ISSUE: April 2024

Liquid Immersion Cooling For Higher Power Density In Server Power Supplies

by Sam Abdel-Rahman and Ashish Ekbote, Infineon Technologies, El Segundo, Calif.

The scaling of cloud-based Internet services and artificial intelligence (AI) have driven significant growth of the processing power in data centers. At the same time, the total cost of ownership (TCO) for data centers is rising, resulting in increasing demand for highly efficient, reliable, and compact server systems. Moreover, with the rapid increases in 5G and edge data centers, edge servers are being deployed closer to cities and other population centers, where energy and real estate costs are much higher.

Traditional fan cooling, widely used in server switch mode power supply (SMPS) racks, is limited by thermal capacity due to poor heat exchange through air. In contrast, liquid cooling is a much more effective method to exchange the racks' heat outside the data center without wasting a significant share of the energy on cooling. In this article, we present the findings from a recent study that validate the potential benefits of immersion cooling for a server SMPS. They emphasize system re-design to capture the highest value of immersion cooling—optimizing energy efficiency and power density—as well as identify reliability advantages.

After a quick look at projections for power requirements in data centers, we explain the principles of two-phase immersion cooling and quantify the power supply efficiency benefits of liquid cooling versus forced-air cooling over a range of power density levels. We then discuss the results of an experiment on a 2.7-kW SMPS that was liquid cooled using two-phase immersion cooling. The data compares component temperatures, power supply efficiency and power supply output obtained with liquid cooling versus fan cooling.

Finally, we review two real-world applications of liquid immersion cooling (two phase and single phase) in the data center by Alibaba and Microsoft. These results highlight performance benefits at both the server power supply and data center levels.

Transitioning The Data Center

The incredible growth in the amount of data created each year (from 2 zettabytes in 2010 to 44 zettabytes (44 trillion gigabytes) in 2020) shows no signs of stopping.[1] In fact, the demand for data center services is predicted to grow strongly driven by media streaming and emerging technologies such as artificial intelligence (AI), machine learning (ML), virtual reality (VR), 5G, and blockchain.[2]

To handle this increasing amount of data processing, an increasing amount of power is required. In the U.S., data center power consumption is projected to reach 35 GW by the end of the decade, almost double the 2022 level.[3] Today's 10 to 14 kW per rack power consumption in existing data centers is likely to rise to 40 to 60 kW for racks designed to handle AI. Globally, energy demand from data centers and data transmission networks currently accounts for 1% to 1.5% of the total electricity use.

According to the International Energy Agency (IEA), "As the efficiency gains of current technologies decelerate (or even stall) in upcoming years, more efficient new technologies will be needed to keep pace with growing demand."[1] Transitioning from air cooling to liquid immersion cooling is an ideal way to handle the increased power requirements with increasingly low levels of available space in data center hot spots.

Introduction To Liquid Immersion Cooling For Server SMPS

A fundamental approach to improving efficiency and power density is transferring dissipated heat more effectively. Most server SMPS system components are limited thermally—they have the potential to deliver more power only if the system's thermal capacitance is enhanced. Immersion cooling enables thermal capacity enhancement, increasing power delivery and density—much higher than what is possible with fan cooling.

Fig. 1 shows the concept of two-phase immersion cooling (2-PIC). With components immersed in a dielectric liquid, heat from components boils the fluid, vapor rises and condenses due to the tubes at the top of the tank and drips down. With room temperature water pumped through these tubes, heat transfers to facility water that flows outside of the data center.

In contrast to fan cooling, immersion cooling solutions would support higher growth in computational power, while reducing costs and delivering across-the-board improvements in energy efficiency, rack density, cooling capacity, and reliability.

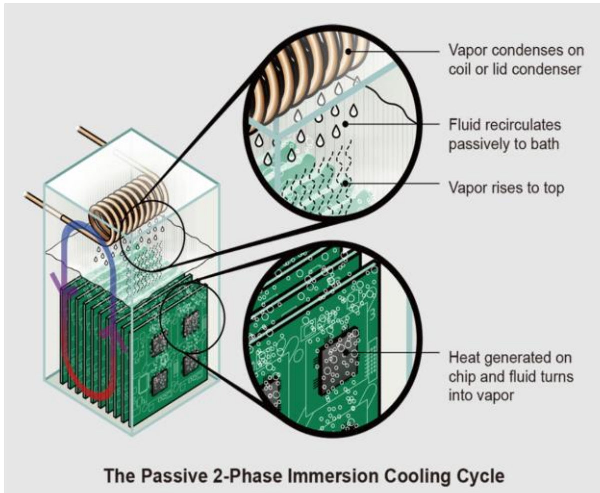


Fig. 1. The passive two-phase immersion cooling cycle includes boiling and condensing.(Image courtesy of LiquidStack.[4])

Efficiency And Density Benefits Using Simulation

Thermal-limited components, such as semiconductors and magnetics, have a positive temperature co-efficient of resistance. As such, an SMPS can either run more efficiently at a defined power level or deliver more power as in the case of liquid cooling. Immersion cooling is an effective way to reduce temperature hot spots among components, offering a clear reliability and lifetime advantage.

To scientifically identify the efficiency and density benefits, Pareto optimization modeling was used to quantify the achievable performance tradeoffs between efficiency and power density for the different cooling environments. It systematically considered all available degrees of freedom to identify the optimal designs positioned on the Pareto-front. Fig. 2 illustrates that compared to a conventional fan-cooled design, an enhanced thermal solution leads to increased efficiency, or power density, or both.

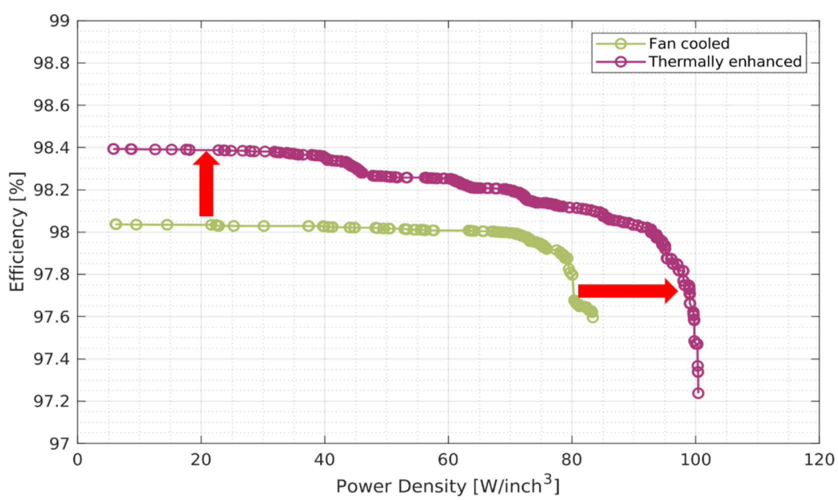


Fig. 2. The benefits of thermally enhanced immersion cooling over fan cooling occur at all power density levels.

Liquid Cooling Tank Setup And Challenges

To verify the results of the simulation, a small-scale experimental setup was built to demonstrate the benefits of two-phase liquid immersion cooling on a server SMPS. As shown in Fig. 3, a quartz tank was used for its high chemical purity and stability. The tank had dimensions of 750mm (l) x 200mm (w) x 300mm (h) which allowed sufficient experimental setup flexibility.

Fig. 3. The components of this liquid cooling experiment included a quartz tank, a condenser, 2-PIC liquid and the unit under test (UUT).

Novec 7100, a proprietary hydrofluoroether (HFE) from 3M, was chosen as the cooling liquid. With a boiling point of 61°C, it clamps heat producing components at this temperature point. It also has a dielectric strength close to 10 kV/mm, which prevents the conduction of any electricity that would result in short circuits.

In this setup, a condenser changes the vapor created due to boiling, back to a liquid. To do this, it needs to have sufficient surface area relative to the heat dissipation in the tank to capture all the vapor and condense it back to liquid to prevent the liquid loss due to evaporation. For this experiment, the condenser was fabricated with a ¼-inch copper tube, that consisted of 13 loops of 7-cm outer diameter (OD) coils.

For the limited space experiment, a cooling system or heat exchanger uses a fan to cool down the fluid exiting the condenser, which warms up during the process, and pumps room temperature fluid back to the condenser.

Experimental Thermal And Efficiency Results Of A Liquid-Cooled SMPS

Fig. 4 shows the experiment with a 2.7-kW power supply unit (PSU) immersed and operating in the tank. Due to the components' heat dissipation, the liquid boils, and a two-phase change occurs (changing from liquid to vapor back to liquid).

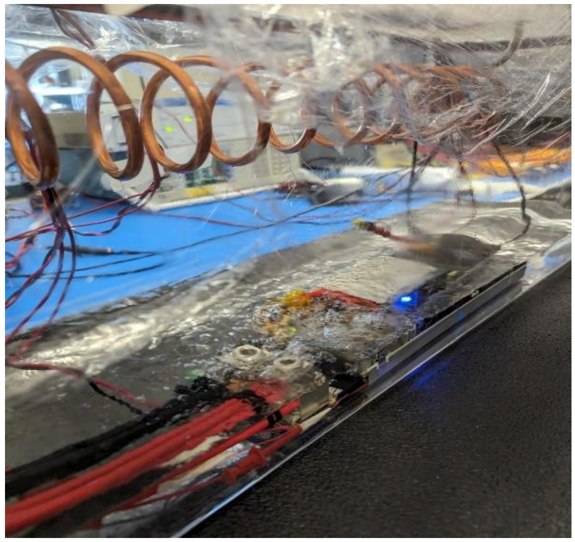
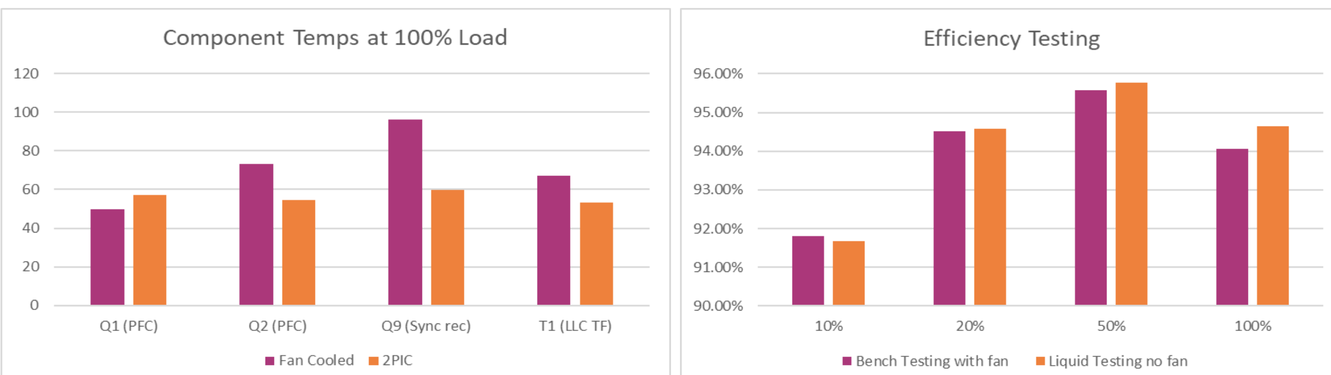


Fig. 4. Experiment with an immersed 2.7-kW PSU.

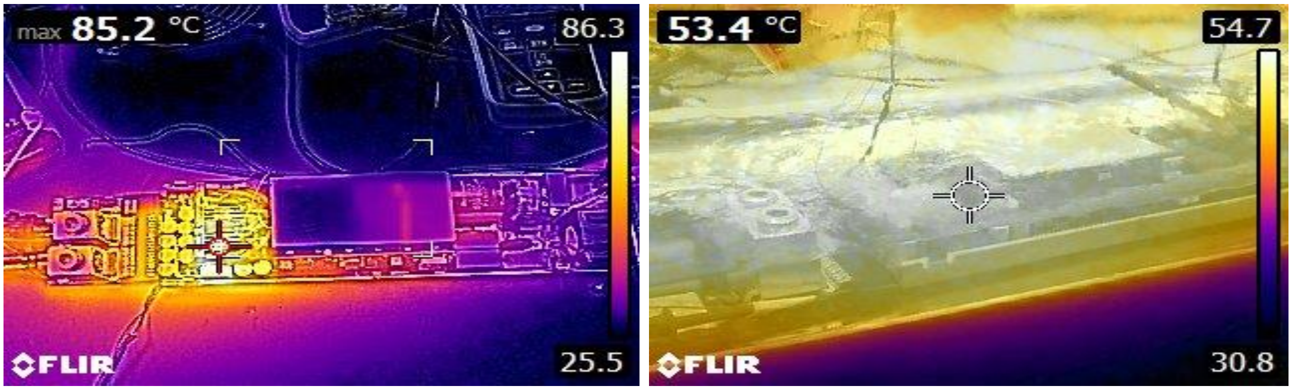
Fig. 5a shows a temperature comparison between fan cooling and two-phase immersion cooling for some critical components. Liquid cooling significantly reduces the operating temperature of these components and therefore improves their reliability. The efficiency comparison between fan cooling and 2-PIC in Fig. 5b shows approximately a 0.6% increase at full load.



(a) *(b)*

Fig. 5. Comparing the performance of fan cooling versus two-phase liquid immersion cooling (2-PIC) for a server power supply. Shown here are comparisons of component temperatures (a) and power supply efficiency across the load range (b).

Fig. 6a is a thermal image of an LLC transformer with fan cooling on the bench, and Fig. 6b is a thermal image of an LLC transformer with 2-PIC. It is very clear that the 2-PIC reduces and unifies temperature, removing hot spots.



(a) *(b)*

Fig. 6. Thermal imaging depicts temperature reductions and uniformity of temperatures achieved by liquid cooling versus fan cooling. Shown here are LLC transformer hot spots observed with fan cooling (a) and the significantly lower temperatures and elimination of hot spots seen with 2-PIC (b).

To further validate the benefits of two-phase liquid immersion cooling on system cost savings and size reduction, a second unit was assembled but without the heat sinks for the PFC and LLC switches. This unit was operated at close to 150% of nominal load to simulate the increase in power density and demonstrate safe operating temperatures even at overload conditions.

The results in Fig. 7 show that the temperatures of some of the critical components are well below the device maximum limits even under these operating conditions.

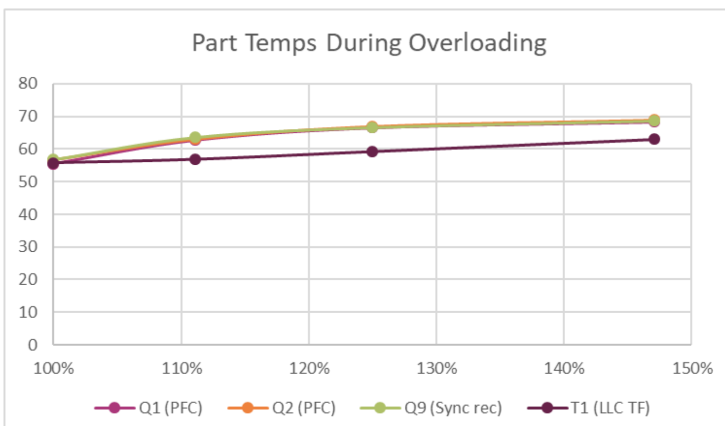


Fig. 7. Temperature comparison of SMPS components with 2-PIC and assembly operating up to 150% of nominal load.

Fig. 8 shows that the efficiency of the unit is maintained by the cooling environment even in the absence of heat sinks. In summary, the unit 2 results show that two-phase liquid immersion cooling technology opens up possibilities of designing a power supply unit with innovative packaging techniques to achieve a much higher power density.

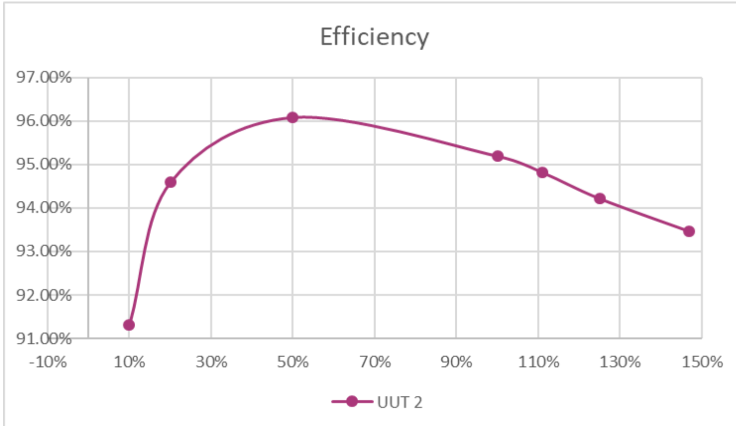


Fig. 8. Even without heatsinks, the SMPS' efficiency remains high while operating up to 150% of nominal load.

Real World Use Cases

Case 1. To increase the efficiency in its next-generation, high-density data centers, Alibaba has been working on immersion cooling for well over six years.[5] At APEC 2023, data was presented based on single-phase (1-ph) liquid cooling implemented at the Alibaba Renhe Data Center, the largest immersion cooling production facility in the world.[6] With 1-ph liquid immersion cooling, the fluid stays in liquid phase and does not boil and evaporate as it does with 2-PIC. However, a 1-ph cooling design is good up to about 100 kW and it costs less than a two-phase approach.

Compared to air-cooling, the temperature of key magnetic, semiconductor, and capacitive components in the 1-ph liquid cooling was reduced by about 40%. In addition to the reduced temperatures that increase reliability and product lifetimes in the data center (for example, up to 10 years longer aluminum electrolytic capacitor life expectancy based on a 10°C lower operating temperature), other benefits of not using a fan for air flow include

- No fan-based vibration to reduce reliability
- No need to provide protection from blowing dust
- No fan noise.

With the 1-ph design, the annual mean power usage effectiveness (PUE) number (the inverse of data center infrastructure efficiency, where an ideal PUE is 1.0) is 1.09 compared to typically achieved PUE values of 1.2 or less.[7] Using a 2-PIC design, it is expected that the PUE can be reduced to 1.07 or even lower.

Case 2. Working with a data center IT system design and manufacturing company, Microsoft implemented a two-phase immersion cooling solution in its data center in Quincy, MA.[8] Using two-phase immersion cooling in the production environment is part of Microsoft's long-term plan to keep up with demand for faster, more powerful data center computers.

When its engineers investigated liquid immersion as a cooling solution for high-performance computing applications such as AI, their findings revealed that two-phase immersion cooling reduced power consumption for any given server by 5% to 15%. Since liquid cooling is a waterless technology, this will also help Microsoft meet a commitment to replenish more water than it consumes by the end of this decade.

With these industry leaders providing the pioneering examples, it is expected that other data centers will follow and implement liquid immersion cooling in the future.

Conclusion

In data centers, traditional air-cooling methods have high costs in energy consumption, water use, real estate, carbon footprint, and more. In contrast, immersion-cooled data centers are scalable, faster, and more energy-efficient. With growing power consumption in data centers requiring higher power densities, two-phase liquid immersion cooling provides the most efficient way of rejecting heat and enabling power densities up to 250 kW per server rack.

In addition to lowering the PUE number and eliminating or mitigating failure points such as fan or overheating of semiconductors and electrolytic capacitors, 2-PIC experiments have demonstrated thermally stable and efficient operation at close to 150% loading.

At Infineon, planned future work includes engaging with customers to define next-generation packaging and product ideas to take advantage of liquid immersion cooling and participating in efforts to standardize the immersion liquid.

References

1. "Data Centres and Data Transmission Networks," International Energy Agency.

2. "Newmark: US data center power consumption to double by 2030" by Matthew Gooding, Data Center Dynamics, January 15, 2024.

3. "Ranked: Top 50 Data Center Markets by Power Consumption" by Julie R. Peasley, Visual Capitalist, January 10, 2024.

4. "The Passive 2-Phase Immersion Cooling Cycle," drawing on LiquidStack website.

5. "Alibaba Immersion Cooling at Scale Research" by Patrick Kennedy, ServeTheHome, April 9, 2018.

6. "Immersion-Cooling Power Supply Unit for Data Center" by Cai Youzhun, presentation at APEC 2023.

7. "High-Performance Computing Data Center Power Usage Effectiveness," NREL.

8. "To cool datacenter servers, Microsoft turns to boiling liquid" by John Roach, Microsoft, April 6, 2021.

About The Authors

Sam Abdel-Rahman joined Infineon Technologies in 2011 and currently serves as a system architect responsible for developing the application roadmap of server/data center SMPS and renewable applications. Sam has experience in the power and semiconductor industry with a focus on system architecture, topologies and control. He received a PhD degree in power electronics from the University of Central Florida.

Ashish Ekbote joined Infineon in 2022 as a lead system application engineer focused on developing solutions for the server SMPS, residential solar and ESS applications. He has more than 15 years of experience developing products and applications in diverse industries such as semiconductor applications, LED lighting, EV power electronics and consumer electronics. Ashish holds an MSEE with a focus on power electronics.

For more on thermal management in power electronics, see How2Power's Design Guide, locate the "Design area" category and select "Thermal Management".