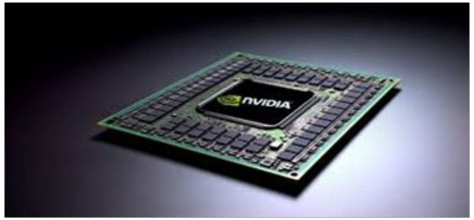
FEA Modeling for Extreme Cooling on Nvidia’s Blackwell GB200

By Michael Berhane, Kyara Gandara, Justin Law, Kevin Medici

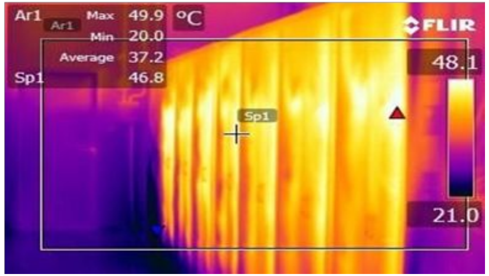
Background: Servers, AI chips, and Heat Problems

- Massive rooms filled with racks of high-powered servers.
- Thousands of machines working 24/7 = huge energy demand.
- A single server rack today can draw **>30kW** of power.

- Modern AI chips like Nvidia Blackwell GB200 are **extremely powerful**.
- Each chip consumes **~1000W** — as much as a microwave oven!
- Clusters of GPUs generate record levels of heat in tight spaces.



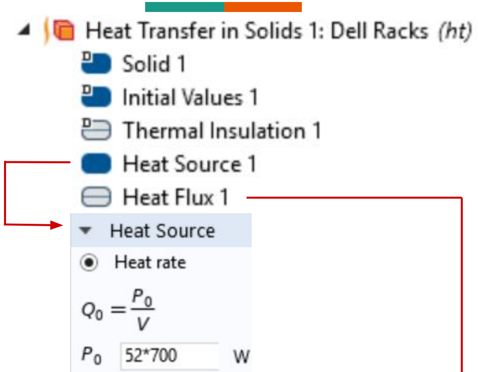
- traditional cooling (fans, air conditioning) cannot keep up.
- **Over 40%** of total data center energy today goes to cooling (Uptime Institute, 2023).
- Hot spots can cause server failures, massive energy waste, and downtime.



Problem Definition: Need for Next-Gen Cooling

- Cooling an AI server farm can **consume as much water as 300 Olympic swimming pools per year** using traditional methods.
- Air and liquid-to-air systems **cannot support extreme power densities** (10–100x higher than legacy data centers).
- Next-generation cooling must:
  - **Cut water usage dramatically** (immersion = closed loop)
  - **Handle up to 120 kW per rack safely**
  - **Improve energy efficiency by 30–50%** over air cooling
- Mission: Build a scalable, immersion-based cooling system that protects Nvidia Blackwell GB200 clusters.

Physics



Q0 = Heat generated per unit Volume
(Watts per cubic meter) [W/m³]
Amount of thermal energy added inside the material per second. Thus, 12,526 Watts of heat per second.

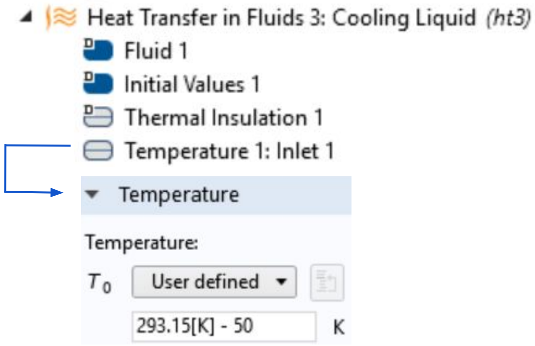
$$q = h_{\text{fluid}} \times A_{\text{server}} \times \Delta T$$

q = heat transfer rate (W)

h_{fluid} = convective heat transfer coefficient (e.g., 200–500 W/m²·K for Fluorinert)

A_{server} = surface area of the server touching fluid (m²)

ΔT = temperature difference between server surface and fluid (K)



Temperature:
FC-40 Can be up to 233.15 K (-40 C).
It is set at -30 C.

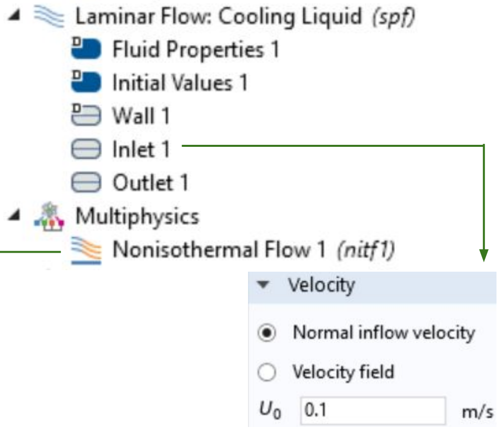
$$Q = \dot{m} \cdot c_p \cdot \Delta T_{\text{fluid}}$$

Q = heat power absorbed (W)

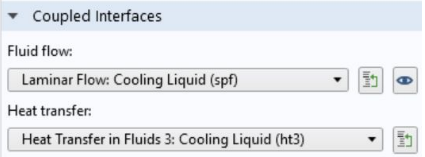
$\dot{m}$ = mass flow rate of fluid (kg/s)

c_p = specific heat capacity of fluid (J/kg·K)

ΔT_{fluid} = temperature rise of the fluid (°C)



Velocity:
Typical Velocities are 0.1 – 0.25 m/s, but due to inconsistencies, 0.1 m/s was the realistic it could be.



Nonisothermal Flow:
Temperature affects fluid.
Flow affects temperature.
Moving liquid carries heat away from hot GPUs.
Connects velocity field (u) into heat equation, and temperature (T) into fluid density/viscosity

Coupled Interfaces :
Two-way coupling between: Heat ↔ Flow

Blackwell GB200 Server Specifications

Component	Value
NVIDIA GB200 Chip (W)	700
Total Chips per Rack	72
Total Rack Power Consumption (W)	50,400.00
Cooling Efficiency (COP)	2.5
Energy Needed for Cooling (W)	28.8
Total Cooling Energy (kW*h)	3,024.00
Dell PowerEdge R740 server	
Server Height (cm)	8.68
Server Width (cm)	48.2
Server Length (cm)	43.4
Server Volume (cm^3)	18,157.52
Coolant Type	Fluorinert FC-40
Coolant Flow Rate (L/min)	12,660.43
Coolant Specific Heat Capacity (J/g°C)	1.1
Temperature of fluid (°C)	-10
Temperature of the server (°C)	100
Temperature Difference (delta T) (°C)	110
Cooling Power via Fluid (W)	18,568,625.34
Total Cooling Energy via Fluid (kW*h)	1,114,117.52

Chassis dimensions

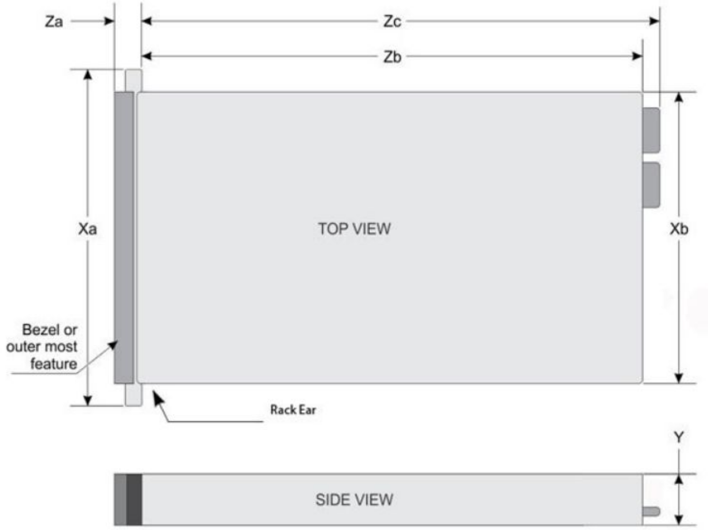
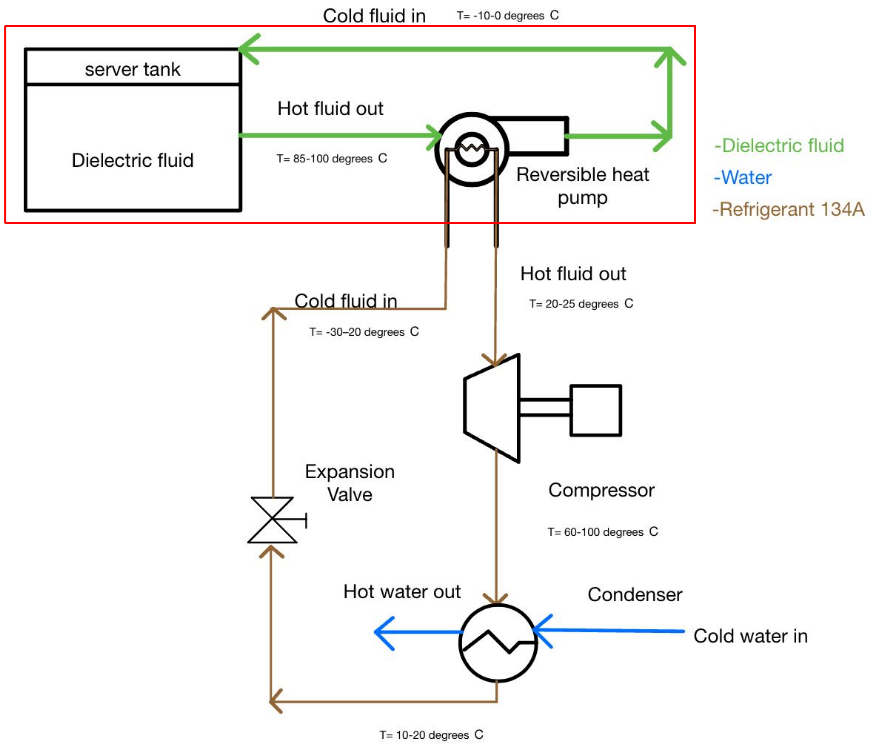


Table 17. Chassis dimensions

Chassis dimensions (cm)						
Xa	Xb	Y	Za bezel	Za without bezel	Zb	Zc
482.0 mm	434.0 mm	86.8 mm	35.84mm	22.0 mm	678.8 mm	715.5 mm

Process Flow Diagram

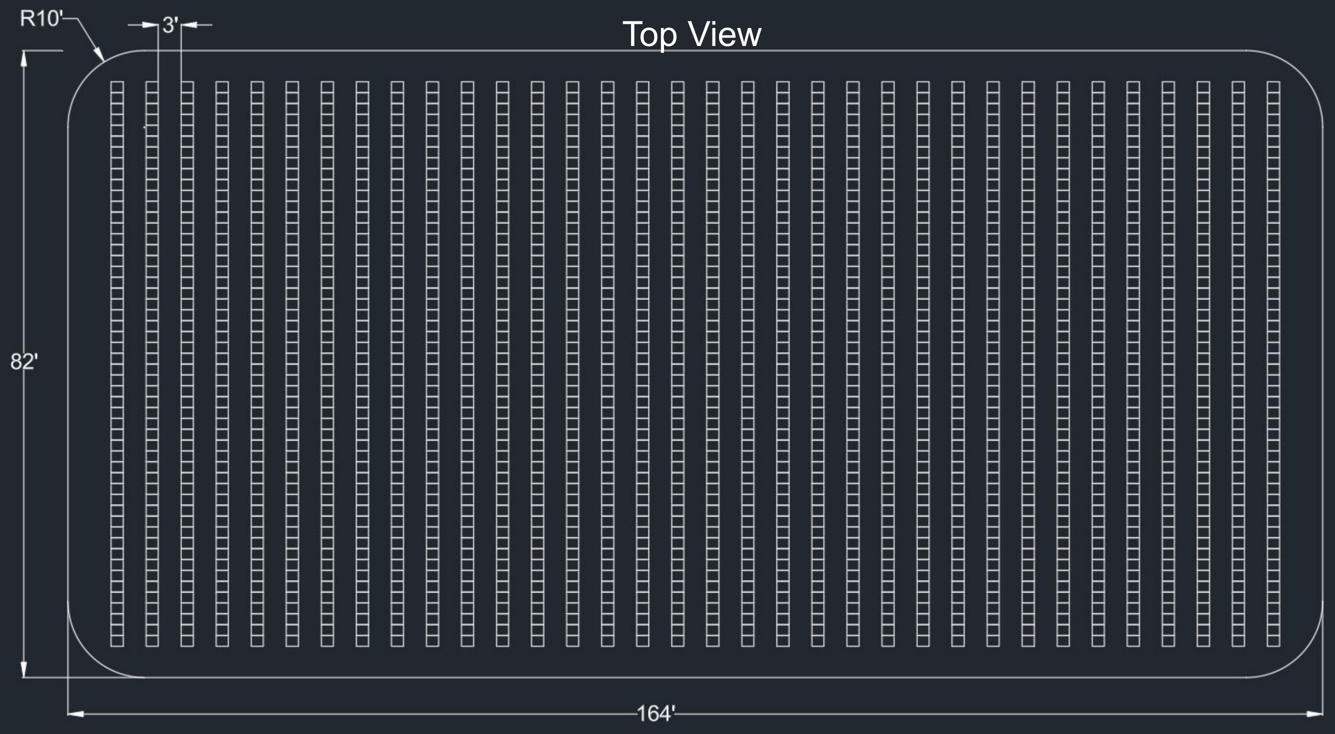


Component	Function
Server Tank	Absorbs heat from servers
Reversible Heat Pump	Moves heat out of system
Compressor	Pressurizes refrigerant
Condenser	Releases heat from refrigerant
Expansion Valve	Drops refrigerant pressure to cool

Energy and Cooling Calculations

Total Power per tank (kW)	Total Power (kW) servers total	Total Power per year(kWh) servers total
30,139.20	8,108,755.20	71,032,695,552.00
Flow Rate (gal/min)	work done by compressor (kW)	work done by condenser (kW)
2,922.92	15,889.81	4,000.00
work done by pump (kW)	energy for dielectric tanks (kW)	Total energy use per tank(kW)
17.90	2,742,667.20	19,907.72
total cooling power (kW)	total cooling power per year (kWh)	COP
3,000,000 - 5,000,000	26,280,000,000 - 43,800,000,000	1.6-2.7

Single Tank Schematic



34 ROWS

52 SERVER RACKS PER ROW

1,768 SERVER RACKS PER TANK

18 SERVERS PER SERVER RACK

31,824 SERVERS PER TANK

DIMENSIONS OF OLYMPIC POOL

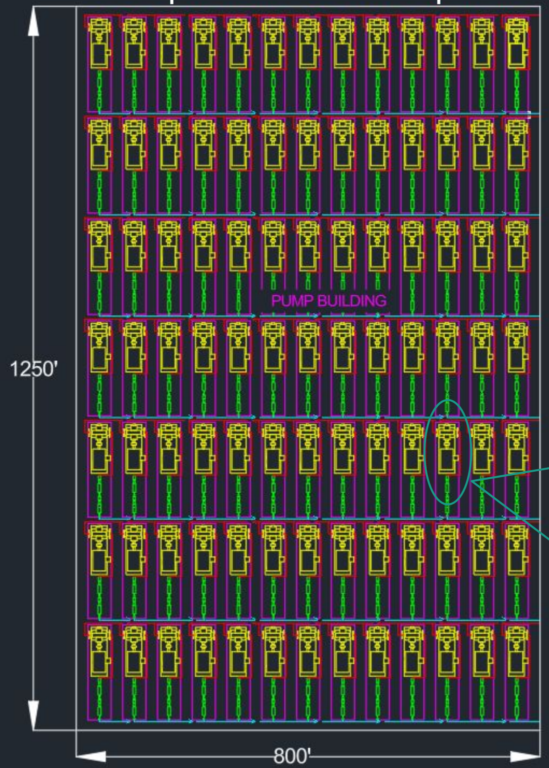
3' BETWEEN ROWS FOR MAINTENANCE

~13,500 SQFT TOTAL

~6,067 SQFT OF SERVERS

Server & Pump Buildings Schematic

Top View Heat Pump Room



91 TOTAL PUMP STATIONS

1 MILLION SQFT TOTAL

~663,000 SQFT FOR PUMP STATIONS

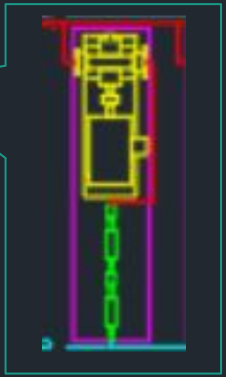
~7,300 SQFT PER PUMP STATION

1 PUMP & 2 REVERSIBLE HEAT PUMPS PER STATION

20' BETWEEN COLUMNS FOR MAINTENANCE / PIPING

11' BETWEEN ROWS FOR PIPING

66% FLOOR SPACE EFFICIENCY (ROOM FOR EXPANSION)



YELLOW - PUMP
GREEN - COMPRESSOR
CYAN - COLD FLUID OUT
RED - HOT FLUID IN
MAGENTA - PUMP STATION OUTLINE

Top View Server Farm

91 TOTAL TANKS

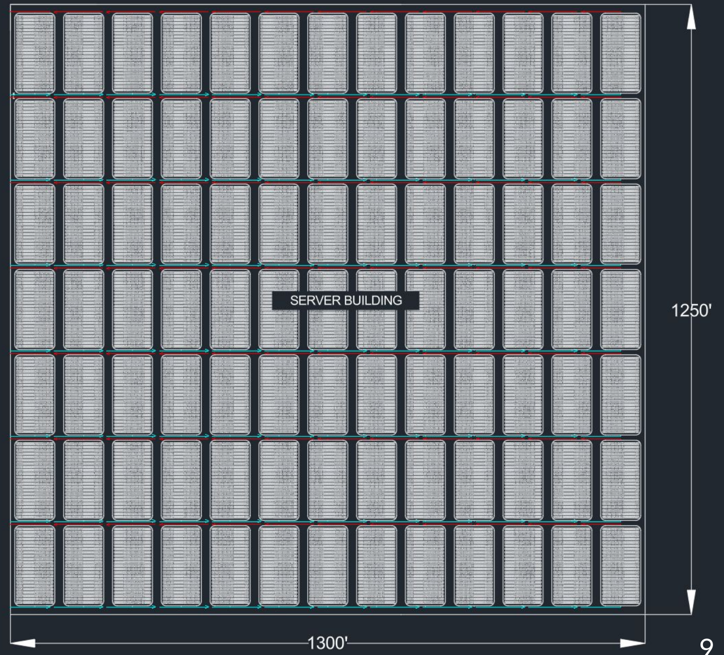
1.625 MILLION SQFT TOTAL

1.23 MILLION SQFT FOR TANKS

18' BETWEEN COLUMNS

11' BETWEEN ROWS FOR PIPING

76% FLOOR SPACE EFFICIENCY

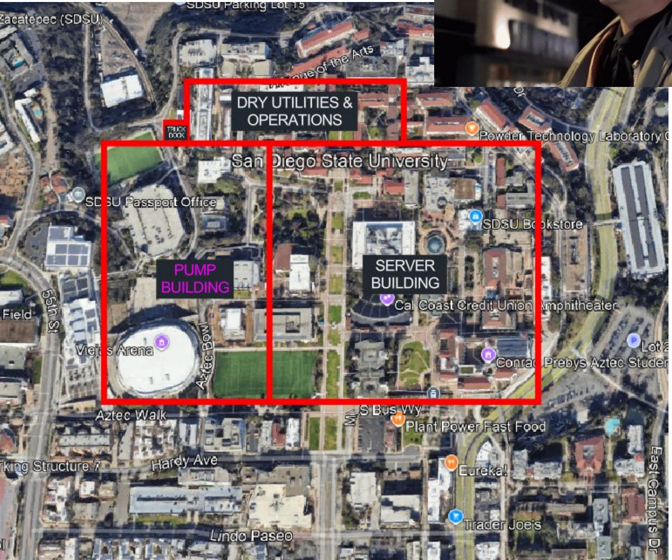


Full Building Schematic

Total of 2.9 Million SQFT

Building Dedicated to Operations & Dry Utilities (Electricity/Internet/Telecommunications)

Top View SDSU vs. Data center



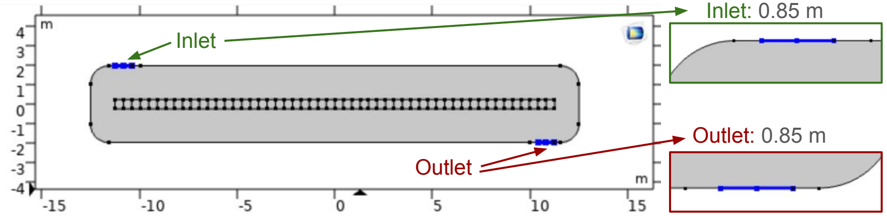
Top View Full Building

Comsol Setup: Cooling Model Details

Dimensions:

Tank (LxWxH): 24.99 x 3.92 x 3 [m] (81.99 x 12.86 x 9.84 [ft])

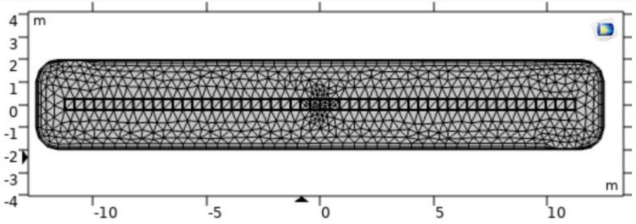
Rack (LxWxH): 0.433 x 0.482 x 0.0868 [m] (1.42 x 1.58 x 0.285 [ft])



Convergence Parameters:

Mesh Size: Extremely Coarse

Discretization: Quadratic Lagrange



Materials:

- Materials
  - 3M™ Fluorinert™ Electronic Liquid FC-40
  - Silicone [solid] *(mat5)*

Coolant Type:

3M™ Fluorinert™ Electronic Liquid FC-40

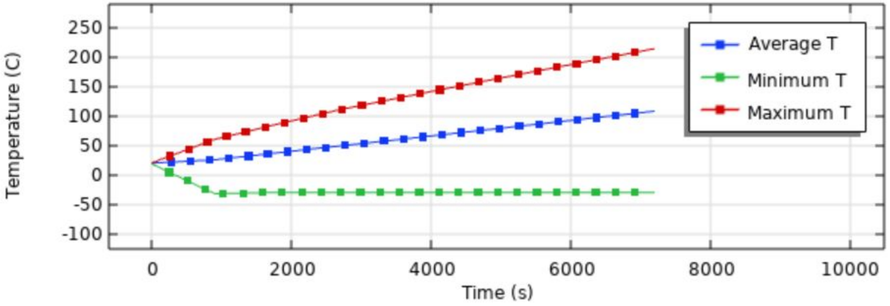
Property	Variable	Value	Unit
Dynamic viscosity	mu	0.0025	Pa·s
Heat capacity at constant pres...	Cp	1100	J/(kg·K)
Density	rho	1850	kg/m³
Thermal conductivity	k_iso ;...	0.065	W/(m·K)
Ratio of specific heats	gamma	1	1

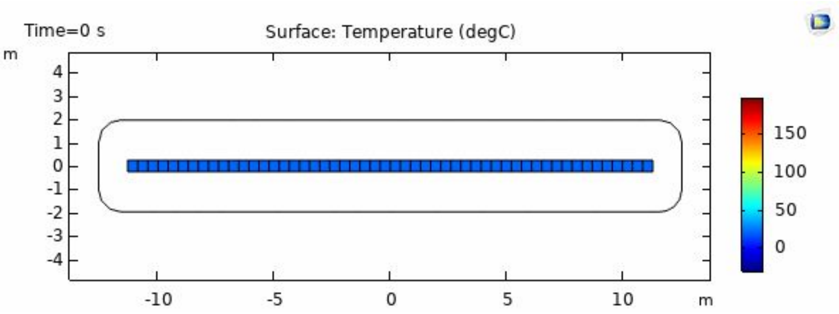
Comsol Results: Temperature Distribution

Server Racks

Temperature (C) vs Time (s)

Note: Simulation ran for 2 hours

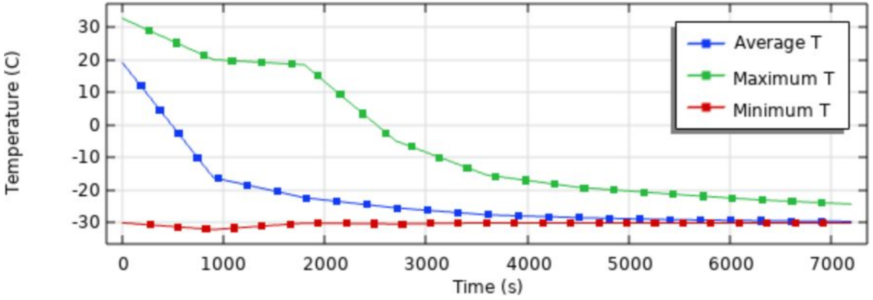


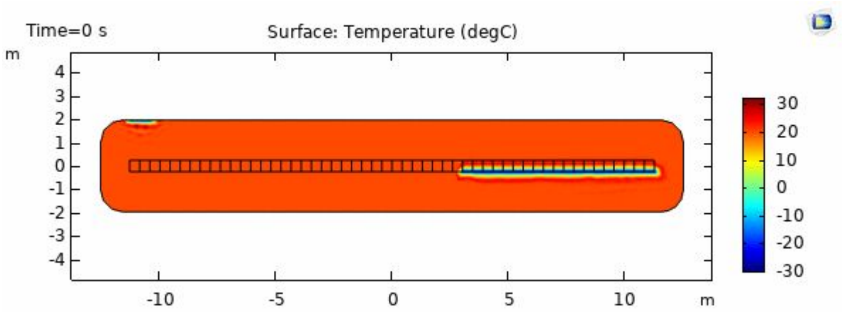


Coolant Liquid

Temperature (C) vs Time (s)

Note: Simulation ran for 2 hours





Cooling Validation: Energy and Fluid Flow Results

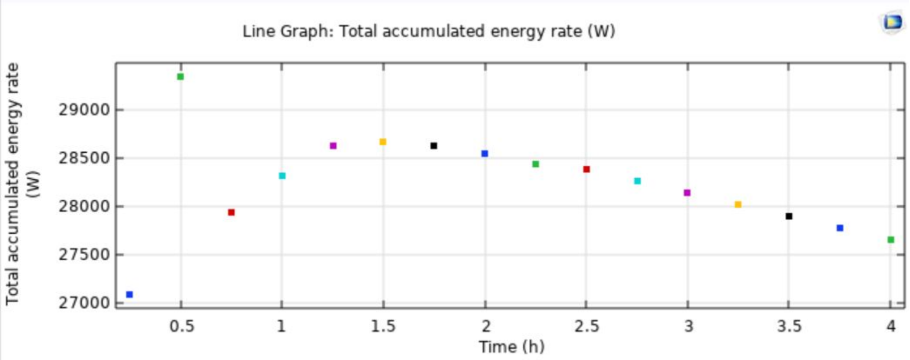
Reduction in Internal Energy Rate= 28,658.44 W− 27,655.53 W =1,002.91 W

This corresponds to a 3.50% reduction over 4 hours.

Reflects on the cooling system's effectiveness in gradually removing heat and approaching thermal steady-state.

The system is removing thermal energy faster than it accumulates, resulting in a net decline in internal energy.

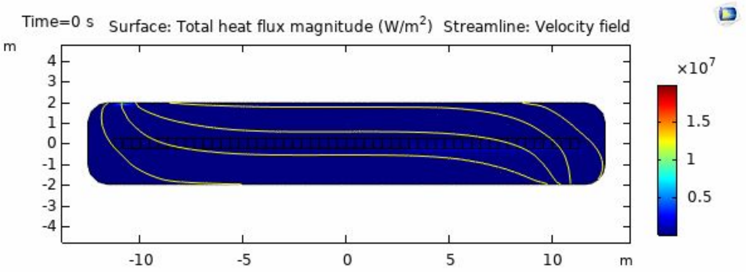
Accumulated Energy Rate vs. Time:
Evaluating System Cooling Performance over 4 Hours



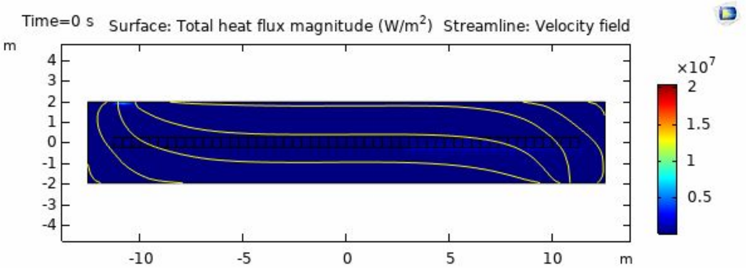
Average thermal energy removal rate = (28,658.44 W +27,655.53 W)/2 = 28,156.99 W

Total energy removed over 4 hours = 28,156.99 W × 4 hr = 112,628 Wh

With Fillets **Note:** Simulation ran for 2 hours



Without Fillets **Note:** Simulation ran for 2 hours

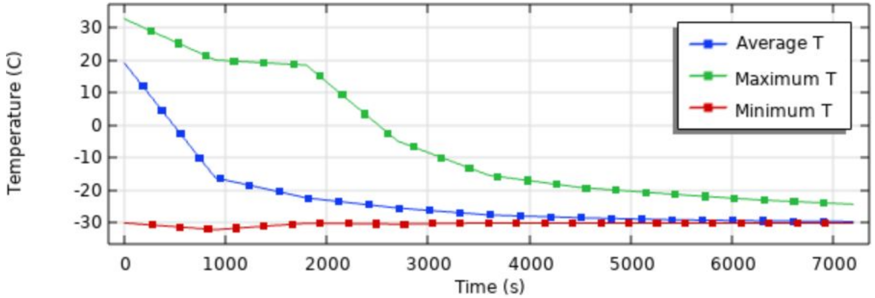


Conclusion

- We designed a cooling system using single-phase immersion cooling with server tanks to manage the extreme thermal loads of the NVIDIA Blackwell GB200 Servers.
- Our FEA models confirmed that the system successfully reduces the total energy buildup inside the server racks.
- Temperature distribution analysis shows that the servers remain within safe operating temperatures when submerged in the cooling tanks.
- The system provides effective and scalable cooling for future high-density AI data centers.

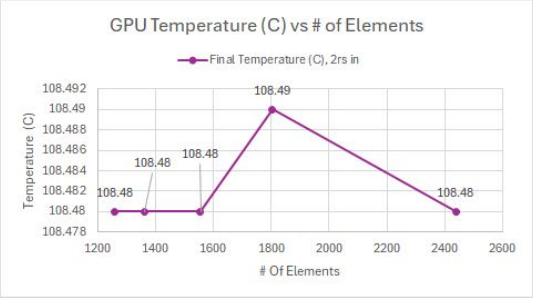
Coolant:

Temperature (C) vs Time (s), 2 Hours in



Convergence Test:

Final Temperature (C) vs # of Elements, 2 Hours in



Future Work and Improvements

<u>Spacing of servers:</u>
Adding spacing can prevent hot spots that may develop in tightly packed arrays.

Coolant can convect more efficiently between and around racks.

<u>Placement of Inlet and Outlet:</u>
Compared to diagonal setups, side-inlet/side-outlet configurations reduce stagnant zones or hot pockets near the corners.

Flow is more predictable and symmetric, making it easier to optimize.

<u>Insulated Piping and Fluid Reservoir</u>

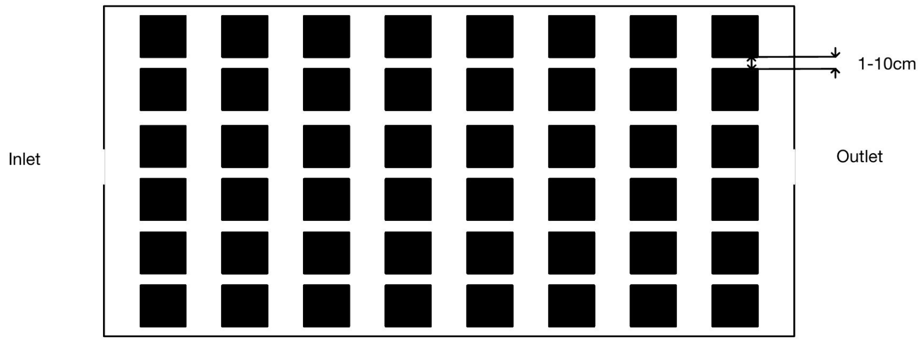
1) Vacuum-Insulated Pipes or Foam-Insulated Tubing

- Minimizes heat gain from the room.
- Prevents condensation or frost on pipes.

2) Low-Temperature-Compatible Pump and Reservoir

- Must handle high-viscosity fluid at −40 °C without stalling.
- <u>Materials:</u> Stainless steel or fluoropolymer-compatible plastics.

Top View modified Server Tank



References

[1] 3M, 3M Fluorinert Electronic Liquid FC-40 Product Information, 3M Company, 2002. [Online]. Available: https://multimedia.3m.com/mws/media/64888O/3m-fluorinert-electronic-liquid-fc-40.pdf

[2] Dell Technologies, PowerEdge R740 Server Specification Sheet, Dell Inc., 2017. [Online]. Available: https://i.dell.com/sites/csdocuments/shared-content_data-sheets_documents/en/poweredge-r740-spec-sheet.pdf

[3] SPRSUN, Simple Diagram & Operation of Heat Pumps – Pump up Your Heat Pump Knowledge, SPRSUN Heat Pump Manufacturer, 2022. [Online]. Available: https://sprsunheatpump.com/Simple-Diagram-Operation-of-Heat-Pumps-Pump-up-Your-Heat-Pump-Knowledge-id41518747.html

[4] Lumen Learning, Applications of Thermodynamics: Heat Pumps and Refrigerators, Lumen Learning Physics Course. [Online]. Available: https://courses.lumenlearning.com/suny-physics/chapter/15-5-applications-of-thermodynamics-heat-pumps-and-refrigerators/

[5] NVIDIA, *NVIDIA Blackwell Architecture Datasheet*, NVIDIA Corporation, 2024. [Online]. Available: https://nvdam.widen.net/s/wwnsxrhm2w/blackwell-datasheet-3384703

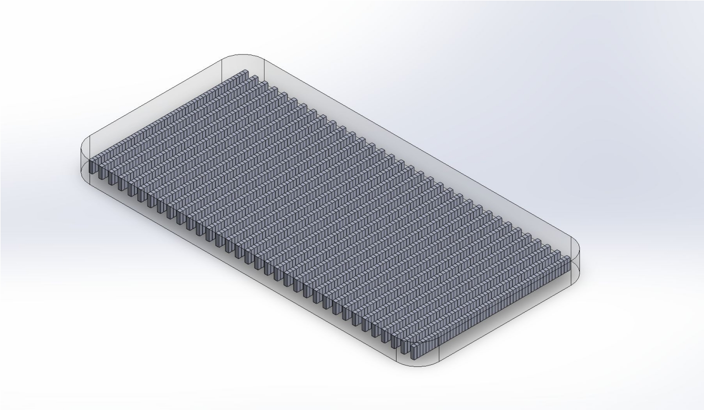
[6] S. Al-Shehri and H. H. Saber, "Experimental investigation of using thermoelectric cooling for computer chips," *Department of Computer Science and Engineering and Department of Mechanical Engineering, Jubail University College*, Saudi Arabia, 2020. [Online]. Available: [local file]

[7] OpenAI, ChatGPT, San Francisco, CA, USA, Apr. 29, 2025. [Online]. Available: https://chat.openai.com

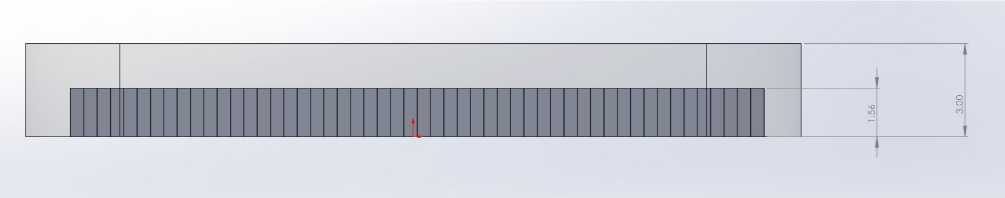
Appendix 1 Cooling Decision Matrix

Characteristics	Chiller	Dry Cooler	Cooling Tower	Reversible Heat Pump
Energy Efficiency (COP)	2	3	5	3
Initial Capital Cost	2	3	3	3
Water Usage	5	5	1	5
Climate Flexibility	5	2	3	4
Maintenance Requirements	3	4	2	3
Environmental Sustainability	4	5	2	4
Scalability for Hyperscale DC	4	4	5	3
Modularity	3	4	3	5
Precision Temperature Control	5	3	2	5
total	33	33	26	35

Appendix 2 3D Model of Server Tank



Server Tank Isometric View (Units in meters)



Server Tank Right View (Units in meters)