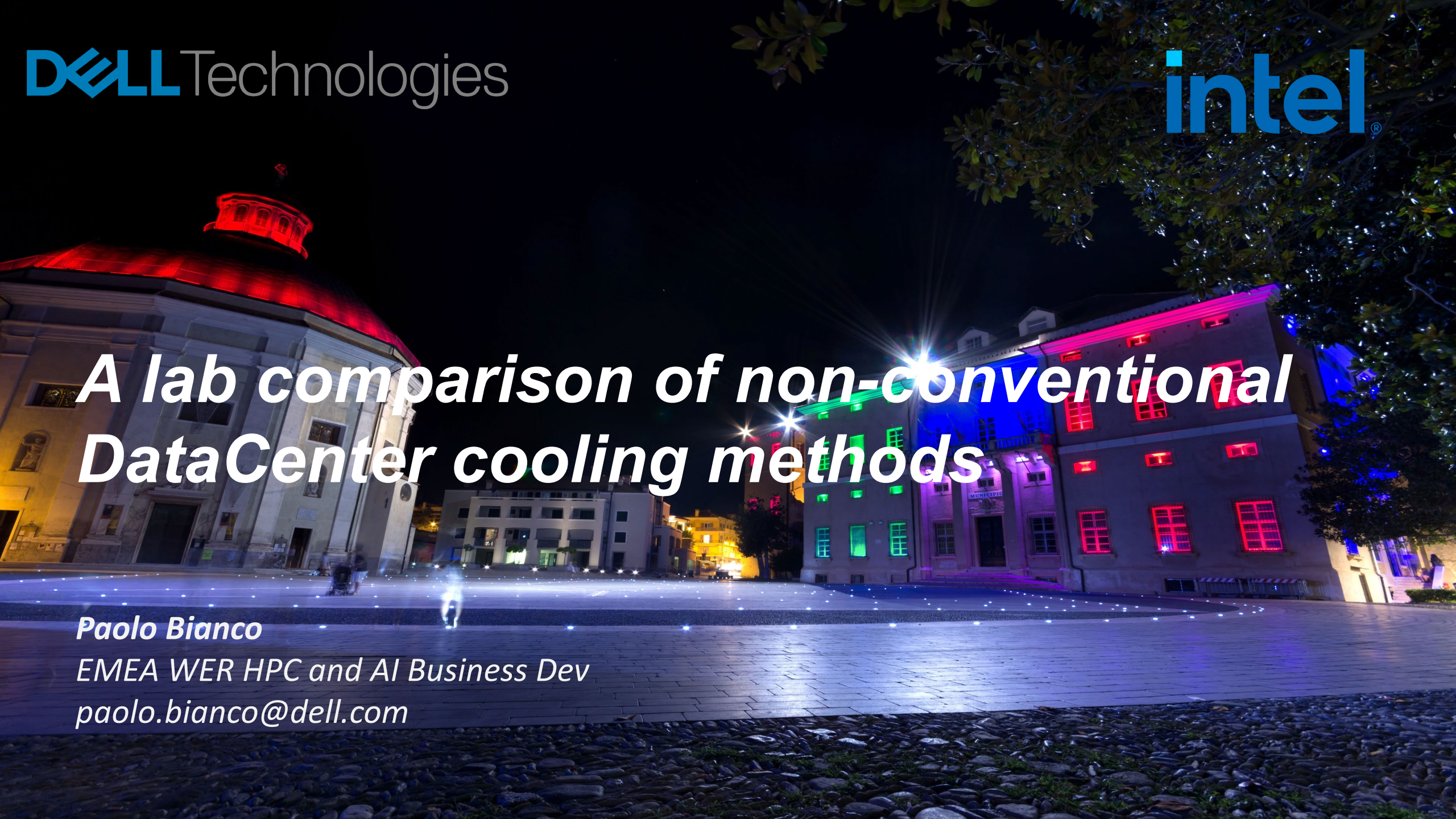
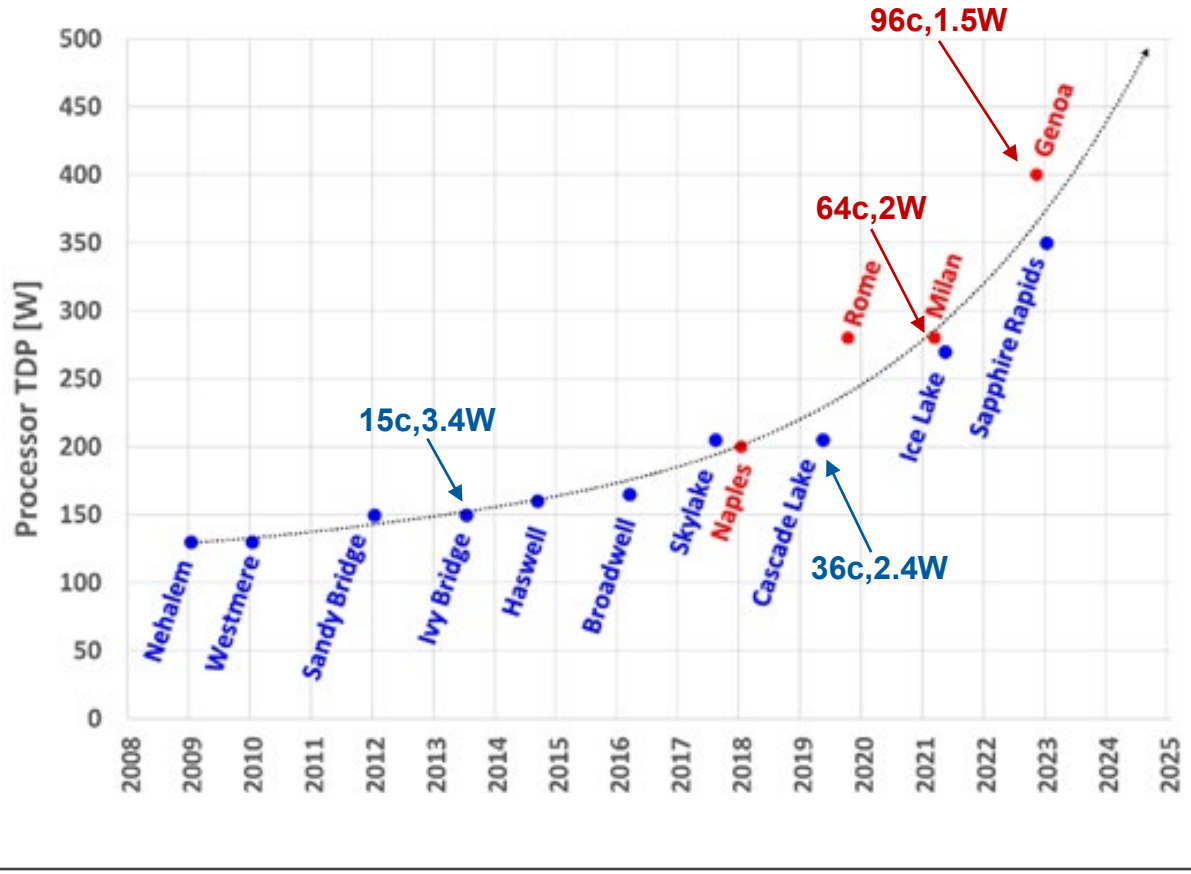


A lab comparison of non-conventional DataCenter cooling methods

Paolo Bianco
EMEA WER HPC and AI Business Dev
paolo.bianco@dell.com

Thermal Design Power Trends

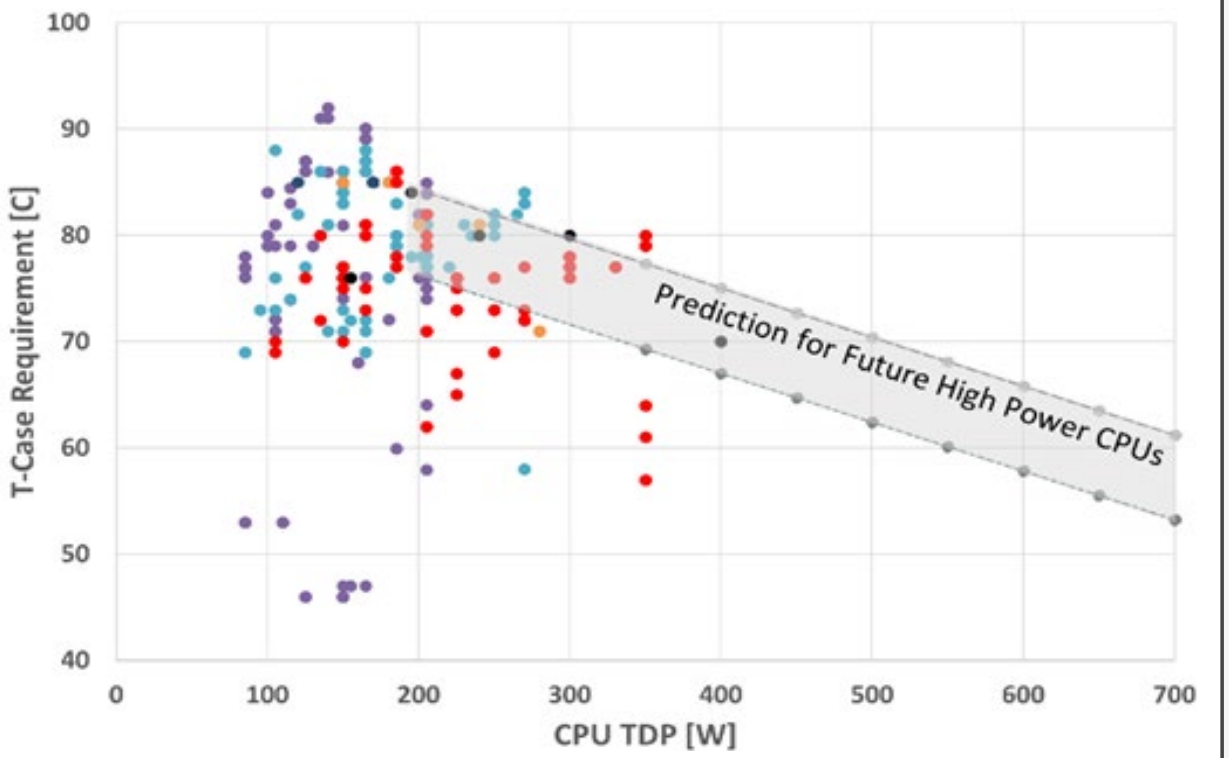
Processor thermal design powers trending to 500 W by 2025



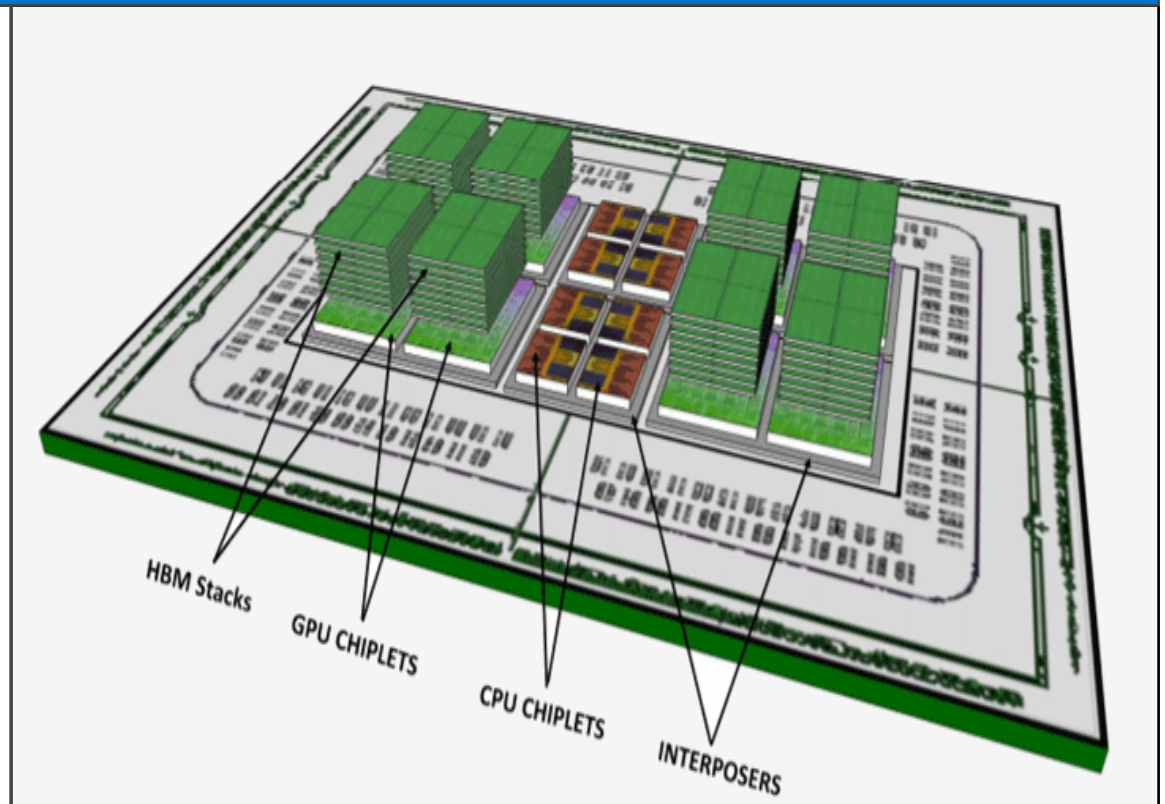
- Competition in the CPU & GPU markets ("Power war") will continue to drive up power
  - Higher TDPs
  - Higher core number
- Increased Memory count, capacity and speed all adding power
- Accelerated adoption of NVMe, high speed I/O and accelerators also contributing more power
- **Result:** extended air-cooling causing challenges within data center

Further constraints

Silicon limits driving lower case temperatures to deal with higher powers



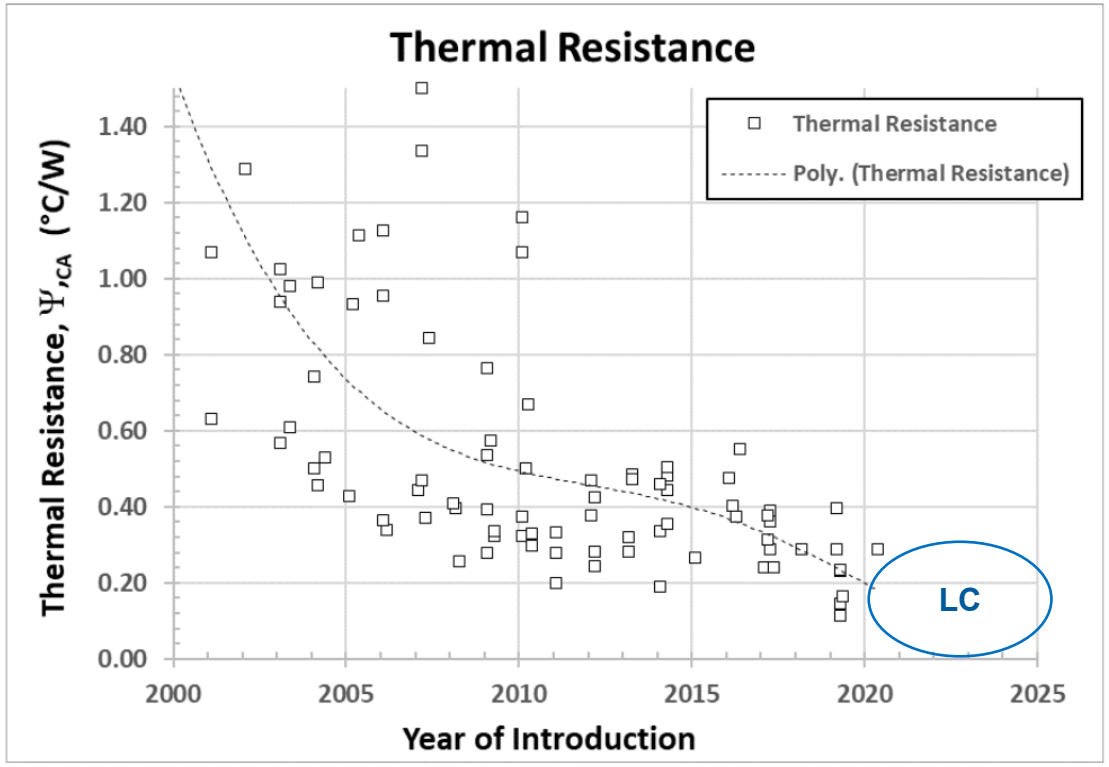
Realistic law: T_{case} = 80 °C at 250 W TDP, decreases at 0.046 °C/W TDP thereafter

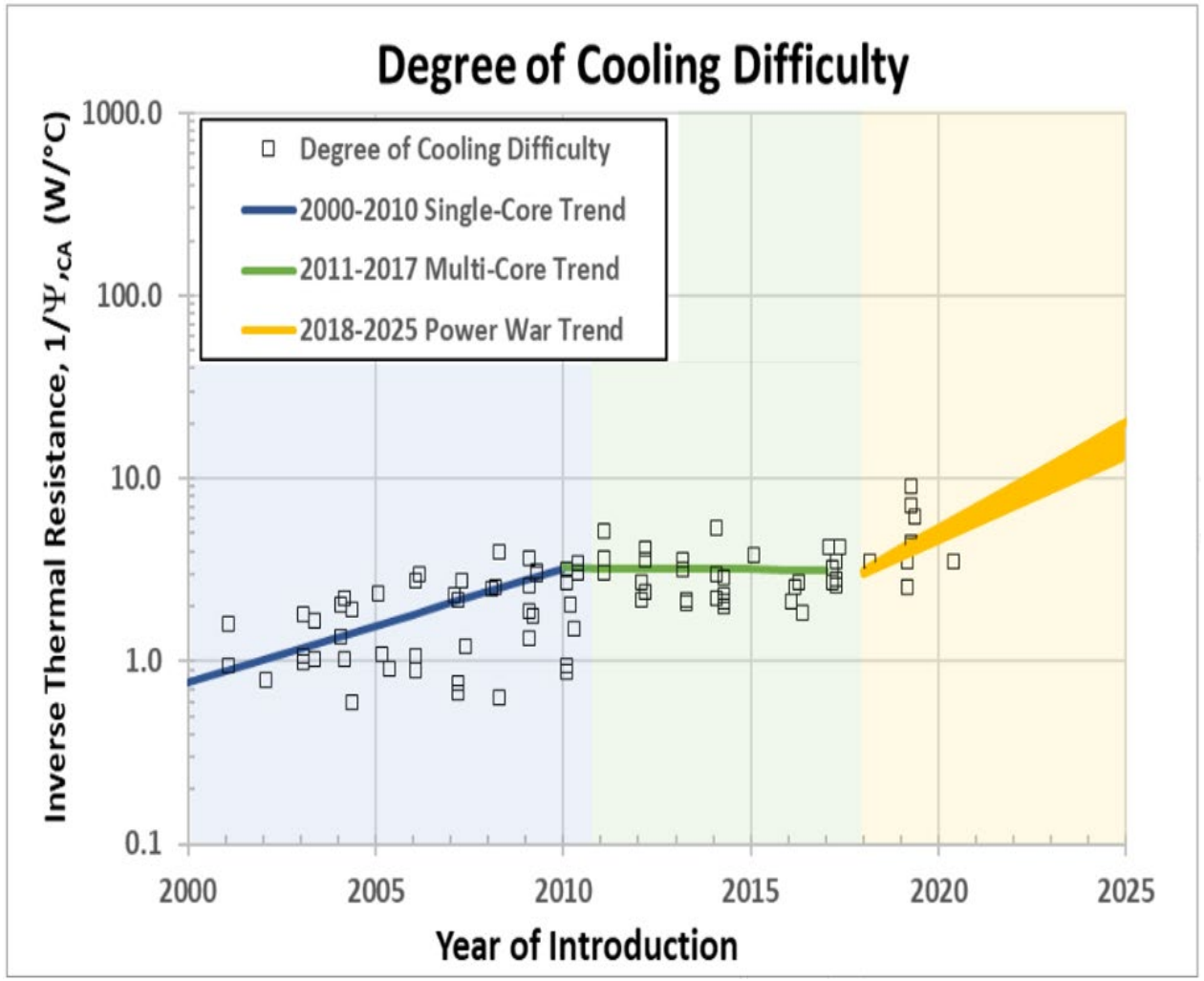


Stacked structures increase thermal resistance and force package temperatures lower

A steep change ahead

- DoC difficulty ramping up exponentially
- Entering the LC Thermal Resistance area

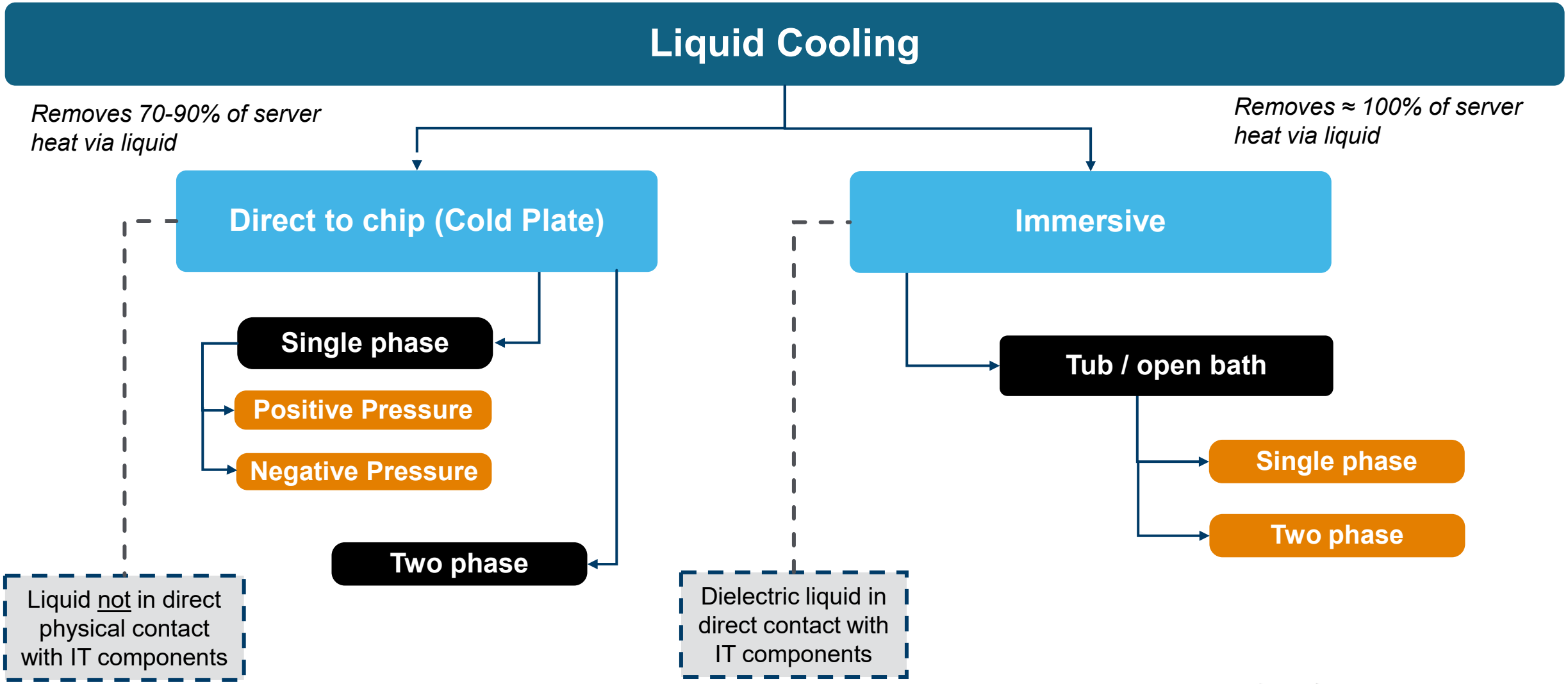




Source : ASHRAE – Emergence and Expansion of Liquid Cooling in Mainstream Data Centers

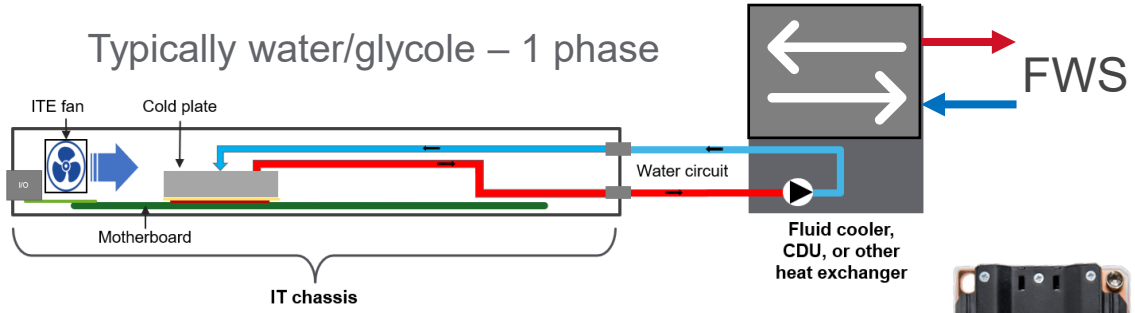
Main technologies for non-conventional (liquid) cooling

From The Green Grid White Paper 70* and WP 265: Liquid Cooling Technologies for Data Centers and Edge Applications



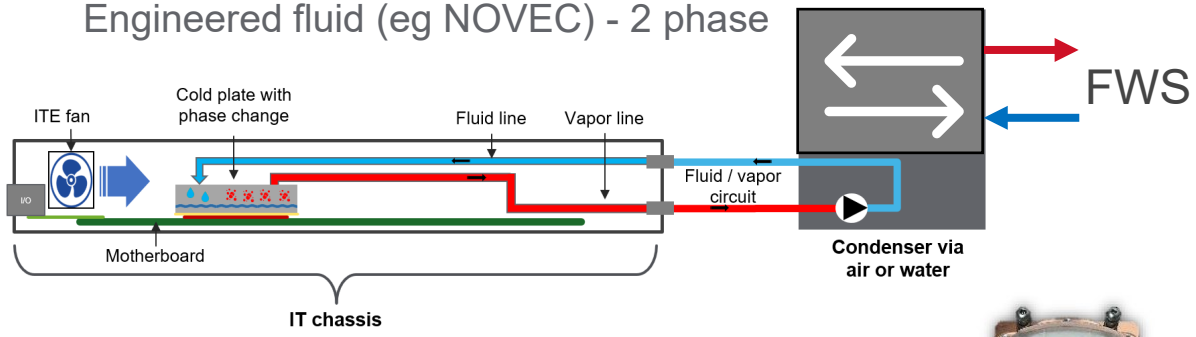
Direct to chip(DCLC)

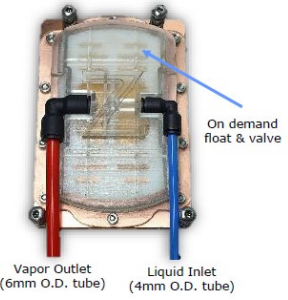
Typically water/glycole – 1 phase



- Easily retrofitted into air cooled server chassis
- Removes ~ 80% of IT heat
- Brings water into the chassis. Risks can be mitigated via Leak Prevention System (LPS) or use of a dielectric
- Positive pressure & negative pressure

Engineered fluid (eg NOVEC) - 2 phase



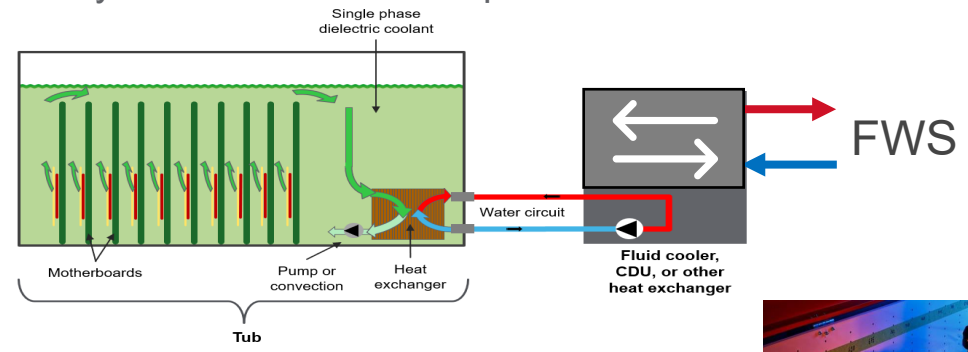


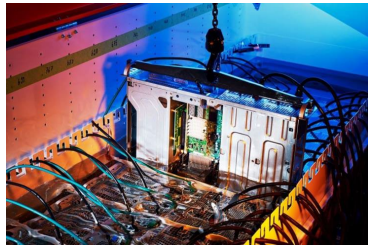
- Easily retrofitted into air cooled server chassis
- Removes ~ 80% of IT heat
- Leverages phase change for heat removal, less engineered fluid required
- No water in server
- Engineered fluid transfers heat directly outside to condenser or indirectly via water loop

Source & Credits: Schneider Electric

Immersed

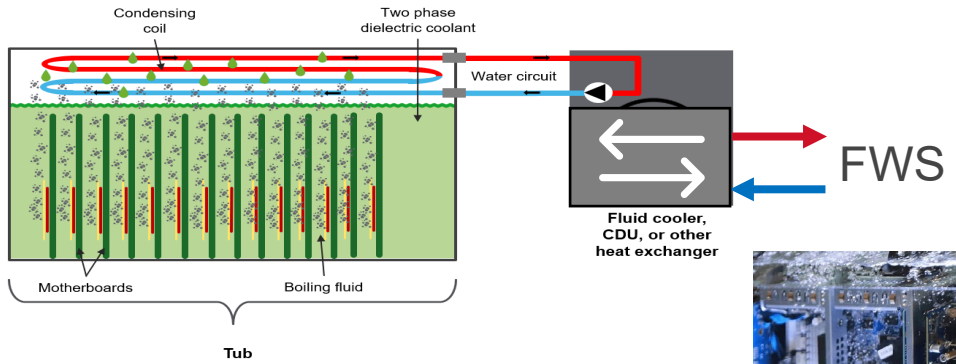
oil / hydrocarbon based – 1 phase

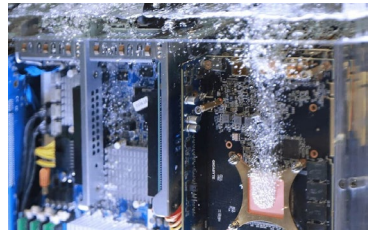




- Less scalable – grow in larger chunks
- Removes > 95% of IT heat
- Requires more fluid than chassis based – cost effective oil-based dielectric is typical
- No fans. Can use pumps or convection to move dielectric fluid

Engineered fluid – 2 phases





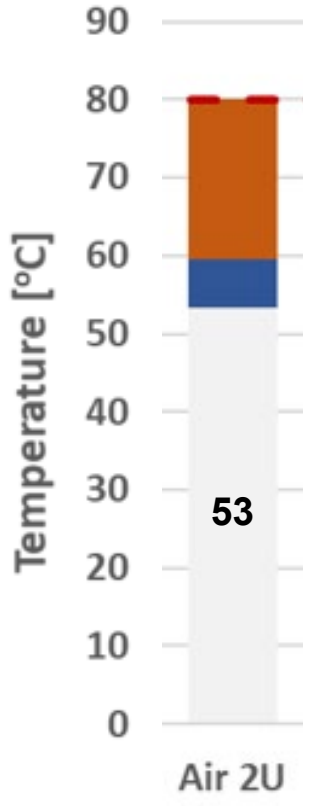
- Less scalable – grow in larger chunks
- Removes > 95% of IT heat
- Requires more fluid than chassis based
- No fans and typically no pumps. Boiling phase change moves the fluid

Cooling water temperatures for technologies

- End-to-end temperature difference used to indicate the maximum allowable facility water temperature
- Stack-up shows how temperature differs throughout the systems
- Modeling results captured for rack server with two processors
  - TDPs ranging from 250 – 500 W

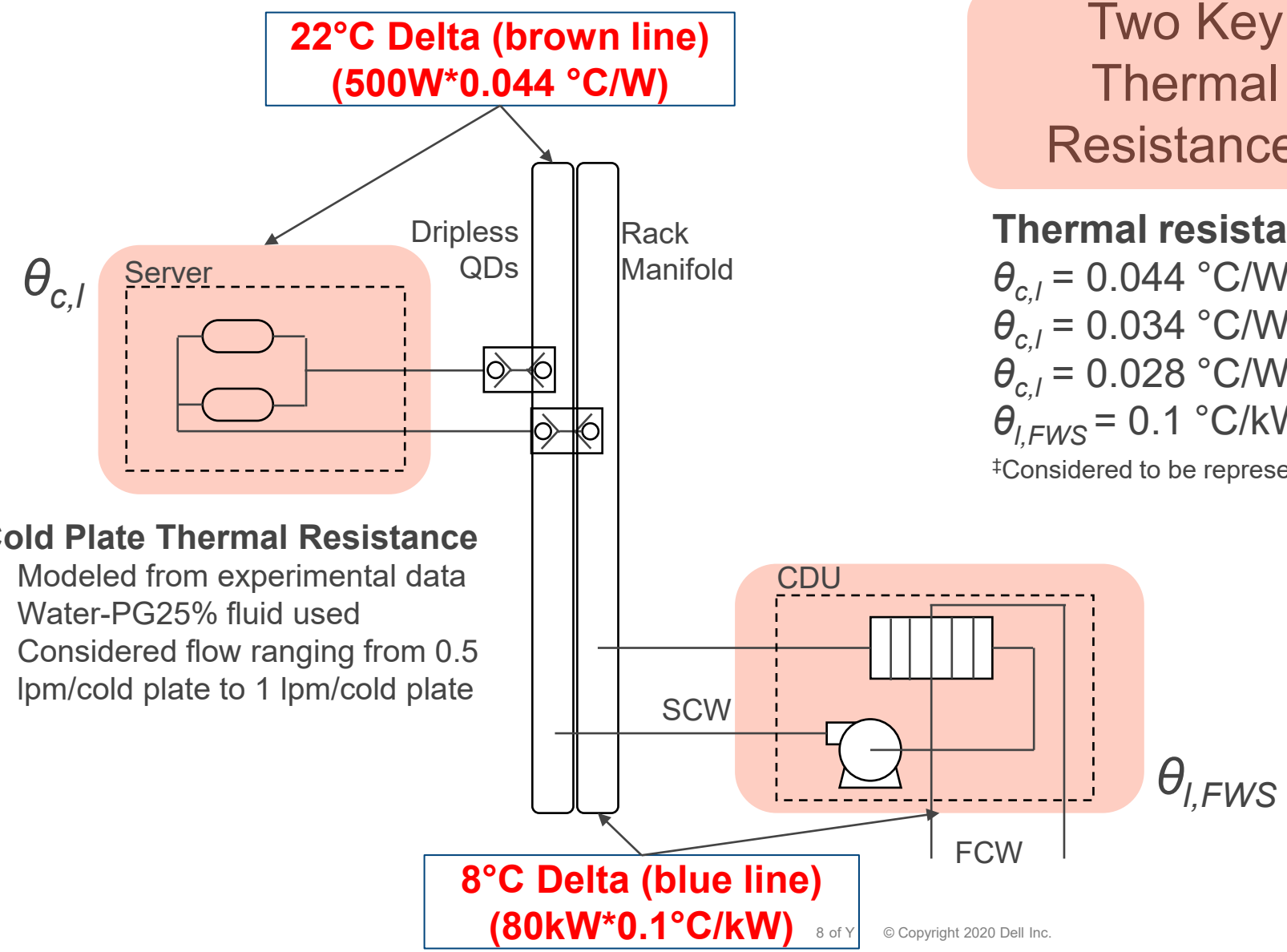
Example: 2U Air Cooling

32 server rack of dual 250 W processors



- - -	Maximum processor case temperature
	Temperature difference between case and secondary coolant
	Temperature difference between secondary coolant and facility water (FWS)
	Maximum FWS temperature (indicated on chart)

Example: Single-phase DLC @500W CPU,80kW/rack



Cold Plate Thermal Resistance

- Modeled from experimental data
- Water-PG25% fluid used
- Considered flow ranging from 0.5 lpm/cold plate to 1 lpm/cold plate

Two Key Thermal Resistances

Thermal resistances‡

$\theta_{c,l}$ = 0.044 °C/W (0.5 lpm/cold plate, 1 lpm/server)
$\theta_{c,l}$ = 0.034 °C/W (0.75 lpm/cold plate, 1.5 lpm/server)
$\theta_{c,l}$ = 0.028 °C/W (1 lpm/cold plate, 2 lpm/server)
$\theta_{l,FWS}$ = 0.1 °C/kW rack

‡Considered to be representative. Not indicative of a specific system.

CDU Thermal Resistance

- Estimated heat exchanger resistance based on available manufacturers' data

Two-phase DLC

Three Key Thermal Resistances

Thermal resistances‡

$\theta_{c,l} = 0.05TDP^{-0.116}$ °C/W

θ_l = variable

$\theta_{l,FWS}$ = 0.24 °C/kW rack

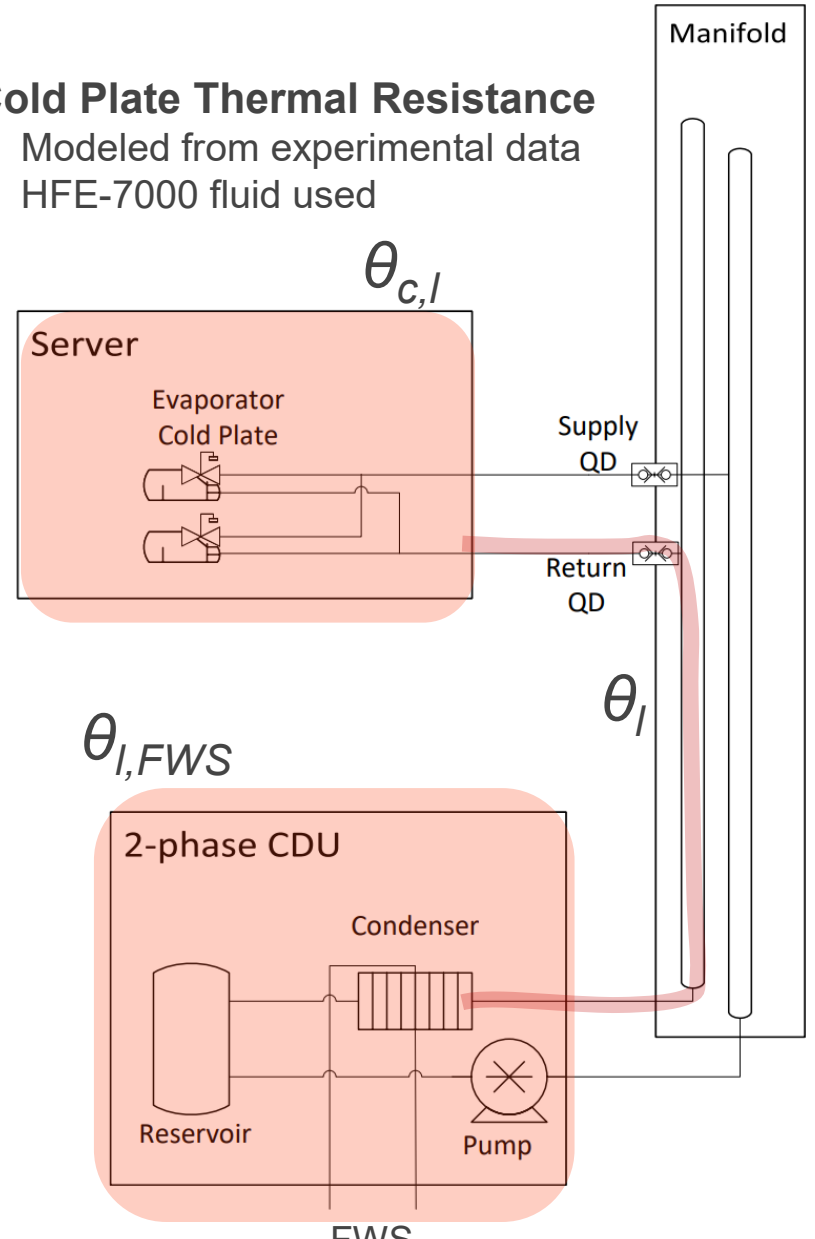
‡Considered to be representative. Not indicative of a specific configuration or design.

CDU Thermal Resistance

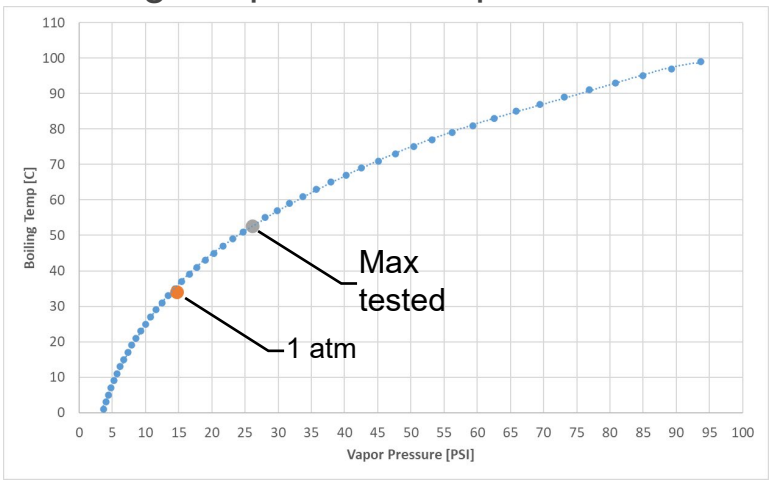
- Estimated resistance using effectiveness-NTU heat exchanger analysis[15]

Cold Plate Thermal Resistance

- Modeled from experimental data
- HFE-7000 fluid used



Boiling temperature vs pressure

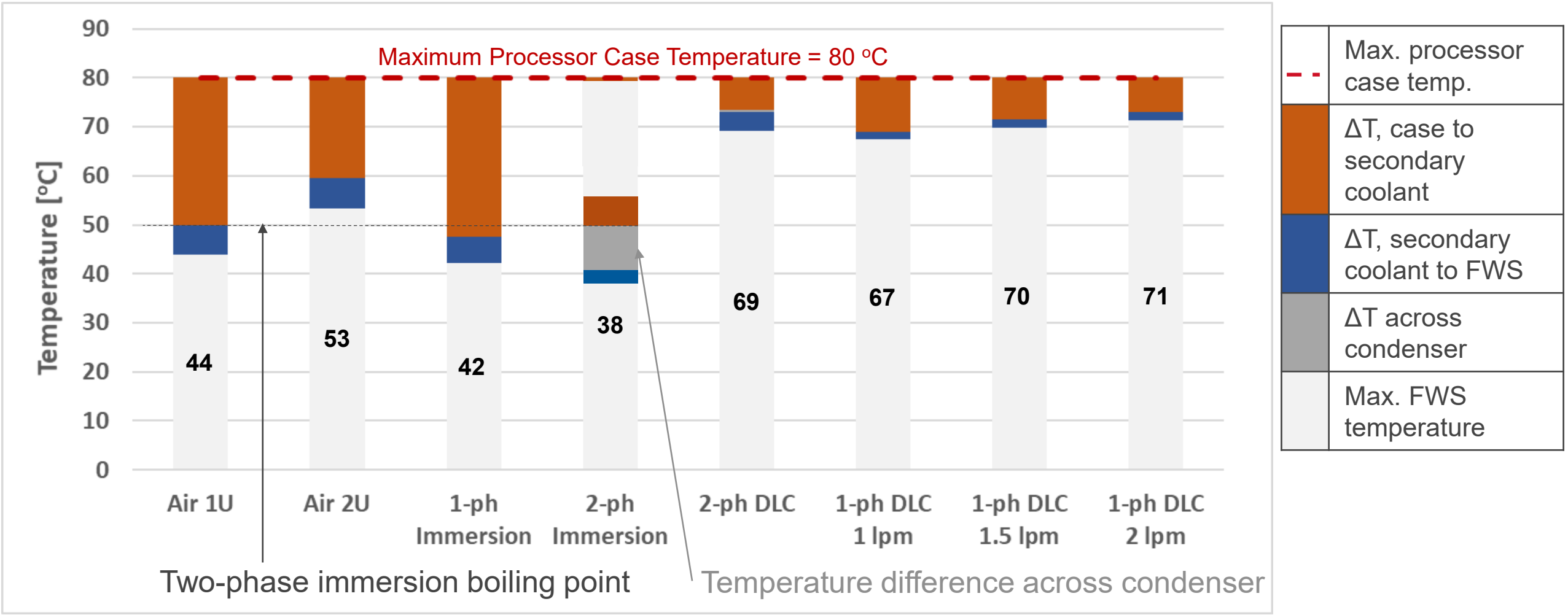


Flow Thermal Resistance

- The boiling point is directly related to pressure as shown above
- The pressure difference between the evaporator and the condenser increases with heat load as additional vapor is generated
- This acts as an additional thermal resistance that increases with increasing heat load

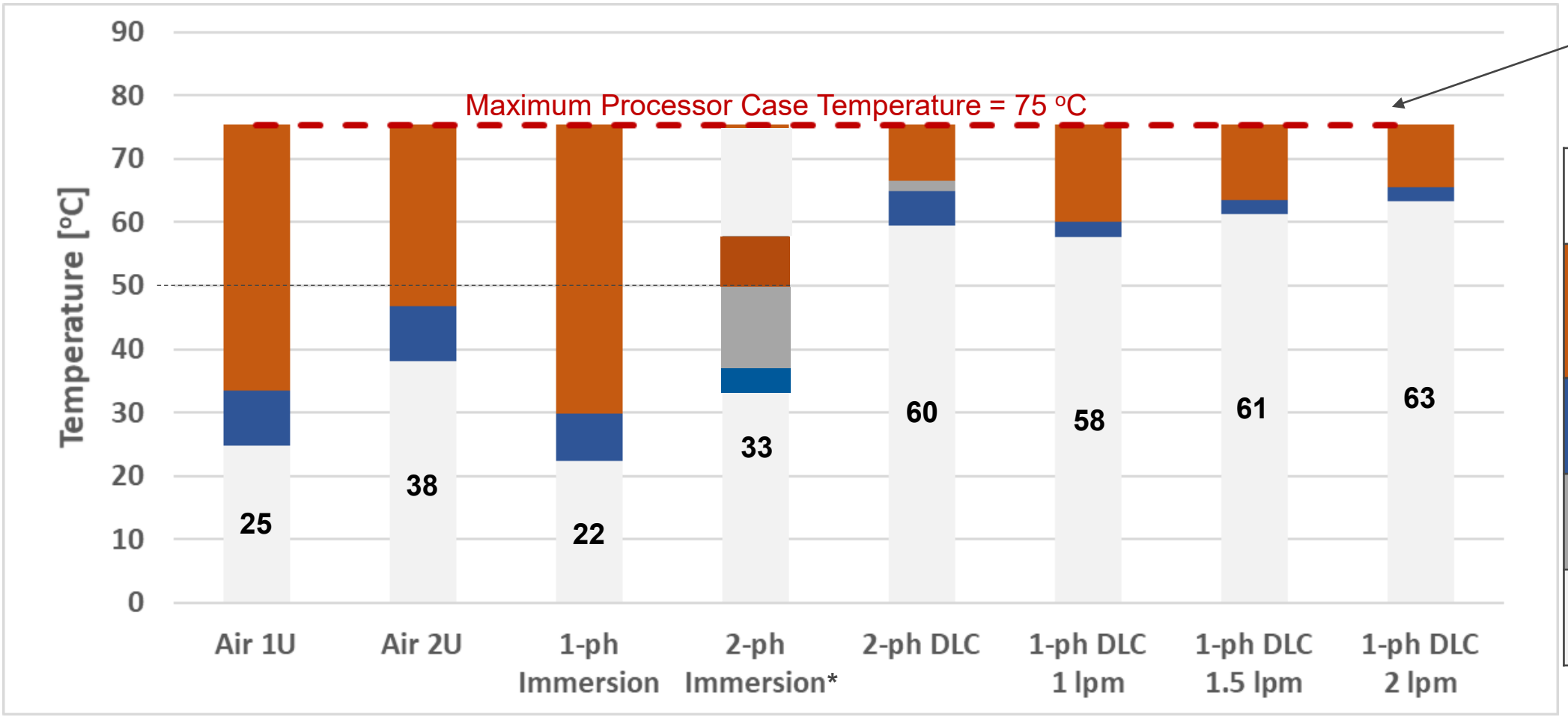
Cooling water temperatures - 250 W processors

32 server rack of dual 250 W processors – 16 kW total rack load



Cooling water temperatures - 350 W processors

32 server rack of dual 350 W processors – 22.4 kW total rack load



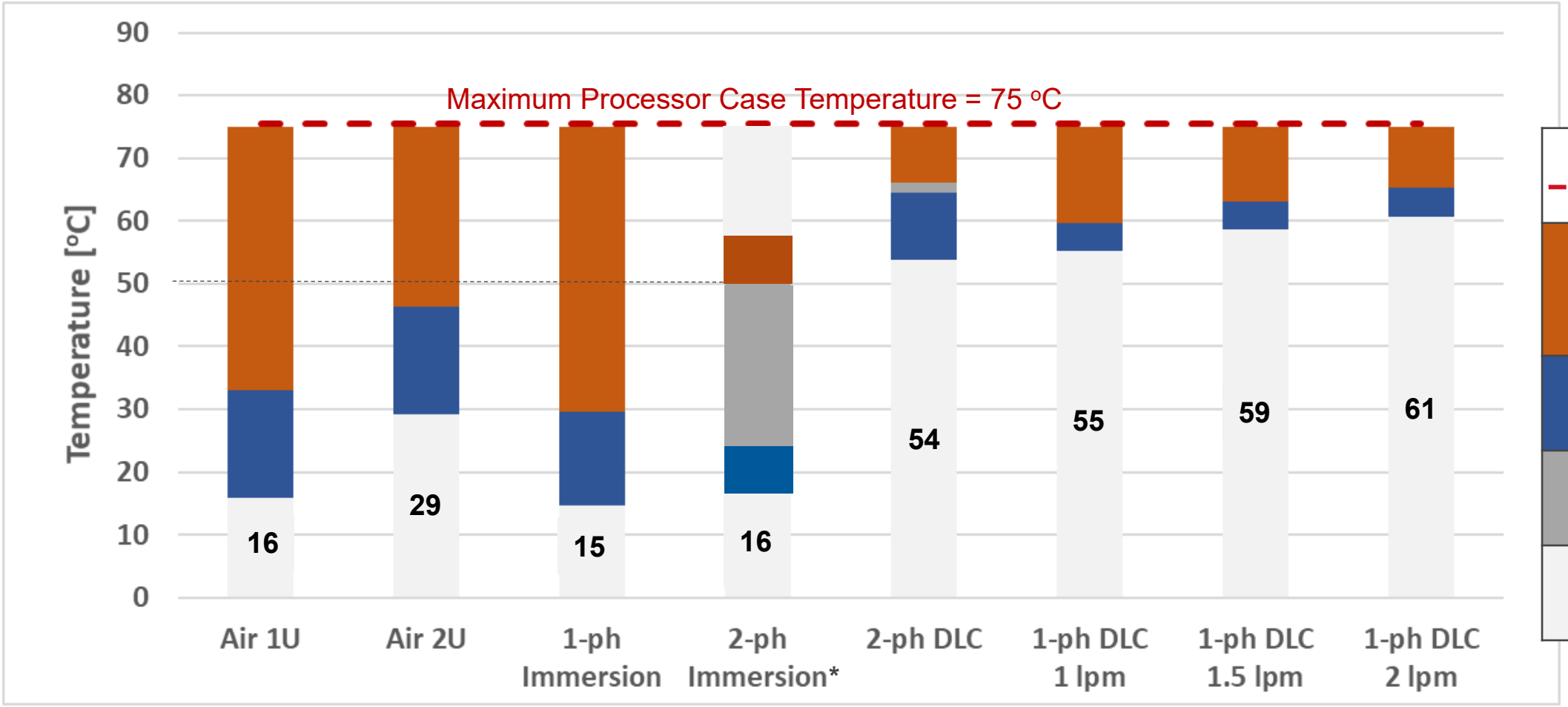
Note that processor case temperature falls with increasing processor TDP

- Max. processor case temp.
- ΔT, case to secondary coolant
- ΔT, secondary coolant to FWS
- ΔT across condenser
- Max. FWS temperature

*Boiling point 50 °C

Impact of doubling rack load - 350 W processors

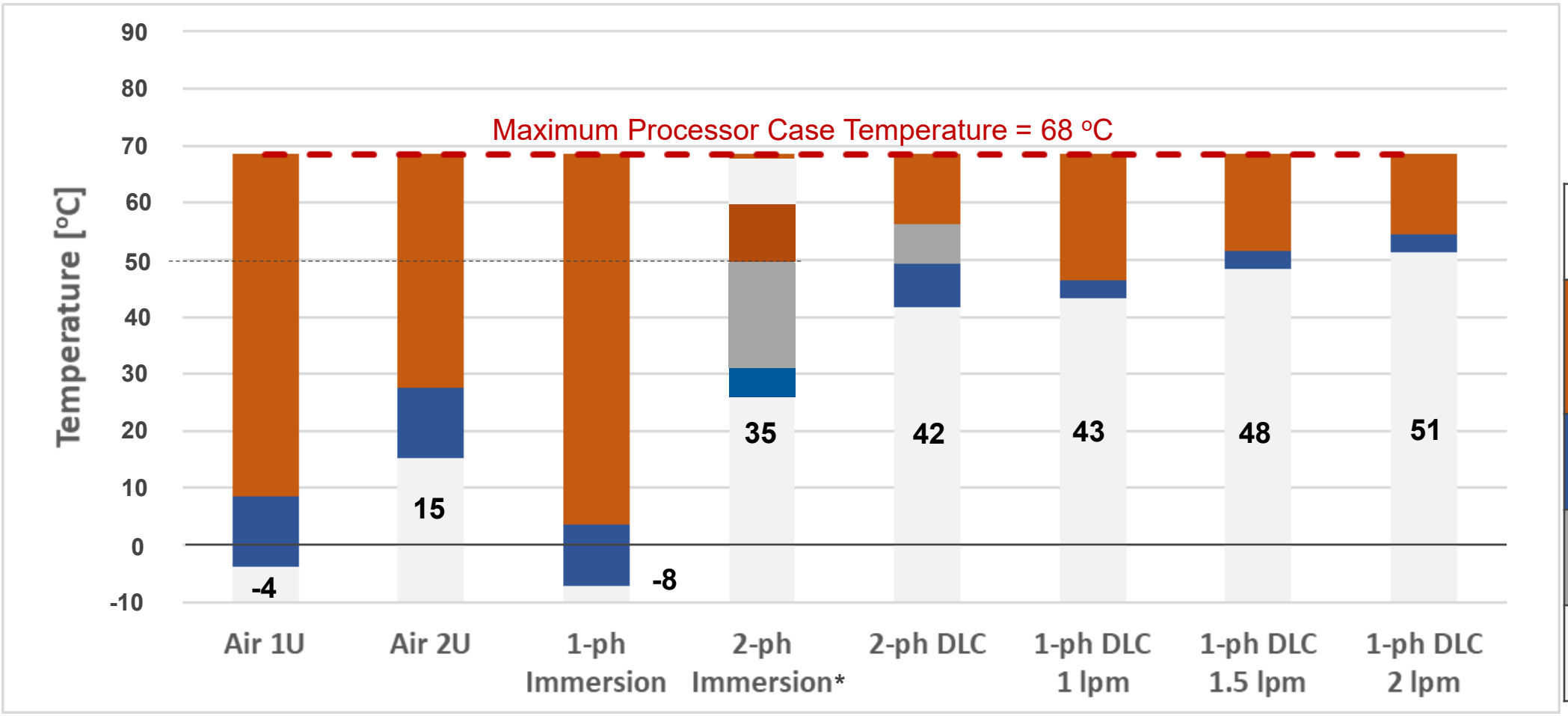
64 server rack of dual 350 W processors – 44.8 kW total rack load



*Boiling point 50 °C

Cooling water temperatures - 500 W processors

32 server rack of dual 500 W processors – 32 kW total rack load



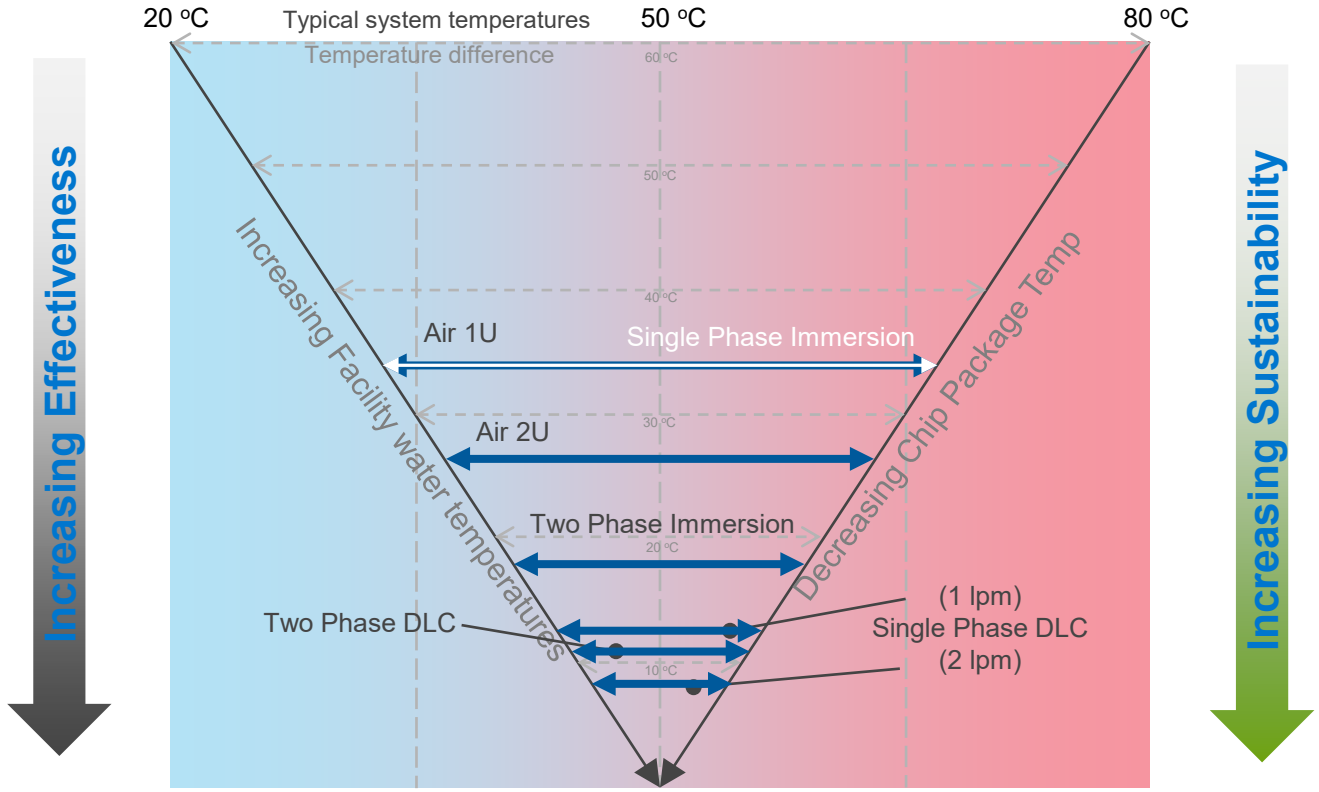
*Boiling point 50 °C

End-to-end technology comparison

- End-to-end temperature difference is used as the comparison metric for each technology
- Temperature difference is visualized in a squeeze plot
- Modeling results captured for server rack with two processors
  - TDPs ranging from 250 – 500 W

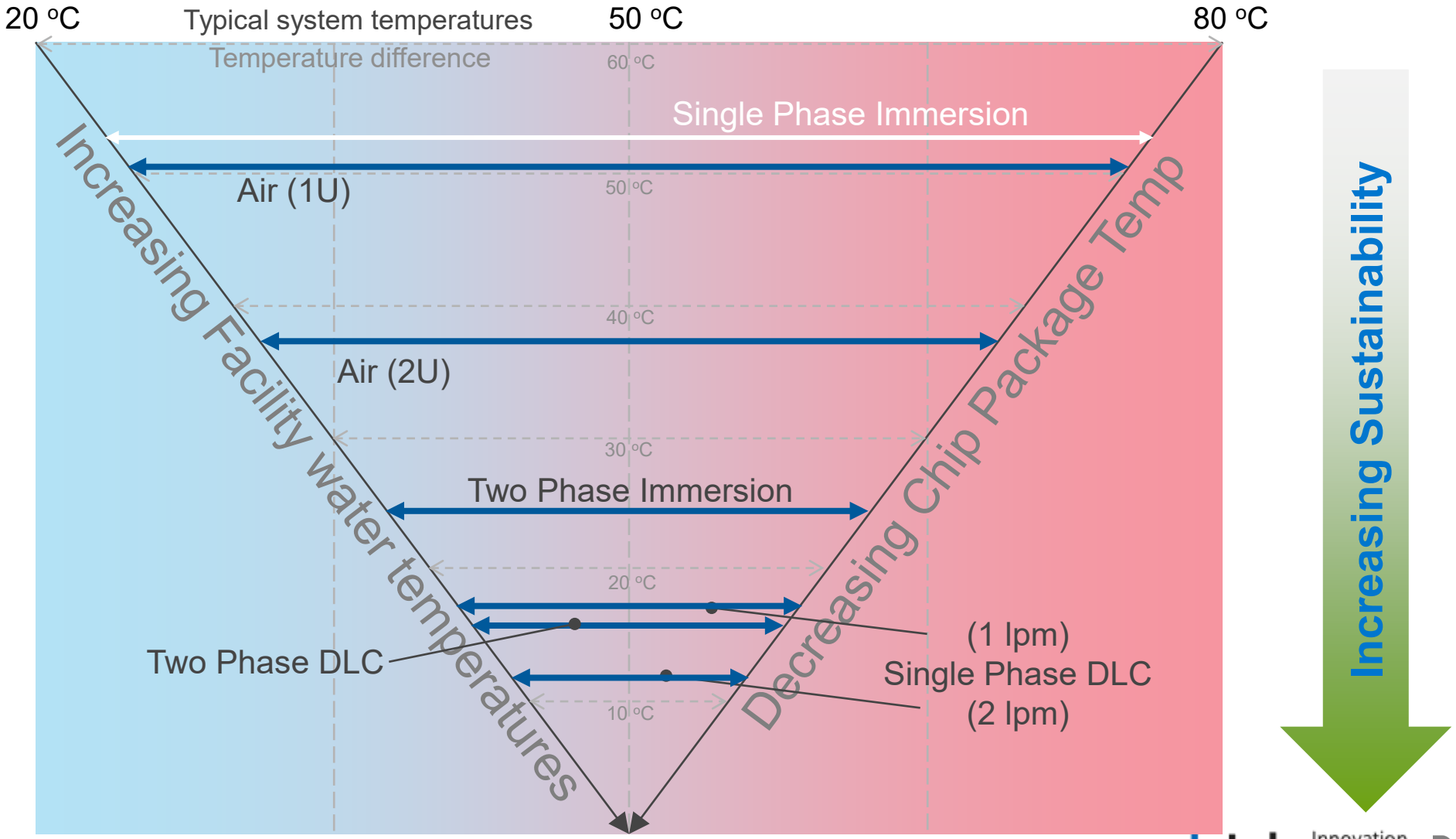
Example: 2U Air Cooling

32 server rack of dual 250 W processors



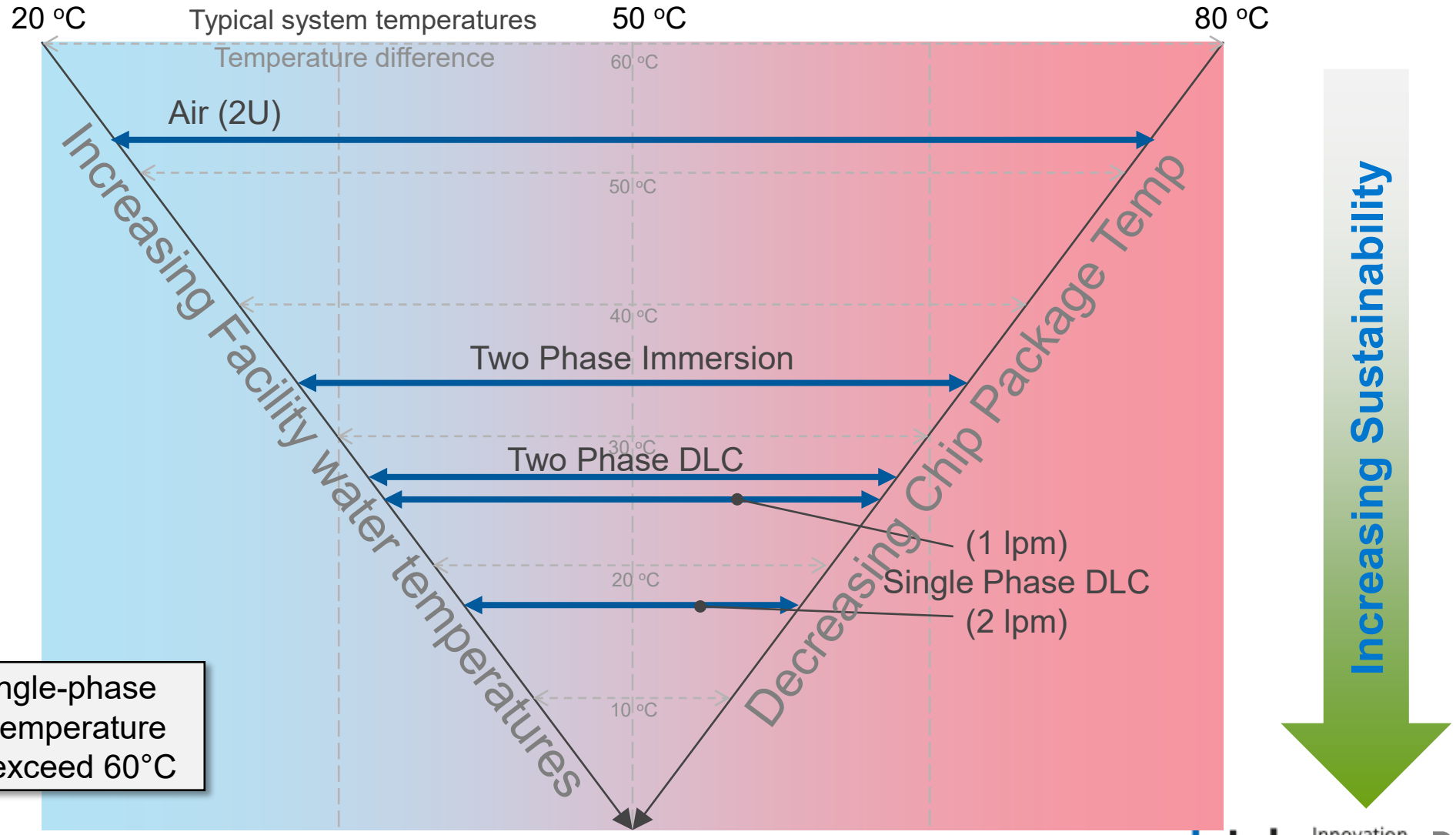
Temperature Squeeze – 350 W processors

32 server rack of dual 350 W processors



Temperature Squeeze – 500 W processors

32 server rack of dual 500 W processors



Air 1U & single-phase immersion temperature differences exceed 60°C

Considerations - 1

- **Air cooling** is limited in 1U chassis (but can still achieve cooling without compressors up to 270 W processors); performance and energy consumption improve significantly in 2U
  - Approximate air-cooled limits are 350 W TDP processors in 1U, 500 W TDP processors in 2U (not hard limits, but good guidelines)
- **Single Phase immersion** is already on its last generation of high-end processors
  - Will require compressor chilled water after about 270 W in most climates
- **Two-phase immersion** is limited to high case temperature packages (70 ^{o}C and up) up to about 600 W
  - Boiling temperature is about 50 ^{o}C.
  - Processor package temperatures below 70 ^{o}C present a challenge to immersion fluids
  - Known working fluids are currently in line for regulation or limitation

Considerations - 2

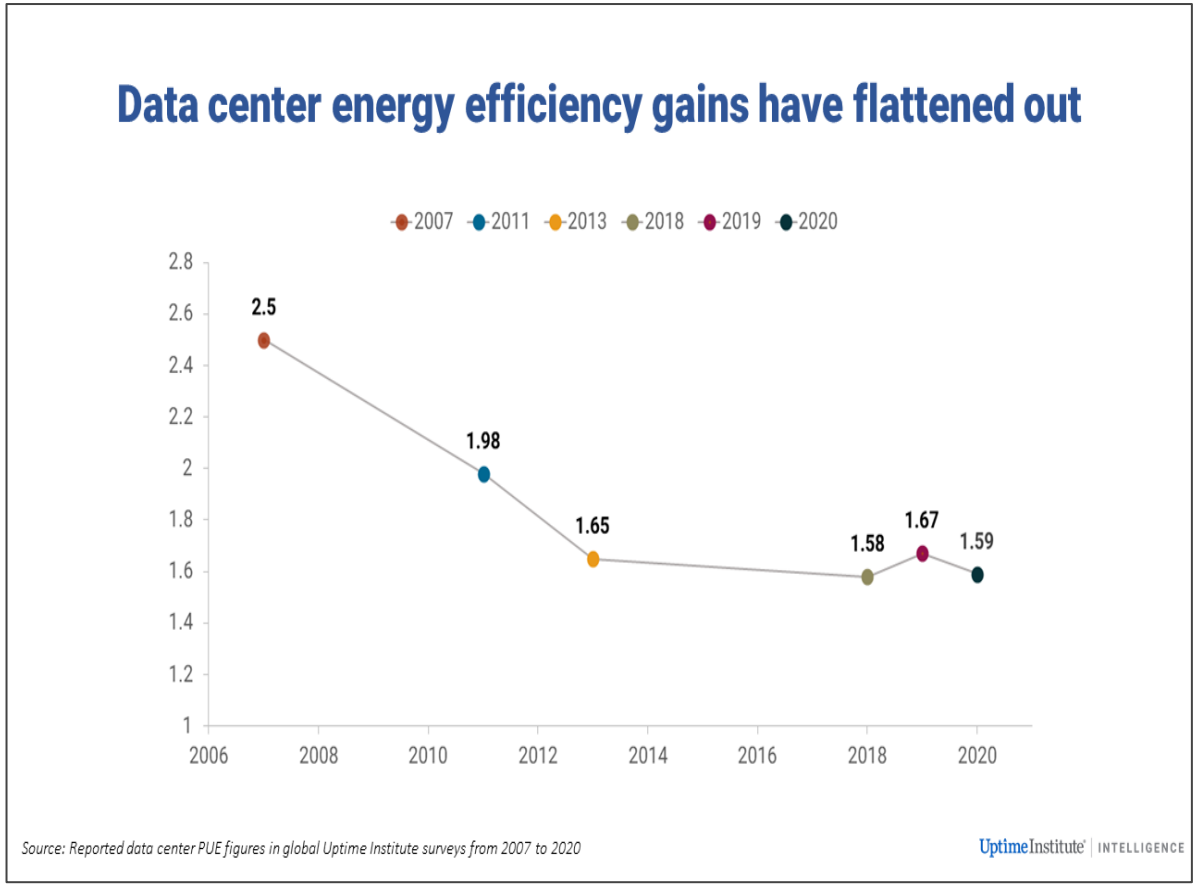
- **Two-phase DLC** has about the same performance as water DLC at 1 lpm
  - Thermal resistance is package- and rack-load dependent
  - Dielectric fluids are in the regulation/elimination crosshairs
  - Case temps are going to be function driven, not heat flux driven
    - Two-phase heat transfer is heat flux driven
  - Service may be expensive with constant liquid top-ups
  - Still some uncertainty about long-term material compatibility
- **Single-phase DLC** can take us out beyond 1500 W packages
  - Dell will offer water DLC factory installed for select platforms
  - Most energy efficient
  - Couples well with existing data center architectures
  - Most serviceable
  - Leaks can be designed out (see automobiles, medical equipment, power electronics)
  - Very flexible and efficient heat rejection options from liquid-liquid CDUs to liquid-air CDUs

HPC goal: High heat capture through water

- Generally, it's assumed that more server DLC touch points equates to highest efficiency operation
  - "Full-liquid cooled" systems are most efficient
  - 90% DLC is better than 80% DLC, which is better than 70% DLC, regardless of cost
- However, you have to look at the details:
  - Claims for > 90% overall heat capture proven to be really closer to 80%
  - Recent testing proves more than 10% travels down into the board and is cooled by air
- Why not 100% Direct Liquid Cooling, with a lower overall PUE?

100% Direct Liquid Cooling: does it worth?

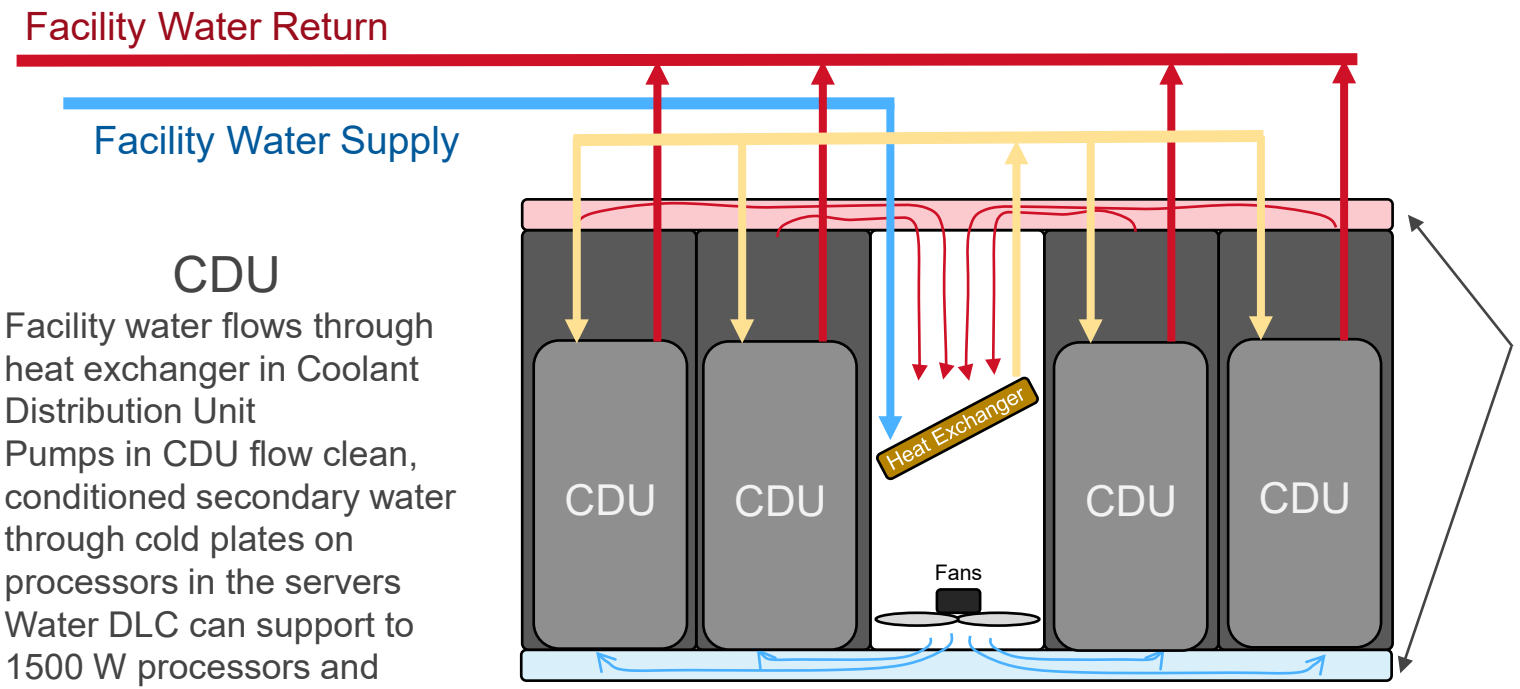
- Today's DataCenters (mean) PUE
  - Best in the world:1,05 ÷ 1,07
  - Mean:1,8
  - Good Air-Cooling:1,15 ÷ 1,6
  - Good Liquid Cooling:1,08 ÷ 1,35
  - Hybrid cooling systems (AIR+DLC): a (weighted) combination of the above values
- Energy costs comparisons between Hybrid and 100%Water Cooled systems
  - PUE_water:1,08, PUE:Air:1,18
  - *2000x2S nodes (2x270W),16 DIMMs,2xSSD,100GbE*
    - *Total Nodes Power: 1,5MWh, Total DC Power: 1,67MWh*
    - *Delta Power Consumption: +85kWh,*
    - *@0,3€/kWh around* ***+ 6% of the Total Computing HW investment***
  - *2100x2S nodes (2x270W+4x700W GPU),16 DIMMs, 2xSSD,100GbE*
    - *Total Nodes Power: 9,4MWh, Total DC Power: 10,4MWh*
    - *Delta Power Consumption: +650kWh,*
    - *@0,3€/kWh around* ***+5% of the Total Computing HW investment***



Side Considerations

- Cost and complexity should be a consideration in DLC design choices
- The addition of memory cooling in today's offering only adds about 10% heat capture but would at least double the cooling solution cost in-server.
  - The cost of capturing all non-DLC heat through a closed coupling + DLC system would be about the same as a memory cooling addition but would boost to 100% capture
- Time to service a CPU in a high-density Compute node (e.g. 4N2U) is under 5 minutes
- Time to service CPU plus ie memory design would add 10s of minutes and would require fixturing for removal/install

“Close Coupled” Air Cooling + DLC



CDU

- Facility water flows through heat exchanger in Coolant Distribution Unit
- Pumps in CDU flow clean, conditioned secondary water through cold plates on processors in the servers
- Water DLC can support to 1500 W processors and beyond
- Available through Dell since 14G

Rack Enclosures

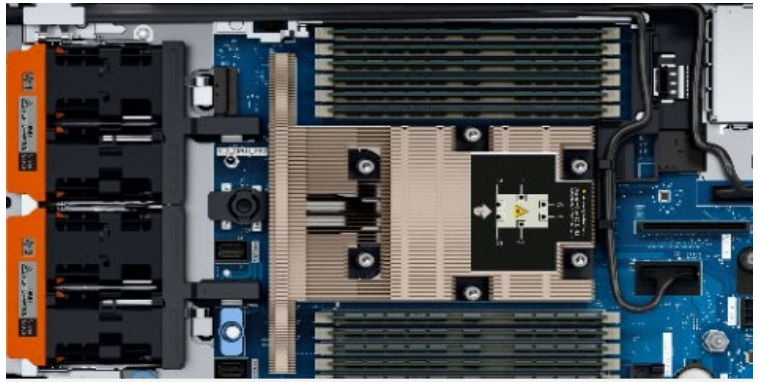
- Keep air circulating through the IT gear in the racks, not in the data hall
- 100% heat collection to facility water
- Very warm air may be used

In-Row Cooler

- Facility water flows through a heat exchanger
- Fans draw air through the heat exchanger
- One in-row cooler may cool multiple racks of IT equipment

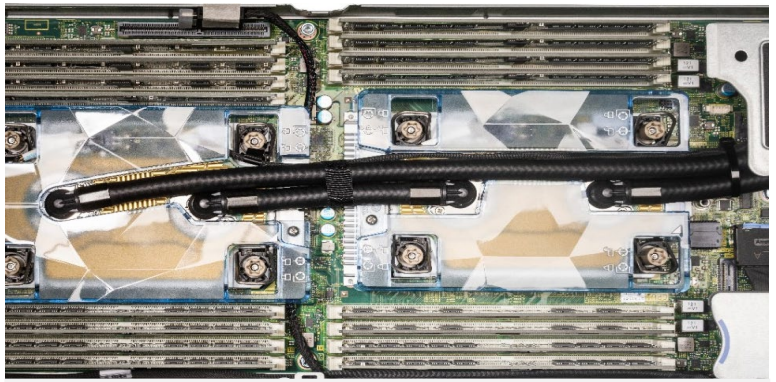
- **Same efficiency but much less complex than 100% LC technologies**
- **About the same CapEx as oil immersion**

Key Takeaways



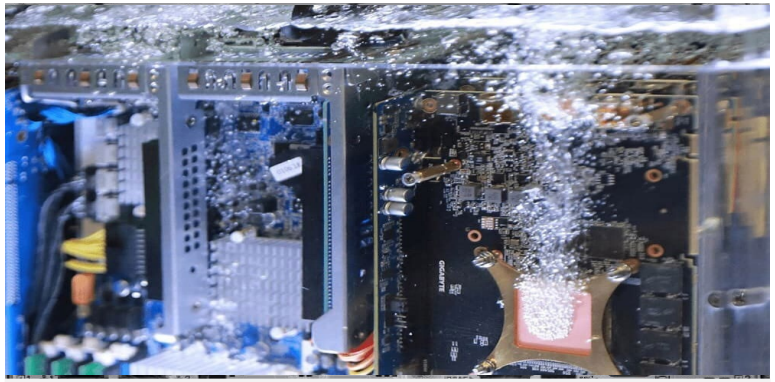
Air Cooling

- Models indicate increasing challenges for air cooling at high TDPs
- Advances in heat sinks and air movers may push limits further
- Air cooling is often limited by the impact of processor heating on the other components in the chassis



Direct Liquid Cooling (DLC)

- Single-phase DLC capable of cooling well beyond 500 W TDPs
- Two-phase DLC capable of cooling high TDPs, though the flow resistance of the vapor return path must be addressed
- Advances in designs or fluids may improve two-phase DLC



Immersion Cooling

- Models indicate increasing challenges for immersion cooling at high TDPs
- Advances in heat sinks and circulation could push single-phase limits further
- Two-phase immersion limited by fluid boiling points and condenser performance

- At today, use Liquid Cooling only **IF** necessary and **WHERE** necessary
- **Plan however for bringing water to the rack**
- Consider ROI of 100% Liquid Cooling before embarking in complex cooled systems acquisition

Innovations in processor design and cooling systems will continue to improve today's technologies

intel. Innovation Built-In DELL Technologies

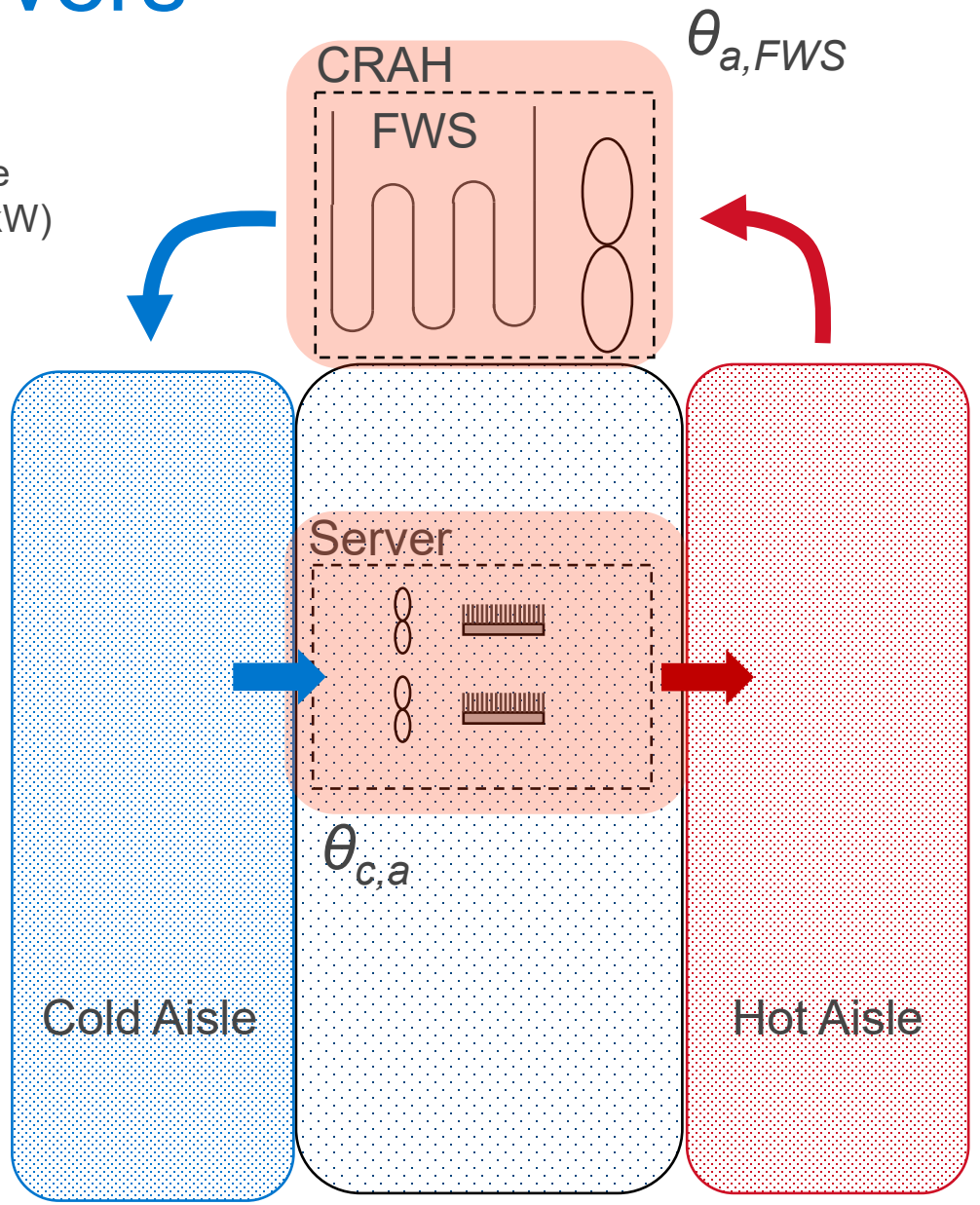

DELL Technologies

Backup

Air Cooled Servers

Computer Room Air Handler
- Modeled from publicly available manufacturers' data (100 cfm/kW)

Air Cooled Heat Sink
- Modeled from experimental data for typical 1U heat sinks and air flow (100 cfm/kW)
- Actual values may vary by ±10% or more depending on geometry



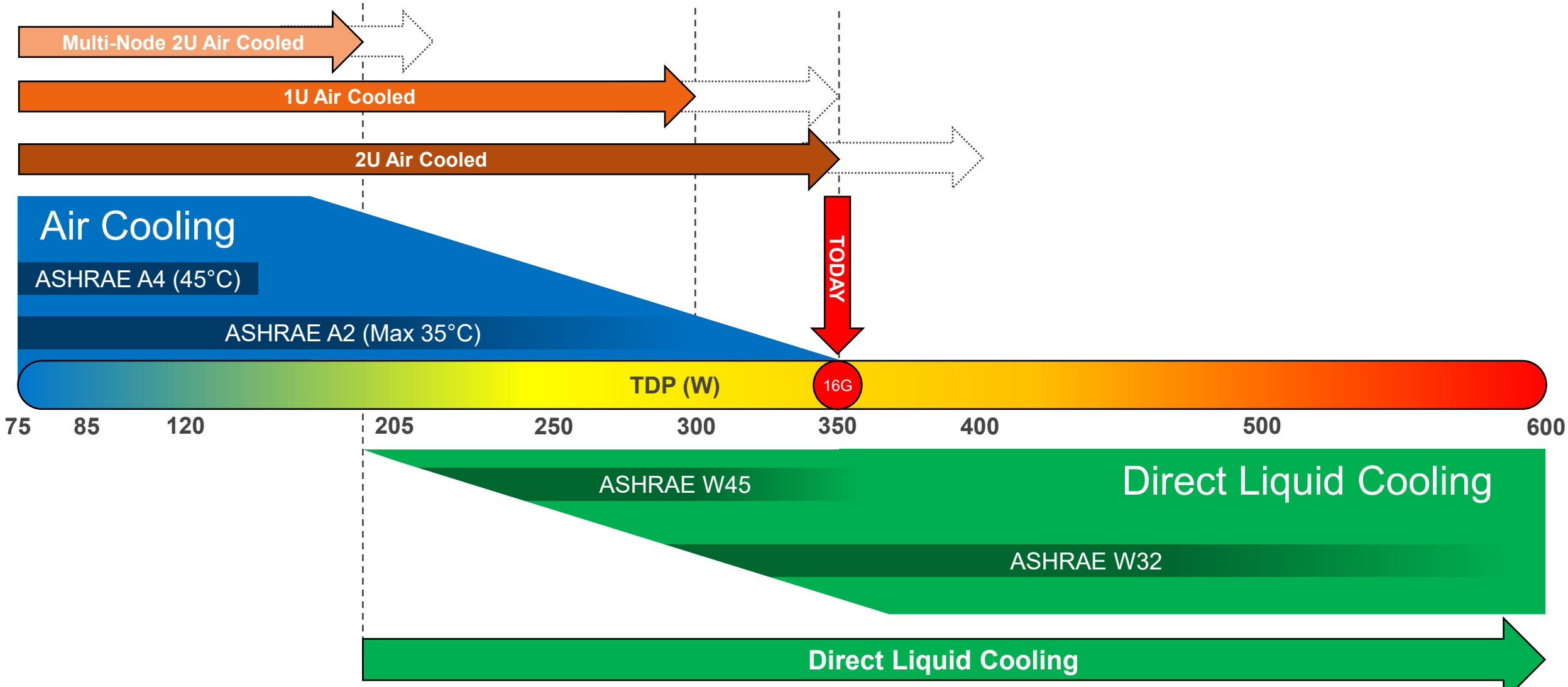
Two Key Thermal Resistances

Thermal resistances‡

$\theta_{c,a}$ = 0.12°C/W, 1U server*
$\theta_{c,a}$ = 0.082°C/W, 2U server*
$\theta_{a,FWS}$ = 0.23°C/kW rack

‡Considered to be representative. Do not reflect one particular server, heat sink, or CRAH.
*1U=1.75", 2U=3.5"

Air vs. Liquid Cooling Thresholds by Form Factor



Source : ASHRAE – Emergence and Expansion of Liquid Cooling in Mainstream Data Centers

intel. Innovation Built-In DELL Technologies

Additional notes and assumptions for thermal models

- Assumed heat pattern and size of heat source approximately equal to Intel Sapphire Rapids
  - Specific processors will not provide the same thermal resistance as the uniform heat flux used here
- Typical TIM resistance is included in $\theta_{c,l}$ values
- Assumed heat generation is uniform such that one case temperature is sufficient to specify processor thermal limits
- Assumed 100% of processor heat is collected by the coolant
- Arbitrary relationship between processor TDP and maximum case temperature assumed: T_{case} = 80 °C at 250 W TDP, decreases at 0.046 °C/W TDP thereafter
  - These values are representative and realistic, but do not represent exactly any known specifications of processors.

Investigated cases and comparison metric

- Three assumed TDPs and case temperatures were used with the thermal resistances to determine temperature differences across the systems
  - TDP of 250 W has a $T_{case,max}$ = 80 °C
  - TDP of 350 W has a $T_{case,max}$ = 75 °C
  - TDP of 500 W has a $T_{case,max}$ = 68 °C
- Comparisons performed with end-to-end temperature differences