Bitcoin 2-Phase Immersion Cooling
and the Implications for High Performance Computing

Alex Kampl
Allied Control Company
alex@allied-control.com

Abstract

Recently, Bitcoin and Bitcoin mining have aroused international interest. This article describes the cooling challenges faced by Bitcoin miners and discusses the implications on the future of cooling for supercomputers, HPC clusters and other high density electronics. We explain the inner workings of the first commercial open bath immersion cooling systems and how Bitcoiners seemingly refine the technology used to create high powered ASICs, and keep them cool.

Keywords

Immersion, cooling, 3M Novec Engineered Fluids, open bath, passive 2-phase, Bitcoin, high performance computing, HPC, ultra high density

Bitcoin

Bitcoin is a peer-to-peer payment system and cryptographic currency introduced as open-source software in 2009 [1]. Users install a free client (the wallet) onto their computer, which allows them to send and receive payments using Bitcoin addresses. At the time of this writing, more than 50,000 payments are sent on a daily basis. Participants known as "miners" verify transactions into blocks and broadcast them to the public block chain by employing a proof-of-work concept using a secure hashing algorithm (SHA-256). Miners get rewarded with 25 new Bitcoins in return for each block. To stabilize the block creation rate at one block about every 10 minutes, the network self-adjusts the difficulty of the hashing calculations. The cost of verifying transactions is the capital cost plus the power consumption of the hardware in use. The more computational power is employed to do the "hashing", the bigger the share of the reward that goes to the miner.

Miners initially employed conventional CPUs to perform the SHA-256 hashes. When GPUs (graphics processing units) were found to be more efficient, the cost to run CPUs quickly exceeded the profits that could be made with them. As Bitcoin grew in popularity, gamers realized that they were in possession of "money making machines", and the arms race for the most efficient mining hardware was on. What is now known as the "Bitcoin mining industry", moved through several hardware generations within two short years. This included FPGA educational kits, custom FPGA hardware with "beefier" power supplies, low-end ASIC chips working in parallel, to arrive at high-end 28nm and 20nm ASIC chips we see today. Power consumption per chip now exceeds that of high-grade gaming GPUs and supercomputer CPUs and accelerator cards (Table 1).

TABLE 1. Evolution of Bitcoin Mining Hardware

Type	Make/Model	Year	MH/s	W
CPU	Xeon E5530 2.40GHz	2009 Q4	7.14	80
GPU	ATI Radeon HD 5830	2010 Q1	245	192
GPU	ATI Radeon HD 6990	2011 Q1	800	450
FPGA	Xilinx Spartan-6	2012 Q1	245	10
FPGA	Xilinx Spartan-6 (4)	2012 Q2	850	40
ASIC	ASIC 130nm (32)	2013 Q1	12K	75
…	…	…	…	…
ASIC	ASIC 28nm (single)	2014 Q1	500k	500
ASIC	ASIC 20nm (single)	2014 Q2	750k	???

Today, mining is often conducted by industrial mining farms, with densely packaged hardware that becomes outdated within months of coming online as electricity costs exceed profits. The rapidly evolving power density of mining clusters coupled with their short life cycle makes deployment in conventional air cooled data centers challenging.

FIGURE 1. A 28nm mining device to be connected to an external power supply.

2-Phase Immersion

2-phase immersion is viewed by many as a viable technology for meeting the power density and energy efficiency needs of the high performance computing market. Power densities 100 times higher than a typical air cooled server have been cooled this way with efficiency superior to direct water cooling [2]. Though new technical approaches [3] simplify its adoption, 2-phase immersion cooling has been slow to catch on in the risk-averse IT industry despite numerous demonstrations in many countries with just about every kind of device imaginable.

A primary hurdle to its adoption is the lack of energy-dense hardware. The cost and weight of fluid needed to fill a server chassis that is 95% air or more is prohibitive. Yet commodity hardware does not lend itself to density increases. Air cooling constrains the area of a motherboard and DIMMs, capacitors and connectors restrict the pitch between boards. For this reason, technology demonstrations to date have only been able to showcase the energy efficiency merits of the technology. Hardware sufficiently dense to showcase the density and resulting cost advantages has not been available except in specialized applications such as Bitcoin mining.
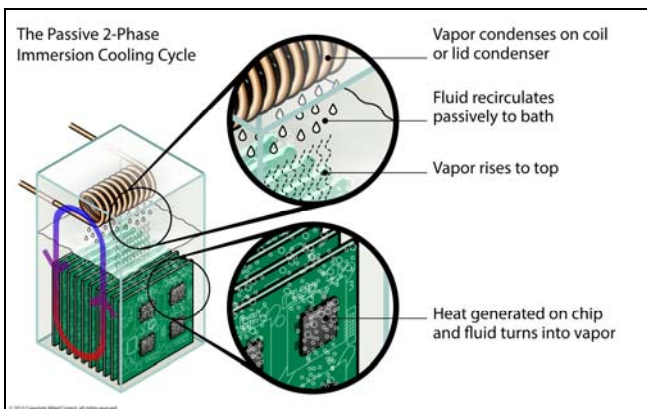

FIGURE 2: The passive 2-phase immersion cooling cycle.

Heat Transfer Challenges

Keeping hardware cool has proven to be a problem right form the early days of custom Bitcoin mining hardware. FPGA chips used at or over the manufacturer specified limits down-clocked when overheating, resulting in lower hash rates. A number of commercial mining hardware manufacturers were forced to make last minute changes, such as adding unplanned additional fans and heatsinks, before delivering hardware to customers [1]. While the market has somewhat matured in the meantime, this remains a continuing trend.

FIGURE 3. A 500W 28nm ASIC mining chip, cooled by a PC water cooling system with a waterblock, pump, custom radiator and external cooling fans.

As they are approaching the limits of traditional cooling methods, the latest high powered ASIC devices face a new problem altogether. Power trains and mining chips are exceeding heat dissipated by commercially available computer hardware, creating a significant amount of engineering overhead, and hardware often ships with throttled performance due to thermal constraints. Removing these thermal bottlenecks, and reducing the associated engineering overhead, is central to the value proposition of 2-phase immersion. Full immersion allows manufacturers to take the artificial lid off and deliver faster machines.

Immersion-2 500kW Bitcoin Mining Cluster

Hong Kong, with its hot and humid climate, is a particularly challenging environment for data centers. With average PUEs of 2.52 [4] being the norm and pollution on the rise, companies need to be creative when it comes to IT cooling. The city is home to Immersion-2, and it comes as no surprise that its first large-scale implementation was a 500kW Bitcoin mining cluster.

FIGURE 4. 3 tanks per 19-inch rack.

Immersion-2 has a footprint smaller than a shipping container and is comprised of 20 conventional 19-inch racks, each housing 3 separate tanks or "baths", as shown in Figure 4**.** The computational hardware is comprised of 80W boards originally cooled with a large aluminum heatsink to which two fans would be attached. When deployed in the insulated stainless steel immersion tanks, 92 boards are stacked vertically on a backplane with less than 10mm board-to-board pitch (Figure 5). To reduce space and fluid usage, boards are arranged in pairs with component sides facing each other and through-hole components (ie. fan connectors) are left unpopulated during production. The weight of the original 92 heatsinks is roughly the same as the weight of an empty tank that contains about 2.75liters of fluid per kW during operation.
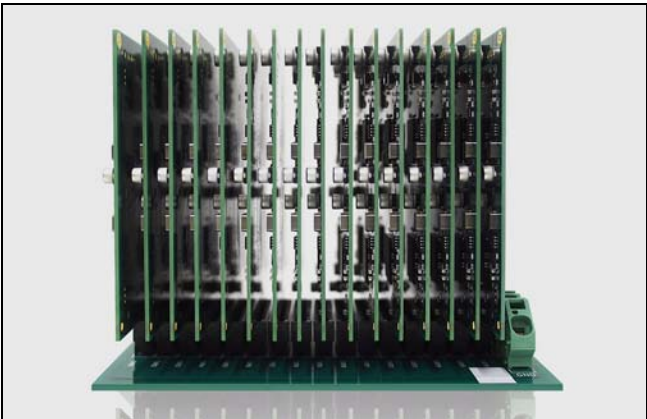

FIGURE 5. 92 boards on similar backplanes with less than 10mm distance board to board.

Condenser for Hot Water Cooling

When the system was put into operation, the 92 boards and 4 power supplies in each tank dissipated about 8kW, but the condenser in each bath is designed to accommodate 25kW at Hong Kong's hottest temperatures (Table 2)**,** making it simple for the client to move through hardware generations. With possible water temperatures of 50°C or higher, the system can be run year-round with simple free cooling equipment. There are ports for two more condensers in each bath, allowing up to 75kW per tank or 225kW per rack. Baths of this density have already been demonstrated [5], as have boards packaged at up to 4kW per liter [3]**.**

This scalability is central to the value proposition of 2-phase immersion. With the addition of more commodity dry tower hardware outside of the datacenter, the mechanical infrastructure can accommodate newer mining hardware with higher power densities. With boiling enhancement, emerging 1000W ASICS can be accommodated with the same fluid boiling temperature $T_f=60C$, allowing the facility to continue operating at PUE<1.02 in Hong Kong's hot summers. This efficiency too is central to the value proposition, for its direct effect on income and hardware lifespan.

TABLE 2. Immersion-2 Condenser Specs at 30 liter/min water flow

Q (kW)	Tin (°C)	Tout (°C)	DP (PSID)
30.22	32	46.58	<0.5
26.09	36	48.6	<0.5
21.96	40	50.62	<0.5

The custom tank condensers use enhanced tubes borrowed from the refrigeration industry. Each is fed facility water directly from a four module dry tower sized for Hong Kong's hot and humid weather (Table 3). The data center requires only a small air-conditioning unit for comfort cooling when people are present. With the only fans in the facility's electrical switchboard cabinet, the data center has been described as being as silent as a public library.

TABLE 3. 1500 Watt cool hundreds of Kilowatt on 30 January 2014

Outdoor Temperature:	24°C
Relative Humidity:	64%
Water-In Temperature:	39.9°C
Water-Out Temperature:	45.4°C
Room Temperature:	27.6°C
Radiator Power:	0.4kW
Pump Power:	1.1kW

Facility Management

Since the cooling system is operating isolated from the indoor environment, and the hardware is submerged in a constant temperature fluid, facility management and maintenance are reduced to a minimum. Without chillers or air handlers, there is no raised floor or dust that needs to be taken care of, and no static electricity that needs counter measures such as humidification. Environmental monitoring is carried out merely for statistical reasons. Instead vital water-in and out temperatures are monitored.

Though the hydrofluoroether working fluid is non-flammable, the buildings original water sprinkler system was kept in place for regulatory reasons. Should the sprinklers go off, the hardware is shielded from water, whereas the "outside" electrical wiring is well protected. Data centers using water sprinkler systems traditionally have to replace hardware after a catastrophic event. Ironically, after a fire in one of Google's data centers in

2010, it was reported that the sprinkler system did more damage than the fire itself [8].

Power Delivery and Cabling Challenges

Some of the challenging aspects of the first installations were unrelated to cooling. In Bitcoin mining, power hungry ASICs are often restrained by thermal limitations in the power delivery portions of the system. Concerned "it would go up in flames", the primary manufacturer was unwilling to provide a backplane with sufficient power density. As with many other components for such an unusual application, the backplane was eventually designed and assembled in-house. Tasks like LAN wiring can become an art form when 92 Ethernet cables need to go in the smallest space possible without obstructing the way in and out for hardware. Standard RJ45 connectors suddenly seem oversized, leading to the idea of stacking boards face to face and removing the outer cable jacketing. It is these small take-home lessons that will prove invaluable when designing dense immersion hardware for HPC or cloud computing. Large-size connectors and cables don't seem to be the way forward, creating an opportunity for new connector and power solutions.

Material Compatibility

When Immersion-2's predecessor went into operation in 2012, it consisted of 24 tanks totaling 70kW of FPGA hardware, and relied on power and USB cables for its operation. Each tank contained over 2 kilograms of off-the-shelf, polyvinylchloride (PVC)-clad USB cables. Upon dismantling the system the following year, it was observed that the dioctyl phthalate (DOP) oil used to plasticize the insulation had been extracted. This caused the cables to become thinner and more rigid in the vapor region above the fluid, although no failures were ever observed. 20 grams of DOP were absorbed by passive carbon filters inside each tank, and a similar amount of DOP was later extracted from the fluid, with no degradation of boiling performance observed during the system's lifespan.

Full immersion has a positive impact on hardware reliability, due to hardware not being exposed to temperature fluctuations, uniform temperature across components, and elimination of the risk of electrostatic discharge in dry air. Without doubt, hardware failure will never be completely eliminated. Designing power supplies and backplanes without heat constraints can be less of a headache, but practical considerations need to be taken into account: the rules of hot swapping, redundancy, and systems protection are changed, due to the increased number of nodes.

Electronic Failure

After the deployment of Immersion-1, a board in one of the tanks failed, causing the power supply to shut down. With no way to identify the affected board, the power supply had to be power cycled repeatedly until geyser-like boiling and smoke could be observed to locate the failing hardware. Though the PCB board and failed devices clearly burned, neighboring hardware was not affected, despite the

8.5mm pitch between boards. Eventually, more intelligent backplanes were developed.

Servicing the hardware meant arms deep in the tank and the head inches above the opening for at least 30 minutes. Though no person working on the system felt sick or unwell, this brought to light concerns about thermal decomposition of the fluid and the possibility of toxic by-products. Engineers studied typical component failures, including high current MOSFETs and buck regulators. Identical boards were driven to failure in both air and fully immersed in fluid, concluding that PCBs arcing and catching fire in air are far more dangerous than the same boards failing within a sealed tank of fire extinguishing fluid. After two years of operating systems 24/7 and working within the very same rooms, there was no single incident of skin or respiratory irritation. This is to be expected with a fluid that is routinely used in vapor degreasers and aerosol cleaners.

FIGURE 6. Working fluids are non-flammable, non-toxic, eco-friendly and clean. Boards come out dry.

Hard Drives and Immersion

Another challenge has been building systems with the requirement for external storage. Traditional HDDs and immersion systems don't get along well due to the mechanical structure of HDDs. In prototype setups the easy fix has been the use of solid state drives (SSDs) or leaving storage devices in adjacent air cooled enclosures. Experiments have been performed with keeping spinning hard drives in the head space of immersion tanks. The latest advances in HDD development may solve this challenge altogether. The first generation of hermetically sealed Helium-filled hard drives, aiming for increased capacity and speed, became available at the end of 2013.

The Path Forward (Supercomputers)

Though the hardware cycle tends to be faster in the mining world, there are strong parallels between this market and the high performance computing market. Tianhe-2, the world's fastest supercomputer, is comprised of 16,000 nodes, each containing two Intel™ Xeon™ Ivy-Bridge processors and three 250W Xeon™ Phi co-processors. It consumes 17.8MW with an additional 6.4MW reported for cooling (not including rack fans?) resulting in a calculated PUE of 1.36. If this could be reduced to 1.05, Tianhe-2 would save 48 million kWH per year with the potential to boost its 40th place standing in

the Green 500 [7]. The resources that went into the development, validation and deployment of the closed loop water-air hybrid cooling system could have been avoided [6]. If conventions were established, Tianhe-3 might be deployed more rapidly with the same mechanical infrastructure and increased reliability. Immersion systems have already been built that show the functionality of high speed copper and optical interconnects in the immersion environment.
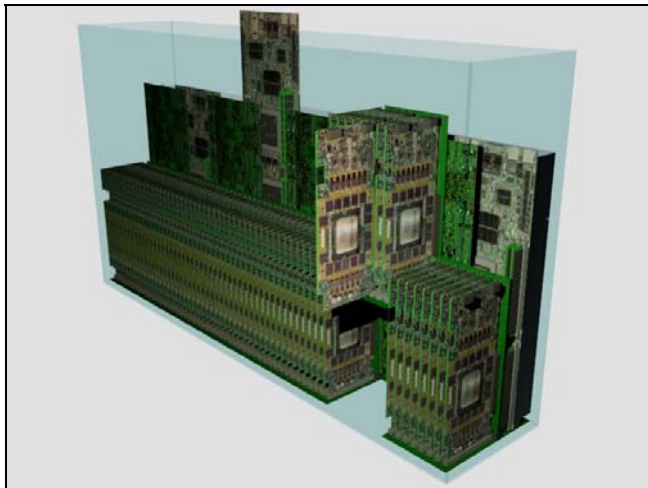

FIGURE 7. 2013 Immersion cooling concept for 64 Intel® Xeon Phi™ and 8 Intel® HPC mainboards with Infiniband® in a 3U "suitcase" form-factor (25kW)

Changing the Rules

Bitcoin is perhaps the perfect incubator for 2-phase immersion. With a basic system, developers have the luxury of quietly (both literally and figuratively) refining the technology on their own turf while exploring the reliability, serviceability and safety of the technology. It might seem unlikely that Bitcoin miners would be the ones to redefine thermal limits, but there's an interesting precedent in the Bitcoin hardware industry. Professor Michael Bedford Taylor, who heads the Center for Dark Silicon at the University of California at San Diego (UCSD), is the author of a paper describing how Bitcoiners have disregarded all traditional rules of chip design, "leading to the development of machines that had orders of magnitude better performance than what Dell, Intel, NVidia, AMD or Xilinx could provide" [1]. Taylor, who was among the developers of the TILERA many-core computing architecture, says this path was "unheard of in modern times, where last-generation chip efforts are said to cost $100 million or more and the number of ASIC starts drop yearly."
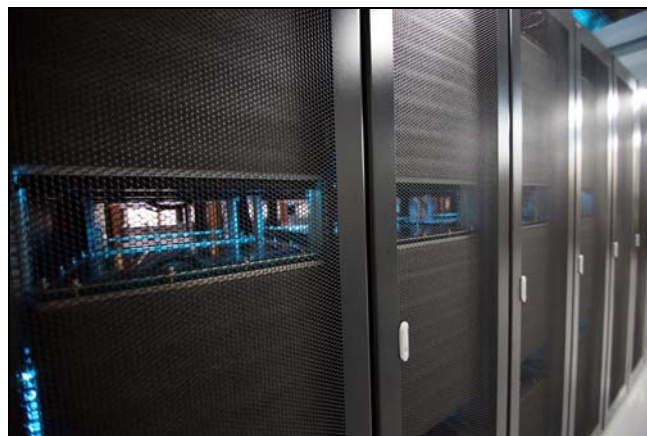

FIGURE 8. Immersion-2 Rack Platform

Author Info
Alex Kampl is VP of Engineering at Allied Control. His company provides engineering and consulting services to clients worldwide. Alex is co-founder of the privately owned company and brings over 20 years of industry experience in the areas of electronic design and manufacturing. Born into a family of product engineers in Austria, Alex has worked and lived in Europe, the USA, and now in Asia.

Full disclosure
Alex and his company have no horse in the Bitcoin race, other than building systems and consulting for a client in the Bitcoin industry.

Contact Details
Alex Kampl
+852 31450055
alex@allied-control.com

Immersion-2 Project Video on YouTube
http://www.youtube.com/watch?v=oZavKweMrP4

References
1. Taylor, M.B., "Bitcoin and the age of bespoke Silicon", 2013 International Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES), Sep. 29-Oct. 4, 2013
2. Campbell, L. and Tuma, P., "Numerical Prediction of the Junction-to-Fluid Thermal Resistance of a 2-Phase Immersion-Cooled IBM Dual Core POWER6 Processor," Proc. 2012 IEEE Semitherm Symposium, San Jose, CA, March 18-22, 2012
3. Tuma, P. E., "The Merits of Open Bath Immersion Cooling of Datacom Equipment," Proc. 26th IEEE Semi-Therm Symposium, Santa Clara, CA, USA, pp. 123-131, Feb. 21-25, 2010
4. Hong Kong Data Centre Survey by Campos for Digital Realty, March, 2012 (http://goo.gl/QNT3hr)
5. Polzer, S., "Design Passive Open Bath Immersion Cooling High Performance Computing (HPC) Test Platform," presentation IMAPs ATW on Thermal Management, Los Gatos, CA, USA, Nov. 12-14, 2012
6. The Green 500 List - November 2013 (http://goo.gl/kCgf3m)
7. Dongarra, J., "Visit to the National University for Defense Technology Changsha, China", June 3, 2013 (http://goo.gl/wkseyc)
8. Brad T. Stilwell, "Mythbusting Clean Agent Fire Protection", Fire Suppression, Fall 2013 Issue