Liquid Cooling Technologies for Data Centers and Edge Applications

White Paper 265
Revision 0

by Tony Day
Paul Lin
Robert Bunger

Executive summary

Increasing IT chip densities, a focus on energy efficiency, and new IT use cases like harsh edge computing environments are driving the interest and adoption of liquid cooling. In this paper we present the fundamentals of liquid cooling, describe the advantages over conventional air cooling, and explain the 5 main direct to chip and immersive methods. To help guide the selection of the appropriate liquid cooling method for a given need, we explain the key attributes that must be considered.

RATE THIS PAPER ★★★★★

Introduction

IT equipment technology change has always been a primary driver in the development of infrastructure cooling solutions. Although liquid cooling has been deployed for many years in mainframes and high-performance computing (HPC), today's demands of cloud, IoT, AI, and edge applications are once again resulting in IT technology changes which is forcing a renewed look at liquid cooling and the development of new technologies. Increasing focus on data center energy efficiency and sustainability is also placing increased pressure on the data center industry to develop and adopt efficient cooling infrastructure like liquid cooling. White Paper 279, *Five Reasons to Adopt Liquid Cooling*, describes the reasons to consider this technology.

Data center owners and designers often have fundamental questions about liquid cooling, such as:

- Why is liquid cooling better than air cooling in transferring heat energy?
- What are the types of liquid cooling solutions?
- What are the benefits and drawbacks of each method of liquid cooling?
- What kind of critieria should I use to choose between different liquid cooling technologies?

In this paper, we answer these questions, and provide guidance in selecting an appropriate liquid cooling method for your application.

Air cooling vs. liquid cooling

The predominate way to remove heat from IT equipment is by airflow through the chassis of the equipment. For typical servers, 70-80% of the heat is generated by the CPU, with the remaining heat from peripherals like memory, power supply, hard drives, SSD, etc. The increasing use of GPUs has further increased the heat generated within the IT chassis. A GPU can generate over 400 watts, but high core count CPUs, like the latest Intel Xeon processor are now also at 400 watts.

Liquids have a much greater capacity to capture heat by unit volume. This allows liquid cooling technologies to remove heat more efficiently and allows the chips to work harder (i.e. increased clock speed). Additionally, the heat is rejected to the atmosphere either via dry coolers or, in the case of hotter environments, cooling towers. Sometimes the heat may be reused elsewhere such as district heating.

In the **Appendix**, we provide a detailed comparison between the heat transfer capabilities of water and air.

Liquid cooling methods

Liquid cooling is not new to data center applications. You can trace it back to the 1960s when it was used in IBM mainframes to solve the cooling challenges for solid state devices which were densely packed and had lower allowable operating temperatures. But, the introduction of complementary metal oxide semiconductor (CMOS) technology in the early 1990s replaced bipolar semiconductor technology, which reduced power consumption. As a result, convective airflow cooling again became the default cooling option for IT equipment.

Convective airflow cooling is still dominant in data centers, but there is broader adoption of liquid cooling in gaming, blockchain mining, and HPC applications, which requires greater compute capacity with special servers. Liquid cooling hasn't seen broader adoption across the data center industry, largely because the compute

Life Is On | Schneider Electric

demand has been met with increasing the number of logical cores that stay within reasonable power limits. Additionally, the data center industry, in general, is conservative, and new technologies and architectures have a slow adoption rate.

There are 2 main categories of liquid cooling – direct to chip (sometimes called conductive or cold plate) and immersive. From these two categories come a total of five main liquid cooling methods, as the diagram in **Figure 1** depicts (the orange boxes). In this section, we'll describe and illustrate each method. The Green Grid White Paper #70, *Liquid Cooling Technology Update* also does an excellent job of classifying the current technologies.

Figure 1
Liquid cooling approach classification
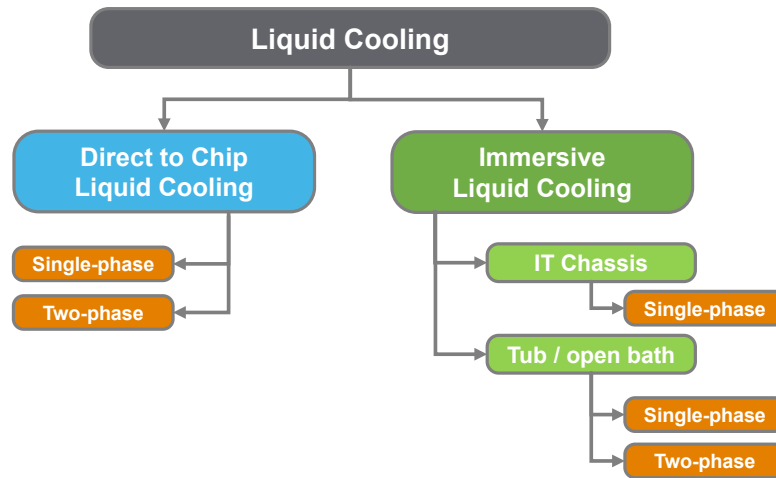
Direct to chip liquid cooling – single-phase

Direct to chip is where the liquid coolant is taken directly to the hotter components (CPUs or GPUs) with a cold plate on the chip within the server. The electronic components of the IT are not in direct physical contact with the liquid coolant (see **Figure 2**). Some designs also include cold plates around memory modules. With this method, fans are still required to provide airflow through the server to remove residual heat. This means that conventional air-cooling infrastructure is reduced but still needed.

Figure 2
Diagram of direct to chip liquid cooling – single-phase
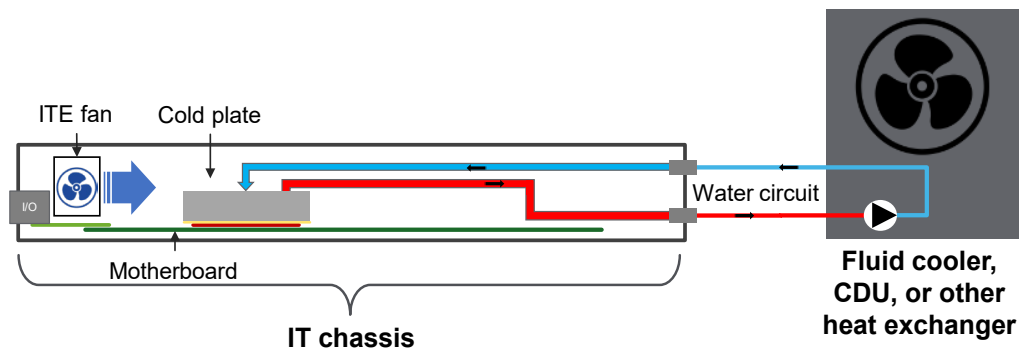
Water or a dielectric liquid can be used as the coolant to the cold plates. Water introduces a downtime risk if there is leakage, but solutions such as leak prevention systems (LPS) keep the water loop at slight vacuum to mitigate the risk of leaks within IT equipment.

Fluid manifolds installed at the back of the rack are used to distribute fluid to the IT equipment, analogous to a rack PDU for electricity. The interface between the

server and the manifold is typically achieved via a non-spill & non-drip coupling which ensures cleanliness and safety of the installation. Single-phase means that the fluid doesn't change state while taking away the heat. For direct to chip, mainly single-phase water-based coolants are used, but some designs use engineered dielectric fluids.

Direct to chip liquid cooling – two-phase

This method is like the previous method, except that the fluid is two-phase which means the fluid changes from one state to another – i.e. from liquid to gas in taking away the heat. Two-phase is better than single-phase (in terms of heat rejection) but requires additional system controls to ensure proper operation. For two-phase direct to chip liquid cooling, engineered dielectric fluid is used. This eliminates the risk of water exposure to the IT equipment. The dielectric vapor can be transported to a condenser outside or reject its heat to a building water loop. **Figure 3** illustrates this method.

Figure 3
*Diagram of direct to chip liquid cooling – two-phase*
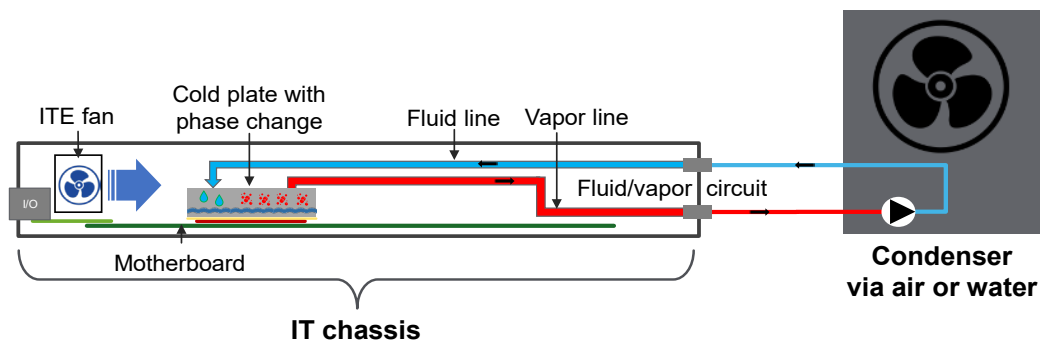
Immersive liquid cooling – IT chassis – single-phase

With immersive liquid cooling, the liquid coolant is in direct physical contact with the IT electronic components. The servers are fully or partially immersed in a dielectric liquid coolant covering the board and the components, which ensures all sources of heat are removed. This approach uses a single-phase dielectric. With immersive liquid cooling, all fans within the server can be removed, and all electronics are placed in an environment which is inherently slow to react to any external changes in temperature and renders it immune to the influence of humidity and pollutants. Since there are no fans, this approach operates in near silence.

Figure 4 illustrates the IT chassis single-phase approach to immersive liquid cooling. The server is encapsulated within a sealed chassis and can be configured as normal rackmount IT or standalone equipment. The electronic components are cooled by the dielectric fluid either passively via conduction and natural convection, or actively pumped (forced convection) within the servers, or a combination of both. Heat exchangers and pumps can be located inside of the server or in a side arrangement where the heat is transferred from the dielectric to the water loop.
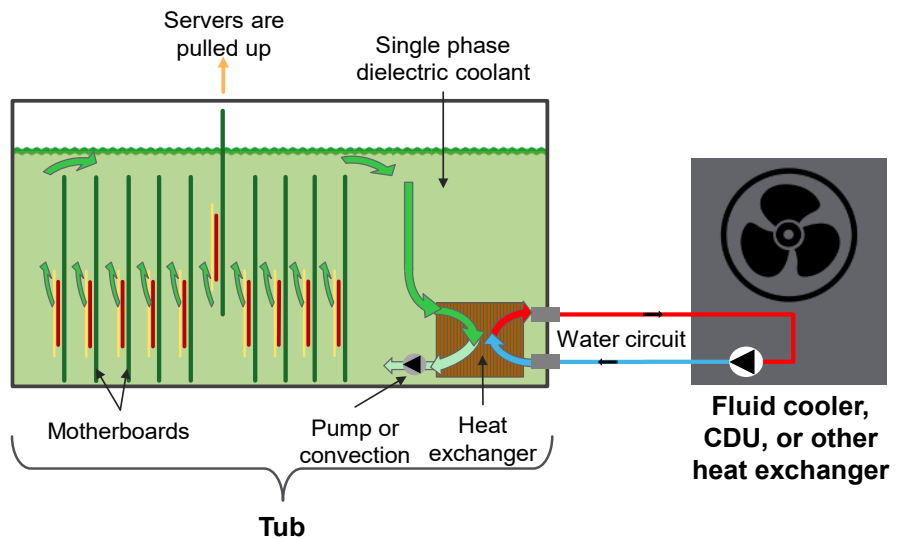
Figure 4
Diagram of immersive liquid cooling – IT chassis – single-phase

Life Is On   Schneider Electric
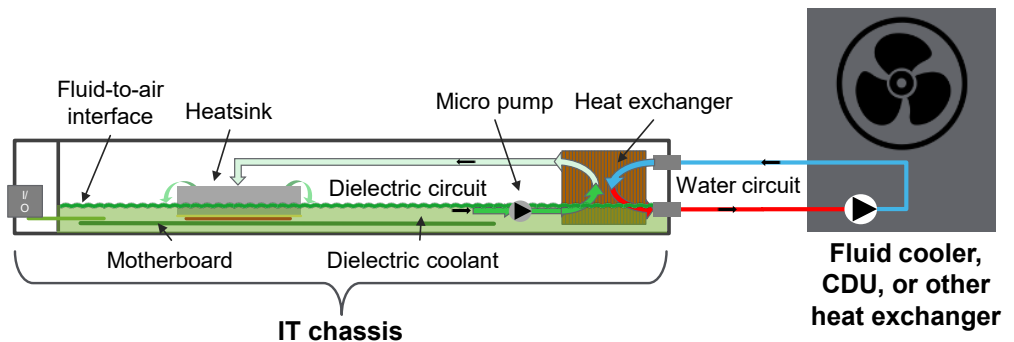
Immersive liquid cooling – Tub – single-phase

With the tub method (also referred to as open bath), the IT equipment is completely submerged in the fluid. With traditional IT racks, the servers are horizontally stacked from the bottom to the top of a rack. However, because this method uses a tub, it's like laying a traditional rack of servers on its back. Instead of pulling servers out on a horizonal plane, tub immersive servers are pulled out on a vertical plane. **Figure 5** illustrates a diagram of this method (orange arrow shows the direction the servers are pulled out of the tub). Many times, this method incorporates centralized power supplies to provide power to all the servers within the tub. The heat within the dielectric fluid is transferred to a water loop via heat exchanger using a pump or natural convection. This method typically uses oil-based dielectric as the fluid. Note that the heat exchanger may be installed inside or outside the tub.
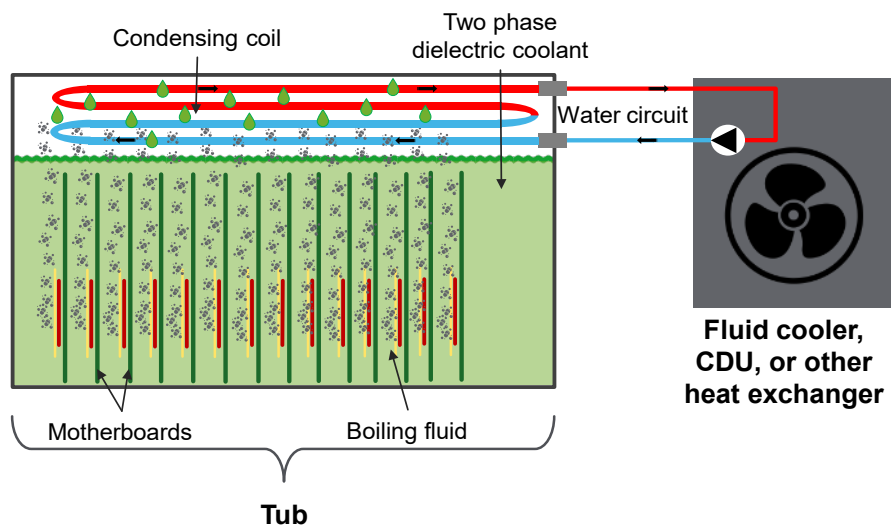
Figure 5
An example of immersive liquid cooling – tub – single-phase

Immersive liquid cooling – Tub – two-phase

Just like the single-phase tub method, the IT is completely submerged in the fluid. The difference here is the two-phase dielectric coolant. Again, this means the fluid changes from one state to another – i.e. from liquid to gas in taking away the heat. **Figure 6** illustrates this architecture. This must be an engineered fluid because of the phase change required.

Figure 6
An example of immersive liquid cooling – tub – two-phase

Life Is On   Schneider Electric

11 attributes to consider

Understanding the benefits and drawbacks of the five major liquid cooling methods can help data center owners select an appropriate solution for their data centers. In this section, we describe 11 important attributes to consider when deciding on the best liquid cooling method for a particular business need.

1. Capital cost

When evaluating liquid cooling, the cost of the whole facility and IT must be considered. When a facility is greenfield and optimized around liquid cooling, leveraging warm water and direct rejection via fluid coolers, capex savings can be achieved over air cooling. If liquid-cooled IT is installed into an existing air-cooled facility, the retrofit cost can be higher. However, in cases where the data center has stranded power and space capacity[1], liquid-cooling could free up that stranded power and space.

- **Direct to chip:** 50-80% of IT heat capture can happen via liquid cooling. There is an increased cost to bring water to each rack and distribute it to each server, but this is offset by a reduction in traditional chillers, CRAHs, and the supporting power system equipment including transformers and switchgear.
- **Immersive:** Over 95% of the heat is removed via liquid, allowing for a drastic reduction in traditional cooling systems. The IT will see an increase in cost due to the fluids, especially engineered fluids, so it's important to understand that tradeoff depending on the immersed technology deployed.

2. Energy cost

Liquid cooling methods have long been known to provide excellent energy efficiency when compared to air cooling. Hyperscale operators have been able to achieve excellent PUE for air-cooled data centers, but this is typically done in favorable climates along with significant amounts of water usage. Another consideration for liquid cooling is the reduction in IT fan energy. This can be a 4-15% savings, which might lead to worse PUEs despite the fact that the overall electric bill decreases. Immersive has a slight advantage over direct to chip because all IT fans can be eliminated. In addition, with immersive, you need less air cooling and thus less CRAH fans.

Liquid cooling can leverage 45°C/113°F water for cooling, allowing for compressor-less cooling, most of the year in many climates. For edge applications where air-cooled DX systems are generally used, the energy savings for liquid cooling is more dramatic.

3. Serviceability

Data center operators are very familiar with air-cooling systems as it has been used for decades, however liquid cooling is new for most of the operators. Although the facility staff can benefit from a reduction in power and cooling equipment to service and maintain, the IT staff must implement new procedures to maintain the IT equipment.

- **Direct to chip:** This is similar to an air-cooled server as most components are accessible in the same manner. A key to making these serviceable is having dripless connectors to ensure the servers can be racked out and worked on like traditional air-cooled servers.

---

[1] Stranded power and space occur when the cooling system reaches 100% load but the UPS load is much lower than 100% and the space is also below 100% utilization.

Life Is On   Schneider Electric

- **Immersive:** This requires new procedures and sometimes new equipment. Oil baths are the most problematic due to the difficulty of containing the oil as IT equipment is removed and worked on. Engineered fluids is less difficult than oil, but care must be taken to ensure fluid cleanliness and to minimize loss / evaporation due to the higher fluid's cost. Chassis-based immersion cooling solutions aim to deliver liquid cooling in familiar form factors and service procedures like those used with standard air-cooled technology. Simple service tasks can be done while the fluid is still in the chassis. Major work requires that you pump down the fluid. Considerations here are around how easy is it to deploy, operate, and maintain the IT equipment. Note that IT equipment reliability can increase with this method because IT components are subjected to a more benign environment.

4. Rack density / compaction

Achieving average rack densities above 20 kW/rack is possible with air cooling, but significant engineering and cost is required to do so. Both direct to chip and immersive can easily handle 20 kW/rack and have the capability to go over 100 kW/rack. Immersive can achieve higher densities since no air movement within the IT is needed.

Both direct to chip and immersive can provide significant compaction over air cooling. Direct to chip and chassis-based immersion can be stacked vertically in a traditional rack form factor.

5. Water usage

Many air-cooled data centers rely on evaporative cooling with cooling towers to achieve low PUEs. This consumes high volumes of water. In many parts of the world this has become problematic. Since liquid cooling can use warm water – 45°C/113°F and sometimes higher, evaporative cooling can be eliminated or drastically reduced, while still achieving high efficiency.

Since immersive liquid cooling removes a greater percentage of the heat compared to direct to chip, this method has the advantage of less water usage in the data center.

6. Harsh environment

Immersive liquid cooling does not require any airflow and is sealed from the outside elements and it provides a capability to be placed almost anywhere. This is an advantage over air cooling and will likely drive immersive liquid-cooled IT at the edge in harsh environments. This is discussed as a key reason for adoption of liquid cooling in White Paper 279, *Five Reasons to Adopt Liquid Cooling*.

7. Fan noise / air movement

Anyone who has been inside an operational data center understands the noise of both the IT and the CRAHs. For some IT applications like occupied office spaces or clean rooms, quite operation and the elimination of air movement is an important attribute. Direct to chip requires much less airflow, lowering IT and CRAH fan speed and therefore noise. Immersive liquid cooling requires no fans, allowing for near silent operation indoors with only pumps needed to move the dielectric.

8. Room layout

Because immersive liquid cooling does not require airflow through the IT equipment, it brings greater flexibility within the data center white space as well as at the edge. Hot / cold aisle arrangement is no longer required. Back-to-back rows are possible.

This allows IT to be placed in locations where air cooling might not be feasible. This can be an important benefit in space-constrained facilities.

Direct to chip still requires a traditional layout since fans remain a part of the IT. Additionally, anything in a chassis form factor more easily fits into an existing data center layout, which might be advantageous in some cases.

9. Ability to retrofit IT equipment

From a manufacturing standpoint, IT manufacturers interested in creating liquid-cooled servers are tasked with varying degrees of design work. Direct to chip is ideal for adapting existing air-cooled servers to remove heat from the chips. Minor alterations to the IT are required. This allows the existing supply chain to remain virtually unchanged, with just the addition of the cold plates and tubing. However, designing an immersive liquid-cooled server from the ground up requires a lengthier product development project. None the less, this type of server affords IT equipment designers many more degrees of freedom because they aren't constrained by the limitations of air-cooled components.

From a field retrofit standpoint, immersive liquid cooling requires a bath or a new chassis for the IT, so conversion is costlier. If chassis-based immersion cooling is brought into an existing data center, it can operate alongside direct to chip and air-cooled systems very easily. Immersive liquid-cooled IT is not yet readily available for many configurations.

10. Scalability

Both direct to chip liquid cooling and IT chassis-based immersion liquid cooling have the ability to scale in smaller increments. Tub-based immersion liquid cooling requires the deployment of the entire tub and fluid, although IT can be deployed incrementally within the tub. Another consideration is understanding single points of failure in the overall design.

11. Fluid tradeoffs

The type of fluids used in a liquid cooling method is an important consideration in determining its applicability to a deployment. These are the fluids that remove the heat directly from the IT. The three main categories are:

- Water based
- Hydrocarbon based oils (dielectric)
- Engineered fluids (dielectric)

Characteristics among these vary, including thermal performance, cost, safety, material compatibility, lifespan, maintainability, and sustainability. **Table 1** illustrates which fluids are appropriate for the 5 liquid cooling approaches described in this paper.

Table 1
Fluid compatibility with liquid cooling architectures

Liquid cooling method	Water based	Oil based	Engineered fluids
Direct to chip (single-phase)	Yes	No	Yes
Direct to chip (two-phase)	No	No	Yes
Immersive – chassis (single-phase)	No	Yes	Yes
Immersive – tub (single-phase)	No	Yes	Yes
Immersive – tub (two-phase)	No	No	Yes

Material compatibility: In any water system, it's important to minimize corrosion and maintain water quality. Some materials are incompatible and can lead to early failures. Condenser water systems have been designed and operated for decades, so these heat rejection systems are well known. For direct to chip methods, filtration is important as many cold plates have micro channels that can get clogged due to poor water quality. For immersive methods, compatibility with IT components is important. Paper labels on boards coming off, and plasticizers leeching out of cabling have been issues with some oils and engineered fluids.

Cost: For dielectrics, oil is much cheaper than engineered fluids. The volume of dielectric needed will vary by technology providers.

Life: Oil tends to have a shorter life than engineered fluids. How often a fluid needs to be changed will have an overall effect on TCO.

Safety: Flash point, fire point, autoignition point, and toxicity are important considerations, especially when considering insurability of the facility. For example, mineral oils are flammable and require safety precautions to prevent fires.

Environmental: Ozone depletion potential (ODP) and global warming potential (GWP) are two main factors to consider. These values cannot be viewed by themselves, as the method used, and rate of vaporization affect release to the environment. Sealed vs open system, recyclability, etc. all play a role.

Conclusion

While data centers and edge environments today are primarily air cooled, we are seeing a growing interest and value in adoption of liquid cooling. The applications of cloud, IoT, AI, and edge are driving the continuous increase of chip and rack power density. There's also a continued focus on energy efficiency and cost. For many applications, liquid cooling is the optimal cooling solution. Direct to chip and immersive liquid cooling are two liquid cooling categories which have benefits for data center owners compared to air cooling. In this paper, we explain the differences between these methods.

- For retrofit sites, rack-based solutions like direct to chip and chassis immersive liquid cooling provide the easiest transition.
- For new sites, and those in harsh environments, immersive liquid cooling is a better solution because it can capture all the heat and isolate the IT from the surrounding air.

Further efforts are necessary for liquid cooling to have broader adoption in the data center industry, but we believe this technology will have a place in data centers and edge applications in the coming years.

✎ About the authors

Tony Day is a Director of Business Development within the IT Division's Office of the CTO at Schneider Electric. He is responsible for working with clients to develop integrated data center design solutions. He is a Chartered Architect and mechanical engineer with experience in both industrial engineering and construction industries. Prior to joining Schneider Electric, he worked both in professional private practice and within commercial organizations with a focus on construction of highly serviced environments including manufacturing, telecommunications, computer rooms, financial trading floors and data centers.

Paul Lin is a Research Director at Schneider Electric's Data Center Science Center. He is responsible for data center design and operation research and consults with clients on risk assessment and design practices to optimize the availability and efficiency of their data center environment. Before joining Schneider Electric, Paul worked as the R&D Project Leader in LG Electronics for several years. He is now designated as a "Data Center Certified Associate", an internationally recognized validation of the knowledge and skills required for a data center professional. He is also a registered HVAC professional engineer. Paul holds a master's degree in mechanical engineering from Jilin University with a background in HVAC and Thermodynamic Engineering.

Robert Bunger is a Project Director within the CTO office at Schneider Electric. In his 21 years at Schneider Electric, Robert has held management positions in customer service, technical sales, offer management, business development & industry associations. While with APC / Schneider Electric, he has lived and worked in the US, Europe, and China.   Prior to joining APC, he was a commissioned officer in the US Navy Submarine force. Robert has a BS in Computer Science from the US Naval Academy and MS EE from Rensselaer Polytechnic Institute.

RATE THIS PAPER ★★★★★

Life Is On | Schneider Electric

Resources

Five Reasons to Adopt Liquid Cooling
White Paper 279

**Browse all
white papers**
whitepapers.apc.com

**Browse all
TradeOff Tools™**
tools.apc.com

Contact us

For feedback and comments about the content of this white paper:

Data Center Science Center
dcsc@schneider-electric.com

If you are a customer and have questions specific to your data center project:

Contact your Schneider Electric representative at
www.apc.com/support/contact/index.cfm

Liquid Cooling Technologies for Data Centers and Edge Applications

Life Is On | Schneider Electric

Appendix

Table A1
Comparisons between air and water

The thermal conductivity of air, in comparison to liquids like water, is not impressive. In fact, air is a good insulator in a non-convective environment. Water has a higher density and greater thermal capacity per pumped unit volume. **Table A1** provides the thermal conductivity and heat capacity comparison between air and water, which shows that water is a better conductor of heat by a factor of more than 23 compared with air, and can hold far more heat than air, around 3,243 times more by volume.

Characteristics	Air	Water
Thermal conductivity[2] under 25°C (W/(m·K))	0.026	0.6089
Density (kg/m^3)	1.29 (1atm, 0°C)	1,000 (4°C)
Specific heat capacity - C_p (kJ/(kg·K))	1.004	4.2
Thermal capacity per unit volume[3] (kJ/(m^3·K))	1.30	4,200
Comparison	Benchmark	3,243

Non-pure water is electrically conductive, so it cannot be used to directly cool IT components. Alternatively, electrically non-conductive dielectric materials such as mineral oils and engineered dielectric fluid coolants, can be in direct contact with operating electrical components and provide excellent heat removal capability. As an example, oil has been used in electrical transformers for many years and allows for effective heat transfer from the internal coil and core to the outer casing.

The following heat transfer formula describes how heat is removed from an object, like the surface of a CPU or GPU:

$$Q = hA(T_s - T_{coolant})$$

- h is the heat transfer coefficient
- A is the available heat dissipation surface area
- Ts is the temperature of surface dissipating heat
- $T_{coolant}$ is the temperature of the coolant used for cooling

As related to this formula, a heat sink's purpose is to increase the heat dissipation area. The heat transfer coefficient of air under forced convection is around 100 W/(m^2·K), and the value for water is around 3,000, which means that compared with air, water can transfer much more heat by a factor of 30 times. We also know that water has a much higher thermal capacity than air, which also means when we transfer the same amount of heat from IT devices, the flowrate for liquid is much lower than air, which can save a significant amount of energy.

---

[2] https://en.wikipedia.org/wiki/Thermal_conductivity

[3] Thermal capacity per unit volume equals density times specific heat capacity.

Life Is On | Schneider Electric