Runtime Power Feasibility for Small Modular Nuclear Reactor-Powered Datacenters

BYU

Shania Ziegler (Brigham Young University) Rohan Basu Roy (Advisor; University of Utah)

THE UNIVERSITY OF UTAH

THE PROBLEM

AI datacenters need firm power at gigawatt scale, and operators are contracting small modular reactors (SMRs) for on-site generation. On-site generation removes the grid as a buffer. The reactor itself has to follow datacenter demand, and two reactor-specific limits decide whether a requested power change is possible at all.

① Ramp limit. Reactor power moves at 1 to 5 percent of rated per minute. A step in demand takes tens of minutes to meet.

② Xenon pit. After a power reduction, xenon-135 builds up and can block any power increase for hours.

WHAT XENON DOES

Fission produces iodine-135, which decays into xenon-135 with a 6.6 hour half-life. Xenon-135 absorbs neutrons strongly, and at steady power the flux burns it out as fast as it appears. Drop the power and burnout slows at once while iodine keeps decaying, so xenon accumulates and peaks hours later, and until it decays the reactor cannot climb back. Operators call this the xenon pit.

FEASIBILITY TEST

a	fuel age, 0 at a fresh core to 1 at end of cycle
$\rho_{Xe}(t)$	reactivity worth of xenon-135, in pcm
$R_{cap}(a)$	reactivity still commandable at fuel age a, falls as fuel burns [1]
σ_m	fixed safety margin, 100 pcm
G	grid-import cap, largest fraction of rated power imported during cheap-price hours
pcm	percent mille, 10^{-5} Δ k/k, the unit of reactivity

$$H(t) = R_{cap}(a) - \rho_{Xe}(t) - \sigma_m$$

Feasible if H(t) ≥ 0 at every t in the lookahead horizon. A check at the current instant may pass even when the projected trajectory later exceeds the limit.

METHODOLOGY

All results are from simulation. No reactor or datacenter hardware is involved.

- Workload. Google Borg cells A-H public traces [3], one month, 1-min steps
- Power. RAPL-measured 208 to 524 W per node, 4,000 nodes, PUE 1.2
- Reactor. Iodine–xenon point kinetics $R_{cap}^{(a)}$ snapshots = 8,000, 3,500, 2,700. 5% /min ramp, SCRAM and restart.
- Prices. 2025 PJM hourly day-ahead prices and generation mixes; imports capped at G = 0.4.
- Water. NREL generation factors [2] for on-site cooling and the hourly grid mix
- Stress tests. Demand spikes plus scheduled surprise grid outages

REFERENCES

[1] Shiny Choudhury, Michael Davidson, and George Tynan. "Physics-Informed Unit Commitment Framework for Nuclear Reactors" *IEEE Access*, vol. 14, pp. 33744-33755, 2026.

[2] Jordan Macknick, Robin Newmark, Garvin Heath and Kathleen 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

[3] Muhammad Tirmazi, Adam Barker, Nan Deng, Md E. Haque, Zhijing Gene Qin, Steven Hand, Mor Harchol-Balter, and John Wilkes. "Borg: The Next Generation." In *Proceedings of EuroSys*, pp 1-14, 2020. doi:10.1145/3342195.3387517

SMALL MODULAR NUCLEAR REACTOR LIMITS

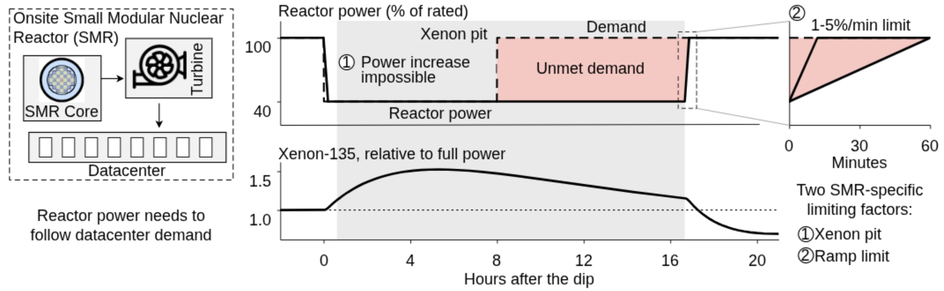


Figure 1. Left, the site. Center, one demand trough. The reactor dips, xenon-135 builds, and power cannot be raised for about sixteen hours, so demand goes unmet. Right, once the increase is permitted the ramp limit still leaves demand short. Traces are representative, produced by the iodine–xenon model in Methodology at a 40% dip.

REACTOR POLICIES AND SCHEDULER

The reactor side decides whether a dip is allowed. The datacenter side decides whose work is dropped when power still falls short. The two are independent.

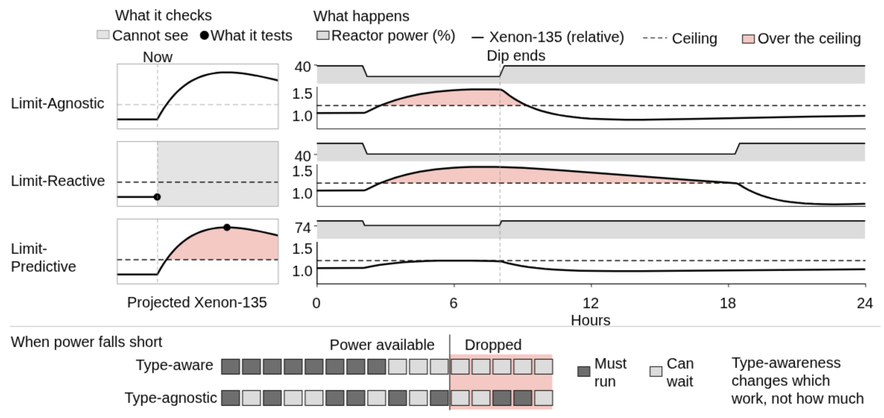
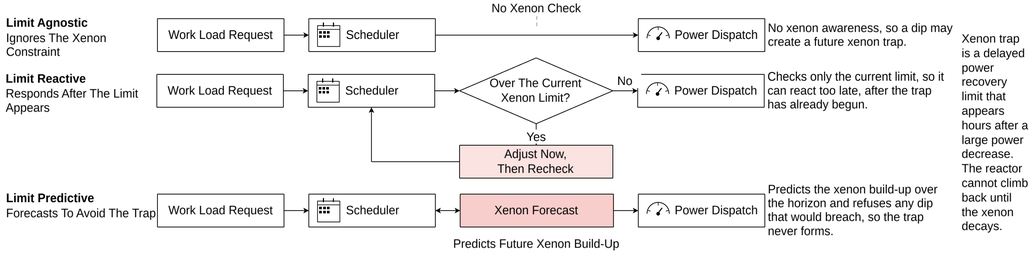


Figure 2. All three test the same projected xenon curve and differ only in how much of it they enforce the ceiling. Limit-Agnostic ignores the ceiling. Limit-Reactive tests the present, passes, and is trapped once the pit forms. Limit-Predictive tests the projected peak and refuses the dip. Bottom, the scheduler. Both drop the same amount of work, so type-awareness changes only which work. Same model as Figure 1.



MEASUREMENT METRICS

Unserved energy.	Work demanded but not delivered, summed over the month.
Reactor shutdowns	Protective shutdowns after the reactivity limit is exceeded
Reactivity headroom.	Distance from the modeled xenon limit.

BEHAVIOR

Limit-Agnostic and Limit-Reactive have nearly identical power trajectories, showing that checking only the present reactor state does little to prevent the delayed xenon constraint. Limit-Predictive projects the future xenon trajectory and limits the earlier maneuver before the constraint is reached.

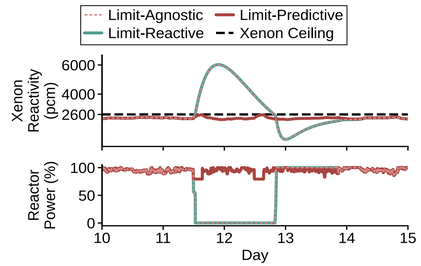


Figure 3. Five representative days. At end of cycle, with a 40% grid-import cap. The dashed line marks the xenon ceiling. Limit-Predictive stays below it; Limit-Reactive overshoots and the gate prevents recovery.

COST OF GETTING IT WRONG

At low reactor margin, workload awareness alone does not prevent xenon failure. Across Borg cells A-H, Limit-Agnostic and Computing-Only leave 183 and 172 MWh unserved and average 7.4 shutdowns, while Limit-Predictive and Consolidated reduce the loss to 264 and 169 kWh and zero shutdowns.

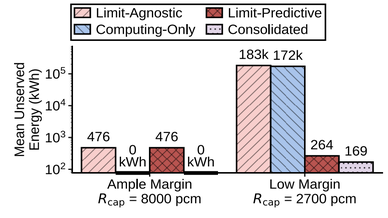


Figure 4. Mean unserved work across Borg cells A-H. The largest separation between policies appears at low reactor margin.

Consolidated	Workload information and projected reactor feasibility
Limit-Agnostic	No reactor feasibility check
Computing Only	Workload/queue information, no reactor state
Limit-Predictive	Projected reactor feasibility

WHERE THE LOSS HAPPENS

Across Borg cells A–H, Limit-Agnostic and Limit-Reactive accumulate similar losses through recurring reactor shutdowns across the month. Limit-Predictive avoids these shutdowns, leaving only a small residual loss.

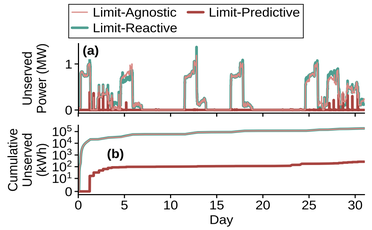


Figure 5. A-H mean instantaneous (a) and cumulative (b) unserved energy over the month. Panel (b) uses a log scale.

GRID FLEXIBILITY

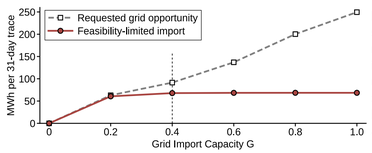


Figure 6. Grid-capacity sweep showing how much cheap-price grid import remains feasible under the projected xenon constraint. Mean across Borg cells A-H.

SMR WATER FOOTPRINT

Water use depends on both where power is generated and how much useful work is served. At low reactor margin, feasibility-aware policies keep the reactor online and serve more work while shifting supply from grid imports toward on-site generation. In our modeled accounting, this increases system-wide water use per unit of work served.

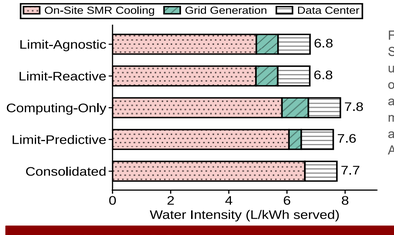


Figure 7. System-wide water use per unit of workload served at low reactor margin, averaged across Borg cells A-H

KEY FINDINGS

- Reactor margin determines when xenon becomes limiting. At ample margin, all evaluated policies remain shutdown-free.
- Instantaneous reactor state is insufficient at low margin. Limit-Reactive performs similarly to Limit-Agnostic because both respond too late to the delayed xenon constraint.
- Forward reactor feasibility changes the outcome. Limit-Predictive and Consolidated avoid shutdowns observed with non-projective approaches across Borg A-H.
- Feasibility can expose usable flexibility. The partial-grid extension limits grid-driven reactor turndown to the amount projected to remain xenon-feasible.