reactive synthesis tools. This export functionality enables progression to the next step of our approach: synthesizing discrete mode switching behavior from the formalized requirements.

3.2 $(TemporalLogic \wedge ReactiveSynthesis) \rightarrow DiscreteAutomata$

Reactive synthesis is an active research field in computer science 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 of possible states and transitions constitutes a *discrete automaton*. Discrete automata can be represented graphically as a series of nodes that are discrete states, with traces indicating transitions between states. From the automaton graph, it becomes possible to fully describe the dynamics of the discrete system and develop intuitive understanding of system behavior. Hybrid systems naturally exhibit discrete behavior amenable to formal analysis through these finite state representations.

We will employ state-of-the-art reactive synthesis tools, particularly Strix, which has demonstrated superior performance in the Reactive Synthesis Competition (SYNTCOMP) 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 graphical representation provided by the automaton enables inspection and facilitates communication with controls programmers who may not have formal methods expertise.

We will use discrete automata to represent the switching behavior of our hybrid system. This approach 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*. Correctness of the switching controller is paramount to this work. 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 with capabilities that extend beyond written procedures. Autonomous control lacks this adaptive advantage. Instead, we must ensure that autonomous controllers replacing human operators will not make switching errors between continuous modes. By synthesizing controllers from logical specifications with guaranteed correctness, we eliminate the possibility of switching errors.

3.3 $(DiscreteAutomata \wedge ControlTheory \wedge Reachability) \rightarrow ContinuousModes$

While discrete system components will be synthesized with correctness guarantees, they represent only half of the complete system. Autonomous controllers like those we are developing exhibit continuous dynamics within discrete states, as described by $f(\cdot)$ in Equation 1. This section describes how we will develop continuous control modes, verify their correctness, and address the unique verification challenges of hybrid systems.

The approach described for producing discrete automata yields physics-agnostic specifications that represent only half of a complete hybrid autonomous controller. These automata alone cannot define the full behavior of the control systems we aim to construct. The continuous modes will be developed after discrete automaton construction, leveraging the automaton structure and transitions to design multiple smaller, specialized continuous controllers.

The discrete automaton transitions are key to the supervisory behavior of the autonomous controller. These 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 based on discrete mode transitions:

1. **Stabilizing:** A stabilizing control mode has one primary objective: maintaining the hybrid system within its current discrete mode. This corresponds to steady-state normal operating modes, such as a full-power load-following controller in a nuclear power plant. Stabilizing modes can be identified from discrete automata as nodes with only incoming transitions.
2. **Transitory:** A transitory control mode has the primary goal of transitioning the hybrid system from one discrete state to another. In nuclear applications, this might represent a controlled warm-up procedure. Transitory modes ultimately drive the system toward a stabilizing steady-state mode. These modes may have secondary objectives within a discrete state, such as maintaining specific temperature ramp rates before reaching full-power operation.
3. **Expulsory:** An expulsory mode is a specialized transitory mode with additional safety constraints. Expulsory modes ensure the system is directed to a safe stabilizing mode during failure conditions. For example, if a transitory mode fails to achieve its intended transition, the expulsory mode activates to immediately and irreversibly guide the system toward a globally safe state. A reactor SCRAM exemplifies an expulsory continuous mode: when initiated, it must reliably terminate the nuclear reaction and direct the reactor toward stabilizing decay heat removal.

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. Instead, we can use local behavior information at transition boundaries. To ensure continuous modes satisfy their requirements, we will employ three main techniques: reachability analysis, assume-guarantee contracts, and barrier certificates.

Reachability Analysis: Reachability analysis computes the reachable set of states for a given input set. While trivial for linear continuous systems, recent advances have extended reachability to complex nonlinear systems. We will use reachability to define continuous state ranges at discrete transition boundaries and verify that requirements are satisfied within continuous modes. Recent advances using neural network approximations of Hamilton-Jacobi equations have demonstrated significant speedups while maintaining safety guarantees for high-dimensional systems, expanding the practical applicability of these methods.

Assume-Guarantee Contracts: 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, enabling subsequent reachability analysis without requiring global system analysis.

Barrier Certificates: Finally, we will use barrier certificates to prove that mode transitions are satisfied. Barrier certificates ensure that continuous modes on either side of a transition behave appropriately. Control barrier functions provide a method to certify safety by establishing differential inequality conditions that guarantee forward invariance of safe sets. For example, a barrier certificate can guarantee that a transitory mode transferring control to a stabilizing mode will always move away from the transition boundary, rather than destabilizing the target stabilizing mode.

Combining these three techniques will enable us to prove that continuous components of our hybrid controller 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 as a liquid-salt cooled reactor design 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 essential for verifying continuous controller performance under realistic plant dynamics. The synthesized hybrid controller will be implemented on an Emerson Ovation control system platform, which is representative of industry-standard control hardware deployed in modern nuclear facilities. 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 interfacing with a realistic reactor simulation, providing assessment of computational performance, real-time execution constraints, and communication latency effects. By demonstrating autonomous startup control on this representative platform, we will establish

both the theoretical validity and practical feasibility of the synthesis methodology for deployment in actual small modular reactor systems.

This unified approach addresses a fundamental gap in hybrid system design by bridging formal methods and control theory through a systematic, tool-supported methodology. By translating existing nuclear procedures into temporal logic, synthesizing provably correct discrete switching logic, and developing verified continuous controllers, we create a complete framework for autonomous hybrid control with mathematical guarantees. The result is an autonomous controller that not only replicates human operator decision-making but does so with formal assurance that switching logic is correct by construction and continuous behavior satisfies safety requirements. This methodology transforms nuclear reactor control from a manually intensive operation requiring constant human oversight into a fully autonomous system with higher reliability than human-operated alternatives. More broadly, this approach establishes a replicable framework for developing high-assurance autonomous controllers in any domain where operating procedures are well-documented and safety is paramount.

3.4 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 (SMRs), particularly for hyperscale datacenters requiring hundreds of megawatts of continuous power. Deploying SMRs at datacenter sites would minimize transmission losses and eliminate emissions from hydrocarbon-based alternatives. However, the economics of nuclear power deployment at this scale demand careful attention to operating costs.

According to the U.S. Energy Information Administration's Annual Energy Outlook 2022, advanced nuclear power entering service in 2027 is projected to cost $88.24 per megawatt-hour [1]. Datacenter electricity demand is projected to reach 1,050 terawatt-hours annually by 2030 [2]. If this demand were supplied by nuclear power, the total annual cost of power generation would exceed $92 billion. Within this figure, operations and maintenance represents a substantial component. The EIA estimates that fixed O&M costs alone account for $16.15 per megawatt-hour, with additional variable O&M costs embedded in fuel and operating expenses [1]. Combined, O&M-related costs represent approximately 23-30% of the total levelized cost of electricity, translating to $21-28 billion annually for projected datacenter demand.

This research directly addresses the multi-billion dollar O&M cost challenge through implementations of high-assurance autonomous control. Current nuclear operations require full control room staffing for each reactor, whether large conventional units or small modular designs. These staffing requirements drive the high O&M costs that make nuclear power economically challenging, particularly for smaller reactor designs where the same staffing overhead must be spread across lower power output. By synthesizing provably correct hybrid controllers from formal specifications, 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 can oversee multiple autonomous reactors rather than manually controlling individual units.

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 the discrete mode-switching logic and the continuous control behavior, this research will

produce controllers with mathematical proofs of correctness. These guarantees enable automation to safely handle routine operations—such as startup sequences, power level changes, and normal operational transitions—that currently require human operators to follow written procedures. Operators will remain in supervisory roles to handle off-normal conditions and provide authorization for major operational changes, but the routine cognitive burden of procedure execution shifts to provably correct automated systems that are much cheaper to operate.

SMRs represent an ideal deployment target for this technology. Nuclear Regulatory Commission certification requires extensive documentation of control procedures, operational requirements, and safety analyses written in structured natural language. As described in our approach, these regulatory documents can be translated into temporal logic specifications using tools like FRET, then synthesized into discrete switching logic using reactive synthesis tools, and finally verified using reachability analysis and barrier certificates for the continuous control modes. The infrastructure of requirements and specifications is already complete as part of the licensing process, creating a direct pathway from existing regulatory documentation to formally verified autonomous controllers.

Beyond reducing operating costs for new reactors, 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, where similar economic and safety considerations 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 the technical feasibility and regulatory pathway for broader adoption across critical infrastructure.

References

[1] U.S. Energy Information Administration. Levelized costs of new generation resources in the annual energy outlook 2022. Report, U.S. Energy Information Administration, March 2022. See Table 1b, page 9.

[2] Environmental and Energy Study Institute. Data center energy needs are upending power grids and threatening the climate. Web article, 2024. Accessed: 2025-09-29.