

Part 2: ITE Thermal Design

Changing Landscape Of Data Centers

BY DONALD L. BEATY, P.E., FELLOW ASHRAE; DAVID QUIRK, P.E., MEMBER ASHRAE; JEFF JAWORSKI

The first modern computers using integrated circuits were simple machines, and air-cooling was sufficient. As compute density increased, heat dissipation grew to be a concern for reliability and functionality, and air-cooling gave way to liquid-cooling. The invention of complementary metal-oxide semiconductors (CMOS) dramatically reduced heat dissipation, and the need for liquid-cooling all but vanished, for a time. But as the limits of compute performance and density are probed, designers continue to study thermal design and the art of thermal management to provide innovative and efficient ways to keep computers cool.

The takeaway from this history lesson should not be that thermal management is now well understood or time-tested, but rather that disruptive change has happened before, and may very likely happen again.

Consider *Figure 1*. If a data center had been constructed in 1985 for the current technology, five years later that data center would have been grossly inadequate to handle the rapidly increasing heat production by the then-current technology. On the other hand, if a facility was built in 1990 expecting that the rate of production would only continue its 10-year trend of growth, the facility would have been grossly

overcooled by 1995, after the introduction of CMOS-based processors.

Therefore, when designing the HVAC for information technology equipment (ITE), it is prudent to consider designs and techniques to reduce the risk of obsolescence due to industry change of the transistors and related parts that make up servers. To do this properly requires at least a working knowledge of thermal management to understand how the IT equipment can affect the needs of the facility on the whole.

Donald L. Beaty, P.E., is president, David Quirk, P.E., is vice president, and Jeff Jaworski is an engineer at DLB Associates Consulting Engineers, in Eatontown, N.J.

Thermal Considerations for ITE Components

Thermal management is the process by which a piece of ITE regulates itself to enable the delivery of reliable, consistent performance. Throughout the past 20 years (the CMOS era), the science of thermal management of the data center has evolved quite rapidly. This was due to the continuous increases in compute density and overall heat dissipation. To understand the whole picture, it's best to start by understanding the component level of the ITE and abstract the implications from there.

Most ITE, such as servers and storage arrays, consists of similar components used to achieve different functions. These components typically include CPUs, memory, support logic, and storage. Each of these component groups has different behaviors as they relate to thermal design, and thus have different thermal specifications. It is these sets of thermal specifications that must be adhered to by the overall piece of ITE, which may greatly affect its design.

Each component may have different thermal specifications, but are typically characterized based on their:

- Reliability limit;
- Functional limit; and
- Damage limit.

The reliability limit is the temperature at which the component will perform reliably in the long term, and thus is a target maximum for the component during normal operation. Short durations at or just above the reliability limit should not significantly impact reliability.

The functional limit is the temperature at which the component may cease to operate correctly, so exceeding this temperature is never recommended and can have unexpected results. Finally, the damage limit is the temperature at which permanent damage may occur. Therefore, never exceeding this limit is of paramount importance.

Understanding the limits of each component is a critical ingredient that influences the overall design of the ITE, since the overall reliability of the whole is at best only as good as the least reliable component. For example, it would be an ineffective ITE design if, during normal operation, a single component would rapidly approach its damage limit, while all of the other components were well within their reliability limits.

Each of the three thermal limits is typically provided for most processors, whereas most memory specifies

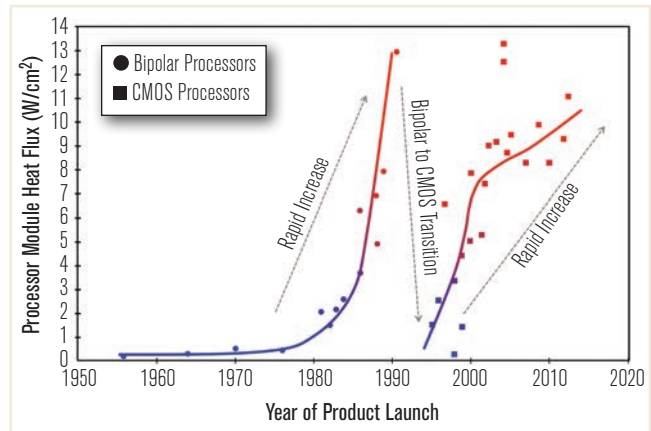


FIGURE 1 Evolution of processor module level heat flux in high-end servers.

only a single temperature that is both the reliability and functional limit. Interestingly, for some DRAM memory, this limit may increase while functioning in an extended temperature range mode, which doubles the refresh rate to essentially halve the time that the data must be stored without corruption. This consumes some additional energy, but may result in an overall decrease in energy by slowing a fan or allowing more efficient heat removal.

To design the ITE to efficiently address each of the components' specifications, we must understand what drives the temperature of each component. Regardless of the cooling medium, each component is affected by the inlet temperature of the cooling medium, preheating of the medium from the inlet to the component, and the heat generated by the component itself.

Since the preheating is due to the heat produced by other components in the ITE, the overall layout of components within the ITE is critical. However, there will be trade-offs associated with any layout optimization. For example, if the greatest sources of heat (CPU, memory) are distributed across the front of the inlet so they don't have downstream impacts on each other, it may mean decreased available depth for full-sized PCIe cards in certain form factors. Figure 2 demonstrates this by comparing two possible layouts.

Thermal limits and layout trade-offs are only two components of the overall design equation. In addition, each component may have various ways it can report its state, and also regulate its state. In particular, CPUs are the most tightly controlled components within the design.

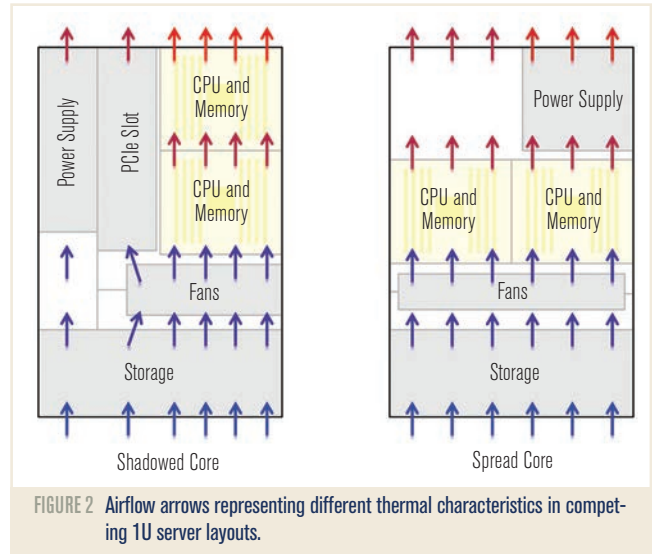
Whereas high performance and energy efficiency may sound like opposite needs, there have been increasing pressures to do both. This is often accomplished through the use of processor states, performance states, and thermal states.

Processor states are essentially forms of sleep that the CPU can dip into for various durations to save energy and consequently regulate the amount of heat produced. While not sleeping, the CPU can also adjust its performance state, which typically adjusts voltage and frequency to regulate its throughput and heat. Finally, thermal states are essentially a time-out for the CPU, completely halting operation for brief periods of time only for the purpose of regulating the heat generated.

Thermal Solutions for ITE Components

Armed with a basic understanding of just some of the multitude of variables affecting the heating of ITE components, we can begin to discuss how they can be cooled.

At their essence, all cooling techniques are about the



removal of heat. In the simplest case, this may be just cool air absorbing heat from the surface of a component, and in the most complicated it may be a multistage removal process involving heat sinks, engineered fluids, and chilled water directly in the ITE chassis. But

Advertisement formerly in this space.

each method transfers heat from the ITE components into a medium that is then directed away from the IT equipment.

Air is the most obvious cooling medium since it is all around us. Thus, air-cooling has always been an attractive, affordable, and comparatively simple option. Take components, add a fan to move some air, and start cooling.

However, air does not have a very high specific heat, nor is it very dense. This, combined with air having a low thermal conductivity, means a lot of air must pass over a large surface area to remove a moderate amount of heat at a reasonable rate. To enable this, thermally conductive aluminum or copper heat sinks may be used to transfer the heat away from a hot component across a large surface area so it can be absorbed by passing air.

To complicate matters further, the amount of energy required to power the fan(s) can grow quite rapidly when the volume of air must be large or if it must overcome a large static pressure. According to the fan laws that describe fan performance, airflow is proportional to fan speed, and power is proportional to fan speed cubed. Therefore, to increase the airflow of a given fan by two times might take an eight times increase in power.

When the volume of air required and energy consumed to move it are too much to overcome, liquid cooling offers solutions. In contrast to air, liquids such as water are significantly denser and have a higher specific heat, which combine to produce a cooling capacity of around 4,000 times that of air by mass. In practical terms, it means much less water is needed to move the same amount of heat than air.

Unsurprisingly, liquid cooling introduces its own set of unique challenges, such as adding additional weight and having the potential to leak. Each of the different approaches to liquid cooling presents further challenges and solutions.

From a volumetric standpoint, the solution with the least liquid is a closed cooling loop within the ITE. In this scenario, the liquid is used to transport heat away from the components before being rejected to air in a water-to-air heat exchanger. This allows for efficient cooling of the components while not necessarily requiring design changes from a rack or facility perspective.

Advertisement formerly in this space.

Adding more liquid to the mix, another solution may use a cooling loop within the ITE that connects to an external technology cooling system (TCS) loop. Typically this loop is separate from the facility water, to which it transfers heat in a liquid-to-liquid heat exchanger. This is done because facility water and TCS water have different characteristics and needs, such as temperature range, water treatment, and pumping requirements.

A third option that incorporates even more liquid into the equation is immersion cooling, in which electronics are completely immersed in a dielectric fluid. While these arguably have the highest potential for a huge mess, they guarantee nearly 100% transfer of heat from the ITE to the cooling fluid as opposed to the facility air. Even more options exist that are hybrids or combinations of these techniques.

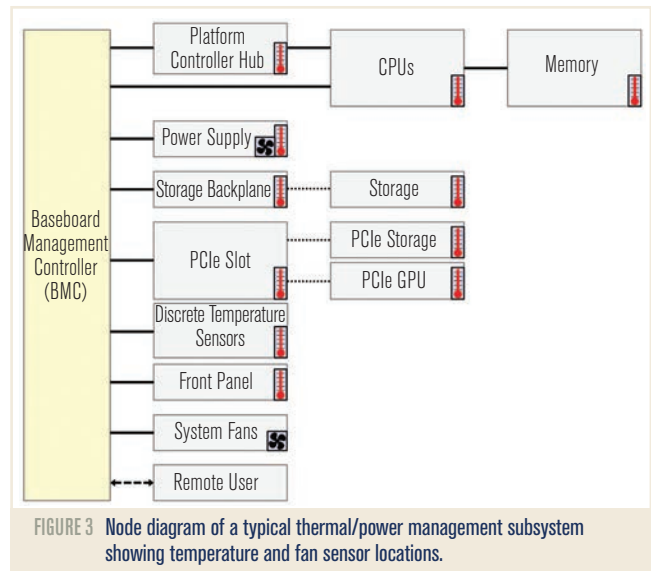
ITE Thermal Management Overview

Thermal management is the process by which a piece of ITE regulates itself, to enable the delivery of reliable, consistent performance. This means the thermal management of the system must optimize performance, efficiency, and even acoustics in the process of keeping the component temperatures within their limits.

The general thermal management process begins by collecting sensor data from throughout the ITE. This may include temperature, power, fan speed, air pressure, and other system activities. This data comes from a variety of sensors that are included in CPUs, RAM, power supplies, hard drives, GPUs, PCIe cards, physical chassis (i.e., at the air inlet), etc., shown in *Figure 3*.

The baseboard management controller (BMC) collects this sensor data over standardized bus protocols, and uses algorithms to generate system responses in the form of fan speeds, CPU performance states, and other power settings. Based on the components within the ITE, these algorithms can be quite different, as the primary heat drivers may vary greatly, as can the locations and availability of sensors. These algorithms are sometimes further affected by boot options or system settings that allow the customization of the ITE to its specific application and environmental conditions.

Although it can be defined somewhat easily, thermal management turns out to be an incredibly complex optimization problem in that it isn't just a single optimization. Rather it is a continuous optimization across a wide



spectrum of conditions that can include fan failures, extreme CPU use, swappable components, and various environmental conditions ranging from overcooled to undercooled to facility cooling failure.

Closing Comments

From the smallest component to the largest ITE chassis, thermal design plays an incredibly important role and influences design throughout the data center industry. It is not a straightforward problem of keeping equipment cool, but rather an optimization of keeping equipment within acceptable temperatures while enabling the right blend of performance and efficiency.

For the HVAC design engineer, the most important focus is the providing the right “entering” thermal conditions (air or liquid) to the ITE. For the ITE manufacturers, it is an optimization problem involving a lot of different pieces of equipment and trade-offs.

As the data center industry rapidly evolves based on changing need and advancing technology, thermal design must constantly adapt to these changing conditions. In fact, across the last 50 years, the biggest constant within the industry has been change itself.

Will heat dissipation continue to grow upward? Or will another new technology revolutionize the CPU or memory, dramatically decreasing heat dissipation?

For the HVAC design engineer, these are challenging questions to answer. So regardless of the speculation, one safe bet is to assume that some form of scalability is necessary in the HVAC system design. ■