

COMPUTING

Cooling Considerations for High-End Server Blades

Application Note

Executive Summary

AdvancedTCA® (ATCA®) technology has become the industry standard for defining and building communications equipment. Different processor architectures have been implemented on ATCA compliant blades. The Intel® Xeon® processor family is one of the most popular architectures successfully used on ATCA server blades today. With a huge software base, a large community of supporting developers and ever growing processing capabilities, the architecture plays a major role in applications such as DPI, NFV/SDN or application virtualization on application servers and gateways. Along with the increasing performance footprint goes the need for higher power provisioning and tighter thermal specifications.

As a result, cooling becomes more challenging for board and system designers, particularly when products need to comply with the more stringent physical requirements of telecommunications equipment. When done right, designs can make use of the full performance capabilities.

With Artesyn's ATCA-7480, the latest ATCA server and packet processing board based on Intel architecture, Artesyn once again proves that a well-engineered server blade can exploit the available processing performance even under stringent thermal conditions as defined by NEBS.

PERFORMANCE ACHIEVEMENTS ON ATCA

The Intel® Xeon® processor family provides server capabilities selectable from different performance classes. Dual socket CPU designs with SMP architecture can be most efficiently implemented on the ATCA form factor. Due to continuous architecture enhancements the performance gains over the recent years are significant.

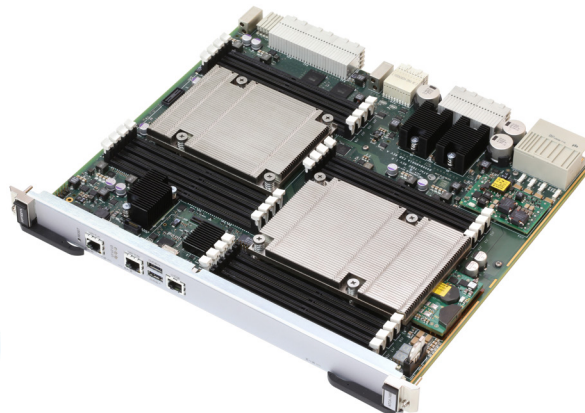
An ATCA server blade introduced in 2010 typically had a capacity of 64 to 96GB main memory, while today's products, such as Artesyn's ATCA-7480, can support up to 512GB of memory. The 2010 blade had 6 cores per CPU. With the latest Intel Xeon EP2600 v3 family, the usable core count on an ATCA blade can be as high as 14 cores per processor. Even more cores can be expected with future processor generations. The size of the main CPU cache grows along with the number of available cores.

Intel has also enhanced the CPU instruction set and introduced application accelerating functions such as encryption for security or 256-bit integer vector instructions for image processing and signal analysis.

The throughput of data paths inside CPUs, between CPUs and to the main memory were greatly enhanced by using wider or faster-clocked interconnects. Main memory technology went from DDR2 and DDR3 to DDR4, which introduced lower power consumption at higher data rates. Also, the I/O connectivity was extended by the use of PCIe Gen 3 and integrating the host controllers directly into the CPUs, as it had been done before with the memory controllers.

The only performance factor that remains relatively stable over the various processor generations is the average clock frequency. However, performance optimization is also possible by clocking individual processor cores at higher frequencies, if the application can benefit from this approach.

All these enhancements are available on ATCA server blades and have helped to increase performance per ATCA slot and optimize energy efficiency. As a result, CapEx and OpEx measured per subscriber or per network packet can be greatly improved.



POWER AND THERMAL REQUIREMENTS

Larger cache sizes, more parallel units switching, higher clocking rates; all these contribute to the overall power consumption. As a result of the enhancements, power consumption of the processor and subsystems increases. In order to compensate for the increased power usage, processor manufacturers introduce new technologies that allow for the shrinking size of silicon dies. Reducing the inner geometries of microprocessors, such as smaller transistor gates and thinner signal traces, is an inevitable method for reducing switching and leakage currents. Also more sophisticated techniques, such as reducing or disabling clocks and supply voltages for idle circuitries, have been invented.

Together with lowering of the general operating voltages applied to silicon, these techniques allow for significant reduction of power losses. Another positive effect of die shrinking is the improved die yield which helps reduce the cost of the silicon. Lower cost and power are mostly exploited with new classes of processors intended for mobile and appliance applications. For processors targeted for high performance servers, the positive effects are eaten up by adding more capabilities (such as the ones listed above) with the next processor generation. In fact the average power consumption of the Intel Xeon processor family has been increased over the past years.

Power efficiency could have gotten worse; however the performance has improved much faster than the average power consumption in the same time period, so that power efficiency as a function of performance versus power has risen.

The power increases in the last four years are in the range of 15 to 45W per Intel® Xeon® processor. This doesn't sound much, but if a given cooling environment did not improve in the same timeframe, the increases can be significant.

For example, consider a scenario where compute blades are replaced in the field without upgrading the enclosure's cooling subsystem. Insufficient cooling could then result in lower achievable CPU performance. If the desire is to also exploit higher clocked CPUs, the processor wattages can grow even more. This puts a lot of pressure on board designers.

Another concern for designers can be the silicon's thermal specifications. Silicon manufacturers typically define the maximum temperature a silicon device is allowed to operate at under normal conditions. This is either specified for the surface of the device package (T_{Case}) or for the silicon substrate ($T_{Junction}$). Thermal specification of silicon devices with high power dissipation such as CPUs can be tight and designers are forced to pay special attention when designing the processor cooling solution. On ATCA, where active CPU cooling is not an option due to reliability concerns, the cooling solution becomes a major design challenge.

DOING IT WRONG?

Board real estate and cooling capabilities of the target enclosures set the stage for board designers. Typically, they have to find the right balance between board functionality and achievable performance. The ideal processing board has hundreds of processors cores, runs at 5 GHz, features multiple disk drives, provides lots of default I/O and still has sufficient real-estate for modular I/O extensions.

In reality, if performance is maxed out, board functionality and configuration flexibility can be limited. By the same token, adding a high degree of configuration flexibility — such as mezzanine cards or storage devices — limits the size of the cooling solution on remaining board real estate. This results in lower achievable performance.


Both the high performance board and the multifunctional board have their advantages, but they can't co-exist satisfactorily on the same design. Board designs that ignore this fact will create problems when they are integrated into a given shelf solution. The promised performance range may then not be fully achievable.

Processors running at full pace that are exposed to higher ambient temperature and higher software load will eventually overheat, which is not acceptable. It questions the stability of operation, reduces the component's life time, exceeds air exhaust temperature above the allowed limits, and in the extreme, can cause serious damage.

In order to prevent this, Intel CPUs typically contain a mechanism called clock throttling. The device measures the on-die temperature and, when certain limits are reached, periodically gates the processor clock frequency for some time. This allows the device to dissipate less power and stay within the thermal limits. There is a further protection available that ultimately terminates operation if even the temporary clock gating doesn't bring down the temperature.

While these are useful protection mechanisms, they are counter-productive to the overall system performance. The developer could proactively reduce clock frequency or use fewer of the available cores. The effect remains the same; the performance provided with the processor variant can only be partially exploited by the application.

Board products that pretend to have both a high degree of flexibility and promise high performance should be treated with caution. Integrators should carefully determine whether the promised performance is achievable in the target shelf environment under all circumstances, such as maximum ambient temperature or high software load. If processing performance is not crucial, it may be the right product. If excellent performance is paramount, it is probably the wrong product decision.



DOING IT RIGHT!

In recent years, new shelf products have become available that provide very strong cooling capabilities in order to support high performance applications. Such shelves are necessary for supporting blades with 400W or even higher power dissipation and enable the ultimate performance experience. Having said this, not every new compute blade that is delivered into communications markets will be integrated into such high performance shelf. In fact, there are thousands of installed shelves that have served at key locations in communications networks for many years. From an economic standpoint, it is understandable that service providers replace equipment as needed and extend their infrastructure capabilities with the growing demand for bandwidth and service capabilities.

Upgrading individual payload blades is a much leaner approach compared to a complete shelf replacement. Shelves are therefore often kept in service for many years; they likely don't have the cooling capabilities the new shelf generation is able to provide. Such shelves often comply with the CP-TA B.4 cooling class (now managed by PICMG®), which defines airflows of up to 40 CFM per ATCA slot. The Artesyn Centellis® 4440 is an example of this product class. As such shelves are still commonly in use, it is important that the latest server blades provide satisfactory performance in such installations.

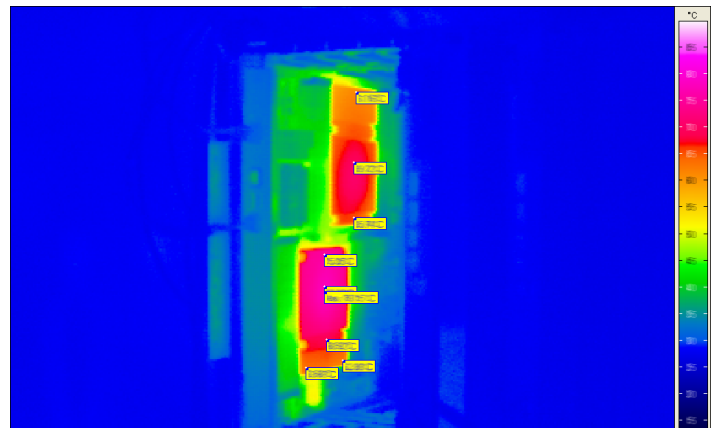
At Artesyn, we carefully define and design ATCA products for communication applications. A key aspect is the design for NEBS Level 3 compliance. The NEBS requirements cover different aspects of the design such as safety, EMC compliance, earthquake or thermal requirements. NEBS requirements define a maximum ambient air temperature of 40 °C during normal operation and up to 55 °C ambient air temperature during exceptional operation for a limited amount of time per year (such as during a loss of the room air conditioning). As CP-TA B.4 shelves are still commonly in use, it is a primary goal for us that our ATCA blade products fit in these environments while providing outstanding performance. It is also paramount that there is no degradation of performance across the entire NEBS L3 temperature range.

The Artesyn ATCA-7480 packet processing blade is based on the most recent Intel Xeon E5-2600 v3 generation. It can host two processors with extended temperature range capability and allow the blade to be fitted into a CP-TA B.4 compliant shelf under NEBS L3

conditions. With 12 cores per socket and up to 512GB of main memory, the blade is optimized for excellent performance.

Furthermore, high power CPUs (120W) are supported by the blade design. This means the blade can be integrated into high performance shelves such as the Artesyn Centellis 8000 family, and enables an outstanding level of data center class performance on ATCA. Processor derivatives with up to 14 cores per processor or up to 2.5 GHz clock frequency can also be supported.

Artesyn undertook extensive thermal design and simulation efforts to enable the performance envelope for the different cooling environments. The cooling solution of the ATCA-7480 supports the operation of the board in a maximum ambient air temperature of 55 °C as defined by NEBS, when installed in a CP-TA B.4 compliant shelf. The design goals have been proven by thermal qualification during design verification and testing. The selected thermal solution makes the product sufficiently robust in shelves with CP-TA B.4 air cooling without compromising the available compute performance. It also adds sufficient headroom for squeezing the ultimate performance out of the product when installed in a shelf with enhanced airflow.



WORLDWIDE OFFICES

Tempe, AZ U.S.A.	+1 888 412 7832	Hong Kong	+852 2176 3540
Paris, France	+33 1 60 92 31 20	Shanghai, China	+86 21 3395 0289
Munich, Germany	+49 89 9608 2430	Tokyo, Japan	+81 3 5403 2730
Tel Aviv, Israel	+972 9 9560361	Seoul, Korea	+82 2 3483 1500



www.artesyn.com