ACM STUDENT RESEARCH COMPETITION, SC26

Design the Plant and the Control Together

Three coupled studies take one exascale cooling plant from measurement, to a validated digital twin, to a design rule. Control alone recovers 27.8 percent of annual cooling energy on hardware already installed. Counting carbon and capital over the service life selects a different plant than energy does.

Shrenik Jadhav, Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University | jadhavs2@my.erau.edu
Zheng Liu University of Michigan-Dearborn | zhengtl@umich.edu
Vidhyashree Nagaraju (advisor), Embry-Riddle Aeronautical University, Daytona Beach Florida | nagarav1@erau.edu

Papers

27.8% annual cooling energy saved under real actuator rate limits

92.4% of the unconstrained theoretical optimum recovered

611 plant topologies simulated over 49,353 ten-minute intervals

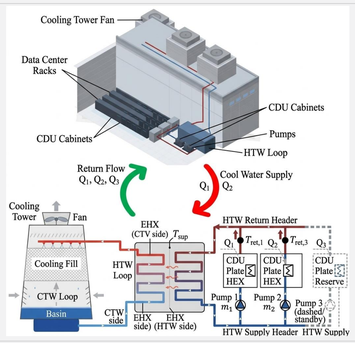
2.67% Worst subloop error against ASHRAE Guideline 14

3,320.7 t CO2e life cycle optimum, 35.7 t below the plant as built

01 Digital Twin-Based Cooling System Optimization for Data Center

How much of a theoretical saving survives the actuator limits of a real plant?

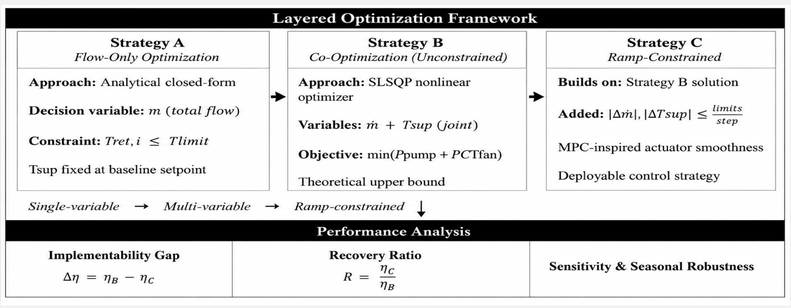
A Modelica twin of the Frontier high temperature water loop: three parallel subloops, each with a plate heat exchanger and a variable speed pump. It is validated against one calendar year of ten minute telemetry, 47,186 records, under ASHRAE Guideline 14. The test is non circular because the twin carries its own specific heat, 3,500 J per kg K, against the 3,709 the dataset assumes.



The plant the twin reproduces: cooling tower and condenser water loop at the left, the high temperature water loop at the right, three duty subloops with their own pumps, and a fourth pump in reserve.

Validated, full year
Return temperature error per subloop against a 5 percent target

1.96%	1.97%	2.67%
Subloop 1	Subloop 2	Subloop 3



The layered framework. Strategy A optimizes flow alone, Strategy B adds the supply temperature setpoint, and Strategy C holds both to real actuator ramp limits.

Annual energy saved
Against a measured baseline of 1,814 MWh, of which tower fans draw 73 percent

20.4%	30.1%	27.8%
Strategy A	Strategy B	Strategy C

WHAT WE FOUND

The cooling towers are the plant. Their fans draw 1,325 of the 1,814 MWh spent on cooling in a year.

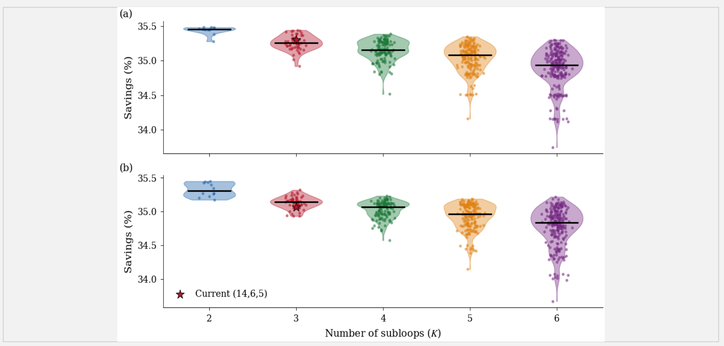
Frontier pumps a median 1.5 times the thermally required flow, which is a 3.4 times overspend in pump power under the cubic affinity law.

Flow control alone saves 20.4 percent, and adding supply temperature reset, a mean rise of 2.7 C in 97.9 percent of intervals, reaches 30.1 percent. Holding the optimizer to 50 kg per second and 1 C per ten minutes keeps 27.8 percent, so 92.4 percent of the theoretical saving survives real actuator limits.

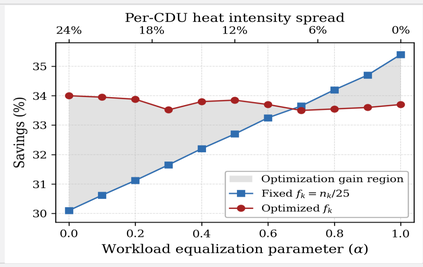
02 Co-Design Optimization for Data Center Cooling System via Digital Twin

How many subloops, which units on each, and how should the flow be split?

Three layers over the same validated physics. Layer 1 enumerates every integer partition of 25 coolant distribution units into 2 to 6 subloops, 611 in all. Layer 2 sets the flow fraction of each subloop. Layer 3 co-optimizes total flow and supply temperature at each of 49,353 timesteps, under a 42 C return limit on every subloop. Enumeration certifies global optimality across the discrete space; sequential least squares programming with analytical gradients solves the continuous part.



Strategy C savings across all 611 partitions, grouped by subloop count, under balanced (a) and worst case (b) unit assignment. The star marks the plant as built, partition (14, 6, 5).



Optimizing the flow fractions removes the penalty for an uneven plant. Held fixed, the saving collapses from 35.4 to 30.1 percent as the per unit heat spread widens to 24 percent. Optimized, it holds near 34 percent across the range.

Topology, decided
The best of 611 partitions against the plant as built

35.48%	0.18 pt	93%
Best, (19, 6)	Gap to as built	Penalty removed

WHAT WE FOUND

The energy optimum is two subloops, partition (19, 6), at 35.48 percent. The three subloop plant as built reaches 35.30 percent, so the topology decision is worth 0.18 points.

Savings fall and spread widens as subloops multiply, from a mean of 35.44 percent at two to 34.89 at six, with the range growing from 0.20 to 1.55 points.

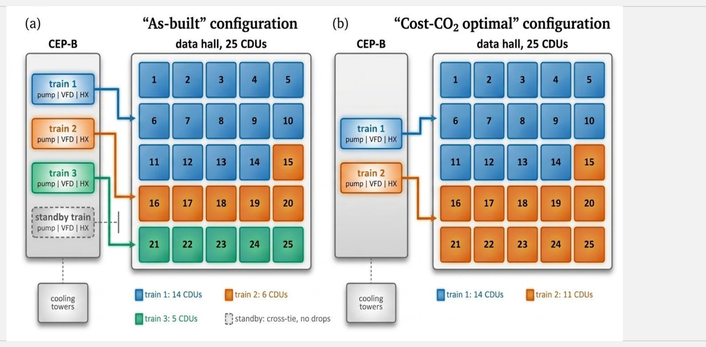
Flow fraction control is what buys robustness. It cuts the unit assignment penalty by 93 percent, from 96 to 7 MWh per year.

Frontier already runs independently controlled pumps, so near optimal operation is a software change on installed hardware.

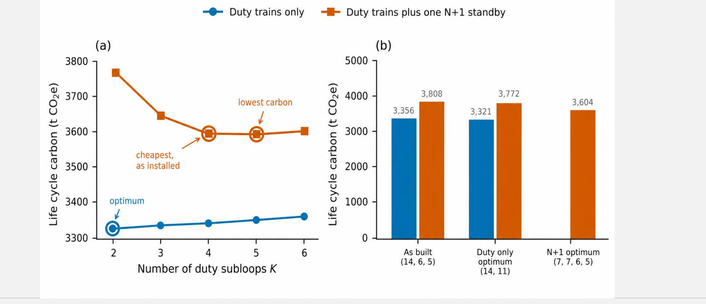
03 Life Cycle Co-Design of the Cooling Subloop Architecture for Data Centers

If carbon and capital are counted over the service life, which plant should be built?

Every topology is now priced over seven years. Operational energy comes from the Layer 3 optimizer. Embodied carbon and capital come from a bill of materials, with bores selected on ASME B36.10M schedules and checked against ASME B31.9. Maintenance comes from a task list of 17 hours per train per year plus scheduled replacements and expected unplanned downtime from a component model on IEEE Std 493 data. The boundary and the stage labels follow EN 15978.



The decision, drawn. The plant as built runs three duty trains carrying 14, 6 and 5 units, with a standby cross tied to the header. The cost and carbon optimum runs two, carrying 14 and 11.



Effect of pricing standby-train redundancy on life-cycle carbon.
(a) Best partition with and without an N+1 standby. The optimum shifts from (K=2) (no standby) to (K=4)–5 (with standby).
(b) Including redundancy moves the optimum to four or five duty subloops, while the as-built configuration remains suboptimal due to its oversized spare.

WHAT WE FOUND

Counting duty trains only, the optimum is two subloops, partition (14, 11), at 3,320.7 t CO2e and 3.99 million dollars over seven years. That is 35.7 t and 63 thousand dollars below the plant as built.

The difference is embodied, not operational. More loops replicate trunk and header runs, and standard pipe schedules quantize the bore.

The energy optimum and the life cycle optimum are different plants. Partition (19, 6) wins on energy by 0.11 MWh per year and loses on carbon by 71.8 t.

Price the standby and the ranking reverses. It is sized to the largest duty branch, so it costs most at low subloop counts. Availability is bought with a spare, not by subdividing the duty.

One chain, three links

01 Measure the plant and prove a model of it → 02 Search every plant the model can evaluate → 03 Price the plant across its service life

Each study supplies the next one with its evaluation function. The validated twin of study 01 is what makes 611 topologies affordable to simulate in study 02, and the per timestep optimizer of study 02 is what makes the operational term of study 03 credible. Alone, each answers a smaller question than the three answer together.

Control first, on hardware you own
27.8 percent at Frontier through setpoints alone, with no new equipment and no change to the plant.

Fewer, larger duty subloops
Grouped contiguously, they lower both cost and carbon, and flow fraction control removes the assignment risk.

Buy availability with a spare
Expected unplanned downtime falls from 4.90 hours per year to below 0.015 once one standby is provided.

Predicted optimal duty subloop count against total peak flow and drop budget, under no redundancy requirement (a), one N plus 1 standby (b), and a standby with hydraulic ride through (c). The redundancy policy changes the answer on half the map, while moving the operating point by a quarter changes nothing. Frontier, Aurora, El Capitan and LUMI are marked.

Robustness of the life cycle optimum. Across a 40 percent range on the embodied carbon factor and a factor of two on grid intensity, partition (14, 11) stays ahead of every design with three or more subloops. The star is the central case.

The Three Studies

1. S. Jadhav and Z. Liu. Digital Twin-Based Cooling System Optimization for Data Center.
2. S. Jadhav and Z. Liu. Co-Design Optimization for Data Center Cooling System via Digital Twin.
3. S. Jadhav, V. Nagaraju and Z. Liu. Life Cycle Co-Design of the Cooling Subloop Architecture for Data Centers.

Additional related work: S. Jadhav and Z. Liu, Machine Learning Guided Cooling System Optimization for Data Center, arXiv:2601.02275 (2026).
Operational record: Sun et al. Scientific Data 11, 1077 (2024), Frontier 2023 telemetry, CC BY 4.0.

Conclusion

1. Software captures most of the savings. Frontier's cooling plant cuts annual cooling energy by 27.8% through setpoints alone, using 92.4% of the unconstrained optimum.
2. Control matters more than topology. Across 611 partitions, the best plant only slightly beats the as-built layout; optimizing flow fractions removes most of the penalty from poor unit assignment.
3. Energy and life-cycle optima differ. The lowest-energy plant is not the lowest-carbon one; embodied carbon changes the answer.
4. Redundancy sets the loop count. Pricing the standby train shifts the optimum from two subloops to four or five, making the built configuration the constrained optimum.

RAISE Lab — Reliability & Assurance in Safety-Critical Software Engineering

Intelligent Manufacturing Lab