Is TSUBAME-KFC a Game-changer?

Kirk W. Cameron, Ph.D.

 *Japan’s prototype marks a significant increase in infrastructure and system efficiencies*

Here’s the pitch: “We would like millions of dollars to build a supercomputer capable of calculating 150 trillion floating point operations per second (TFLOPS). Hundreds of scientists will use the system to investigate the causes of global warming, drugs that may cure cancer, and the origins of the universe. The machine will be built from the most advanced equipment available from NEC, Intel, NVIDIA, Mellanox, and other manufacturers. This machine will be a prelude to a larger system worth hundreds of millions of dollars a few years from now. Oh, <cough> and we plan to submerge the entire system in mineral oil.”

While this may seem contrived, at some point in the not-too-distant past, Japanese researchers proposed a similar scenario to develop a prototype system now known as the TSUBAME-KFC supercomputer. The system combines the computing power of two Intel Xeon ES-2620 processors with four NVIDIA Tesla K20X graphics processing engines per node. The resulting heterogeneous system is capable of more than 150 TFLOPS of computation running the LINPACK benchmark. This supercomputer ranks at #311 in the Top500 List of the world’s fastest supercomputers.

While all supercomputers like TSUBAME-KFC generate excessive amounts of heat, most others use a form of air cooling to reject the heat from the confines of the data center. Cold air is piped in, and hot air is piped out. Air cooling is commonly available and effective, but the infrastructure is somewhat inefficient and costly to operate.

What makes TSUBAME-KFC special is that all 2,720 CPU cores are submerged in a liquid, namely dielectric mineral oil. Though the oil itself is a blend marketed as

GreenDEF by Green Revolution Computing, this substance is not all that different (sans perfume) from the baby oil with which we are all familiar. The GreenDEF liquid can absorb heat up to 1200 times more efficiently than air. By constantly circulating the oil across a supercomputer composed from commodity system racks, components can achieve densities previously difficult without highly efficient, expensive air cooling.

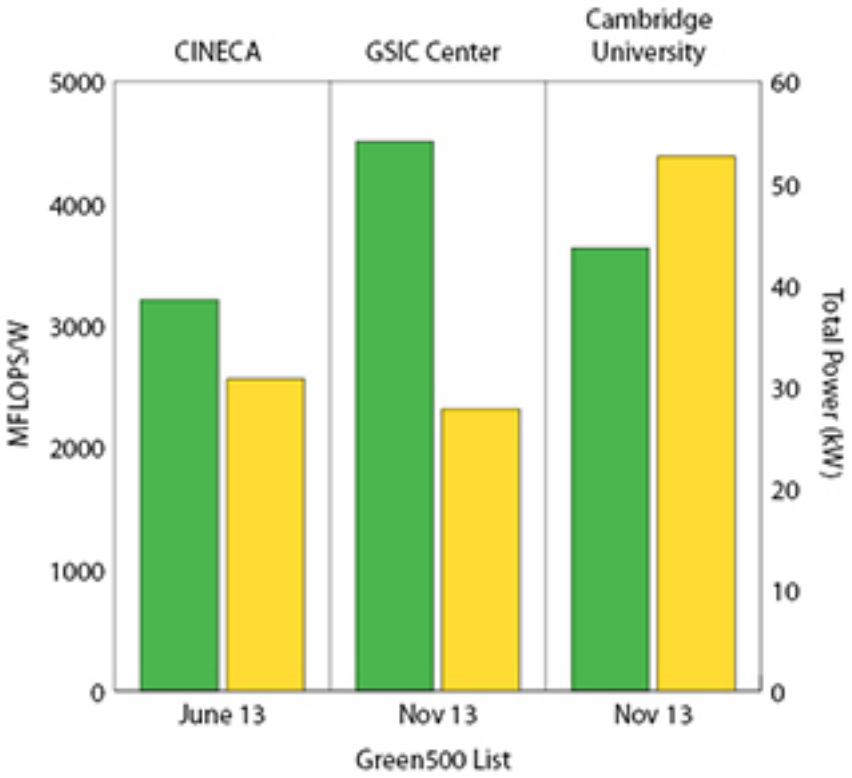
TSUBAME-KFC's power and cooling advantages are well-aligned with those of typical data center operators. Arguments for greening the planet notwithstanding, supercomputer and datacenter operators care less about saving the environment than they do about the number of machines they can support and keep up and running. Environmental initiatives are usually out of their control and typically driven more by the threat of government regulation than by altruism. Operators care less about daily operating costs than they do about supporting the mission of their organization. People may get praise for increasing the bottom line via savings, but while rising energy costs can be absorbed or passed on to clients, people get fired if their organization's mission is jeopardized by a lack of infrastructure support. The operator's mission is to provide value to their customers from consumers to scientists to software- or platform-as-a-service providers.

At a recent conference, a government agency supercomputer operator was asked how much power they would use in future systems. The answer: "All of it." Thus, the real gain for datacenter and supercomputer deployment from liquid cooling is the ability to repurpose reductions in total power from reduced air cooling. There is little disagreement in the broader community that the supercomputers of the future will be power capped by their surrounding infrastructure, whether it is a retrofit room or a new building with a dedicated power substation. Despite the cultural barriers to adoption of liquid cooling, the 1200x efficiency increases equate to more servers in less space using total power comparable to that of their air-cooled counterparts. More servers in the same power envelope means operators can deploy additional infrastructure to support consumers, scientists and businesses.

The TSUBAME-KFC prototype may be a game changer. In November 2013, the system became the first Japanese system to be awarded the top spot in the Green500 List of energy efficient supercomputers. TSUBAME-KFC was the first supercomputer to surpass 4,000 MFLOPS per watt efficiency. The system (see MFLOPS/watt data in Figure 1) is 25 percent more efficient than the #2 ranked system from Cambridge University on the same list; and 40 percent more efficient than the #1 ranked system from CINECA Eurotech Aurora from the list six months earlier.

It's no accident that Japan is now leading energy efficiency in supercomputing. The TSUBAME-KFC system is a prototype to test the feasibility of a larger liquid cooled deployment in a country reeling from systemic power grid limitations. After little more than two years, the Fukushima nuclear plant disaster that followed an earthquake and tsunami has heightened sensitivity and constraints in the island nation. For example, to reduce strain on the country's power infrastructure, 15 percent power reductions have been enforced upon consumers. This foundational shift in Japan's view of energy usage will likely drive innovations in energy efficiency

technologies that might not be readily adopted elsewhere.



CINECA
Eurotech Aurora HPC 10-20, Xeon E5-2687W 8C 3.100GHz, Infiniband QDR, NVIDIA K20

GSIC Center, Tokyo Institute of Technology
TSUBAME-KFC - LX 1U-4GPU/104Re-1G Cluster, Intel Xeon E5-2620v2 6C 2.100GHz, Infiniband FDR, NVIDIA K20x

Cambridge University
Wilkes - Dell T620 Cluster, Intel Xeon E5-2630v2 6C 2.600GHz, Infiniband FDR, NVIDIA K20

Source: Green500.org

Figure 1 The TSUBAME-KFC system could also mark a notable shift in Green500 List system trends. In the early years of the Green500, the most efficient systems were custom designs (e.g. IBM BlueGene) using many lower power processing cores to achieve extremely high performance numbers in a relatively flat and large power envelope. Of late, the Green500 has been dominated by high-density systems primarily using relatively high-power, highly-efficient (e.g. NVIDIA GPGPUs) commodity parts for a relatively small total power footprint (see total power in Figure 1). Furthermore, for the first time, the average power of Green500 List systems has decreased from the June 2013 List to the November 2013 List.

What does this all mean? The MFLOPS/watt metric of the Green500 rewards both machines that drastically increase MFLOPS (the numerator) and drastically decrease watts (the denominator). In early Green500 Lists, much of the gains came from large changes in MFLOPS with somewhat modest increases in power usage for a large power envelope. This biased the top entries toward large, top-performing systems. In more recent lists exemplified by the TSUBAME-KFC system, the bias seems to be shifting towards smaller commodity-based systems within a smaller

power envelope that make progress in both performance and power efficiency.

More data is needed to ascertain whether this is a significant shift or just an anomaly in the current list. While offload engines (e.g. NVIDIA GPGPUs) provided significant efficiency gains in both performance and watts, they were common initially only to the top performing systems in the Top500 List. These same GPGPUs are now used in a clear majority of systems in the Green500 and Top500 Lists. The proliferation of GPGPU efficiencies across more entries offers some explanation for the decrease in average power while performance continues to increase. But, these same efficiency gains could be fleeting, and only time will tell if the trends can continue.

So, the TSUBAME-KFC marks a significant increase in infrastructure and system efficiencies. Liquid cooled systems have come and gone since the early Cray supercomputers. Such solutions are still members of an exotic class of machines. However, a shift in design mentality from building the biggest systems to building the biggest systems possible within a power envelope may make liquid cooling a more palatable option to operators to better use the watts they are given.

Kirk W. Cameron is a co-founder of the Green500 List and a professor of computer science in the College of Engineering at Virginia Tech. He may be reached at editor@ScientificComputing.com.