Foundations for FMU Surrogate Model Design in ExaDigiT: Preliminary Analysis of Three IBM POWER9 Cooling Systems

Yishak Tadele Nigatu
yishaktadele.nigatu@studenti.unitn.it
University of Trento
Trento, Italy

Wesley Brewer
brewerwh@ornl.gov
Oak Ridge National Laboratory
Oak Ridge, TN, USA

Jonathan McConnell
mcconnelljh@ornl.gov
Oak Ridge National Laboratory
Oak Ridge, TN, USA

Valentine G. Anantharaj
vga@ornl.gov
Oak Ridge National Laboratory
Oak Ridge, TN, USA

Sandro Fiore
sandro.fiore@unitn.it
University of Trento
Trento, Italy

Abstract

High-fidelity Functional Mock-up Unit (FMU) simulations of data-center cooling systems are accurate but computationally expensive for real-time control and rapid scenario exploration that sustainable, large-scale, high-performance computing (HPC) systems demand. Replacing these simulators with efficient deep learning surrogates is a key step toward energy-aware digital twins. But effective surrogate design requires a prior understanding of the input-output relationships, temporal dynamics, non-linearities, and spatial coupling. This work addresses that gap through a systematic statistical analysis of FMU simulation data from three IBM POWER9 cooling systems: Marconi100, Summit, and Lassen, all within the ExaDigiT framework. Key findings are (i) a structural decoupling between the primary and secondary loops, as the primary loop is actively managed and substantially non-linear and load-following, whereas the secondary loop is passive with outputs remaining practically constant regardless of inputs; (ii) level-driven dynamics (median rate/level ratio = 0.33) with 54% of pathways exhibiting higher-order transient responses and Autocorrelation Function (ACF) persistence beyond 1,000 lags, mandating recurrent or attention-based temporal architectures; (iii) negligible inter-CDU (Cooling Distribution Unit) thermal propagation that was confirmed by distance-independent correlations and asymmetry ratio $R_{\mathrm{asym}} = 0.95$. Which allows scalable per-CDU modeling; and (iv) near-zero thermodynamic violations with tolerance-calibrated energy conservation constraints. These results yield concrete surrogate design principles for surrogate architecture selection, feature engineering, and physics-regularized training.

CCS Concepts

• **Computing methodologies** → **Neural networks**; **Supervised learning**; *Simulation types and techniques*; • **Hardware** → *Power and energy*; • **General and reference** → Performance.

ACM ISBN 979-8-4007-2174-8/26/04
https://doi.org/10.1145/3802973.3804456

Keywords

ExaDigiT, Surrogate Models, FMU, Data Analysis, Physics Regularization

ACM Reference Format:
Yishak Tadele Nigatu, Wesley Brewer, Jonathan McConnell, Valentine G. Anantharaj, and Sandro Fiore. 2026. Foundations for FMU Surrogate Model Design in ExaDigiT: Preliminary Analysis of Three IBM POWER9 Cooling Systems. In *Workshop on Systems and Methods for Sustainable Large-Scale AI (GreenSys '26), April 27–30, 2026, Edinburgh, Scotland Uk.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3802973.3804456

1 Introduction

Data centers consume an estimated 1–2% of global electricity as of 2018 [9], a figure that has since grown substantially with the rise of artificial intelligence workloads and large-scale model training. Recent projections indicate that data center energy consumption may double by 2030 [5], with some estimates indicating that the sector could account for 8–13% of global electricity demand within the next decade. This growing power density has motivated digital twin frameworks that provide real-time virtual representations of cooling infrastructure [23, 2].

The ExaDigiT project [3] represents a state-of-the-art digital twin for liquid-cooled supercomputers, integrating workload simulation through its Resource Allocation and Power Simulator (RAPS) with Functional Mock-up Unit (FMU)-based thermo-fluid models derived from the Modelica TRANSFORM library [11].

A central limitation of this framework is the computational cost of its FMU simulator: while it delivers high-fidelity predictions of temperatures, pressures, flow rates, and power consumption, its runtime severely limits its application to online decision-making, rapid scenario exploration, and large-scale parametric studies. This computational bottleneck necessitates the development of machine learning-based surrogate models.

Deep learning surrogates can provide exponential speedups while retaining acceptable fidelity. Yet their design choices, such as architecture selection, feature engineering, temporal window sizing, spatial coupling strategy, and physics-informed regularization, demand a rigorous understanding of the data they must reproduce.

This paper addresses this gap with a systematic data analysis of FMU-generated data for three IBM POWER9 [21] systems: Marconi100 [1] (49 CDUs), Summit [24] (257 CDUs), and Lassen [19]

(44 CDUs). Six progressive analysis stages each target a specific surrogate design decision: (i) Direct-effect analysis: examines isolated, pairwise input–output relationships under steady-state conditions. (ii) Multivariate effect analysis: investigates confounding, cross-correlation structures that emerge when inputs act jointly, and temporal memory requirements. (iii) Rate-of-change analysis: determines whether the cooling model responds primarily to input levels or their temporal changes. (iv) Non-linearity characterization: quantifies the degree of non-linear coupling between each input–output pair to determine required architectural depth and activation type. (v) Spatial interaction analysis: tests for inter-CDU thermal coupling to determine whether spatial modeling is required. (vi) Physics consistency analysis: evaluation of conservation-law constraints to calibrate tolerance-based physics regularization terms. To our knowledge, this is the first pre-surrogate data analysis for FMU-based HPC cooling systems across multiple production facilities at scale.

2 Related Work

Digital Twins for Liquid-Cooled Supercomputers: Digital twin frameworks for HPC facilities have evolved from static thermal maps into dynamic, multi-physics simulation environments. The ExaDigiT framework [3] represents the most comprehensive open-source digital twin framework: it pairs RAPS with Modelica-based FMU cooling models from the TRANSFORM library [11] to perform transient simulation of the full cooling hierarchy with faithful replication of cooling parameters under real HPC workloads [3]. [7] extended this foundation with AutoCSM, a template-based auto-generation methodology for constructing facility-specific FMU models at scale. Despite this progress, FMU computational expense remains the central obstacle.

Surrogate Models for Datacenter Thermal Management: The use of machine learning surrogates to accelerate datacenter thermal simulations has grown rapidly. [6] applied deep neural networks to room-scale temperature prediction, while a hybrid Artificial Neural Network—Convolutional Neural Network (ANN—CNN) framework [14] achieved a 300,000× speedup over CFD solvers with 87% mean absolute error reduction. A vision-based 3-D convolutional neural network, U-Net Variants, Fourier Neural Operator, and vision transformer framework [22] demonstrated 20,000× accelerations with generalization across datacenter configurations, real-time hot-spot detection, and workload redistribution, yielding 7% energy savings.

Physics-Informed Deep Learning: Physics-Informed Neural Networks (PINNs) [20], encode governing PDEs into the training loss, producing accurate surrogate construction in data-sparse regimes with built-in satisfaction of conservation laws. [10] demonstrated PINN's broad applicability to fluid mechanics, heat transfer, and coupled multi-physics systems with improved generalization beyond the training envelope. [4] further showed PINNs can solve ill-posed thermal problems, including those with incomplete boundary conditions, beyond conventional numerical methods; and [16] demonstrated hard-constraint boundary treatments for thermodynamic feasibility. These motivate the physics-consistency analysis in Section 4.6.

Neural operators extend this paradigm by learning mappings between function spaces rather than fixed-dimensional input–output pairs: DeepONets [15] established the theoretical foundation, while the Fourier Neural Operator [12] achieved state-of-the-art accuracy on Navier-Stokes equations, a strategy adopted in [22]. These operator architectures naturally accommodate variable-length input histories for the time-series prediction task central to cooling system surrogates, motivating the temporal dependency analysis of Section 4.2.

Temporal Deep Learning for Physical Systems: The correct treatment of temporal dependencies is a decisive factor in surrogate model accuracy. Beyond classical LSTMs, Transformer variants have emerged for long-range dependencies: PatchTST [18] achieves state-of-the-art long-term forecasting via channel-independent patch tokens. iTransformer [13] demonstrated that inverting the Transformer to operate on variate tokens better captures multivariate correlations, and for irregularly sampled physical systems with multi-scale dependencies, the Rough Transformer [17] uses path signatures to augment attention with a lightweight mechanism capturing both local and global temporal structure. Mamba [8] offers transformer-level expressivity with linear scaling, which is attractive for long trajectories where our analysis reveals ACF persistence beyond 1,000 lags. The choice among these architectures cannot be made a priori: it depends on the specific temporal dependency structure, which our autocorrelation and lag analyses (Section 4.2) are designed to reveal.

Data Analysis as a Prerequisite for Surrogate Design: The principled pre-analysis of simulation data is a fundamental prerequisite for surrogate models' generalizability. Without characterizing nonlinearity profiles, ad-hoc depth selection leads to underfitting or overparameterization. Without measuring spatial coupling, practitioners face a dilemma between expensive graph architectures and insufficient independent models. Without assessing physics violations, regularization terms are calibrated blindly. Our six-stage methodology addresses each failure mode systematically, grounding surrogate design in data structure rather than heuristic assumption.

3 Methodology

3.1 System Description

The cooling systems under study are modeled within ExaDigiT using FMU-based thermo-hydraulic simulations. Each system comprises Coolant Distribution Units (CDUs) interfacing between a primary high-temperature water (HTW) loop from a central cooling plant and a secondary rack-level loop circulating coolant through compute cabinets.

Three inputs drive each CDU: IT heat load (Q_{flow}, kW), inlet air temperature (T_{Air}, K), and external wet-bulb temperature (T_{ext}, K). The FMU produces 11 per-CDU outputs: temperatures ($T_{\text{prim,s}}$, $T_{\text{prim,r}}$, $T_{\text{sec,s}}$, $T_{\text{sec,r}}$), pressures ($p_{\text{prim,s}}$, $p_{\text{prim,r}}$, $p_{\text{sec,s}}$, $p_{\text{sec,r}}$), flow rates ($V_{\text{flow,prim}}$, $V_{\text{flow,sec}}$), and pump power (W_{CDUP}).

3.2 Data Generation and Experimental Setup

Simulation data are generated using a custom `FMUSimulator` interfacing ExaDigiT's RAPS module, combining operational scenario modeling with systematic exploration.

- Operational Scenarios: Heat load profiles (Q_{flow}) encompass normal operation (40-60% utilization, ±1-3% noise), edge cases (near-idle, near-peak, traffic spikes), and fault conditions, parameterized via Latin Hypercube Sampling (50% normal, 45% edge, 5% fault). Inlet air temperatures (T_{Air}) follow a first-order thermal model with CDU-specific LHS (Latin Hypercube Sampling) parameters. External temperatures (T_{ext}) are sourced from historical weather data via RAPS, smoothed with a Savitzky-Golay filter.
- Systematic Exploration: 270 scenarios—steady-state grids, step responses, ramp sweeps, sinusoidal patterns, and combined stress tests—are sequenced via greedy nearest-neighbor ordering. Each simulation begins with a 2-4 h stabilization phase (using sliding-window variance detection of threshold 0.1) before data recording, with cosine-interpolated transitions (15-60 s ramps) between scenarios.

4 Experimental Analysis and Results

4.1 Direct-Effect Analysis

4.1.1 Correlation Analysis. Pearson correlation (r) and mutual information (MI) were computed for all 33 input-output pairs (Table 1; all $p < 0.01$). $Q_{\text{flow}} \rightarrow p_{\text{sec}}$ yields the highest coefficients across systems (Summit r=0.988, Marconi100 r=0.832, Lassen r=0.799), followed by $Q_{\text{flow}} \rightarrow V_{\text{flow,sec}}$ (r=0.779, 0.753, 0.721). T_{ext} correlates most strongly with $T_{\text{prim,s}}$ (up to r=0.327) and $p_{\text{prim,s}}$ (up to r=0.279), while T_{Air} produces weak direct correlations ($|r| \leq 0.182$).

Table 1: Direct Isolated Effects Analysis for All Systems (r = Pearson correlation, MI = Mutual Information). Bold values indicate strong relationships. Return and supply values are shown as r,s (return, supply) for paired variables. (Sys(System): M=Marconi100, S=Summit, L=Lassen)

Sys	Outputs	Q_flow		T_Air		T_ext	
		r	MI	r	MI	r	MI
M	T_prim (r,s)	-0.022, -0.019	0.121, 0.121	-0.073, -0.059	0.124, 0.139	0.033, **0.255**	0.201, 0.211
	T_sec (r/s)	0.042, 0.023	0.133, 0.129	-0.097, -0.098	0.134, 0.135	0.058, 0.058	0.181, 0.181
	V_flow_prim	0.008	0.136	-0.048	0.125	0.025	0.205
	V_flow_sec	**0.753**	0.213	0.100	0.100	-0.116	0.130
	W_flow	0.021	0.128	-0.098	0.135	0.058	0.180
	p_prim (r/s)	-0.077, -0.013	0.131, 0.122	-0.029, -0.088	0.124, 0.119	0.025, **0.279**	0.212, **0.222**
	p_sec (r/s)	**0.832**, **0.832**	**0.321**, **0.321**	0.007, 0.007	0.098, 0.098	-0.002, -0.002	0.131, 0.132
S	T_prim (r/s)	0.047, 0.028	0.035, 0.030	-0.009, 0.018	0.032, 0.029	0.003, **0.327**	0.045, **0.111**
	T_sec (r/s)	0.108, 0.078	0.034, 0.031	-0.010, -0.011	0.032, 0.033	-0.009, -0.009	0.033, 0.033
	V_flow_prim	-0.091	0.029	-0.022	0.028	-0.077	0.077
	V_flow_sec	**0.779**	0.176	0.031	0.023	0.026	0.023
	W_flow	0.075	0.031	-0.011	0.032	-0.009	0.032
	p_prim (r/s)	0.066, 0.066	0.036, 0.036	0.009, 0.009	0.032, 0.033	**0.214**, **0.216**	**0.120**, **0.121**
	p_sec (r/s)	**0.988**, **0.988**	**0.582**, **0.582**	0.182, 0.182	0.018, 0.018	0.024, 0.024	0.026, 0.026
L	T_prim (r/s)	0.198, 0.042	0.175, 0.183	0.080, 0.133	0.166, 0.171	-0.047, 0.146	0.165, 0.163
	T_sec (r/s)	0.035, -0.005	0.188, 0.194	-0.027, -0.028	0.173, 0.171	-0.027, -0.028	0.155, 0.163
	V_flow_prim	-0.067	0.178	-0.029	0.197	-0.027	0.193
	V_flow_sec	**0.721**	**0.227**	0.124	0.138	-0.010	0.133
	W_flow	-0.011	0.185	-0.028	0.172	-0.027	0.160
	p_prim (r/s)	-0.033, 0.041	0.184, 0.175	0.034, -0.029	0.176, 0.170	-0.004, 0.024	0.206, **0.219**
	p_sec (r/s)	**0.799**, **0.799**	**0.309**, **0.309**	0.145, 0.145	0.151, 0.151	-0.039, -0.039	0.131, 0.130

4.1.2 Response Surface Analysis. Response surfaces (second-degree polynomial with interaction terms on a 30×30 grid) were fitted for Marconi100 (Table 2). Three behavioral categories emerge: (i) secondary loop variables ($V_{\text{flow,sec}}$, p_{sec}, W_{CDUP}) are effectively invariant (with response magnitudes below 1% of their mean values), confirming fixed-speed secondary pump design; (ii) temperatures exhibit planar, additive responses: $T_{\text{prim,s}}$ driven by T_{ext} (spanning a 27°C range), while secondary temperatures showed greater sensitivity to T_{Air}; and (iii) primary flow rate emerged as the dominant control variable, displaying the largest dynamic range (0–108 GPM) with saddle-shaped, strongly interactive response surfaces over (Q_{flow}, T_{ext} pairs).

Table 2: Response surface analysis summary: Marconi100 CDU outputs. Sensitivity levels: H=High, M=Moderate, L=Low, VL=Very Low, Ng=Negligible. Non-lin.=Non-linearity; Int.=Interaction strength; P=Primary, S=Secondary.

Output	Loop	Range	Q_{flow}, T_{Air}, T_{ext}	Non-lin.	Int.
$V_{\text{flow,prim}}$ (GPM)	P	0-108	H, M, H	Strong	Strong
$V_{\text{flow,sec}}$ (GPM)	S	317.46 ± 0.01	Ng, Ng, Ng	None	None
$T_{\text{prim,s}}$ (°C)	P	13.5-40.5	L, L, H	Weak	Weak
$T_{\text{prim,r}}$ (°C)	P	25.0-43.0	M, M, H	M	M
$T_{\text{sec,s}}$ (°C)	S	33.6-48.0	M, H, M	Weak	Weak
$T_{\text{sec,r}}$ (°C)	S	33.6-50.4	M, H, M	Weak	Weak
$p_{\text{prim,s}}$ (psig)	P	58.5-67.5	M, M, M	Strong	Strong
$p_{\text{prim,r}}$ (psig)	P	35.0-53.0	H, M, H	Strong	Strong
$p_{\text{sec,s}}$ (psig)	S	29.70 ± 0.006	Ng, Ng, Ng	Weak	Weak
$p_{\text{sec,r}}$ (psig)	S	0.694 ± 0.004	Ng, Ng, Ng	Weak	Weak
W_{CDUP} (kW)	S	4.10-4.18	L, VL, VL	Weak	None

4.2 Multivariate Effect Analysis

4.2.1 Partial Correlation Analysis: is computed via residualization with significance threshold $\alpha = 0.05$ and edge threshold $|r_{xy \cdot Z}| > 0.1$, reveal system-specific direct-effect structures.

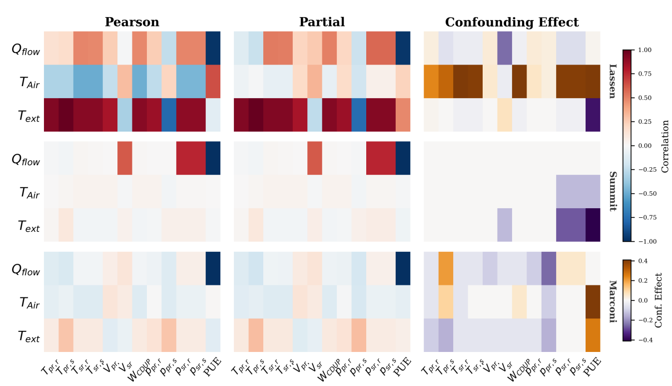

Figure 1: Partial Correlation Heatmaps for Marconi100, Summit, and Lassen

On Lassen, T_{ext} retains partial correlations of 0.81-0.98 with temperature outputs after controlling for other inputs. The confounding difference Δ_{conf} for T_{Air} on Lassen reaches 0.41 for pressure and flow variables, indicating that the corresponding Pearson correlations were substantially inflated. On Summit, the structure is sparser: Q_{flow} has strong partial correlations with pressure variables (≈0.73) and PUE (−0.96), while T_{Air} partial correlations are near zero for most outputs. Confounding differences on Summit are below 0.05. On Marconi100, T_{ext} maintains moderate partial correlations (0.10-0.31) with temperature and flow outputs and Q_{flow} exhibits a strong negative partial correlation (−1.00) with PUE. Confounding differences are typically below 0.05.

Table 3: Cross-correlation analysis for Marconi100, Summit and Lassen; PCC=Peak Cross Correlation, Lag=lag at PCC

	Lassen						Marconi100						Summit					
	Q_{flow}		T_{Air}		T_{ext}		Q_{flow}		T_{Air}		T_{ext}		Q_{flow}		T_{Air}		T_{ext}	
Input	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag	PCC	Lag
$T_{pr,r}$	0.22	−259	−0.27	426	0.45	−107	0.14	−441	−0.28	−494	0.35	−103	0.07	1313	0.07	−1465	0.22	−118
$T_{pr,s}$	−0.14	1398	−0.36	744	0.53	−62	0.21	−397	−0.30	−451	0.57	−50	−0.09	−559	0.07	−1447	0.33	−81
$T_{sr,r}$	0.11	850	−0.25	−140	0.16	−658	0.09	395	−0.27	−1382	0.18	−1059	0.06	1296	0.04	−86	0.09	−916
$T_{sr,s}$	0.11	850	−0.25	−138	0.16	−655	0.09	395	−0.27	−1380	0.18	−1056	0.06	1294	0.04	−86	0.09	−917
V_{pr}	−0.11	1500	−0.31	−217	0.23	−982	−0.10	1060	0.23	1113	0.17	998	−0.08	−916	0.07	509	0.13	−614
V_{sr}	0.34	13	0.35	−439	0.34	−103	0.21	−450	−0.21	402	0.38	−116	0.66	15	0.10	995	0.13	−133
W_{COUP}	0.11	1040	−0.25	−139	0.16	−659	0.09	395	−0.27	−1380	0.18	453	0.06	1294	0.04	−86	0.10	−917
$p_{pr,r}$	−0.10	1500	−0.31	−209	0.22	−982	−0.11	1357	−0.30	−1500	0.22	607	−0.08	−1037	0.08	509	0.13	−614
$p_{pr,s}$	0.10	1500	0.32	−225	−0.23	−983	0.16	−394	−0.32	−443	0.41	−47	0.09	−915	−0.07	508	−0.13	−614
$p_{sr,r}$	0.32	−3	0.33	−436	0.44	−97	0.20	−448	−0.27	−528	0.48	−119	0.73	1	0.12	982	0.19	−129
$p_{sr,s}$	0.32	−3	0.33	−436	0.44	−97	0.20	−448	−0.27	−528	0.48	−119	0.73	1	0.12	982	0.19	−129
PUE	−1.00	0	0.20	736	−0.25	−248	−1.00	0	0.25	912	−0.29	347	−0.96	0	0.10	715	0.11	356

4.2.2 Cross-Correlation Analysis: CCFs (cross-correlation functions) were computed for all input-output pairs over ±1,500 lags, measuring the correlation between an input at time t and an output at time $t + k$. Table 3 details the complete result.

Marconi100 and Lassen exhibit negative T_{Air} correlations with thermal outputs (−0.27 to −0.36), while Summit shows near-zero values. Lag structures vary substantially: Lassen shows shorter response delays (−100 to −450 steps for T_{Air}), Marconi100 lags over 1,300 steps, and Summit shows intermediate, highly variable lag patterns. The T_{ext} effect on primary flow rate peaks at positive lags (+998 to +1, 113 steps).

4.2.3 Lagged Effect Analysis. Distributed linear lag models (up to 1,200 steps / 20 min, 10-step increments, 5-fold CV) are implemented to quantify system memory (Figure 2). Marconi100 shows the strongest memory (T_{ext} lags up to 1,200 s, R^2 improvement up to 0.93); Lassen is moderate (T_{ext}: 865-1,150 s, up to 0.88); and Summit has the weakest (Q_{flow}/T_{Air} improvements ≤0.04). Across all sites, T_{ext} consistently requires sequence-based inputs, while W_{CDUP} and PUE contributes negligible lag-based improvement.

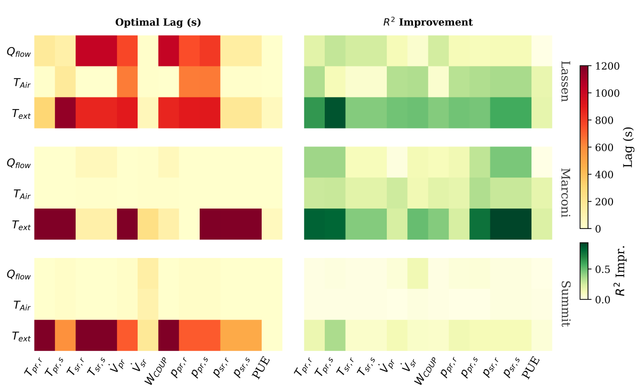

Figure 2: Lagged Effect Analysis: Optimal Lag & R^2 Improvement

4.2.4 Autocorrelation Analysis: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to characterize temporal dependencies within individual variables. The results reveal that all 11 outputs display slow, near-linear ACF decay beyond 1,000 lags while remaining positive, with PACF showing a dominant lag-1 spike (≈1.0) followed by rapid decay.

4.3 Rate-of-Change Analysis

For each input-output pair (u_i, y_j), the level effect (r_{level} and rate effect (r_{rate}) was computed as follows:

$$r_{\text{level}} = \text{Pearson}(u_i(t), y_j(t)) \ , \ r_{\text{rate}} = \text{Pearson}(\dot{u}_i(t), y_j(t))$$

Pairs were classified as level-dominant when $|r_{\text{level}}| > 1.2 \cdot |r_{\text{rate}}|$ and rate-dominant for the converse.

From 39 input-output pairs examined, 27 (69%) are level-dominant, 6 (15%) are rate-dominant, and 6 (15%) are mixed. The median rate/level ratio is 0.33. By input: Q_{flow} is 12/13 level-dominant; T_{Air} is 9/13 level-dominant; T_{ext} is split 6/6/1 (level/rate/mixed).

Impulse response classification is performed via first-order exponential fitting $y(t) = y_{\text{final}}(1 - e^{-t/\tau})$. The response order classification reveals that *54.2% of input-output pairs exhibit higher-order dynamics* that cannot be adequately represented by simple exponential decay models, with only 1.0% qualifying as true first-order systems and 38.5% as approximate first-order. This prevalence of higher-order responses indicates that the FMU incorporates cascaded thermal and hydraulic processes requiring deep learning model architectures with sufficient representational capacity for complex temporal patterns.

4.4 Non-Linearity Characterization

Polynomial regression models of degrees 1-5 with ridge regularization ($\alpha = 1.0$) quantify nonlinearity strength as the R^2 improvement over the linear baseline. Three consistent patterns emerge: T_{ext}: predominantly linear ($\Delta R^2 \approx 0$) for most outputs, except $V_{\text{flow,prim}}$. T_{Air}: pervasive non-linearity (>60 pp for most outputs), requiring degree-5 representations. Q_{flow}: strongly non-linear for primary variables (85-100 ≈ pp) but linear for secondary pressures, consistent with fixed secondary pump operation.

Each system had notable differences: Marconi100 shows the most uniform non-linearity; Summit displays more moderate patterns in pressure outputs; Lassen presents an intermediate profile. These findings warrant a unified deep learning architecture with non-linear activations and 4+ layers, with system-specific fine-tuning for hydraulic subsystems.

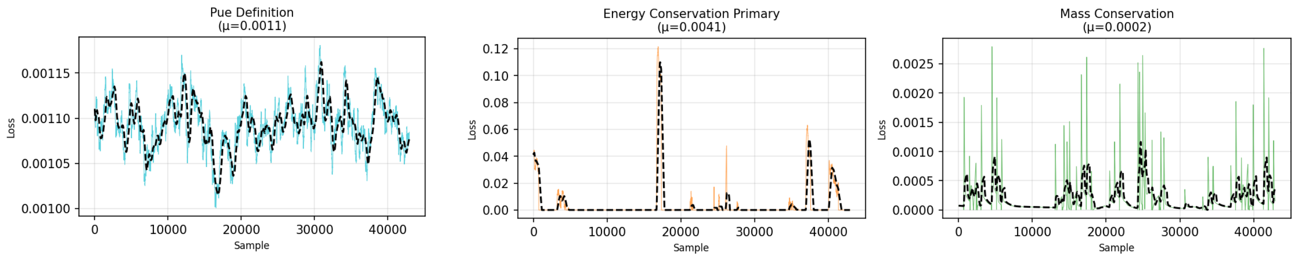

Figure 3: Individual physics loss components over samples. Components with all-zero values are not shown.

4.5 Spatial Interaction Analysis

This analysis tests whether thermal events propagate between CDUs via shared infrastructure, using Summit (257 CDUs) with 10 hotspot CDUs at 80-95% capacity and 247 at a 20 kW baseline. The results confirm *negligible inter-CDU thermal propagation.*

4.5.1 Temperature Gradient Analysis. Inlet air temperature gradients exhibit a mean of 8.41°C, a maximum of 27.16°C, and a standard deviation of 6.10°C, while primary return temperature gradients are substantially smaller (mean 1.15°C, maximum 13.07°C, standard deviation 1.38°C); this sevenfold attenuation between inlet and outlet reveals that the shared primary cooling loop effectively homogenizes thermal conditions across CDUs. Near-unity persistence values ($\bar{\rho}_{\text{inlet}} = 0.996$, $\bar{\rho}_{\text{outlet}} = 1.000$) further confirm that these gradients are *static spatial features* rather than dynamically propagating disturbances.

4.5.2 Propagation Asymmetry: The asymmetry analysis examined whether thermal effects preferentially propagate in one direction along the CDU array (Figure 4). The asymmetry ratio $R_{\text{asym}} = 899,799/943,401 = 0.95 \approx 1$ confirms *no preferential propagation direction*, with symmetric positive/negative gradient means (±0.54°C).

4.5.3 Load Propagation Analysis. Lagged cross-correlations between all CDU pairs for Q_{flow} confirm that: (i) inter-CDU heat load correlations are weak ($|r| < 0.15$ for the majority of pairs, Figure 5), with no distance-dependent decay; (ii) correlation strength shows no systematic relationship with distance; and (iii) optimal lag values are randomly distributed, ruling out any characteristic propagation speed. These results confirm that each CDU can be modeled independently without spatial graph architectures.

4.6 Physics Consistency Analysis

Ten physics loss components (energy/mass conservation, temperature ordering, HX feasibility, approach temperature, PUE consistency) are evaluated on Summit's systematic data (43,000+ samples). The FMU is largely consistent (mean total loss = 0.0113): temperature ordering, HX feasibility, approach temperature, and secondary energy conservation show *zero violations*(see Figure 3).

The primary energy conservation loss ($\mu = 0.0041$) is the dominant contributor, exhibiting periodic spikes at transient operating conditions. Mass conservation shows minor numerical discrepancies ($\mu = 0.0002$), and PUE definition exhibits a small systematic offset ($\mu = 0.0011$).

5 Discussion

The analysis yields five actionable surrogate design principles: (i) There is a fundamental asymmetry between the near-invariant secondary loop and the dynamically controlled primary loop. In addition, all temperature outputs require only low-order representations, while primary hydraulic variables ($V_{\text{flow,prim}}$, p_{prim}) demand non-linear architectures, which overall suggests a hybrid surrogate strategy. (ii) Near-unit-root output persistence (ACF positive beyond 1,000 lags), extended cross-correlation delays (up to 1,300 steps), and 54% higher-order dynamics require recurrent or attention-based architectures with history windows of ≥1,000 seconds. (iii) The system is predominantly level-driven (median rate/level ratio = 0.33), so explicit derivative features are not mandatory for most pathways. However, selective rate augmentation for T_{ext} driven pathways may improve transient fidelity without inflating the input dimension for all channels. (iv) Negligible inter-CDU coupling (as $|r| < 0.15$, no distance-dependent attenuation and the observed random lag distribution) enables independent per-CDU models with shared weights, scalable to arbitrary CDU counts without graph-based or spatially aware architectures. (V) Zero-violation temperature ordering and HX feasibility constraints can be enforced as hard constraints. Tolerance-based energy conservation penalties (calibrated to the observed mean loss of 0.0041) accommodate structural FMU limitations without over-constraining the surrogate during training.

6 Conclusions and Future Work

This paper presented a systematic data analysis of FMU-based datacenter cooling simulations across three IBM POWER9 systems (Marconi100, Summit, and Lassen), establishing data-driven foundations for deep learning surrogate design within the ExaDigiT framework. The analysis revealed a fundamental structural asymmetry between the primary and secondary loops; the dominance of level-driven dynamics (median rate/level ratio = 0.33), with 54% of input-output pathways exhibiting higher-order transient responses and ACF persistence beyond 1,000 lags; inter-CDU thermal propagation is negligible (confirmed by the distance-independent correlations and an asymmetry ratio of 0.95); and the FMU exhibits near-zero thermodynamic violations. This characterization of temporal dynamics, non-linearity profiles, and physics consistency produces concrete specifications for architecture selection, feature engineering, and physics-regularized training. Future work will leverage these findings to develop and benchmark physics-informed deep learning surrogates targeting real-time deployment within ExaDigiT.

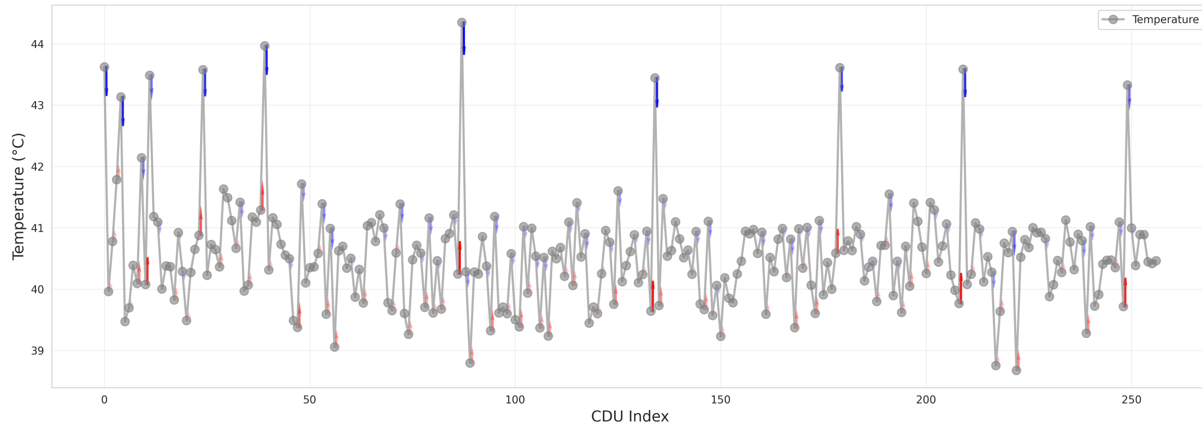

Figure 4: Gradient vector field with the designated hotspot CDUs' significant positive(Red → Heating) and negative(Blue → Cooling) gradients.

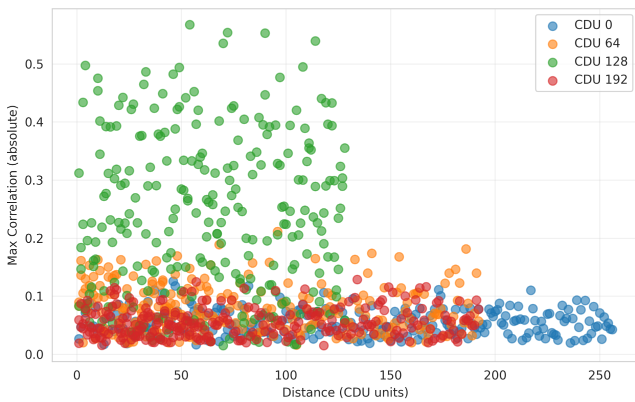

Figure 5: Load propagation analysis: Maximum cross-correlation magnitude vs. CDU separation distance for four representative source CDUs.

Acknowledgement

This research was partially sponsored by and used resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility at the Oak Ridge National Laboratory (ORNL) supported by the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

References

[1] Mohsen Seyedkazemi Ardebili et al. 2022. Rule-based thermal anomaly detection for tier-0 hpc systems. In *International Conference on High Performance Computing*. Springer, 262–276.

[2] Jyotika Athavale et al. 2024. Digital twins for data centers. *Computer*, 57, 10, 151–158. doi:10.1109/MC.2024.3436945.

[3] Wesley Brewer et al. 2024. A digital twin framework for liquid-cooled supercomputers as demonstrated at exascale. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis* (SC '24). IEEE Press, Atlanta, GA, USA. doi:10.1109/SC41406.2024.00029.

[4] Shengze Cai et al. 2021. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143, 6, 060801. doi:10.1115/1.4050542.

[5] Sophia Chen. 2025. Data centres will use twice as much energy by 2030 — driven by AI. *Nature*, (Apr. 10, 2025). Bandiera_abtest: a Cg_type: News Publisher: Nature Publishing Group. doi:10.1038/d41586-025-01113-z.

[6] Lele Fang et al. 2024. Temperature prediction in data center combining with deep neural network. *Applied Thermal Engineering*, 244, 122571. doi:10.1016/j.applthermaleng.2024.122571.

[7] Scott Greenwood et al. 2024. Thermo-Fluid Modeling Framework for Supercomputer Digital Twins: Part 2, Automated Cooling Models. In 208–217. doi:10.3384/ecp207208.

[8] Albert Gu et al. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *ArXiv e-prints*. doi:10.48550/arXiv.2312.00752.

[9] Nicola Jones. 2018. How to stop data centres from gobbling up the world's electricity. *Nature*, 561, 7722, (Sept. 12, 2018), 163–166. Bandiera_abtest: a Cg_type: News Feature Publisher: Nature Publishing Group Subject_term: Energy, Engineering, Research data, Computer science. doi:10.1038/d41586-018-06610-y.

[10] George Em Karniadakis et al. 2021. Physics-informed machine learning. *Nature Reviews Physics*, 3, 6, 422–440. doi:10.1038/s42254-021-00314-5.

[11] Vineet Kumar et al. 2024. Thermo-Fluid Modeling Framework for Supercomputer Digital Twins: Part 1, Demonstration at Exascale. *Modelica Conferences*, 199–207. doi:10.3384/ecp207199.

[12] Zongyi Li et al. 2021. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*. https://openreview.net/forum?id=c8P9NQVtmnO.

[13] Yong Liu et al. 2023. Itransformer: inverted transformers are effective for time series forecasting. In https://openreview.net/forum?id=JePfAI8fah.

[14] Yuce Liu et al. 2025. An ANN–CNN hybrid surrogate model for fast prediction of 3D temperature fields in large datacenter rooms. *Buildings*, 15, 22, 4042. doi:10.3390/buildings15224042.

[15] Lu Lu et al. 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3, 3, 218–229. doi:10.1038/s42256-021-00302-5.

[16] Seyedalborz Manavi et al. 2023. Enhanced surrogate modelling of heat conduction problems using physics-informed neural network framework. *International Communications in Heat and Mass Transfer*, 142, 106662. doi:10.1016/j.icheatmasstransfer.2023.106662.

[17] Fernando Moreno-Pino et al. 2025. Rough transformers: lightweight and continuous time series modelling through signature patching. In https://arxiv.org/abs/2405.20799 arXiv: 2405.20799 [stat.ML].

[18] Yuqi Nie et al. 2022. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. (2022). https://openreview.net/forum?id=Jbdc0vTOcol.

[19] Tapasya Patki et al. 2021. Monitoring large scale supercomputers: a case study with the lassen supercomputer. In *2021 IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, 468–480.

[20] Maziar Raissi et al. 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. doi:10.1016/j.jcp.2018.10.045.

[21] Satish Kumar Sadasivam et al. 2017. Ibm power9 processor architecture. *IEEE Micro*, 37, 2, 40–51. doi:10.1109/MM.2017.40.

[22] Soumyendu Sarkar et al. 2025. Fast 3D Surrogate Modeling for Data Center Thermal Management. *ArXiv e-prints*. doi:10.48550/arXiv.2511.11722.

[23] Angira Sharma et al. 2022. Digital Twins: State of the art theory and practice, challenges, and open research questions. *Journal of Industrial Information Integration*, 30, 100383. doi:10.1016/j.jii.2022.100383.

[24] Woong Shin et al. 2021. Revealing power, energy and thermal dynamics of a 200pf pre-exascale supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–14.