Runtime Power Feasibility for Small Modular Nuclear Reactor Powered Datacenters

Shania Ziegler (Undergraduate Student)
Brigham Young University
Provo, Utah, USA
sziegle@byu.edu

Rohan Basu Roy (Advisor)
University of Utah
Salt Lake City, Utah, USA
rohan.basuroy@utah.edu

SMR-Powered datacenters. AI buildouts have pushed single-campus datacenter demand toward the gigawatt scale, and grid interconnection at that size now takes years. Operators have responded by siting generation on the campus itself, and small modular reactors (SMRs) lead the field because they supply firm low-carbon power at facility scale.

A grid-connected datacenter can use external generation to absorb load mismatches, but an SMR-backed site with limited interconnection cannot assume that the grid will absorb every maneuver. The reactor must therefore follow part of the datacenter's changing demand while respecting its own dynamic limits. Reactor output does not move the way a grid tie does, since its ramp rate caps power changes at 1 to 5 percent of rated per minute and a slower second limit appears hours after the operator acts. Fission produces iodine-135, which decays into xenon-135 with a 6.6 hour half-life, and xenon-135 absorbs neutrons strongly enough that steady operation burns it out as fast as it appears. When power drops that burnout slows at once while the existing iodine keeps decaying, so xenon accumulates, peaks several hours after the original command [1], and blocks any increase until it decays. This is the xenon pit.

Challenges. Datacenter power management treats available power as an external signal to follow [6, 7], and nuclear load-following work stays on the power plant side [3], so neither exposes reactor state to the scheduler that requests a maneuver. An on-site SMR makes that omission expensive, because supply carries multi-hour state the present power level does not reveal, and two dips that look identical at the moment of decision can differ by hours in when the reactor may next raise output. The margin also shrinks over the fuel cycle as reactivity falls with burnup [2], so a dip that is safe on a fresh core can leave an infeasible recovery at end of cycle.

Our Approach. A power maneuver is feasible when the reactor can carry it out and still recover to full output once demand returns. We expose that check through a feasibility layer below the workload scheduler: the scheduler requests a power trajectory, and the layer determines whether the reactor can carry it out before the maneuver is committed. Xenon headroom is $H(t) = R_{\text{cap}} - \rho_{\text{Xe}}(t) - \sigma_m$, where R_{cap} is the available reactivity ceiling, $\rho_{\text{Xe}}(t)$ is the predicted xenon-135 worth in pcm, and σ_m is a fixed 100 pcm physical margin. Predictive enforcement uses an additional 20 pcm buffer below that ceiling. The maneuver clears when headroom remains nonnegative over a 10 hour lookahead horizon. The check forward-integrates the iodine–xenon dynamics from the current inventory under the proposed profile, and when headroom goes negative it bisects on dip depth for the deepest floor that clears, or refuses the dip.

We design three reactor policies that differ in the reactor information they consult. Limit-Agnostic never evaluates $H(t)$ and takes every dip the price signal offers. Limit-Reactive evaluates $H(t)$ at the present instant, so it permits a dip while headroom is positive and blocks power changes once headroom has gone negative, which covers the increase that would end the transient. Limit-Predictive evaluates every point of the horizon before committing and declines the dip that creates the future breach.

We separate reactor information from workload information with two additional runtime controllers. Computing-Only uses queue urgency, workload state, price, and grid availability without reactor state. Consolidated uses the same workload information together with projected reactor feasibility. The comparison isolates what workload information and reactor-state information each contribute to the resulting power trajectory.

Methodology and Evaluation. We simulate at one-minute resolution. The reactor model carries iodine–xenon point kinetics, a reactivity ceiling R_{cap} evaluated at fixed operating snapshots, a 5 percent per minute ramp, and SCRAM with restart. We evaluate ample, intermediate, and low reactor margin at R_{cap} = 8,000, 3,500, and 2,700 pcm, respectively.

On the load side, we replay eight public Google Borg cells A–H [4], spanning variation primarily in workload flexibility and also in utilization and burstiness. Cell A provides the representative mechanism trace, while the primary comparison reports means across all eight cells. The site remains grid connected, with imports up to 40 percent of rated power. Grid operation uses January 2025 PJM hourly day-ahead prices and generation-mix data [8]. We additionally exercise demand spikes, scheduled grid outages, surprise grid outages, and compound stress events.

At low reactor margin, Limit-Agnostic leaves 182.9 MWh unserved and Computing-Only leaves 172.0 MWh, averaged across Borg cells A–H, and both average 7.4 reactor trips. Limit-Predictive reduces unserved work to 264 kWh with no trips, while Consolidated reduces it to 169 kWh with no trips. The large separation therefore appears when reactor state enters the decision: workload information alone reduces some loss but does not prevent the low-margin xenon failure mode. Predictive and Consolidated solutions both avoid that failure, with their relative ordering varying across individual cells.

Fixed recovery delays do not reproduce that behavior. At ample margin, fixed 2, 4, and 6 hour delays leave 18.0, 31.3, and 40.4 MWh unserved, respectively. At low margin they still leave approximately 181 MWh unserved and average 7.0–7.4 trips. Waiting for a fixed interval therefore strands work when reactor margin is available without identifying when the xenon trajectory is actually feasible.

The amount of future reactor information matters primarily when margin is tight. At R_{cap} = 2,700 pcm, Limit-Predictive with zero lookahead averages 132.1 MWh unserved and 5.25 trips across A–H. A 2 hour projection reduces that to 269 kWh with no trips,

and the result changes little beyond 4–6 hours. At $R_{\text{cap}} = 3{,}500$ pcm, the same horizon sweep produces no trips, which shows that future nuclear reactor state becomes important as the physical margin becomes binding.

The enforcement buffer provides robustness near that boundary. With no additional buffer, Limit-Predictive averages 107.4 MWh unserved and 4.38 trips at low margin, while Consolidated averages 12.6 MWh and 0.62 trips. Every tested buffer from 10 to 100 pcm eliminates trips across A–H; the canonical controller uses 20 pcm. These values are enforcement sensitivities rather than calibrated measurement uncertainty.

We also separate the controller's reactor model from the simulated plant. Across one-at-a-time sensitivity perturbations of the iodine and xenon decay constants by ±5 percent, including the alternative half-lives reported by the DOE handbook, and reactor flux by ±10 percent, both projection controllers remain trip-free across all eight cells at low margin. The largest tested mismatch consumes 15.3 pcm of the 20 pcm enforcement buffer. The controller re-projects each minute from the observed reactor state, so parameter error is corrected at each subsequent decision rather than accumulating over the full horizon.

The feasibility calculation remains small relative to the one-minute control interval. Computing-Only takes 0.53 ms per decision on average, while Consolidated takes 1.70 ms on average. The reactor-state projection accounts for approximately 1.01 ms of the mean Consolidated decision time.

The original Cell-A experiment provides a separate restricted offline reference for the reactor-policy comparison. Limit-Predictive leaves 271 kWh unserved against 211 kWh for that reference. Water use includes on-site water required for reactor cooling, datacenter consumption, and off-site generation for grid imports [5], so system-wide water depends on both where power is generated and how much workload is served.

Overall, the following are our key contributions.

- A runtime feasibility test coupling iodine–xenon dynamics to datacenter workload management through one headroom query below the workload scheduler.
- An eight-cell Borg evaluation separating workload awareness from reactor-state awareness under ample, intermediate, and low reactor margin.
- Evidence that reactor-aware operation reduces low-margin unserved work from hundreds of MWh to hundreds of kWh while eliminating reactor trips in the evaluated cases.
- A comparison against fixed 2–6 hour recovery delays, together with lookahead, enforcement-buffer, and reactor-model sensitivity tests that identify when projected reactor state is necessary and how robust the feasibility decision remains.
- Grid-connected evaluation using PJM price and generation data, including outage stress tests, water accounting, and millisecond-scale runtime feasibility decisions.

We *open-source* the simulator, the parameterized reactor model, and the designed policies so that researchers can attach reactor feasibility to their own schedulers without rebuilding reactor physics for SMR-aware datacenter work. Code: https://github.com/Shania-Ziegler/smr-datacenter-feasibility.

References

[1] U.S. Department of Energy. *DOE Fundamentals Handbook: Nuclear Physics and Reactor Theory, Volume 2.* DOE-HDBK-1019/2-93, 1993.

[2] S. Choudhury, M. Davidson, and G. Tynan. Physics-Informed Unit Commitment Framework for Nuclear Reactors. *arXiv:2507.18150*, 2025.

[3] U. Baker, M. Garrouste, S. Choi, G. J. Soto, R. Snuggerud, B. Kochunas, and B. Lindley. Simulation of the NuScale SMR and Investigation of the Effect of Load-Following on Component Lifetimes. *Nuclear Technology*, 210(1):1–22, 2024. doi:10.1080/00295450.2023.2216973.

[4] M. Tirmazi, A. Barker, N. Deng, M. E. Haque, Z. G. Qin, S. Hand, M. Harchol-Balter, and J. Wilkes. Borg: The Next Generation. In *Proceedings of EuroSys*, pages 1–14, 2020. doi:10.1145/3342195.3387517.

[5] J. Macknick, R. Newmark, G. Heath, and K. C. Hallett. *A Review of Operational Water Consumption and Withdrawal Factors for Electricity Generating Technologies.* NREL/TP-6A20-50900, National Renewable Energy Laboratory, 2011. doi:10.2172/1009674.

[6] B. Acun, B. C. Lee, F. Kazhamiaka, K. Maeng, U. Gupta, M. Chakkaravarthy, D. Brooks, and C.-J. Wu. Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters. In *Proceedings of ASPLOS*, pages 118–132, 2023. doi:10.1145/3575693.3575754.

[7] A. A. Chien, R. Wolski, and F. Yang. The Zero-Carbon Cloud: High-Value, Dispatchable Demand for Renewable Power Generators. *The Electricity Journal*, 28(8):110–118, 2015. doi:10.1016/j.tej.2015.09.010.

[8] PJM Interconnection. *Data Miner: Day-Ahead Locational Marginal Price and Generation Data*, 2025.