JUNIPer
NETWORKS

Engineering
Simplicity

DAY ONE GREEN: OPTIMIZED THERMAL DESIGN

Keeping the power consumption of high-performance, high-bandwidth networking equipment as low as possible is a critical design requirement of Juniper Networks.

By Attila Aranyosi, Rebecca Biswas, and Gautam Ganguly

Day One Green

Optimized Thermal Design

Keeping the power consumption of high-performance, high-bandwidth networking equipment as low as possible is a critical design requirement imposed by Juniper's customers (facility / data center owners) for reduced OpEx (operational expenses) which is another way to say a reduced carbon footprint. We design lower-power systems that can help our customers reach their network's Scope 3 carbon targets. That's because efficient thermal design of our high-bandwidth products contributes to reduced system power consumption and we optimize the thermal management solutions at each level: component, board, and system level.

To reduce power consumption at the component level, thermal engineers work closely with the ASIC team to evaluate different floorplan options and identify the best arrangement that meets both electrical and thermal requirements with the lowest possible heat flux levels while mitigating hot spots and reducing leakage currents. Figure 1 shows examples of a thermally inefficient and an optimized MCM (Multi-chip Module) floorplans.
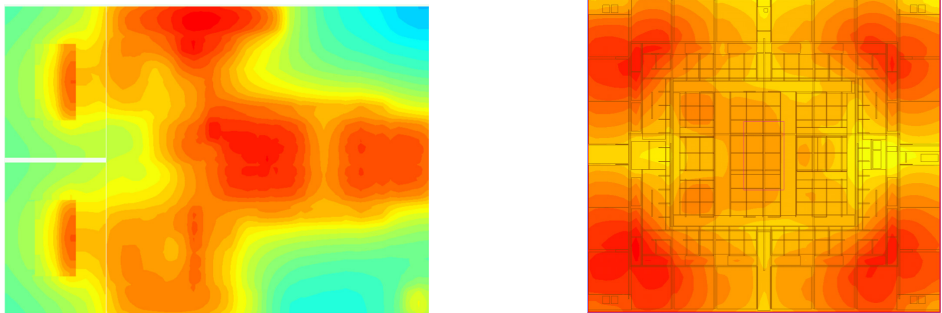

Figure 1   *A Thermally Inefficient MCM Floorplan Compared to an Optimized One*

ASIC power efficiency is continuously improving (See *Day One Green: Improving Network Power Consumption with ASIC Architecture and Technology*) by moving to new technology nodes but ASIC power density keeps increasing too, because the ASIC and system bandwidths advance faster than the efficiency improvements. To keep pace with these trends and to keep ASIC junction temperatures below their long-term reliability limits with the lowest possible power consumption of the cooling subsystem (in air-cooled systems, fans), Juniper uses lidless, a.k.a. bare die ASICs and MCMs. These eliminate a high thermal resistance element, the TIM1, or thermal interface material between the chip and the lid of lidded packages, from the conduction heat transfer path. Furthermore, Juniper uses the highest-performance TIM2s between the chips and their heat sinks to maximize cooling efficiency.

Besides the ASICs and MCMs, high-bandwidth 400G and 800G pluggable optical modules also consume significant power (between 12W and 25W per module), which in an 1RU, 36-port line card can translate into 900W total optics power, about 40% of the total line card power. We work closely with module vendors to influence the thermal design of modules so that the transceivers can be cooled with the least amount of energy. The main options to achieve thermally efficient module designs are:

- Optimized conduction paths from the main heat-dissipating components to the module case where heat is ultimately removed via integrated and riding heat sinks.

- Tight flatness specification for the top surface of the module housing to reduce the thermal contact resistance between the module and its riding heat sink

- And choosing DSPs (digital signal processors, the highest-power component of optical modules) that have efficient package thermal design, with low junction-to-case thermal resistance and higher junction temperature limit. Figure 2 depicts the surface temperature maps of thermally inefficient and optimized high-power optical modules.
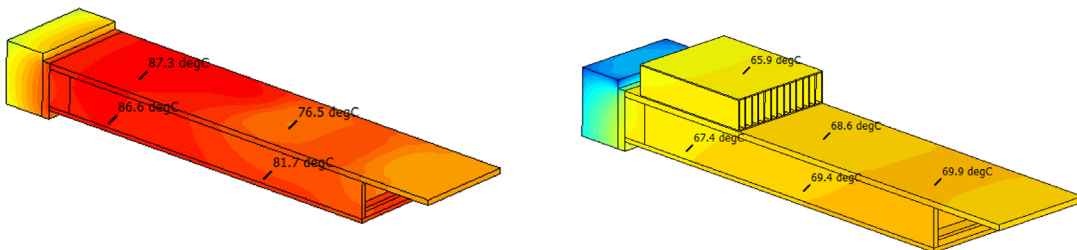

Figure 2   Surface Temperature Maps of Thermally Inefficient and Optimized High-power Optical Modules Under the Same Boundary Conditions

In many systems, retimers are needed to meet stringent signal integrity performance targets. However, retimers dissipate a significant amount of heat. To reduce system power consumption, flyover cables may be used to replace retimers although trade-offs between power consumption and cost should drive such decisions.

At the board level, Juniper supports HW and SI teams, and carries out thermal feasibility analyses to optimize the board layout and heat sinks and keep component temperatures below their respective long-term reliability limits. We reduce leakage power as much as possible while balancing component thermal margins to keep fan speeds at their lowest levels, further minimizing fan power consumption. Figure 3 shows a vapor chamber main heat sink with a secondary, floating heat sink, which thermally isolates lower-power components with lower temperature ratings from the high-power ASIC. We achieve efficient cooling of the DC-DC power converters (POLs) and a reduced amount of Joule heating in the printed circuit board (PCB) via efficient heat spreading in the power and ground planes.
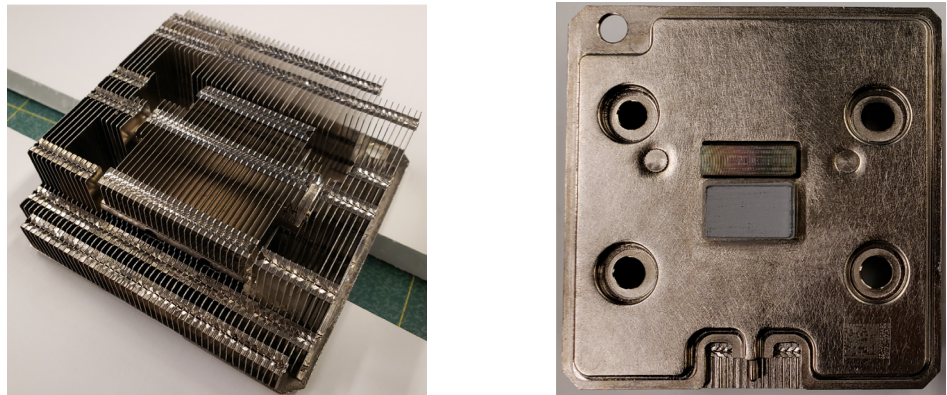

Figure 3   *Vapor Chamber Main Heat Sink with a Secondary, Floating Heat Sink*

At the system level, air-cooled equipment is still dominant in the networking industry, and here we select high-efficiency (50-55%) fans which operate in the high-efficiency range against the back pressure imposed by the system. Fan efficiency is defined as the ratio of pumping power (the product of air pressure and airflow rate) and electrical input power. Figure 4 illustrates fan efficiency and aero-performance (P-Q) curves. In the example shown in Figure 4, the maximum efficiency (~48%) is achieved at 105 CFM airflow rate and 5.5 in. w.g. pressure.

With the proper fans and under the worst-case thermal design conditions, power consumption of the cooling system in our equipment, with the fans running at full

speed, is typically 10% of the total system power. That's still a significant amount of energy, but with poor fan selection, this number can be twice as high.
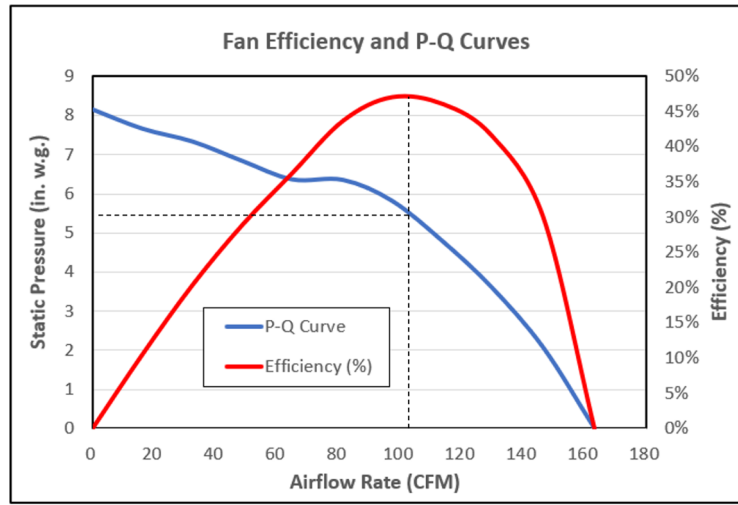

Figure 4   *Fan Efficiency and Aero Performance (P-Q) Curve*

As our systems, most of the time, operate under nominal/normal operating conditions (that is, the ambient conditions are significantly more benign than the worst-case condition), we use fan speed control to reduce power consumption. It can result in massive energy savings, as the power consumption of a fan varies with the cube of speed (rpm) ratio. For example, at 50% fan speed, the power consumption is only one-eighth of that at full speed.

Further significant energy savings can be achieved using liquid cooling. PUE (power usage effectiveness) is a ratio between total facility power and power consumed by the IT load. Efficient, liquid-cooled data centers are expected to go below a PUE of 1.10. Although liquid cooling has been used in high-performance computing for quite some time, it is still awaiting acceptance in the networking industry primarily due to reliability concerns. However, realizing that with current power dissipation and power density trends we are approaching the limits of air cooling, there is significant traction in the industry to introduce some form of liquid cooling (cold-plate based or immersion) in the very near future. The other main driver to do so is the potential for huge (50-80%) energy and space savings. In recent years, the Open Compute Project has made excellent progress in establishing a strong liquid cooling ecosystem with standardizations. Juniper Networks has developed several

liquid-cooled proof of concept systems (single and two-phase) and currently we have been evaluating multiple newer liquid-cooling technologies in preparation for their deployment.

Juniper's thermal team is continuously innovating to enhance our cooling efficiency at the component, board, and system level. Pay close attention to this space for more Day One Green additions.