ML-Guided Cold Plate Design and Thermal Analysis for Liquid-Cooled HPC Servers

Refik Mert Cam*
University of Illinois Urbana-Champaign
rcam2@illinois.edu

Avisek Naug
Hewlett Packard Enterprise
avisek.naug@hpe.com

Andrew E. Shao
Hewlett Packard Enterprise
andrew.shao@hpe.com

Soumyendu Sarkar†
Hewlett Packard Enterprise
soumyendu.sarkar@hpe.com

Abstract

Efficient thermal management is a major bottleneck in scaling high-performance computing (HPC) systems, where cooling accounts for a substantial share of total energy use. Liquid-cooled cold plates are increasingly adopted in data centers and power electronics, yet their design optimization remains costly due to computationally burdensome computational fluid dynamics (CFD) simulations and high-dimensional geometric spaces. We introduce a physics-informed neural network (PINN) framework for rapid thermal analysis and design exploration of parameterized cold plates. Our approach jointly solves the incompressible Navier–Stokes and conjugate heat transfer equations, leveraging a two-stage curriculum that first stabilizes liquid flow field learning before introducing thermal coupling. Once trained, the model produces physically consistent predictions and orders-of-magnitude faster inference than conventional CFD solvers. We demonstrate the framework across multiple cold plate topologies, capturing design-dependent flow patterns and thermal gradients that inform geometry–performance trade-offs. These results establish PINNs as a promising surrogate modeling tool for accelerating liquid-cooling design workflows, with implications for reducing the energy and carbon footprint of HPC infrastructure.

1 Introduction

The growing scale and power density of high-performance computing (HPC) and data centers have made thermal management a first-order design constraint Naug et al. [2024, 2025], Zuo et al. [2021], Guillen-Perez et al. [2025], Sarkar et al. [2023b,a, 2024b,a, 2025]. Cooling can consume 24–60% of total data center power usage, significantly contributing to operational costs and carbon footprint Salim and Tozer [2010]. Liquid-cooled cold plates are becoming increasingly common due to their ability to handle high heat fluxes, outperforming conventional air cooling Xu et al. [2023]. However, the design of cold plates is inherently complex: performance depends on the coupled fluid flow and heat transfer behavior inside geometrically intricate channels, as well as tight constraints on manufacturing and deployment Yuan et al. [2024], Wu [2023].

Traditionally, design evaluation relies on computational fluid dynamics (CFD) simulations, which accurately resolve coupled fluid and heat transfer behavior but incur high computational cost, especially in three-dimensional conjugate heat transfer settings Zhong et al. [2025], Song et al. [2024].

*Work done during internship at HPE Labs, Hewlett Packard Enterprise
†Corresponding Author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: ML for Systems.

This cost hinders rapid iteration and large-scale design space exploration, slowing the deployment of advanced cooling technologies Zhong et al. [2025], Song et al. [2024]. Surrogate models based on machine learning offer a potential remedy, but black-box regression approaches often lack physical consistency and fail to generalize Karniadakis et al. [2021], Cam et al. [2024].

Physics-informed neural networks (PINNs)Raissi et al. [2019] offer a mesh-free alternative for solving coupled fluid flow and heat transfer PDEs, enabling fast parameter sweeps and integration into optimization workflows. Unlike regression-based surrogates, which often lack physical consistency and struggle to generalize, PINNs embed governing equations directly into the learning process, ensuring solutions remain faithful to underlying physics. Despite this promise, applying PINNs to realistic 3D conjugate heat transfer geometries remains challenging due to complex boundary conditions, numerical stiffness, and training cost Sharma et al. [2023].

In this work, we present:

1. **Parameterized geometry construction:** a modular tool to rapidly generate and vary cold plate layouts, supporting systematic exploration of design alternatives.
2. **Coupled PINN formulation:** a unified model that solves incompressible Navier–Stokes and advection–diffusion equations for conjugate heat transfer, implemented in NVIDIA PhysicsNeMo PhysicsNeMo Contributors [2023].
3. **Curriculum training strategy:** a staged learning procedure that first stabilizes flow prediction before incorporating thermal boundary conditions, improving training robustness.
4. **Design-space demonstration:** evaluation across multiple cold plate topologies, capturing design-dependent flow and thermal trends, and highlighting the potential of PINNs as surrogates for accelerated cooling system design.

2 Methodology

2.1 Parameterized Cold Plate Geometry

We construct a 3D parallel-channel cold plate using a constructive solid geometry (CSG) approach with symbolic parameterization of key features. The solid domain comprises an aluminum base plate with top and bottom shells, while the fluid domain includes inlet/outlet manifolds, ports, and interior channels (Fig. 1). Design variables such as channel width and length, manifold lengths, and overall plate length allow systematic variation of geometric proportions for design-space exploration and performance trade-off studies. The baseline geometry is adapted from Wu [2023].

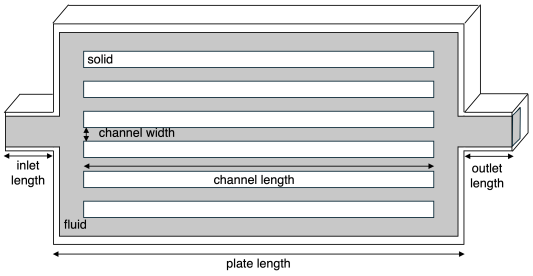

Figure 1: Parameterized parallel-channel cold plate geometry.

2.2 Governing Equations

The coupled steady-state incompressible Navier–Stokes and conjugate heat transfer equations govern fluid flow and thermal transport within the cold plate.

Fluid flow (Navier–Stokes):

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$\rho(\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \mu \nabla^2 \mathbf{u}, \tag{2}$$

where $\mathbf{u} = (u, v, w)$ is velocity, p is pressure, ρ density, and μ dynamic viscosity.

Heat transfer: In the fluid, temperature evolves by advection–diffusion,

$$\rho c_p(\mathbf{u} \cdot \nabla T_f) = k_f \nabla^2 T_f, \tag{3}$$

while in the solid region conduction dominates,

$$k_s \nabla^2 T_s = 0, \tag{4}$$

where T_f and T_s denote fluid and solid temperatures, c_p is specific heat, and k_f, k_s thermal conductivities.

Interface conditions: At the solid–fluid interface Γ_{sf},

$$T_f = T_s, \tag{5}$$

$$k_f \frac{\partial T_f}{\partial n} = k_s \frac{\partial T_s}{\partial n}, \tag{6}$$

ensuring temperature continuity and conservation of normal heat flux.

2.3 Nondimensionalization

Conjugate heat transfer involves large thermal property contrasts (e.g., $k_s \approx 237$ W/m·K for aluminum vs. $k_f \approx 0.6$ W/m·K for water), which cause ill-conditioning when training directly on dimensional PDEs. We therefore solve in nondimensional form using characteristic scales: plate length L^*, velocity U^*, and reference conductivity k^* (fluid conductivity). Temperature is defined as

$$\theta = \frac{T - T_{\text{ref}}}{\Delta T^*}, \tag{7}$$

with ΔT^* a representative inlet–outlet difference. Diffusivities and heat fluxes are scaled by $\alpha^* = k^*/(\rho^* c_p^*)$ and $q^* = k^* \Delta T^*/L^*$, yielding order-unity coefficients and balanced residuals during training.

2.4 PINN Formulation

The coupled system is solved using PINNs. The networks take spatial coordinates (x, y, z) within the cold plate geometry as input and predict the corresponding physical fields. Separate Fourier neural networks Tancik et al. [2020] are used for (i) velocity–pressure and (ii) fluid/solid temperature prediction. The total loss function combines:

1. PDE residuals for Navier–Stokes, advection–diffusion, and conduction equations,
2. boundary losses enforcing Dirichlet and Neumann conditions,
3. interface losses enforcing temperature and flux continuity, and
4. integral constraints ensuring volumetric flow rate conservation.

This formulation enables the networks to approximate steady-state velocity, pressure, and temperature fields directly from spatial queries, while ensuring consistency with governing physics across parameterized cold plate designs.

2.5 Boundary Conditions and Curriculum Training

Flow Conditions At the inlet, a velocity of 0.1 m/s is prescribed, while the outlet pressure is fixed at 0 Pa. A pointwise constraint enforces velocity components, and an integral constraint ensures the correct volumetric flow rate. No-slip conditions are applied on all channel walls.

Thermal Conditions The inlet temperature is set to $25°$C. The outlet is assigned a zero-gradient condition. A localized surface heat flux of $50{,}000$ W/m^2 is imposed on the solid plate, while all other external surfaces are adiabatic.

Curriculum Training Training is staged for stability: (i) **flow-only pretraining**, where the velocity–pressure network is trained using Navier–Stokes residuals, no-slip walls, and continuity constraints; and (ii) **thermal coupling**, where the flow network is frozen and thermal subnetworks are trained with advection–diffusion, conduction, interface continuity, and thermal boundary conditions. This progression mitigates stiffness in the coupled system and prevents flow errors from contaminating thermal learning.

3 Results

We assess PINN predictions against high-fidelity `OpenFOAM` Jasak [2009] simulations for three cold-plate geometries differing only in channel width (5, 6.5, and 8 mm). Two application-relevant metrics are considered:

1. **Pressure drop** between inlet and outlet, measured 0.1 mm from the plate surface, reflecting hydraulic resistance and pumping power requirements.
2. **Maximum temperature** in the solid plate, indicating thermal safety margins for electronic reliability.

Table 1 reports results across designs. The PINN captures geometry-dependent flow and thermal trends, with relative errors of 5–14% compared to CFD.

Table 1: Comparison of pressure drop and maximum temperature between CFD baselines and PINN predictions across three channel-width designs.

Design	Pressure Drop [Pa]			Max Temperature [K]		
	CFD	PINN	Rel. Err. [%]	CFD	PINN	Rel. Err. [%]
5 mm	55.5	62.9	13.5	334.08	318.06	4.8
6.5 mm	42.8	40.9	4.4	334.06	318.42	4.7
8 mm	35.5	30.7	13.3	334.02	319.06	4.5

Both CFD and PINN reveal that maximum plate temperature is nearly constant across channel widths, indicating limited thermal sensitivity under current operating conditions. In contrast, pressure drop decreases systematically with increasing width, reflecting reduced hydraulic resistance and pumping power demand. This highlights a key design insight: wider channels improve cooling efficiency without compromising thermal safety. Figure 2 shows representative pressure and temperature fields for the channel width of 6.5 mm design, while additional comparisons for the 5 mm and 8 mm designs are provided in Appendix 3 and Appendix 4. The PINN reproduces global flow patterns, supporting its utility as a surrogate for rapid design-space exploration.

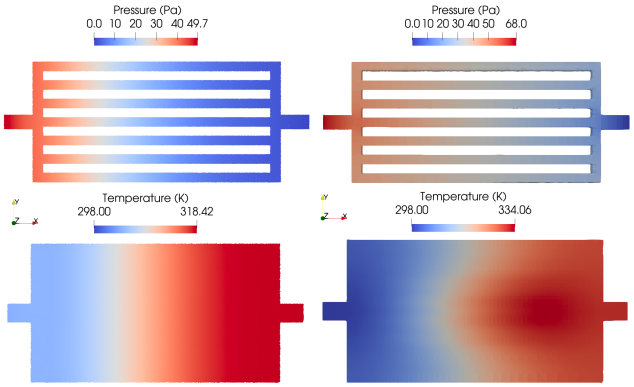

Figure 2: Comparison of PINN (left) and `OpenFOAM` (right) results for the 6.5 mm channel-width design. The first row shows the pressure field at the bottom of the channels, while the second row shows the temperature field at the bottom of the solid plate.

4 Discussion

Our results indicate that the PINN framework can predict flow and thermal fields in realistic cold-plate geometries with good agreement to CFD. In this preliminary work, a separate PINN was trained for each geometry variation. Future efforts will focus on incorporating design parameters (e.g., channel width, spacing, manifold dimensions) directly into the network input along with spatial coordinates. This would allow a single model to generalize across the design space, enabling efficient surrogate modeling and rapid iteration. Beyond parallel-channel plates, future work could also incorporate parameterized geometries such as serpentine, spiral, and bio-inspired topologies. From a computational perspective, each CFD simulation required approximately 4–5 minutes, whereas the trained PINN produced predictions for 500,000 domain points in only 2–3 seconds—nearly two orders of magnitude faster. While PINNs require additional training time, extending them to accept design parameters as input could amortize this cost and make them highly attractive for large-scale design optimization in cooling applications. More broadly, such acceleration of thermal management design supports the development of energy-efficient cooling technologies, which are critical for reducing data center carbon footprints and mitigating climate change impacts.

References

Refik Mert Cam, Umberto Villa, and Mark A Anastasio. Learning a stable approximation of an existing but unknown inverse mapping: application to the half-time circular radon transform. *Inverse Problems*, 40(8):085002, 2024.

Antonio Guillen-Perez, Avisek Naug, Vineet Gundecha, Sahand Ghorbanpour, Ricardo Luna Gutierrez, Ashwin Ramesh Babu, Munther Salim, Shubhanker Banerjee, Eoin H Oude Essink, Damien Fay, et al. Dccluster-opt: Benchmarking dynamic multi-objective optimization for geo-distributed data center workloads. *arXiv preprint arXiv:2511.00117*, 2025.

Hrvoje Jasak. Openfoam: Open source cfd in research and industry. *International journal of naval architecture and ocean engineering*, 1(2):89–94, 2009.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Avisek Naug, Antonio Guillen-Perez, Ricardo Luna Gutierrez, Vineet Gundecha, Cullen Bash, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, Dejan Markovikj, Lekhapriya Dheeraj Kashyap, et al. Sustaindc: Benchmarking for sustainable data center control. *Advances in Neural Information Processing Systems*, 37:100630–100669, 2024.

Avisek Naug, Antonio Guillen, Vineet Kumar, Scott Greenwood, Wesley Brewer, Sahand Ghorbanpour, Ashwin Ramesh Babu, Vineet Gundecha, Ricardo Luna Gutierrez, and Soumyendu Sarkar. Lc-opt: Benchmarking reinforcement learning and agentic ai for end-to-end liquid cooling optimization in data centers. *arXiv preprint arXiv:2511.00116*, 2025.

PhysicsNeMo Contributors. Nvidia physicsnemo: An open-source framework for physics-based deep learning in science and engineering. `https://github.com/NVIDIA/physicsnemo`, 2023. Version released on 2023-02-24.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. 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, 2019.

Munther Salim and Robert Tozer. Data centers' energy auditing and benchmarking-progress update. *ASHRAE transactions*, 116(1), 2010.

Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Ricardo Luna Gutierrez, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, and Vineet Gundecha. Concurrent carbon footprint reduction (c2fr) reinforcement learning approach for sustainable data center digital twin. In *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pages 1–8, 2023a. doi: 10.1109/CASE56687.2023.10260633.

Soumyendu Sarkar, Avisek Naug, Ricardo Luna Gutierrez, Antonio Guillen, Vineet Gundecha, Ashwin Ramesh Babu, and Cullen Bash. Real-time carbon footprint minimization in sustainable data centers with reinforcement learning. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023b.

Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Ricardo Luna, Vineet Gundecha, Ashwin Ramesh Babu, and Sajad Mousavi. Sustainability of data center digital twins with reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21):23832–23834, Mar. 2024a. doi: 10.1609/aaai.v38i21.30580. URL `https://ojs.aaai.org/index.php/AAAI/article/view/30580`.

Soumyendu Sarkar, Avisek Naug, Ricardo Luna, Antonio Guillen, Vineet Gundecha, Sahand Ghorbanpour, Sajad Mousavi, Dejan Markovikj, and Ashwin Ramesh Babu. Carbon footprint reduction for sustainable data centers in real-time. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):22322–22330, Mar. 2024b. doi: 10.1609/aaai.v38i20.30238. URL `https://ojs.aaai.org/index.php/AAAI/article/view/30238`.

Soumyendu Sarkar, Avisek Naug, Antonio Guillen, Vineet Gundecha, Ricardo Luna Gutiérrez, Sahand Ghorbanpour, Sajad Mousavi, Ashwin Ramesh Babu, Desik Rengarajan, and Cullen Bash. Hierarchical multi-agent framework for carbon-efficient liquid-cooled data center clusters. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28):29694–29696, Apr. 2025. doi: 10.1609/aaai.v39i28.35370. URL `https://ojs.aaai.org/index.php/AAAI/article/view/35370`.

Prakhar Sharma, Llion Evans, Michelle Tindall, and Perumal Nithiarasu. Stiff-pdes and physics-informed neural networks. *Archives of Computational Methods in Engineering*, 30(5):2929–2958, 2023.

Zhihao Song, Xintian Liu, Yu Fang, Xu Wang, and Shengchao Su. Topology optimization for cold plate using neural networks as proxy models. *Engineering Optimization*, 56(12):2359–2386, 2024.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Mao-Sung Wu. Multi-objective topology optimization of cold plates featuring branched and streamlined mini-channels for thermal management system of lithium-ion battery module. *Journal of Energy Storage*, 72:108362, 2023.

Sijun Xu, Hua Zhang, and Zilong Wang. Thermal management and energy consumption in air, liquid, and free cooling systems for data centers: A review. *Energies*, 16(3):1279, 2023.

Jifeng Yuan, Zhengjian Gu, Jun Bao, Tao Yang, Huanhuan Li, Yaping Wang, Lei Pei, Haobin Jiang, Long Chen, and Chaochun Yuan. Structure optimization design and performance analysis of liquid cooling plate for power battery. *Journal of Energy Storage*, 87:111517, 2024.

Qixuan Zhong, Liang Gao, Wei Li, and Akhil Garg. Transfer learning based topology optimization of battery cooling channels design for improved thermal performance. *Applied Thermal Engineering*, 263:125400, 2025.

Wangda Zuo, Michael Wetter, James Van Gilder, Xu Han, Yangyang Fu, Cary Faulkner, Jianjun Hu, Wei Tian, and Michael Condor. Improving data center energy efficiency through end-to-end cooling modeling and optimization. *Report for US Department of Energy, DOE-CUBoulder-07688*, pages 1–109, 2021.

A Additional Visual Comparisons of PINN and CFD Results Across Channel Widths

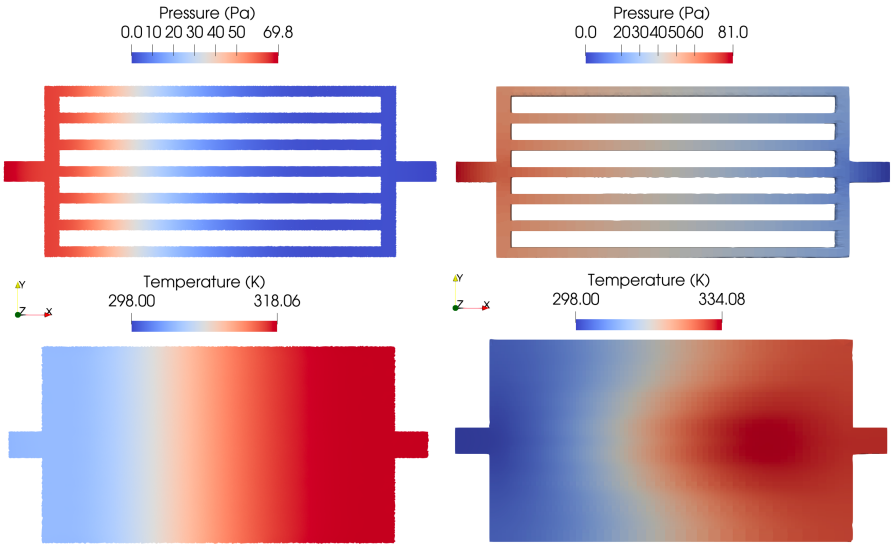

Figure 3: Comparison of PINN (left) and `OpenFOAM` (right) results for the 5 mm channel-width design. The first row shows the pressure field at the bottom of the channels, while the second row shows the temperature field at the bottom of the solid plate.

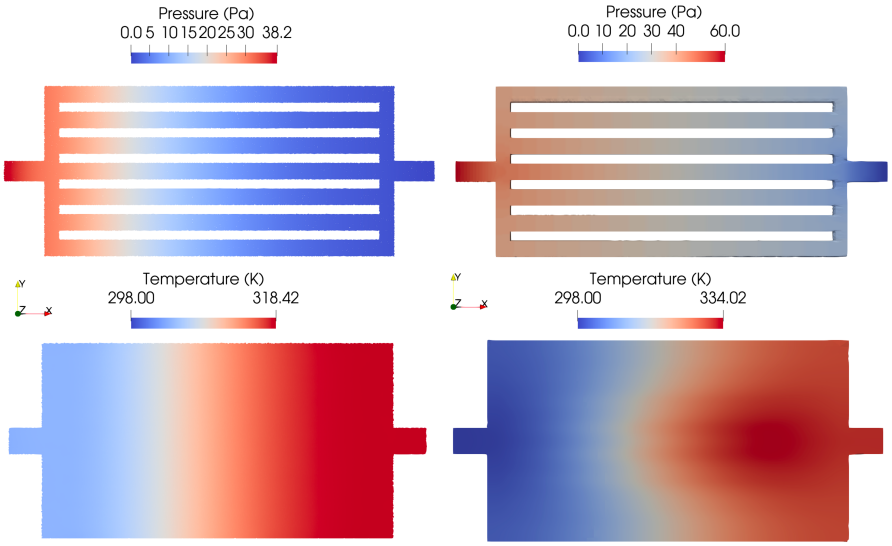

Figure 4: Comparison of PINN (left) and `OpenFOAM` (right) results for the 8 mm channel-width design. The first row shows the pressure field at the bottom of the channels, while the second row shows the temperature field at the bottom of the solid plate.