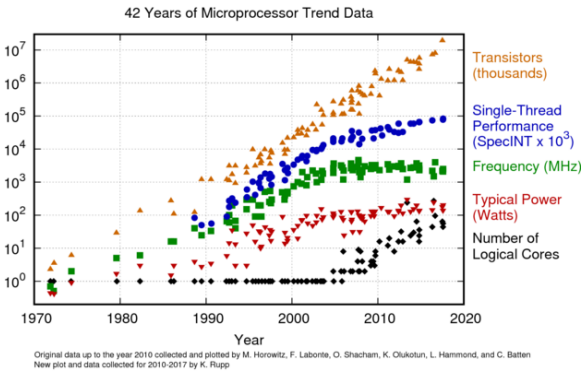
Massively Parallel Liquid Cooling for AI Supercomputing

Deepak Boggavarapu, PhD
LiquidBrain
deepak@liquidbrain.cool
San Francisco, CA

Summary: Our new microchannel liquid cooling system delivers 10X performance improvement over air cooling. Large 12cm x 12cm ceramic microchannel plates that are only 1mm thick remove 300 W/cm2 heat flux while staying below chip junction temperature limits. Massive parallelism allows 10 kW of heat load to be removed in area the size of a human hand. We solve current thermal problems and enable new architectures for future systems.

Introduction

For over 40 years Moore's Law has propelled computing, but in recent years growth has reached a plateau. This is shown in the following graph [1] which shows transistor count increasing in yellow. However frequency is flat over the last decade while the number of cores in black has increased. This is due to thermal limits shown in red which limit operating frequency. It is simply not possible to remove the heat load from the chip.



While Moore's exponential is slowing, a new computing explosion has begun with AI/ML taking off. Artificial Intelligence (AI) is pushing the boundaries of computing systems with ever expanding models and training sets. The next phase of cloud computing growth is being driven by AI. One study shows AI models with 300,000X increase in compute resources over the last 7 years [2]. Direct liquid chip cooling can solve this pressing problem.

The current state of the art in data center cooling is forced air cooling. The fundamental problem with this approach is that HVAC systems and chilled water systems are used at the facilities level but fan based air cooling is used in the building and the rack. Fan cooling can consume up to 25% extra parasitic power. Air is inherently a good insulator and impedes the flow of heat: water is 4000 times more efficient than air in transferring heat. Direct liquid cooling at the chip level significantly improves performance and energy efficiency. With the ultimate goal of computing systems to reach and exceed human brain level performance, liquid cooling is the only possible approach in the forseeable future.

Microchannel Cooling

Here we describe a new massively parallel liquid cooling system that enables large numbers of CPU's, GPU's or other dedicated AI processing chips to operate at very high packing density and at thermal performance far in excess of current technologies.

Microchannel cooling has been used in laser chip cooling and solar energy to achieve very high performance. In this approach, small channels on the order of microns to hundreds of microns in size are fabricated that pass liquid as close as possible to the heat source to remove heat. The purpose is to increase surface area of the channels as much as possible to aid heat transfer

while maintaining flow and pressure at acceptable levels. However, these sophisticated cooling systems have not been adopted in computer chip cooling.

Systems Engineering

Many complex and conflicting engineering requirements must be met when designing cooling systems. These include:

1) Choice of materials and choice of conductor or an insulator. A concern with metals is the possibility of galvanic corrosion in the system. Insulators enable a whole new range of design and architecture options. Materials choice impacts the coolant choice as well.
2) Coefficient of thermal expansion – the cooling system material should match the CTE of the silicon chip. CTE mismatches lead to structural failures of solder bonds after many temperature cycles. Thus matched CTE improves reliability.
3) Temperature uniformity – the spatial temperature variation across the chip should be minimal. Localized hot spots create points of failure that spread.
4) Low fluid pressure drop – the system should be designed with minimal pressure drop to reduce parasitic power draws and improve reliability of seals, materials, and connections.
5) Low flow rate – low flow rate is desirable and increases the effectiveness of the cooling system. Improved effectiveness results in higher fluid exit temperature which can allow the extracted 'waste heat' to be used for other useful work. This also allows the heat exchanger rejecting heat to the environment to be more efficient.
6) Low thermal resistance – the lower the thermal resistance of the cooling system and the entire heatflow path allows more heat to be dissipated while keeping the chip junction temperature at specified levels.
7) Parallelism – parallel flow paths prevent single points of failure and create robust and reliable systems.
8) Large area – chip cooling systems have traditionally addressed the design issues of a single chip at a time. But the future will see new architectures with large arrays of chips or wafer scale integration.
9) Choice of coolant – The coolant impacts storage temperatures, operating temperatures, corrosion, and fouling issues.

These are just some of the many contrary constraints that must be solved as we peer forward to the future of datacenters, supercomputing, and AI.

The Solution

Building on the previous work our team has done in high power laser diode cooling, solar energy systems, and complex systems engineering, we have designed a cooling system to meet the constraints mentioned. Our solution consists of a ceramic microchannel plate that is CTE matched to the chip, is massively parallel, large area, and much more.

The microchannel ceramic plate is designed such that each chip receives fresh coolant and is massively parallel with no serial liquid paths. Within each chip area the liquid flow is also parallel so there no single points of failure and no hot spots. The plate has tens of thousands of liquid flow channels embedded in the material and is likely the largest monolithic microchannel system of its kind (figure 1 and 2). The plate is

12x12cm and 1mm thick and can be configured for much larger areas (30cm x 30cm), different geometries, chip sizes, etc. The system can be configured for large numbers of discrete chips or wafer scale large area cooling with 100% fill factor. The system can dissipate >4kW heat load with heat flux of >250 W/cm2 while keeping junction temp rise below 55C. It is capable of heat loads exceeding 10kW in this same size. Total area is 144 cm2 which is about the area of a human hand.

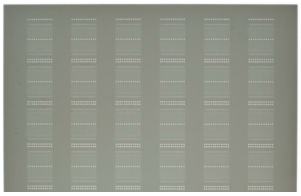


Figure 1. Large area microchannel plate. 12cm x 12cm area, 1mm thick and with tens of thousands of fluid channels.

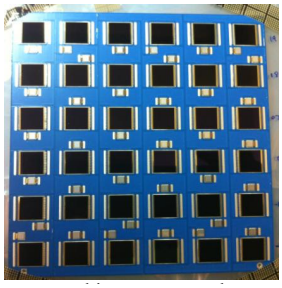


Figure 2. Large area microchannel plate (blue) with 36 chips (black) and interconnects.

These microchannel systems have been tested in harsh outdoor conditions in high concentration photovoltaic power systems under conditions much more challenging than a datacenter (figure 3 and 4). These systems were designed to produce solar electric power with very high efficiency and recover the waste heat with very high effectiveness. In this case the waste heat could be used for water desalination. Recovery and re-use of the waste heat from datacenters is another way to help decarbonize datacenters.

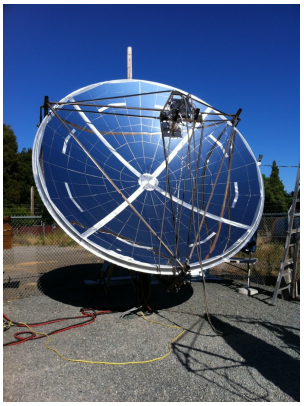


Figure 3. High concentration solar photovoltaic dish system with microchannel liquid cooling in rugged outdoor environments. The liquid cooling system is at the focal point of the dish and keeps triple junction solar cells cool.

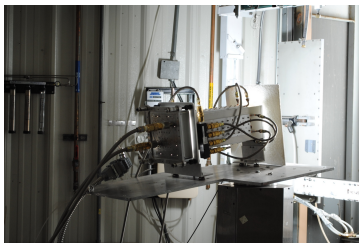


Figure 4. Testing of microchannel liquid cooling system at National Renewable Energy Laboratory that set performance records.

We have met all 9 of the design constraints simultaneously in one product that can dramatically improve compute performance. For example an Intel Xeon 8168 has thermal design power (TDP) of 205 Watts, estimated die size of 7cm2 for heat flux of 30 W/cm2. A typical air cooled server with this chip with ambient temperature of 35C would operate at chip junction temperature of 85C (delta T of 50C). With our microchannel system, the chip would operate at 40C with a 5C delta T. With a typical delta T budget of 55C, our microchannel system allows the chip to operate at 300 W/cm2 heat flux which far surpasses current chip heat flux.

The graph below shows the relative performance our cooling technology (figure 5). Our first gen technology is 10X better than air

cooling. Our second gen technology will have another 2x improvement (LB v2)

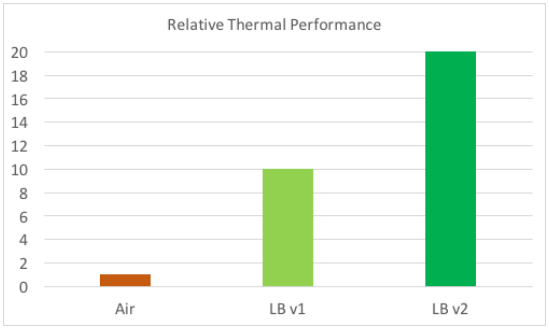


Figure 5. Relative performance of our microchannel cooling technology. Our first gen technology is 10X better than air cooling (green LB v1).

The chart below compares the heat flux several high performance chips. An Intel server chip has flux of 30 W/cm2, and Nvidia GPU operates at 31 W/cm2. Recent advances in AI chip development offer large wafer scale integration as one approach. But such wafer level parts operate at similar average heat flux of 32 W/cm2. As mentioned above, our cooling system enables heat fluxes of 300 W/cm2, a 10X improvement and over large areas.

	Model	TDP (W)	Die Size (cm2)	Heat Flux (W/cm2)	Cooling
Intel	8168	205	7	30	air
Nvidia	GV100	250	8.15	31	air
Cerebras	wafer	15,000	462	32	liquid

Figure 6. Heat flux of high performance chips

The opportunity exists to combine large numbers of chips with novel interconnects and achieve performance levels that exceed other approaches and with high manufacturing scalability. Heterogeneous integration of large arrays of chips each optimized for specific functionality as in figure 2 may far exceed performance of wafer scale chip fabrication approaches.

The Path Forward

Our microchannel technology offers a path forward from solving today's most challenging thermal problems with proven technology to new architectures that get us closer to human brain like performance.

- We cool 300 W/cm2 heat loads with junction temperature below limits. A 10X improvement.
- Large arrays of chips cooled with 10kW heat load in the size of human hand. More than 10X increase in area and heat flux.
- This technology enables new system possibilities with 1000 GPUs or AI optimized chips in a volume of a cubic foot.
- Achieving brain like performance will require a 100,000X increase in compute power in coming decades which will require novel architectures and interconnects. We can take inspiration from biology where neurons, synapses, and capillaries are all integrated and intermixed in the brain. Microchannels are the capillaries of electronic computing from which new architectures will arise.

Optimized thermal technology compliments and makes any chip approach better.

References

[1]https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/

[2]https://openai.com/blog/ai-and-compute/