WARM LIQUID COOLED CLOUD FACILITIES: WHAT DO WE ACHIEVE?

Author (s):

Bharath Ramakrishnan, Microsoft

Husam Alissa, Microsoft

Dennis Trieu, Microsoft

Robert Lankston, Microsoft

Mark Shaw, Microsoft

Zaid Kahn, Microsoft

Christian Belady, Microsoft

Executive Summary

Liquid cooled cloud facilities are widely adopted across the globe for a variety of performance, business, environmental and sustainability reasons. Liquid cooling is the prime driver of densification of server HW and microprocessor power density by solving challenges of heat flux at a chip level and addressing HW reliability and efficiency. Liquid cooling not only impacts the HW of current generation, but also help the industry keep up with the technology roadmap of future generation like 'Heterogenous Integration' and hence would leverage many aspects of advanced semiconductor packaging innovation, fabrication, and additive manufacturing techniques. The impact of liquid cooling as captured in this article is plenty fold as it has a direct impact to the company wide goals and effect the broader community like "Open Compute Project" in innumerous ways like driving sustainable cooling/power solutions and enabling potential waste heat recovery options at a data enter facility level.

Table of Contents

Introduction		4
1	Data Center Energy Usage Trends	4
2	Microprocessor Power Trends	5
3	Heterogenous Integration	6
4	Overclocking the CPUs	7
5	Advanced Packaging and Fabrication	10
6	AI/ML chips of Next Generation	11
7	Implications of Lower Facility Water Temperature	12
8	Conclusion	13
9	References	13
10	License	14
11	About Open Compute Foundation	14

Introduction

Data centers around the world are always expanding to meet the growing demand of internet and information usage for a variety of techno-economic reasons like social, medical, communication, and scientific advancement. With the rise in global data center capacity, the required data center energy consumption [1] goes up too.

1 Data Center Energy Usage and Trends

Looking at the Figure 1 below which reads from 2010 to 2018, there are couple interesting power and energy efficiency trends to be observed carefully. For example, the electricity use in the data centers was reported to be around 194 TWh/year back in 2010 while an estimated electricity use of only 205 TWh/year was reported in 2018. This is roughly equivalent to about 1-1.5% of global electricity usage. But it should be noted that the overall global compute instance (equaling several hundred million) grows up by 550% from 2010 to 2018 while requiring a moderate rise in electricity of only 6%.

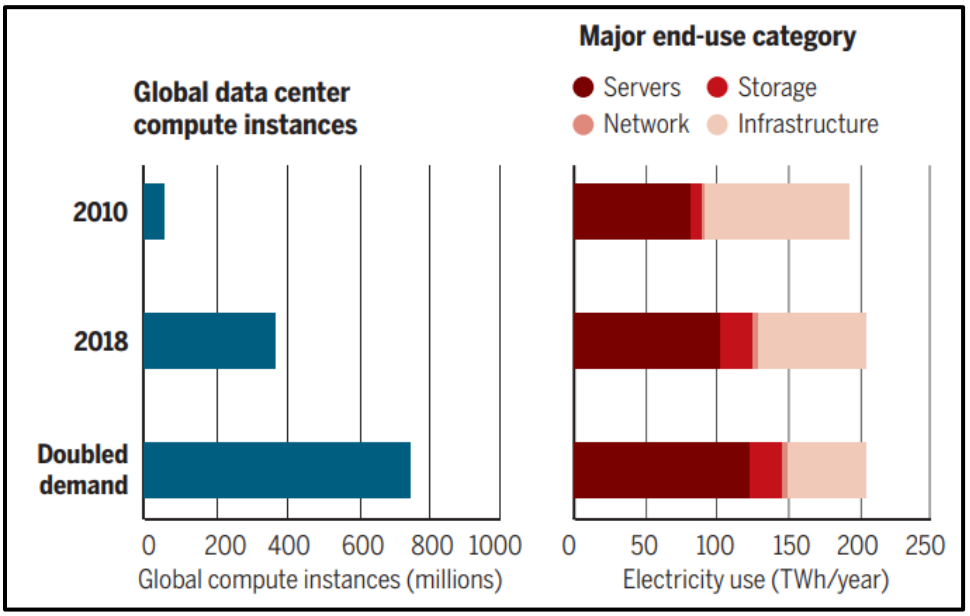

Figure 1. Data center energy use (2010-2018) from Masanet et.al

Densification of data center going from 2010 to 2018 which includes server [A], storage [B] and network [C] devices and the infrastructure [D] improvements including PUE (the total amount of energy used by a data center divided by the energy used by its IT equipment) account for this gain in energy consumption efficiency. Referring to the same chart [1],

[A] A combination of increased server efficiencies (watt-hour per computation) and greater server virtualizations (which reduces the amount of server power required for each computing instance) enabled a '*six-fold' rise in global compute instances* (measured in million) with only a *25% increase in global server energy use* (TWh/year).

[B] During the same time, increased storage drive densities and efficiencies (A 9 % reduction measured in watts per terabyte was reported for the installed storage) enabled a *25-fold increase in the global storage capacity* (measured in exabytes) while requiring only a *3-fold additional global storage energy use*.

[C] Shifting towards faster and more energy efficient network technologies enabled a *10-fold increase in global data center IP* (Internet Protocol) traffic (measured in zettabytes/year) in 2018.

[D] The PUE gains coming from cutting edge cooling solutions and power supply efficiencies enabled large decrease in the energy use of data center infrastructure systems (ie cooling and power provisioning).

In total, the overall energy use of IT devices (servers, storage, and network) has increased from around 92 TWh/year in 2010 to around 130 TWh/year in 2018 thanks to the many technological and operational efficiency gains which enabled a substantial rise in the IT services that has been offered with comparatively smaller growth in energy use. This data [2] considers all data center types including traditional data center, hyperscalers and cloud (non hyperscalers) customers serving IoT, high performance computing (HPC), artificial intelligence (AI), machine learning, 5G, Video conferencing, block chain and social media applications.

Hence it is prudent to conclude that server densification along with other HW and SW advancements enabled such energy efficiency gains. While it was also suggested in [1] that great public investment and funding should be allocated towards advancements of new technologies including increased chip specialization, artificial intelligence for computing resource and infrastructure management, ultrahigh density storage, quantum computing and advanced heat removal technologies including liquid cooling and immersion technologies to extend IT industry's historical efficiency gains well into the future. This article will broadly discuss the two aspects of microprocessor chip specialization paired with the advanced cooling technologies in the coming segments.

2 Microprocessor Power Trends

Microprocessors which are the backbones of Central Processing Units (CPUs), Graphical Processing Units (GPUs), Accelerators (FPGAs), and ASICs constitute a significant portion of the high-density data center server, network, and storage HW which orchestrate different data center services, applications, and workloads accordingly. Looking at the chart on Figure 2 showing 42 years of microprocessor trend data, it is apparent that the number of transistors (measured in million) increases year after year in the last five decades following the ever-famous Moore's law where the number of transistors double every two years through a combination of frequency

increases and transistor increases [3]. But the typical power to the microprocessor (measured in Watts) plateaus at couple hundred Watts especially after mid-2000s. Similarly, the clock frequency stabilized at around 3000 MHz suggesting a single thread performance stall. Even though the chip core counts increase, because of the inability to remove the dissipated heat thereby has led to "dark" or unused Silicon where the increase in core/transistor count are only moderately utilized. In simpler words, 'As microprocessor chip transistor count increases, the inability to remove heat is limiting the number of transistors that can be concurrently active'. Consequently, even though the circuit elements are scaling following Moore's law, the performance scaling is not following the trend. New advanced packaging and cooling technologies including 3D packaging and Heterogenous integration aka 'chiplets' [4,5,6] are needed to combat the decline of Moore's law trajectory.

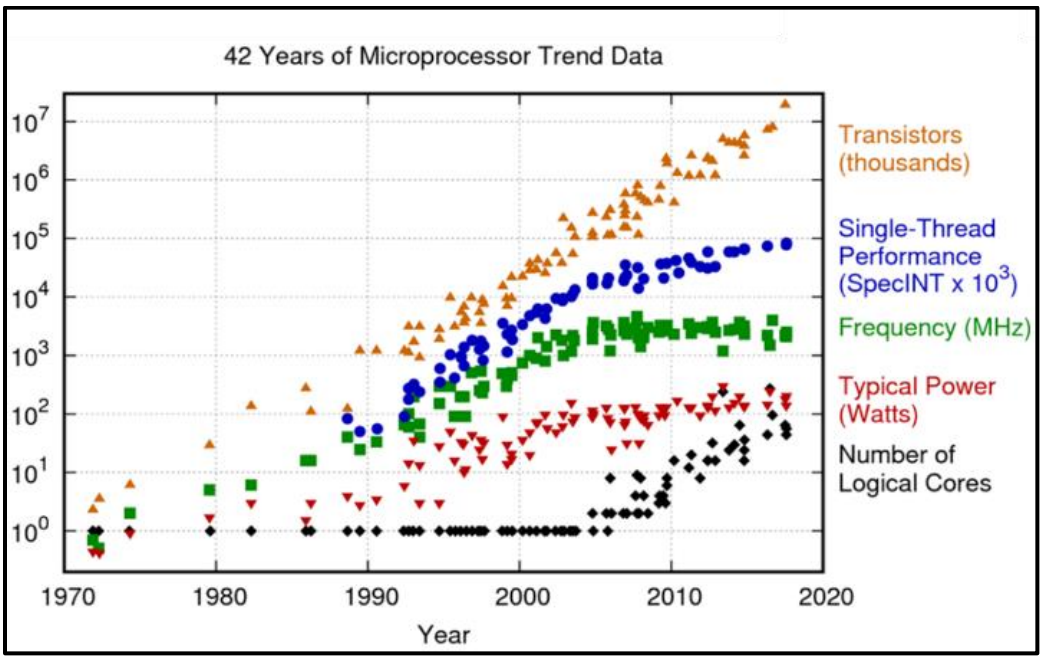

Figure 2. Microprocessor trend data over the past 42 years

3 Heterogenous Integration

Heterogeneous Integration (HI) definition from Dr. Moore [3] at a system perspective: '*It may prove to be economical to build large systems out of smaller functions, which are separately packaged and interconnected. The availability of large functions, combined with functional design and construction, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically*" which in aggregate provides enhanced functionality and improved operating characteristics. In this definition, HI or chiplets could mean to include any components be it an individual die with multiple cores, MEMS device,

assembled package or sub system, that are integrated into a single package. Hence, HI certainly offers a solution for performance scaling. For example as shown in Figure 3, instead of fabricating a single large multicore CPU die, smaller dies can be tiled within a package on an interposer with very short connections in-between the dies to realize the same performance offered by a single large die [7]. After a certain microprocessor power level, air cooling of high-density electronic component, let alone HI devices or stacked 3D devices get increasingly difficult and highly energy inefficient especially considering the form factor, scalability, and sustainability.

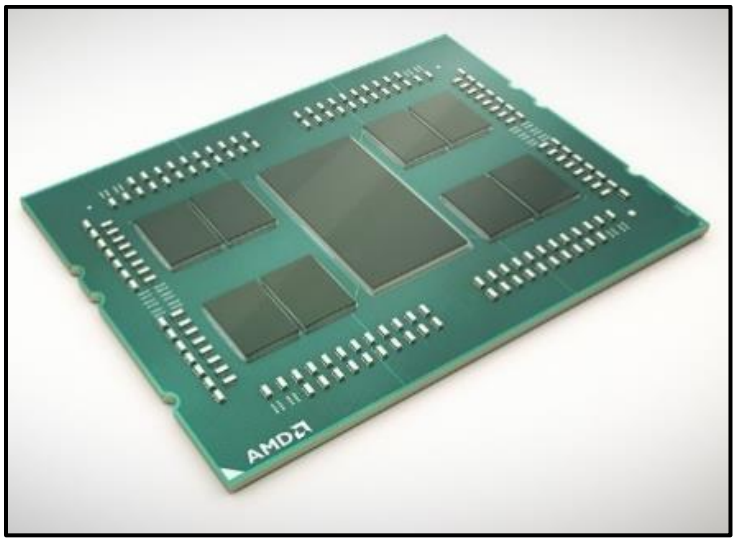

Figure 3. Chiplet from AMD EPYC 7003 Processors [7]

4 Overclocking the CPUs

Liquid Cooling and Immersion Cooling with its innate ability to carry high heat capacities for the given volume of heat transfer fluid completely outpace conventional air cooling especially when it gets to high density chips like 3D Integrated Circuits or stacked dies or ultimately Heterogeneously Integrated devices of the future. For example, the volumetric heat transfer ratio of water vs air is about 1900:1 and the ratio comparing an engineered 2P dielectric fluid vs that of air is even better, a massive 15000:1. A recently concluded study from MSFT [8] compares the thermal performance of a commercially available Intel multi core high density overclockable CPU as shown in Figure 4 (i9-9900K) [9] which was characterized using three different cooling mediums namely, air, water and a two-phase dielectric fluid respectively.

Figure 4. Overclockable desktop class CPU from Intel (i9 9900k)

Three different heat sinks; a traditional air-cooled heat sink, a liquid cooled cold plate and a 2P immersion cooled boiler plate were evaluated for thermal performance experimentally for a variety of operational and system parameters including, CPU core voltage, Core Frequency, Operating temperatures, and operating conditions including thermal interface materials, flow rates being tested at multiple workload types mimicking different data center operation types. As shown in Figure 5, at a cooling medium temperature of 34 ˚C for example, air cooled CPU throttled at a maximum clock frequency of 3300 MHz operating well past its TDP, while water cooled cold plates offered 41% higher clock rates compared to air and 2P immersion cooling yielded closer to 51`% higher clock rates than the air-cooled counterpart at 34˚C. Using 2P immersion in certain scenarios, the CPUs clocked well beyond 5200 MHz. 2P immersion cooling using a higher saturation temperature (aka boiling point) of 50 ˚C likewise yielded similar higher device performance. Also, in the case of 2P immersion cooling, while the corresponding lower junction temperatures (Tj) are well below the maximum allowable operating Tj,max while this also indicates potentially lower leakage current from the chip which is in other terms realized as increased longevity, reliability and serviceability of the device or system in whole. More importantly this rise in CPU core frequency can be realized in a hyperscale data center environment like ‘oversubscription of servers’ and Virtual machine (VM) autoscaling [10].

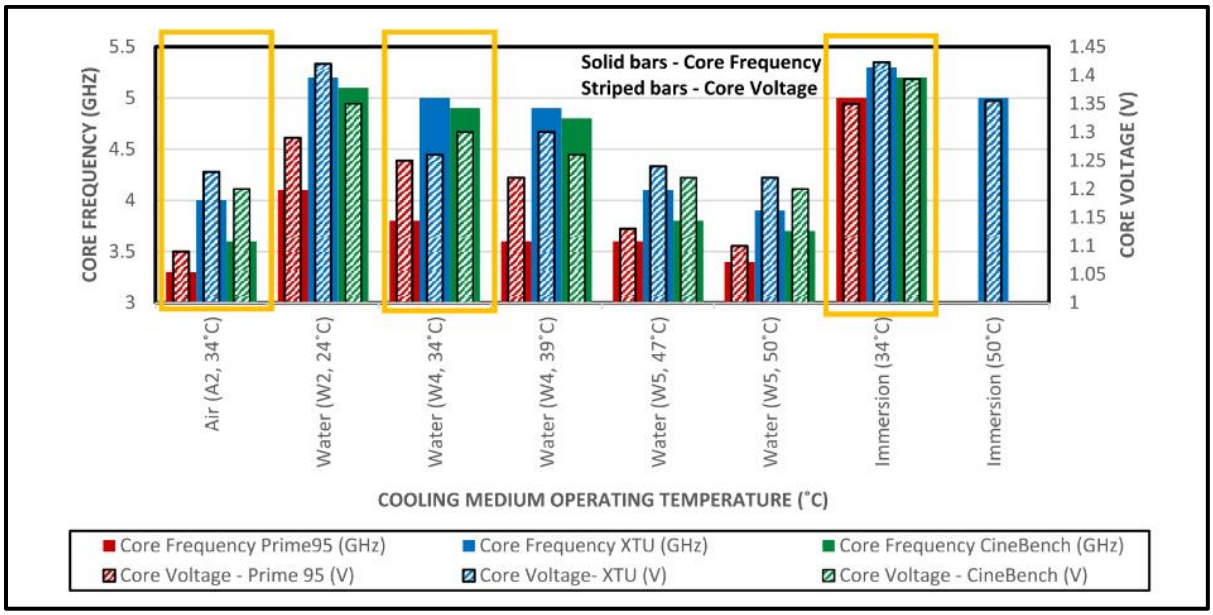

Figure 5. CPU core frequency (GHz) (solid bars in the chart) versus CPU core voltage (V) (striped bars in the chart) for all the cooling configuration while stressing the CPU with different workloads, respectively; Prime95, XTU, and Cinebench.

In the case of using liquid cooling utilizing copper microchannel cold plates in particular, the water temperatures were varied between 24, 34, 39, 47 and 50 ˚C to observe the CPU performance corresponding to different ASHRAE [11] Water Class Temperatures (according to the latest 2021 water classes: W27, W32, W40 and W45 respectively). An important trend to be taken into CPU package design consideration is that the CPU performance drops with increasing water inlet temperature. Similarly using 2P immersion cooling, going from fluids with boiling point of 34 to 50 ˚C had impacted a rise in junction temperature with no significant impact to the device performance. Lower water class temperatures (W27, W32) might require expensive chiller operation to extract necessary peak CPU or HW performance whereas the higher water class temperatures (W40, W45) spectrum from so called 'warm water-cooled data centers' could operate in par or better than air cooling and could untap potential case for waste heat from exit facility water loop. Hence, liquid cooling is deemed as a potential solution in not only enhancing the device densification but also improving the efficiency and performance of the device and weighed in as a serious contender to reliable and noiseless operation (without lot of fans and associated power).

5 Advanced Packaging and Fabrication

Aided by Liquid cooling, 'Packaging of semiconductor devices' hold the major key for further thermal budget to untap potential cooling performance gains at the chip level. For example, the package material, its thermal conductivity (k), and size along with the thermal interface material (TIM) can be optimized further on top of leveraging niche semiconductor fabrication techniques including additive manufacturing at scale. Moreover, by bringing the cooling fluid closer to the die as possible, it is helping to eradicate layers of thermal resistance which would otherwise be present in a conventional packaging to cooling schema which typically includes layers of Die (heat generating source), TIM1 (Typically high k material like Indium), Package case or lid or Integral Heat Spreader (IHS), TIM2 (Typically grease with low k), and finally the cooling fluid. For example as shown in Figure 6, when the package case or IHS is integrated with liquid cooling manifolds, it can allow operating at higher inlet water temperatures above 45 ˚C without impacting the device performance. By going one step further, when the liquid is pushed directly through the Silicon die itself containing etched microchannels (aka Embedded cooling or Microfluidic pin fin cooling), the resistance to heat transfer from chip to heat transfer fluid is reduced substantially. Embedded cooling would be the steppingstone in thermal management of next generation high density 3D stacked dies or HI devices linked by through silicon vias (TSVs). Microfluidic pinfin cooling as demonstrated in a recent Microsoft's work [12] can also let devices operate using fluids with higher inlet temperatures greater than 45 ˚C hence facilitating further heat reuse potential when it gets to DC scale. Although, this step of 3D microfluidic cooling involves tremendous investment and innovation in the semiconductor fabrication facilities and associated supply chain processes, the microfluidic liquid cooling solution offers multiple advantages including latency and bandwidth gains that comes with the reduced interconnect lengths on top of offering drastic sustainability impacts.

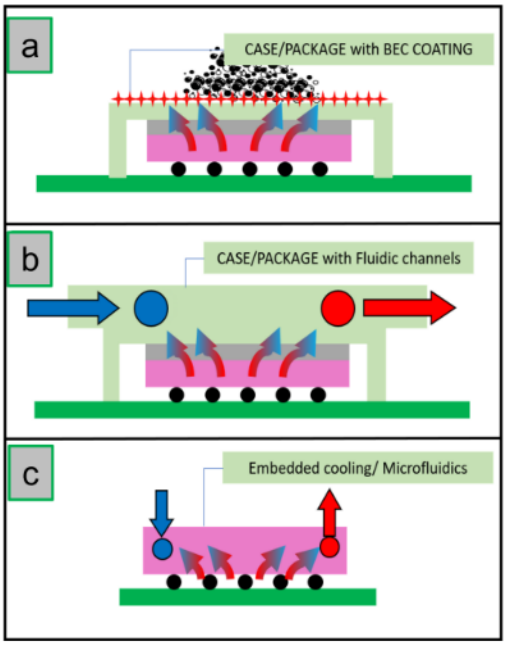

Figure 6. Schematic of modified package designs to gain more thermal advantage using liquid cooling.

6 AI/ML chips of Next Generation

AI/ML (Machine Learning) ASICs are leading the charge toward liquid cooling with very high-power densities and challenging thermal requirements. Current generations of AI/ML chips can exceed 600W and are projected to push past the 1kW mark in the next few generations. Often these high-power ASICs are coupled with High Bandwidth Memory (HBM) in the same die. The 3D stacked architecture of HBM drives significant internal thermal resistance which, when combined with the relatively low memory temperature requirements (typically <85C), can lead to significantly low cooling requirements. This is especially true in lidded packages which can have case temperature requirements of <60C.

As ASIC TDP increases and system thermal solutions become increasingly optimized, chip package resistances consume an increasingly higher percentage of the overall thermal budget, in some cases >50%. Well optimized, lidless packages of >700W can be air cooled, while less optimized solutions may require liquid cooling with low fluid temperatures. To continue to drive higher performance and increased power into ASICs the entire thermal solution, from the internal dies to the external heat rejection, must be optimized in concert to continue to provide sustainable and efficient solutions.

7 Implications of Lower Facility Water Temperature

Sustainability goals of Microsoft battling the 'ecosystem threatening' global climate change include reducing their historic carbon emissions associated with its operation by 2050 (and that includes scope 1, scope 2 and scope 3 emissions) [13] and replenish their facilities with more water than what been used for its services by 2030 [14]. Liquid cooling of power-hungry high density data center strategy aligns well with the Microsoft sustainability goals by eliminating expensive chiller operation or even their adiabatic evaporative based cooling and the associated water usage that goes with conventional air-cooling process. Liquid cooing not only enables the devices of next generation like 3D stacked dies and heterogeneously integrated devices, but also facilitates HW and data center densification, which are all realized as embodied carbon gains through cramming the data center energy, size, and space. Lower inlet water temperature can enable device level performance gains but can take a toll on the energy overhead and may involve a chiller to function. Even though having lower facility water inlet temperature enables reduced flow rate hence lower pumping power expenditure, the overall heat energy that can be captured at the exit gets impacted adversely. Whereas high temperature liquid cooling allows higher inlet temperatures of the cooling fluid (>50 ˚C) which can unlock many likely DC heat reuse or recapture scenarios [15,16] like district heating, space heating, water heating and essentially any process that can share heat like shown in Figure 7.

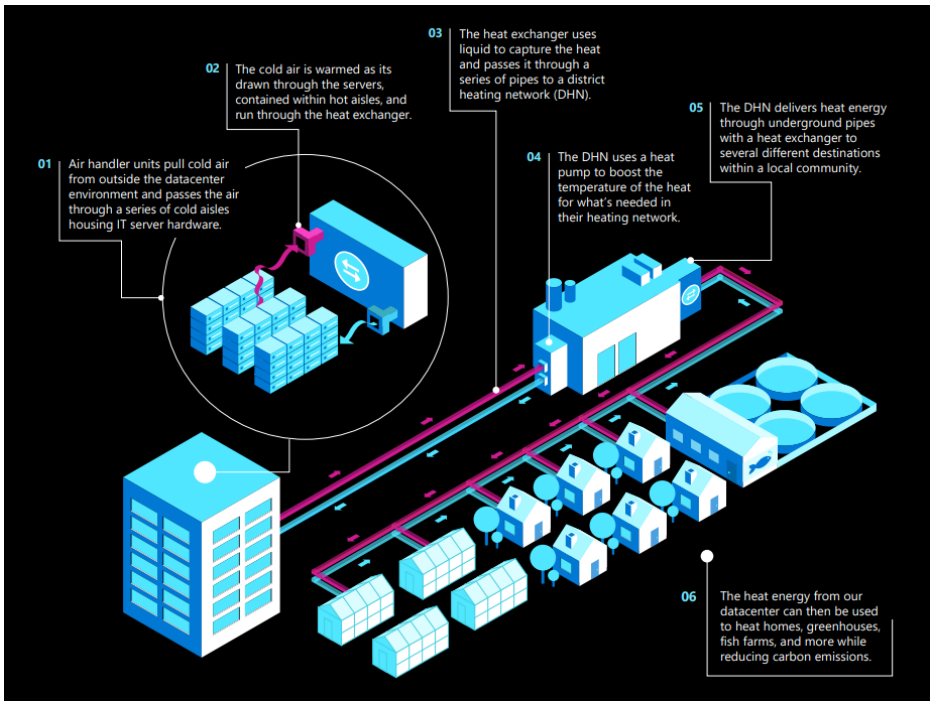

Figure 7. Modern datacenter heat energy reuse

Advanced state of the art data center technologies including fuel cells, direct air capture (DAC), adsorption chillers and sea water desalination plants can also take advantage of process heat coming out of liquid cooled data center facilities. This option of sharing heat could also be a prospective revenue ($) or market option to Microsoft or to any hyperscale provider in general. While following COP26 [17] and other global climate pacts, multi megawatt data centers not just in EU, but anywhere in the globe can capture the high-grade waste heat (>50 ˚C) coming out of its data center and share it with the existing district heating lines and, in that process, helping many traditionally coal fired district heating lines move towards more green energy source repurposed out of the data center heat.

8 Conclusion

Finally, it is good to notice that even the latest generation air-cooled data center fleets [18] are trying to push the inlet air temperature to a higher operating point thereby reducing the strain on the air-cooling energy expenditure. Energy density and cost-efficient liquid cooling and immersion cooling options emerged out to be the favorable choice compared to their air-cooled fleets of earlier generation datacenter facility. But this would require enormous investment, innovation and more importantly adoption from all fronts starting at the semiconductor foundry, architecture & packaging, liquid cooling technology vendors, system integrators, software developers, environmental health, and safety (EHS), utility providers, consumers, and customers. Encouraging sign is that it's not just Microsoft or other cloud service providers, but also the broader global communities comprising other foremost technology organizations and academic institutions, independent establishments like Open Compute Project (OCP) [19], and key energy / utility companies which are not just deeply committed but also heavily invested to looking into accelerating new technologies to market that can fight against the global climate and energy crisis of the current generation. The end sustainability goal involves alignment between a lot of technology stakeholders, associated markets, and corresponding engineering departments and in a way that's what makes this problem more interesting, unique, and inclusive. Having said all that, 'Hot liquid cooling of cloud facilities is here to stay'.

9 References

1. Recalibrating global data center energy-use estimates, BY ERIC MASANET, ARMAN SHEHABI, NUOA LEI, SARAH SMITH, JONATHAN KOOMEY, *SCIENCE*28 FEB 2020: 984-986
2. Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper (virtualization.network)
3. G. E. Moore, "Cramming More Components Onto Integrated Circuits," in *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82-85, Jan. 1998, doi: 10.1109/JPROC.1998.658762. Moore: Cramming More Components Onto Integrated Circuits | IEEE Journals & Magazine | IEEE Xplore
4. IEEE Heterogeneous Integration for HPC and Data Centers: PowerPoint Presentation (ieee.org)

5. A. Bar-Cohen, “Gen 3 ‘embedded’ cooling: Key enabler for energy efficient data centers,” IEEE Trans. Compon., Packag., Manuf. Technol., vol. 7, no. 8, pp. 1206–1211, Aug. 2017.
6. S. H. Fuller and L. I. Millett, The Future of Computing Performance: Game Over or Next Level?. Washington, DC, USA: National Academy Press, 2011
7. AMD EPYC 7003 Processors (Data Sheet)
8. CPU Overclocking: A Performance Assessment of Air, Cold Plates, and Two-Phase Immersion Cooling | IEEE Journals & Magazine | IEEE Xplore
9. Intel Core i99900K Processor 16M Cache up to 5.00 GHz Product Specifications
10. Cost-Efficient Overclocking in Immersion-Cooled Datacenters | IEEE Conference Publication | IEEE Xplore
11. ASHRAE: TC9.9 Emergence and Expansion of Liquid Cooling in Mainstream Data Centers (ashrae.org)
12. Integrated Silicon Microfluidic Cooling of a High-Power Overclocked CPU for Efficient Thermal Management | IEEE Journals & Magazine | IEEE Xplore
13. Microsoft will be carbon negative by 2030 - The Official Microsoft Blog
14. Microsoft will replenish more water than it consumes by 2030 - The Official Microsoft Blog
15. Mikko Wahlroos, Matti Pärssinen, Samuli Rinne, Sanna Syri, Jukka Manner, Future views on waste heat utilization – Case of data centers in Northern Europe, Renewable and Sustainable Energy Reviews, Volume 82, Part 2, 2018, Pages 1749-1764, ISSN 1364-0321
16. Utilization of Waste Heat in the Data Center, A white paper by NeRZ in collaboration with eco - Association of the Internet Industry
17. COP26-Explained.pdf
18. IT@Intel: Green Computing at Scale White Paper
19. Home » Open Compute Project

10 License

11 About Open Compute Foundation

At the core of the Open Compute Project (OCP) is its Community of hyperscale data center operators, joined by telecom and colocation providers and enterprise IT users, working with vendors to develop open innovations that, when embedded in product are deployed from the cloud to the edge. The OCP Foundation is responsible for fostering and serving the OCP Community to meet the market and shape the future, taking hyperscale led innovations to everyone. Meeting the market is accomplished through open designs and best practices, and with data center facility and IT equipment embedding OCP Community-developed innovations for efficiency, at-scale operations and sustainability. Shaping the future includes investing in

strategic initiatives that prepare the IT ecosystem for major changes, such as AI & ML, optics, advanced cooling techniques, and composable silicon. Learn more at www.opencompute.org.