High Temperature Coolant Demonstrated for a Computational Cluster

Egor Druzhinin, Alexey Shmelev, Alexander Moskovsky, Yuri Migal
ZAO "RSC Technologies"
Moscow, Russia
Emails: druzhinin@rsc-tech.ru
alexeysh@rsc-tech.ru
moskov@rsc-tech.ru
migal@rsc-systems.ru

Vladimir Mironov
Lomonosov Moscow State University
Moscow, Russia
Email: vmironov@lcc.chem.msu.ru

Andrey Semin
Intel Deutschland GmbH
Munich, Germany
Email: andrey.semin@intel.com

***Abstract*—Energy efficiency is an important aspect of high performance computing systems design, as the power consumption grows. Free cooling is a widespread technique to lower energy consumption of the cooling subsystem, usually requires expensive equipment to be installed. The standard chilled water temperature is 7 °C may be attained large part of the year in the Europe, and in North America, when atmosphere air temperature is low enough. In case the air temperature is higher, freon compressors should be used to cool down the liquid. In the presented work we report about blade server design and computational cluster installation with direct liquid cooling, that operate at coolant temperatures above 50 °C, which enables 24 hour, year-round free cooling with atmosphere air at the most locations on Earth. The server design is based on an approach to utilize coldplates with channel structure and liquid circulation for heat removal from semiconductor components. We have designed a coldplate with low heat-resistance that ensures effective cooling with only 20-30° temperature difference between coolant and electronic parts of a server. Under stress-test conditions the coolant temperature was up to 65 °C while server operation was undisturbed on the individual server level (based on Intel Grantley platform with dual Intel Xeon E5-2697 v3 processors). We also report first stress test results for the 24-node computational cluster on the same platform, when coolant liquid temperature is set to 50 °C and more.**

***Index Terms*—hot liquid cooling; cold plates; energy efficiency; computational cluster.**

I. Introduction

Modern supercomputers and high-performance computing (HPC) clusters consume large amounts of electric power, so its important to address infrastructure component effectiveness and efficiency, like cooling subsystem. Addressing infrastructure efficiency would be also critical to attain exascale level of performance for the future generation supercomputers, especially given power limits available in existing computing centers (e.g. notorious 20 MW power limit for an exaFLOPS-level machine [1], [2]).

A traditional quantitative measure of how efficiently energy is utilized in datacenter is power usage effectiveness (PUE) developed by The Green Grid consortium [3]. PUE is the ratio of total amount of energy used by a computer data center facility to the energy delivered to computing equipment. The ideal PUE is 1.0 when all energy is consumed by computing equipment and datacenter infrastructure (such as cooling) takes no energy. In a real datacenter a lot of energy is used by infrastructure equipment and PUE value is higher than 1.0.

One of the most important energy consumer in datacenter is the cooling system: even in the optimized liquid cooled datacenters it could utilize as much as 30% of total energy supply [4]. PUE is significantly improved in modern datacenters with the advent of so-called free cooling [5]. They are specially designed for effective use of outdoor environment to remove the heat from servers. However, free cooling is most effective if the datacenter is able to operate at high temperatures.

Nevertheless, supercomputers must be energy efficient due to huge amount of electricity consumed by them. Additional savings are possible by reusing some of the heat, but they may require coolant temperatures to be even higher than a human could tolerate. Usually, higher temperature coolant is considered as a more energy efficient option due to absence of chiller equipment, which reduces capital expenditure for system construction. However, semiconductors operating on higher temperatures may have higher leakage current resulting in degradation of energy efficiency. Interference of these two factors remains nontrivial, especially for the most recent hardware generation and multitude of semiconductors used in server design.

Compactness is uniquely important for HPC design. To be efficient, parallel processor components need communication to be as fast as possible. So, at the end, these components have to be placed as close to each other as possible to minimize communication length. It rules out the use of free airflow cooling methods, that sometimes are applied in a datacenter design (e.g. in [6]). Coldplate-based design seems to be a good alternative there, because it is not only provides efficient cooling but also enables compact system setup.

In the current work we have studied a traditional homogeneous server operating with hot liquid cooling setup. Liquid coolants have much better thermal properties than gases: typical PUE values for water cooled datacenters utilizing free cooling could be less than 1.1. We designed a coldplate with low thermal resistance: temperature difference between CPU top cover and liquid coolant is minimized. That design enabled us to explore server energy efficiency in a wide coolant temperature range.

The initial individual server testing was then supported by the study of computational cluster. The liquid cooled 24-node cluster has been installed at Joint Supercomputer Center of Russian Academy of Sciences. Heat exchange between the coolant and the outdoor environment is based on the use of a dry cooler only. Dry cooler fans may me switched on and off depending on the coolant temperature threshold. The mild 10 °C air temperature enabled us to conduct free cooling experiments with coolant temperatures up to 50 °C, while computational nodes were under test with High Performance Linpack (HPL) [7] benchmark.

II. Details of the Experiment

The coolant temperature impact on performance and energy consumption was studied on an individual server level as well as for the cluster which was based on the same hardware. To study the cooling performance on single-node level the server was thermally insulated from the environment to eliminate its impact on the experiment and to mimic very high density packaging that traditionally found within blade server systems racks. The thoroughness of the isolation was examined with Fluke Ti32 thermal imager. The resulting experimental setup is presented on Figure 1. The multi-node cluster solution has 24 of such servers in a rack. Rack was connected to the cooling system that will be briefly described below.

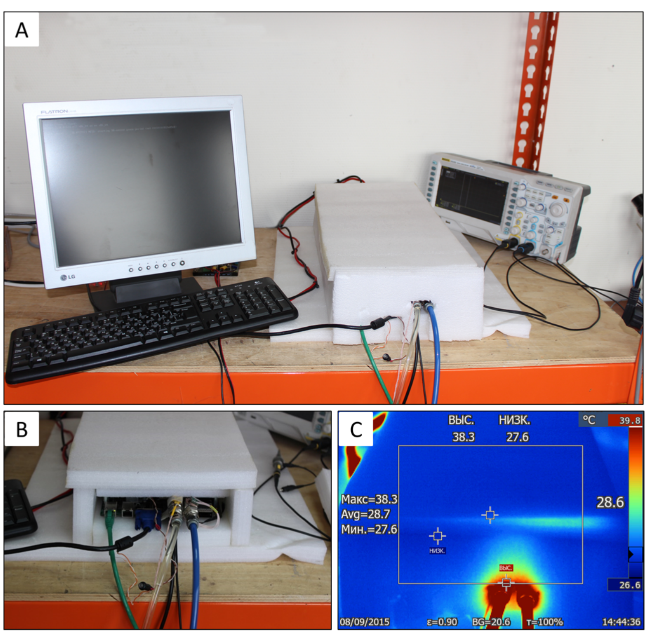

Fig. 1. a) Overview of the experimental setup; b) liquid cooling and system connections; c) Fluke Ti32 view (T_{max} = 38.3 °C, T_{min} = 27.6 °C).

A. Hardware specifications

A server based on RSC Tornado solution was used to study the performance and energy efficiency of our water-based liquid cooling solution on single-node level. It has dual socket Intel ServerBoard S2600KPF with two Intel Xeon E5-2697 v3 (14 cores, 2.6 GHz, 145 W TDP) installed. A latest production BIOS was used. The server was also equipped with 64 GiB (8×8 GiB) of DDR4-2133 registered memory with ECC support and a 120 GB solid state drive.

Temperature of the coolant was controlled by liquid-to-air heat exchanger. It was programmed to activate when coolant temperature reaches experiment threshold. Its efficiency was optimized to avoid liquid temperature decline on fan activation. Measurements of inlet and outlet liquid coolant temperature was performed using thermocouples. Experiments were conducted at 19, 45, 50, 55, 60 and 65 °C inlet water temperatures. The average difference between inlet and outlet water temperature was 3.5 °C. Flow rate sensor was used to measure liquid flow rate through the system. The average rate of water flow in cooling system was 30 ml/s.

Our cooling solution was also studied in larger environment using the fresh installation of computational cluster based on new RSC Tornado technology at Joint Supercomputing Center of Russian Academy of Sciences. The system under study includes 24 nodes. They have similar hardware specification as the one described above and also equipped with Mellanox Infiniband FDR Infiniband interconnect. Currently installed infrastructure includes rack, piping and dry cooler. It is substantial to run 150 compute nodes (≈100 kW heat dissipation). There is also an option for expansion up to three racks (450 nodes, 300 kW) without upgrade of cooling subsystem.

Cluster-level experiments were conducted at 19, 45, 55 and 57 °C inlet water temperatures. Input power consumption was measured with limited precision and was equal to 13±1 kW in all cases.

B. Benchmark details

DGEMM matrix multiplication benchmark was used to simulate stress-test conditions. Intel MKL (version 12.2.3) DGEMM implementation was used in this test. Dimensions of the matrices were selected to be 87936×192×87936 that fit in 58 GiB of memory and provides a stress level similar to hot phase of HPL benchmark.

The DGEMM kernel was run continuously for at least 1 hour for each of different inlet temperatures of the coolant ranging from 19 to 70 °C. Performance information was collected for every DGEMM iteration during the benchmark. Power usage and temperatures of CPU, memory as well as the whole system was monitored using out-of-band telemetry provided by Intel NodeManager. The state of liquid cooling system (temperature and inlet and outlet coolant flow rate) was also controlled during all benchmarks. The data from the first 20-30 minutes of all benchmarks were discarded to ensure stationary temperature of coolant and all server components.

TABLE I
DGEMM PERFORMANCE RESULTS AND POWER CHARACTERISTICS OF THE LIQUID COOLING SOLUTION AT DIFFERENT COOLANT TEMPERATURES

T_{LC} (°C)	Performance (GFLOPS)	System power (W)	Efficiency (GFLOPS/W)	Efficiency decrease (%)
19	974	359	2.72	—
45	985	388	2.54	7.0
50	985	390	2.53	7.5
55	981	395	2.49	9.4
60	976	396	2.46	10.3
65	969	398	2.44	11.5

TABLE II
HPL PERFORMANCE RESULTS ON MULTI-NODE ENVIRONMENT AT DIFFERENT COOLANT TEMPERATURES

T_{LC} (°C)	Performance (GFLOPS/node)	Efficiency (GFLOPS/W)	T_{DTS}: min, max (°C)
20	988	1.82	–
45	988	1.82	-20, -17
55	989	1.82	-6, -3
57	983	1.81	-6, 0

Stress-testing of the 24 node cluster was performed using HPL benchmark [7] which was run for two hours. The problem dimension was N = 603456; the block size was NB = 192.

III. RESULTS

A. Server-level experiment

As one can see on the Figure 2, liquid cooling solution that was used in this study has relatively low thermal resistance. Its estimated value was 0.74 $°C \cdot in^2 \cdot W^{-1}$ that is on-par with commercially available coldplates (eq. Lytron plates [8]). The average difference between CPU (T_{CPU}) and liquid coolant (T_{LC}) temperatures is about 26 °C even at T_{LC} = 60-65 °C.

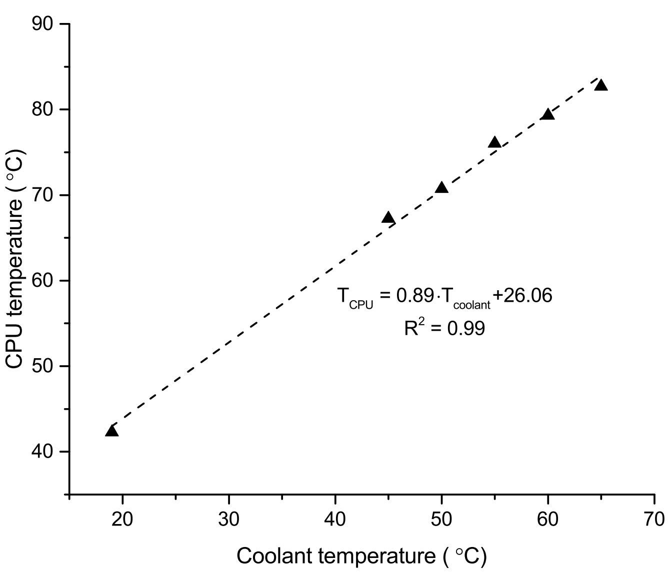

Fig. 2. Dependence of average measured CPU sensor temperature on coolant inlet temperature. T_{CPU} - CPU temperature, T_{LC} - coolant temperature.

The performance results of DGEMM benchmark (Table I) shows no temperature dependence at $T_{LC} \leq 65$ °C and is also stable upon the time. At higher temperatures the CPU clock rate scales down and and performance degradation (>100 GFLOPS at T_{LC} = 70 °C) was observed. The data from distributed temperature sensing (DTS) devices integrated in CPUs [9] also indicate overheat at T_{LC} = 70 °C. For that reason we didn't refer to high (>70 °C) temperature results further.

A modest increase of system power consumption upon coolant temperature growth is observed for the server. We also performed an analysis of power consumption of most important system components. The values for CPU (288–291 W) and memory (21 W over all tests) depend only slightly on temperature. Tests were run with Turbo Boost feature enabled that adapts CPU performance to fit in the power limit. The power consumption of other system components is much more sensitive to the temperature. The largest contribution to residual power consumption is made by chipset components and voltage regulators. They doesn't have such power control and contributes much more energy with temperature growth. In our setup power consumption of chipset linearly increases from 50 to 90 W when the temperature rises from 30 to 65 °C.

While the average system power consumption shows linear dependence on T_{LC}, floating point performance almost does not change. These trends leads to efficiency decrease of ≈2.2% on every 10 °C growth of T_{LC}. The overall change in efficiency was 11.5% upon growth of coolant temperature from 19 to 65 °C (Table I).

B. Cluster-level experiment

We also studied the behavior of the same hardware in a larger environment. At T_{LC} = 57 °C half of the nodes were heated according to data from DTS sensors in CPUs (T_{DTS} = 0 °C). At the same time, the other half of nodes show quite reasonable DTS value (-6 °C) leaving the room for the improvements of this result. CPU overheating eventually leads to the performance decrease (see Table II). However the observed performance changes are very small (about 0.5%) up to T_{LC} = 57 °C. The absolute values of HPL benchmark results are very close to the single-node DGEMM performance and correspond to 85% theoretical peak performance for dual socket Intel Xeon E5-2697 v3 node.

The energy efficiency of the studied system is not as good as for single node (1.8 GFLOPS/W and 2.5 GFLOPS/W respectively). This is due to only 24 nodes was used for the benchmark. However, the full capacity of the cooling system corresponds to the much larger environment. In such case power efficiency of this cooling solution would improve dramatically. According to our estimates, PUE of the studied system would be less than 1.08 for the full rack of 150 nodes. This value can be further improved down to 1.04 if the whole capacity of the cooling system is used (3 racks, 450 nodes).

We should mention, that the precision of power consumption measurements in multi-node experiment does not allow us to capture its change upon T_{LC} growth. Our measurements shows that it was ≈13 kW and this value was used in calculations. For the evaluation of PUE characteristic the estimate of 360-400 W/node from the single-node experiment was used.

IV. Related work

An important problem of hot liquid cooling systems is the possibility of energy reuse. It should be stressed, that the gain from free cooling and energy recuperation/reuse should compensate computational efficiency decrease. Switching to the free cooling already provide notable energy efficiency growth. The problem of heat reuse is much more complex. The most efficient way is to use hot water for heating [10], [11], however, it is very climate and country specific and not always possible. In many cases HPC datacenter designers have to think of possible energy conversion in useful work, however its mount is limited by thermodynamics second law. Energy recovery is most effective at very high coolant temperatures, but there are restrictions imposed by hardware.

We observe a modest decrease (about 10%) of power efficiency when coolant temperature increases from 19 to 65 °C. Similar results has been obtained for other contemporary hot water cooling solutions, namely Aquasar (7% power consumption increase from 30 to 60 °C), CooLMUC (5% power consumption increase from 30 to 50 °C) and iDataCool (7% of efficiency decrease from 49 to 70 °C) [11]–[13]. In the study of Aquasar hot liquid cooling system [11] the exergy analysis showed about 10% of possible energy reuse at 50-60 °C water temperature. Some additional gain is also possible due to free cooling. Thus the use of hot coolant is reasonable in datacenters and would result into reduction of their PUE value. Theoretical PUE value for the liquid cooling solution studied in this work is 1.04, that is already a very good value. Another successful example of hot liquid cooling usage was presented by Eurotech [14]. They reported PUE value of 1.05 at 50 °C coolant temperature, however they didn't performed a detailed performance analysis that includes FLOPS/W metrics.

V. Conclusion

In this paper we explored performance and power profiles of the hot liquid cooling solution with temperature of coolant between 19 °C and 70 °C, designed for RSC Tornado supercomputer architecture. This hardware exemplifies modern server platform with high temperature coolant option available due to appropriate design. Since commodity semiconductor components are used (CPU, memory modules, server board), our results provide insight on the modern hardware in general.

In this article we have demonstrated the performance and energy characteristics of small cluster that was designed on the basis of the RSC Tornado platform. The use of hot liquid cooling technology allows us to drastically simplify infrastructure of the cluster. For example it is cooled only by a single dry cooler and freon compressor cooling machines are not longer necessary. We have achieved a stable performance up to the 55 °C of inlet coolant temperature with a theoretical PUE of 1.04 for a large cluster. Climate-wise, 40-60 °C coolant temperatures enable free cooling on most of the Earth 24×7, except desert areas. It means that the use of hot liquid cooling allows to construct a supercomputer facility with low PUE value in almost any place on Earth regardless of climate condition.

References

[1] P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carson *et al.*, "ExaScale computing study: technology challenges in achieving exascale systems," *Government PROcurement*, vol. TR-2008-13, p. 278, 2008.

[2] D. Wade, "ASC business plan (NA-ASC-104R-15-Vol.1-Rev.0)," Tech. Rep., 2015.

[3] J. Haas, J. Froedge, J. Pflueger, and D. Azevedo, *Usage and public reporting guidelines for The Green Grid's infrastructure metrics (PUE/DCiE)*, The Green Grid, 2009. [Online]. Available: http://www.thegreengrid.org/~/media/WhitePapers/WhitePaper22PUEDCiEUsageGuidelinesfinalv21.pdf

[4] M. P. David, M. Iyengar, P. Parida, R. Simons, M. Schultz *et al.*, "Experimental characterization of an energy efficient chiller-less data center test facility with warm water cooled servers," in *2012 28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*. IEEE, 2012, pp. 232–237.

[5] C. Gough, I. Steiner, and W. A. Saunders, *Energy efficient servers: blueprints for data center optimization*, 1st ed. Apress, 2015.

[6] Yahoo launches second 'computing coop' data center in New York state. [Online]. Available: http://www.datacenterknowledge.com/archives/2015/04/27/second-yahoo-data-center-comes-online-in-new-york-state/

[7] A. Petitet, R. C. Whaley, J. Dongarra, and A. Cleary. (2008) HPL - a portable implementation of the high-performance Linpack benchmark for distributed-memory computers. [Online]. Available: http://www.netlib.org/benchmark/hpl/

[8] The performance of standard cold plate technologies is compared in a graph showing local thermal resistance. Lytron Inc. [Online]. Available: http://www.lytron.com/Cold-Plates/Standard/Performance-Comparison

[9] M. Berktold and T. Tian, *CPU monitoring with DTS/PECI*, 2010. [Online]. Available: http://www.intel.com/content/www/us/en/embedded/testing-and-validation/cpu-monitoring-dts-peci-paper.html

[10] S. Zimmermann, M. K. Tiwari, I. Meijer, S. Paredes, B. Michel *et al.*, "Hot water cooled electronics: exergy analysis and waste heat reuse feasibility," *International Journal of Heat and Mass Transfer*, vol. 55, no. 23-24, pp. 6391–6399, 2012.

[11] S. Zimmermann, I. Meijer, M. K. Tiwari, S. Paredes, B. Michel *et al.*, "Aquasar: a hot water cooled data center with direct energy reuse," *Energy*, vol. 43, no. 1, pp. 237–245, 2012.

[12] N. Meyer, M. Ries, S. Solbrig, and T. Wettig, "iDataCool: HPC with hot-water cooling and energy reuse," in *Supercomputing: 28th International Supercomputing Conference*, 2013, pp. 383–394.

[13] A. Auweter and H. Huber, "Direct warm water cooled Linux cluster Munich: CooLMUC," *Inside*, vol. 10, no. 1, pp. 81–82, 2012.

[14] Hot water cooled supercomputer. Eurotech. [Online]. Available: http://www.eurotech.com/en/hpc/hpc+solutions/liquid+cooling