ADVANCES IN CYBER-PHYSICAL SYSTEMS

Vol. 3, No. 2, 2018

HIGH DENSITY HIGH PERFORMANCE COMPUTING SYSTEMS COOLING

***Anatoliy Melnyk*[1], *Volodymyr Savyak*[2]**

Lviv Polytechnic National University, 12, S. Bandery Str., Lviv, 79013, Ukraine
Authors' e-mail*: aomelnyk@lp.edu.ua*
[2]*Hewlett Packard Enterprise, Kyiv, Ukraine*
Authors' E-mail*: volodymyr.savyak@gmail.com*

Submitted on 01.12.2018

***Abstract*: The paper deals with the cooling approaches and aims at using solutions for the energy efficient and economically viable high density computing system construction and usage. Power consumption of the standalone high density computing systems and supercomputing clusters with different types of cooling has been considered and their characteristics have been estimated. The influence on the overall budget of the system construction and exploitation for 5 years of life time has been analyzed.**

***Index Terms*: supercomputers, high performance computing cluster, high density computing system, cooling, energy efficiency.**

I. INTRODUCTION

Researchers and engineers who decided to build the supercomputer clusters or other high density computing systems, need to pay special attention to power and cooling of the systems. Our experience of delivering HPC systems in almost 27 countries proves that many people didn't realize the importance of the efficient cooling and only reality of large energy bills and dead systems helps to consider this topic as an important one.

Experienced datacenter and system builders understand – 1) energy costs are high and 2) ineffective cooling makes all your super high performance computing system rubbish. Datacenter professionals can appeal to the datacenter best practices and standards according to Uptime Institute, TIA or BICSI but take attention, we are discovering in this work system cooling, not a datacenter construction, otherwise, they are interconnected. We also evaluate the usage of the computing system with the internal (system related) cooling modules in the existing datacenter.

We have found that for many high performance computing systems projects "during the new cluster system design, cooling is often fallen out of major focus of attention, as it is more related to thermal engineers, rather than scientists, who design a supercomputer" [1] but, in fact, it influences the system design heavily.

In the work we also paid attention to the most important hidden problems, which are not obvious at first glance. We had a chance to discover many of them, based on the 15 years of experience of high density computing systems and server room construction.

An understanding what exactly influences the overall energy consumption, how to make the cooling efficient and effective and also relevant in terms of overall system cost and design is the major goal of the current research. We are mostly focused on the high density system but triggering many of the general important aspects and principles should be also taken into account.

II. BAD COOLING MAKES PROBLEMS…

It is important to understand that bad cooling makes problems; it can spoil all the computing system construction. However good your computing system is, it will not work long if it can't be cooled effectively.

Making cooling system we should take into account:

- Effectiveness, to deliver needed capacity with good quality for each of the component.
- Efficiency, consume less and be potentially available for the energy reuse.
- Cost (CAPEX – Capital Expenses, OPEX – Operational Expenses and thus TCO – Total Cost of Ownership).
- Ease of Usage (standard 19" rack equipment support, ease of installation, ease of service, continuous operation, while servicing, resiliency, deployment time, scalability).
- Safety (for engineers, users, environment).

An important point here is that the above mentioned system characteristics should be considered not just about the cooling system itself but the overall computing system and the entire infrastructure influenced.

Once we saw a datacenter (DC) designed and built on the river banks. DC cooling system heat exchanger was initially cooled by the river water. Otherwise, the DC operations where shut down by eco-police and the user had to rebuild the cooling to continue operation with air cooled external heat exchanger and much higher cooling costs vs initially expected. It was not safe for the nature.

Based on the experiments and experience with the number of dense computing systems we can extract the following list of the most important problems, caused by the improper cooling design.

Problems discovered during the design, construction and procurement cycle based on the implementation of different cooling solution:

- Cooling can require building reconstruction in case of the installation in the existing premises, which also brings budget increase to:
    - Walls reconstruction or modification;
    - Raised floor construction;
    - Floor/walls strengthening to increase the loading-carrying ability;
    - Windows shuttering by the strong materials, to protect from light, heat and avoid window blow out by gas of fire suppression system;
    - Multiple holes in the walls to bring in and out heat transfer pipes and cables;
    - Ceiling reconstruction, etc.
- Increasing peak power load for cooling purposes, which requires permit for higher power supply feeder, which is a costly resource in many countries.
- Increasing needs for the costly power equipment, including uninterruptable and reserve power supply, batteries (reserve power supply is not always used in high performance computing systems if the power line is of a high quality).
- Space requirements limitations can cause the issue with some of the cooling solutions implementations.
- Impossible to build a high density computing solution because of inability of the different types of cooling to take out heat effectively.
- Equipment, software and procedures setup to handle emergency situations are required;
- Cooling solution noise for both internal and external cooling modules.
- Heat disposal or reuse sometimes could be a problem but also is an opportunity to get cheap energy which otherwise will be wasted.

Problems discovered during exploitation period, and among them are so-called hidden problems, which are not always easily identified by the users and/or support organizations:

- High energy bill, which is often not only the size of the budget problem but also a type of the budget problem. Many of the HPC systems once bought with the grant or other one-time budget will require large OPEX investments, which are not necessary on place but could reach for the exploitation period of the system acquisition cost.
- Difficult to service and, thus, requires high cost professional services and also can cause longer planned downtimes.
- Often failures of cooling system that cause unplanned system downtime and/or compute equipment death.
- Inability to scale compute and cooling system because of the improper design.
- Computing equipment failing and improper functioning, because of overheat.
- Shortening life of the system (quick ageing), which is not an immediate death but highly increased percentage of the yearly equipment failing.
- Low performance of computing elements (often throttling of compute engines – CPU, GPU), because of ineffective cooling / overheat.
- Increasing power consumption and heat generation because of higher leakage currents.

Analyzing different types of cooling we should check the above given mentioning to be sure we avoid or minimize risks and build effective, efficient and safe solution.

III. DIFFERENT TYPES OF COOLING OVERVIEW

Modern computing systems, based on semiconductors, consume electric power and generate heat. Power consumed goes to the heat except of some noise, light, chemical transformations but they are so much small that can be neglected and in most cases they will also end up with heat. Here and later we will use power consumption as easier to measure value but taking into account that it equals to energy transferred to the heat.

Most of the heat in computing system is generated by the large semiconductors elements, like CPU, GPU, FPGA, chipset and memory, but in the large system it is worth calculating, even heat is generated by cables as a separate value, but for a smaller one it can be neglected. And don’t forget about energy loss (means again heat dissipation) in the power conversion (power supply units). These elements are not separate but integrated typically in the chasses of the computer modules. And we should take into attention that the cooling system should be designed to cool effectively the computing system @ full load.

This research is focused on the system design and construction, based on the standard components, which is a typical approach for the system of below 1 million USD range and quite often used for the systems of 1–10 million USD range.

Let’s take a look for power consumption of the simple computing system with 2 (two) multicore CPU of the latest generation, 384GB RAM in total, Platinum power supply and multiple cooling fans in a closed chasses for the rack installation (Table 1 and Table 2). We have used consumption data from components manufacturer’s specifications as the most reliable source for the information. As discovered below they are similar to what we get on the same system type at 25 ^{o}C inlet air for cooling.

Based on the above given data we see the major power consumers and, thus, heat generators in the systems. We should not only discover how to cool the computing node but also should understand how heat is taken away out of the computing node components. This is needed to insure that the exact elements which are the hottest are cooled well and function properly, as the CPU and GPU overheated will throttle and slow down, and the memory and PSU will fail in case of overheat.

Table 1

Computing node configuration type 1 components power consumption

HPE DL360 Gen10	Type	W	%
Computing node configuration type 1.	In total power consumption=>	552	100
CPU	2x Intel Xeon Gold 6148	300	54
Memory	12 x 32GB DDR ECC	120	22
Chipset	Intel C621 Chipset	15	3
HCA	100Gbit EDR Mellanox Infiniband	15	3
FANs	Single Rotor Fans	30	5
PSU	800W Platinum PSU	44	8
SSD	2x480GB LFF SATA Mixed Use Hitachi	18	3
Other	cables, different internal chips, e.t.c.	10	2

Table 2

Computing node configuration type 2 components power consumption

HPE DL380 Gen10	Type	W	%
Computing node configuration type 2	In total power consumption=>	1240	100
CPU	2x Intel Xeon Gold 6148	300	24
GPU	2xNVIDIA V100 32GB PCI	600	48
Memory	12 x 32GB DDR ECC	120	10
Chipset	Intel C621 Chipset	15	1
HCA	100Gbit EDR Mellanox Infiniband	15	1
FANs	Single Rotor Fans	70	6
PSU	1600W Platinum PSU	92	7
SSD	2x480GB LFF SATA Mixed Use	18	1
Other	cables, different internal chips, etc.	10	1

Table 3

Different kinds of Server/Computing nodes power consumption

Server/Computing node	Power (KW)*	Power Per 42U Rack**
Mining Rig, Celeron G3930, 4GB DDR4-2400 Crucial, 6xNVIDIA ASUS DUAL-GTX1070-O8G, 1xSSD 80GB Kingston, PSU Gold+ 1200Wt	0.9	9
2U HPE ProLiant DL560 Gen10, 4xIntel Xeon-Platinum 8176 (2.1GHz/28-core), 448GB DDR4-2666 ECC, 2x480GB SSD MU, 3x2port 10Gbit Eth, RAID E208i-a SR, 2N PSU 1600W Platinum	0.97	19.4
Mining Rig, Celeron G3930, 4GB DDR4-2400 Crucial, 6xNVIDIA TURBO-GTX1080-8G, 1xSSD 80GB Kingston, 2xPSU Gold+ 1200Wt	2.1	21
1U HPE ProLiant DL360 Gen10, 2xIntel Xeon-Platinum 8168 (2.7GHz/24-core), 128GB DDR4-2666 ECC, 2x480GB SSD MU, 2x2port 10Gbit Eth, RAID E208i-a SR, 2N PSU 800W Platinum	0.56	22.4
4U HPE Apollo 6500 Gen10, 2xIntel Xeon-Gold 6148 (2.4GHz/20-core), 8xNVIDIA V100 32GB SXM2, 768GB RAM, 2x480GB SSD, 2x2port 10Gbit Eth, 2xIB EDR Mellanox, 4x2200W Platinum	3.2	32
2U HPE Apollo r2600 Gen10 4xnodes (2xIntel Xeon-Gold 6148 (2.4GHz/20-core), 384GB DDR4-2666 ECC, 2x480GB SSD, 1x2port 10Gbit Eth, IB EDR Mellanox), 2N PSU 2200W Platinum	2.2	44
1U HPE Apollo sx40 2xIntel Xeon-Gold 6148 (2.4GHz/20-core), 4xNVIDIA V100 32GB SXM2, 384GB RAM, 2x480GB SSD, 1x2port 10Gbit Eth, RAID HBA, IB EDR HBA, 2N PSU 2000W Titanium	1.7	68

* Average power consumption will differ depending on the application/algorithm running. For HPE servers we used company data from HPE Power Advisor [3] for 100 % load metric. For Mining rigs we experimented, and made 3 measures with the available rigs under ETH, x16r, CryptoNightV8 algo with 100 % power capping (typically rigs are used underpowered if air cooled) and used average for the result.

** Industry standard racks are not used usually for mining rigs but we can fit the amount of equipment into the corresponding space. For all types of servers/nodes we only utilize 40 U of the rack and 2 (two) upper units are left for extra equipment, like networking, etc.

According to “HPE ProLiant Gen10 Server Extended Ambient Temperature Guidelines” [2], recommended server working temperature range is 10–35 ºC, but with higher than 30 ºC the server consumption grows. The document also says that server with a specific configuration set with a restriction to use some of the components, mostly high consuming ones, can work at range of 5 ºC to 45 ºC dry bulb temperature, or 28 ºC wet-bulb non-condensingLet’s look also at the power density of the modern computing nodes and elements. We also include into the review crypto mining equipment as this kind of equipment is popular at the time when the article is being written and in fact used to deliver high performance computing of a specific typeAs seen above we have power consumption 0.25 –1.7 KW per rack unit, while implementing modern computing systems, based on industry standard components. Let’s look at the cooling technologies capabilities and check which of the approaches can host the above mentioned high density computing equipment (Table 4).

Table 4

The most popular cooling solutions for computing systems

Cooling Approach	KW/rack*	Pros	Cons
Cooling server room air with the standard consumer air conditioning systems	5**	Low cost Easy maintenance	Low effectiveness Low efficiency Low availability
Cooling with industrial air conditioning systems, raised floor, cold & hot aisle separation	10	Classic well known approach Low space required	Low effectiveness Low efficiency High cost of the solution
Ducted exhaust	15	Moderate to low price with better vs standard cooling capacity by pure air No water used Good for free air cooling	Required special architectural modifications Still low capacity
In Row cooling with separation of hot and cold corridors	20	Moderate price with moderate cooling capacity	Moderate cooling capacity Corridors alignments with expansion
In Row cooling, rack containment	30	Moderate price with moderate cooling capacity	Moderate cooling capacity Rack containment alignments
Rear doors heat exchanger cooling (active and passive)	55	Low price with high cooling capacity	Air flow limitations System service issues Cooling interrupted @ service Constant flow, lower efficiency Pressure drop with passive, thus, higher internal node fan CFM
Closed coupled cooling	80	Moderate price with high cooling capacity. Room neutral	Harder service Higher risks for overheating
CPU, GPU, Memory direct liquid cooling on the component level	80	High cooling capacity Partially room neutral Relatively easy to implement	Requires some additional cooling Requires slightly nodes modification
Full node cold plates direct liquid cooling	100	Highest cooling capacity Room neutral	High price Difficult to develop Exclusive node design
Immersion cooling	100	Highest cooling capacity Room neutral Requires no mechanical node modification Possible to use for non-standard equipment	Requires BIOS modification High weight High Price

* Maximum, based on industry recommendations and examples. Can differ, depending on the equipment type.

** Blowing directly to the front of the rack with cold air can bring 10–20 KW cooling possibility, but this kind of cooling solution requires a lot of space and can’t be efficiently implemented at scale. Is not an industry recommended?

As we see from the above given data, the closer we bring heat transfer material (air, water, oil) to the compute elements, – the more efficiency and cooling capacity we get. Leading marketing companies predict high growth for the liquid cooling market at about 28 % CAGR till 2022 [4].

Let’s take a closer look at several cooling solutions.

IV. FREE-AIR COOLING PROS AND CONS

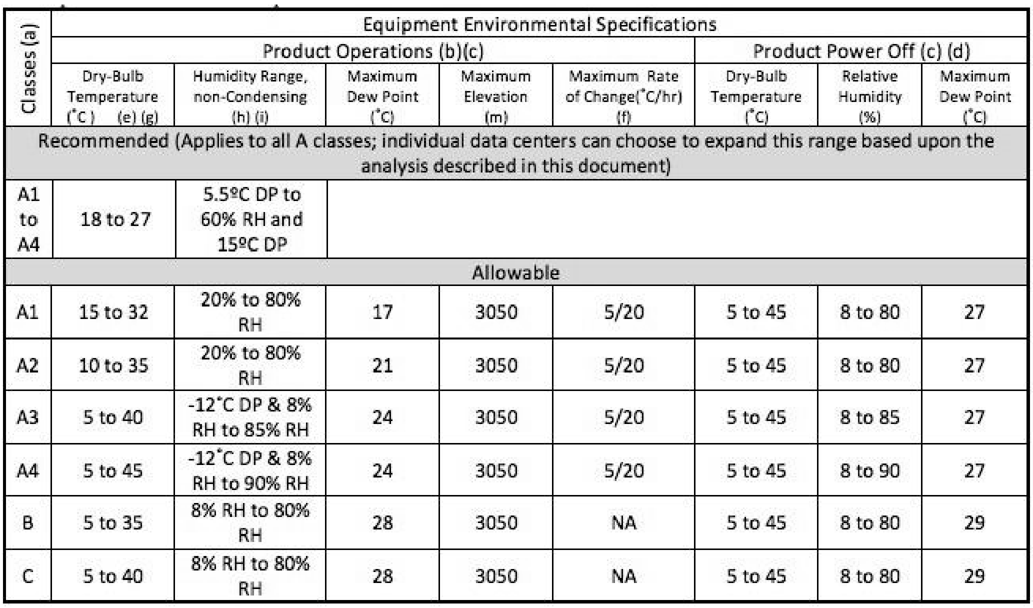
During the last 15 years there were a lot of talks and experiments, using free-air cooling. For the modern high performance computing equipment there are multiple limitations of the incoming air maximum temperature but in most cases high density IT equipment can only efficiently function with Ashrae A1 range, which means up to 32 °C maximum incoming temperature for the system when cooled by air. Otherwise, different equipment has its own operating temperature range, which can be higher, as stated above for the HPE Proliant servers. When we use raised flow air distribution we should understand that the top of the rack equipment will receive air with +2–4 °C higher temperature than the button of the rack one, as we saw during our systems implementation. Keeping the incoming temperature up to 32 °C in case of the usage of the DCL (Direct Liquid Cooling) is quite acceptable based on our practice, but with air cooling on this temperature range we face an increase of the energy consumption and, thus, extra heat generation. This brings an additional energy need not just to power the computing system but also to cool it. As we all know, liquid has higher thermal density than air and can take away more heat with the smaller delta-T (the temperature difference between incoming air/liquid and outcome from the computing node). Below there is a Table 5 with the detail environment parameters with ASHRAE recommendations, which are good guidelines.

For a low cost projects free air cooling with air only can be an interesting option as can be relatively easy organized for low to moderate density installations. This kind of approach is quite standard for computing mining farms. Its combination with low cost consumer or industrial air cooling solution (which can be used as additional cooling source during hot seasons) could be a good option not just for “cold” countries installations but also in the moderate warm climates, like central Europe. Otherwise, industrial type of free air cooling is difficult to organize in many climates. Equipment, which can guarantee high quality effective and efficient air-only-cooling is expensive and its payout can be lengthy.

Another problem appears to be even larger for the extreme low temperature regions during cold seasons. As an example, HPE EcoPOD, a kind of combination of free cooling and evaporative cooling solution, is limited to –19 °C. This kind of limitations could be a show stopper for using free air cooling in some regions. One of the free air (air only) cooling problem solutions is frosting. When humidity in the area of installation is high due to the sea or large river located not far from the installation place, frosting will be a show stopper at low temperatures.

Table 5

Equipment environmental specifications according to ASHRAE [5]



Classes (a)	Equipment Environmental Specifications							
	Product Operations (b)(c)					Product Power Off (c) (d)		
	Dry-Bulb Temperature (°C) (e) (g)	Humidity Range, non-Condensing (h) (i)	Maximum Dew Point (°C)	Maximum Elevation (m)	Maximum Rate of Change(°C/hr) (f)	Dry-Bulb Temperature (°C)	Relative Humidity (%)	Maximum Dew Point (°C)
Recommended (Applies to all A classes; individual data centers can choose to expand this range based upon the analysis described in this document)								
A1 to A4	18 to 27	5.5ºC DP to 60% RH and 15ºC DP						
Allowable								
A1	15 to 32	20% to 80% RH	17	3050	5/20	5 to 45	8 to 80	27
A2	10 to 35	20% to 80% RH	21	3050	5/20	5 to 45	8 to 80	27
A3	5 to 40	-12°C DP & 8% RH to 85% RH	24	3050	5/20	5 to 45	8 to 85	27
A4	5 to 45	-12°C DP & 8% RH to 90% RH	24	3050	5/20	5 to 45	8 to 90	27
B	5 to 35	8% RH to 80% RH	28	3050	NA	5 to 45	8 to 80	29
C	5 to 40	8% RH to 80% RH	28	3050	NA	5 to 45	8 to 80	29

We have made several experiments with air-cooled systems type 1 to observe the consumption and heat generation behavior at different temperatures (Fig. 1).

We see a small difference at 25 °C with the data we have based on manufacturers specification, which is 1–2 % deviation.

One 42U rack, while populating 40U (let’s leave 2 U for communication) can host 40 of the above measured servers (HPE DL360 Gen10, 1U rackmount), so the consumption of the rack will be slightly more than 20 KW.

The more energy is consumed, it will lead server to consume more air to be cooled more effectively. And this pushes us to deliver more air into the Datacenter. So, this will force to spin fans (moving air into datacenter) faster and, thus, they will consume more as the fan power is proportional to (fan speed)^3.

Here comes the amount of air consumed by a single server (the same as above tested for the energy consumption) for the different temperatures (Fig. 2).

The above given data shows us how the energy and air flow of the servers/computing nodes is increasing with the higher temperature. This happens because of the higher consumption of the fans [6] as they start pushing more air and high leakage currents [7, 8].

Small fans in the nodes typically consume more power per air flow than the large fans, which we use for getting cold air into the datacenter. This could be either inlet air blowing fans in case of the free air cooling or fans on the heat exchanges (like InRow units), when water is used to transfer heat in and out of the datacenter. But anyway the large fans will also increase the consumption and we need to take it into the consideration.

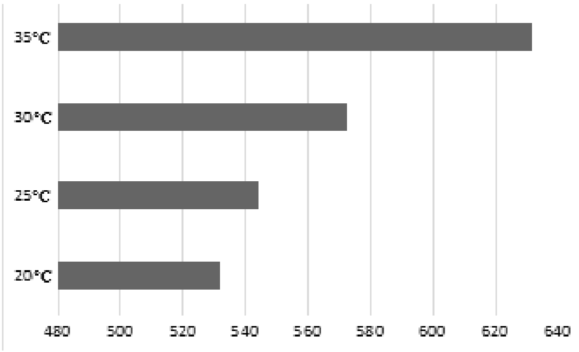


Fig. 1. The energy consumption of the server type 1 at different temperatures

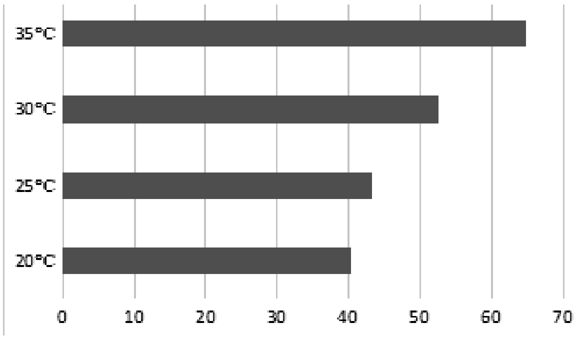


Fig. 2. The air flow of the server type 1 at different temperatures

To calculate the additional energy for the effective delivery of the air to the servers/computing nodes in our calculations, let's use the exact cooling and ventilation system. To simplify, we will not calculate stream degradation for the air filtration, assuming that we use the way of bringing cold to the server room/system with the water and exchange the heat via cooling towers, InRow systems.

The formula to calculate input load for FANs can be the following [6]:

$$P_{fan}=0.040459894+0.08804497*X+$$
$$+0.0729612*X2-0.943739823*X3,$$

where P_{fan}= load power ratio, X=load ratio of fan operation (cfm/design cfm).

According to the above given consideration, let's calculate an additional power consumption of facility fans with a need to push air at the full load for a full equipped compute rack of the system type 1 servers (Table 6). The scenario can be managed with two MaxxAir IF18 3000-CFM 18-Inch Blade Heavy-Duty Exhaust Fan [9]. Each consumes 640 Wt @ 100 % load and delivers 3000-CFM.

One fan is installed to get air in the datacenter and the second – to get it out; this kind of combination allows to control pressure inside of the data center / server room.

Power consumed by ventilators is not the only one that influences the power consumption and, thus, heat generation, as given above. With the temperature increase we should take special attention to the leakage currents [7], which are dramatically increasing with the temperature. With the given example in Fig. 1 we see 15 % power usage increase on the computing node, and only 2–4 % of it is based on the internal servers/nodes fans.

Based on the experiments, made on GPU card [8], we see the following leakage currents representing power consumption increase for the temperature growth of the GPU (Fig. 3).

For the NVIDIA GPU GXT 1060 and GTX 1070, equipped with 2 fans for most of the vendors of the boards, the temperature of the chip 75 °C is reached with standard Power Limit setup, which is 100 %, running Decred, Keccak algo with ~23 °C of the inlet air for the cooling. So with the 25 °C inlet air temperature we will get ~80°C and with 30 °C we are getting ~90 °C if no downclocking is made which leads to ~40 % of current leakage. With 150W per GPU@100 load brings 30–35 % of extra power consumption at 30 °C inlet temperature for the GPU mining system with 6x NVIDIA GTX 1070 boards.

All the above given consideration, will affect our costs of using the system. We will calculate and compare them in the last part of the current article.

The above given consideration, should be taken into account when designing high density system with free air cooling.

Otherwise, when water or other liquids are used as a heat delivery we can extract and transfer heat more effectively, thus, we can keep higher inlet cooling media temperatures for the computing nodes. The idea of free cooling in combination with water becomes more reasonable as discussed in the section below.

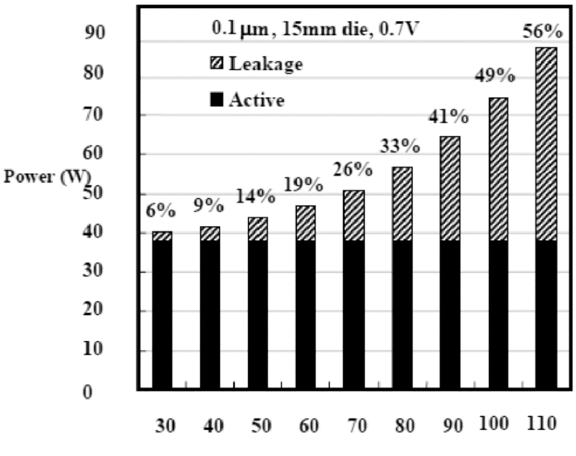


Fig. 3. Leakage currents representing power consumption increase for the temperature growth of the GPU

Cooling with liquid brings better efficiency

Water always makes electrical and so computing engineers to be scared. It can destroy equipment and together with high electric power can also kill people. That is why water cooled solutions are put last in the queue in many even high density computing projects and is used very rarely in most of the enterprise datacenters. But if made right, the water cooled solution brings good results. First of all, it delivers good effectiveness and for a sure good efficiency, bringing cost of usage for the system down.

Table 6

Large fan power changes at load for 1x rack cooling at different temperatures

Temperature, °C	Power 1xRack	CFM 1xRack	FAN @100 % load Wt	FAN 100 % load CFM	x2 Fans (in + out)	% of CFM	P_{fan}	Power, W	% of total power
20 °C	21 280	1 613	640	3000	1280	54 %	0.26	327.06	1,54 %
25 °C	21 760	1 732	640	3000	1280	58 %	0.30	380.20	1,75 %
30 °C	22 894	2 109	640	3000	1280	70 %	0.47	596.86	2,61 %
35 °C	25 256	2 599	640	3000	1280	87 %	0.79	1004.96	3,98 %

Using air to cool the systems is not the most efficient way because of the density and thermal capacity and, thus, the amount of heat it can transfer. Actually, air is a well known insulator. We use air to protect ourselves from freezing during winter times in our houses and in clothes we wear. Funny but sure that the water and other kind of liquids are much better in transferring the heat as they have 100s time higher heat transfer coefficient vs air has with cooper and other materials, which extract heat from silicon inside of the computing nodes. To understand how it works we need a little bit of thermodynamics. And we should also remember that heat transfer also depends on the flow rate of the gas/liquid. A good theory on this topic is given in [10]. And there are Wikipedia articles with good summary [11]. The below formula statements are taken from Wikipedia.

The heat transfer coefficient or film coefficient, or film effectiveness, in thermodynamics and in mechanics is the proportionality constant between the heat flux and the thermodynamic driving force for the flow of heat (i.e., the temperature difference, ΔT).

The overall heat transfer rate for the combined modes is usually expressed in terms of an overall conductance or heat transfer coefficient, U. In that case, the heat transfer rate is:

$$Q = h \cdot S \cdot \Delta T,$$

where S: surface area, where the heat transfer takes place, m^2; ΔT: difference in temperature between the solid surface and surrounding fluid area, K.

The general definition of the heat transfer coefficient is:

$$h = \frac{q}{\Delta T} ,$$

where q: heat flux, W/m^2; i.e., thermal power per unit area; h: heat transfer coefficient, $W/(m^2 \cdot K)$; ΔT: difference in temperature between the solid surface and surrounding fluid area, K.

It can also be stated as:

$$h = \frac{Q}{S \cdot \Delta T} .$$

Heat transfer coefficient depends on multiple factors, like fluid/gas flow, shape and irregularity of the surface and can be calculated with a complex formula. For us, in the current research, because of the nature of high computing system design it is more important to have general understanding of the above mentioned process, and using the predefined approximate values calculate the amount of air and/or water or needed to cool the system.

Based on the data given in the engineering toolbox [12] we see that heat transfer coefficient between water and steel/cooper is about 100 times higher than the one for air to steel/cooper. Based on this we can state that with the goal to extract and transfer the given (generated by system) amount of heat we can use much less water than air and use smaller difference of cooling gas/liquid vs steel/cooper heat extractor (which we typically use on top of CPU/GPU/DIMM silicon).

A simplified formula for the cooling power capacity in J/h [13], when taking into account air and liquid flow is:

$$\acute{Q} = \acute{m}\, Cp\, \acute{V}\, \Delta T ,$$

where m is density of material, kg/m^3; Cp is thermal capacity, J/kg·K; V: is gas/liquid flow, m^3/h; ΔT: Difference in temperature, K.

Let's model how much low can be delta T for the different water (reasonable for real water cooling equipment) flow and how it relates to use of the air for cooling for a racks of servers (let's chose system type 1 server – Table 7).

Table 7

Air, water flow and power needed to make it work

	m, kg/m^3	Cp, J/kg·K	V, m^3/h	ΔT, K	Q, kJ/h	fan/pump Power, W
Air, 25 °C	1,2	1005	5 884*	24	85 149*	760
Water, 25 °C	997	4186	3.10	7	90 564	255
Water, 25 °C	997	4186	2.20	10	91 816	181
Water, 25 °C	997	4186	0.90	24	90 146	74

* As stated above section according to the guidelines of DC design, we are using 2x of the FANs, one is to blow the air in and another to blow the air out to control the balance of the air pressure. This is the actual reason, why power capacity is calculated as ½ of the actual $\dot{Q} = \dot{m} Cp \dot{V} \Delta T$ for the integrated air flow, indicated as V(m^3/h) for air cooling. It is also important to indicate that the given example, otherwise reasonable, is a rough estimation. Relevant datacenter designs should include filters to protect the equipment from the dust and this will increase the consumption. It will also include multiple FANs for redundancy and this will lower the consumption, as each of the FANs will need to deliver less air.

Table 8

Full DLC, partial DLC solutions influence on the computing node power consumption

	Power/node, @full air, W	Water cooled, W	Air cooled, W	Internal fans, W	Power/node, @water, W	% of economy
Full cover cold plate	567	491	15	–	506	11 %
CPU/GPU/Memory	1 274	1 048	156	12	1 216	5 %
CPU/Memory only	567	432	111	6	549	3 %

With this parameters in mind we have a difference between fan/pump power consumption about ~2–3 % of the projected server's consumption in our case, which is actually the level of the economy which influences the overall power consumption model.

The other advantage with water is that we are able to use smaller and to be able to use higher inlet temperature to use more free air cooling to cool external heat exchangers of the water cooled system. This gives us a chance to run external chillers on free air cooling mode longer or even go for dry cooling capabilities.

We should also take into consideration, that DLC (Direct Liquid Cooling) solution will not require powerful internal fans in the server, the power which we need to take away from the server in case of the full size cold plate will be "zero" and almost zero in case of the usage of a partial-cover cold plate. In the case with partial DLC the only fan is needed in the power supply, and small additional fans to take all heat away from the rest of the low power elements which are not covered by cold-plate.

Here comes the calculation for the extra power economy from the usage of low power fans and no fans at all (Table 8). We assume that full DLC model still requires fan for the power supply, but Titanium+ large scale rectifiers are used with efficiency of 97+ % [14]. For partial DLC we keep the same standard Platinum power supplies.

While comparing power per node at full air cooling and water cooling at 25 °C we found that the difference per node consumption is 3–5 % for the partial DLC and about 11 % will be full DLC solution. But as we can see the major power advantage of the full DLC is granted not only by internal coolers but also using large external power rectifier, which is a standard approach for many of the full DLC solutions.

As discussed above, using water cooling the most advantageous could be to minimize and be able to use higher inlet water temperature, leading with the idea of free air cooling and stay in the recommended range of inlet temperatures, to insure stable reliable system operations. And with this approach we should balance the inlet water temperature to avoid larger leakage currents, which can overcome advantages, gained from using higher inlet water temperatures.

A big potential free air cooling advantage is the ability to refuse from chilled water at all. If the climate in the installation zone is cold enough that we can use dry tower, we have also CAPEX economy and in the case we still need some period of time to use chilled water, we will have only OPEX economy. The OPEX economy will come from the much lower need in cooling power, because of 3x or even 4x lower, thus, much smaller cooling capacity needs, using inlet water temperature up to 32 °C, so called warm water cooling.

V. FINANCIAL IMPACT OF THE DIFFERENT COOLING TECHNOLOGIES ON THE HIGH DENSITY COMPUTING SYSTEM PROJECT BUDGET

As the financial impact on the system CAPEX (capital expenses) and OPEX (operational expenses) depends on all the system components, let's analyze which is the portion of all and each component. Based on the analyses made in the current research we will prepare the financial impact model and calculations of the high density computing system with the most effective and efficient cooling solution.

We will not calculate exact CAPEX. We only give the general CAPEX recommendations and comparisons. As an example, "refuse from chiller installation if you can use dry cooler with DLC in your region", as stated above. The reason is that CAPEX is much more dependent on the selection of exact brand and model of the equipment than OPEX, which is more dependent on the type of the solution selection. We will focus on power consumption as the main factor for OPEX, which we can influence while building cooling solution for the high density computing systems.

Analyzing chiller behavior is not easy but there are useful tools on vendor's side. Using the Schneider Electric tool we were able to model chiller energy use, depending on the year in a range of 13 % to 17 % of the cooling power delivered in Kyiv region, excluding power of the pumps. But let's don't miss that this % is given from the total cooling power delivered to Datacenter, not just IT one. We need to cool not just IT but other thermal radiation.

For the annual calculation of the power usage we will focus on two types of cooling technologies @25 °C inlet air/water, which are relatively easy to implement with the usage of the high density industry standard computing nodes/servers. The first one is a technology with the usage of (1) InRow cooling with rack containment and the second one is (2) a direct liquid cooling (DLC) on component level for CPU, GPU, and Memory. These technologies deliver high effectiveness and high efficiency cooling, the 1st is focused on moderate density up to 30KW per rack and the 2nd is focused on up to 80 KW densities per rack as mentioned in Table 4. Both solutions require chiller or other active cooling type. We make our calculations with the usage of a chiller with free air cooling mode option. Before we made the final consumption calculations we should define the % of the heat/power, which should be cooled by air cooling system. We will use Table 1 for the input data.

According to Table 8, from 12 % to 20 % power of the computing system should be cooled by air in partial DLC, depending on the CPU/GPU mix and ~3 % in full DLC cooling solution.

The below tables are calculated for 25 °C of inlet water. Overall consumption of the servers with DLC is lower even if we compare with the same heat transfer ratio of air cooled and water cooled servers, because internal server fans will operate in low consumption mode. The following two tables are made for 20 x computing node/server system type 1 in full load under HPL (High Performance Linpack). The full DLC solution is calculated as no chiller mode, based on full time on dry cooler. This kind of calculation is a little bit misleading, because with higher than 25 °C temperatures will lead to higher consumption, but with water it is relatively easy to make larger water flow and effectively take heat away, keeping the same power consumption and heat transfer as it is with @25 °C. This will definitely take some extra power, but because of the relatively small amount of hours during a day with the temperature higher than 25 °C, this extra consumption could be neglected. According to the temperature data, analyzed during last year we have the following temperature history during last year (Fig. 4).

The 25 ^{o}C+ temperature was measured during 222 measure events, lasting ~444 hours, with even distribution (Fig. 5). And there were also 6 events with temperature in range of 30–32 ^{o}C.

444 hours is ~5 % of all year time and this will lead to 5 % higher energy consumption during the year, which will lead to about 0.2 % energy increase in overall bill. From another side, during ~95 % of time we will have below 25 °C temperature with at least 2 % of lower energy consumption, which will lead to about 2 % decrease in overall bill. To simplify the overall calculations, we will leave these differences out of our calculation balance (Table 9).

For the DLC and full DLC solutions we took into account that in server fans will blow slower and will consume much less energy. The economy was calculated with input load fan formula [6].

While the worst case scenario gives PUE about 2.0, we can see that using the advanced technologies of InRow cooling with rack containment and the combination with DLC brings all datacenter average PUE on the level of 1.19–1.37. Together with this, the large portion of power consumption comes from power distribution and UPS modules. If the power in the server room has high quality and the user can avoid using UPS, use line interactive (low consumption modules) or DC batteries only, PUE will go down significantly (Fig. 6).

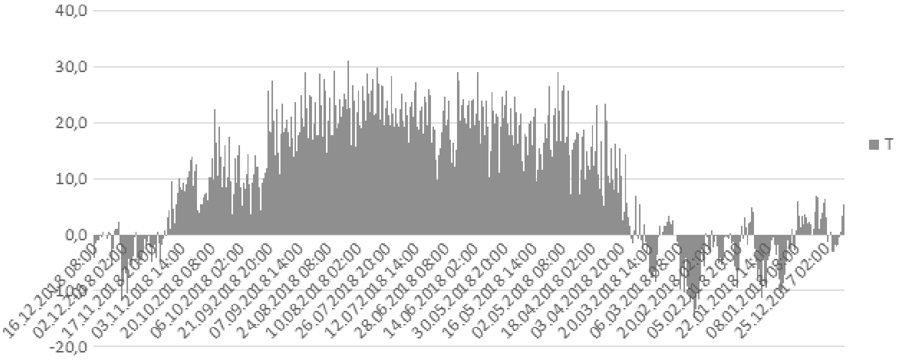


Fig. 4. Hourly temperature in °C measured in Kyiv Zhuliany airport during 2018 (December 2017 – December 2018) [15]

Fig. 5. 25 °C and higher temperature events during 2018

Table 9

Power consumption portion of datacenter/server room components, calculated for full rack, 40 pcs, of systems type 1

Component	Air only	Air, InROw water, on chiller		Air+DLC CPU/Memory, chiller		CPU/Memory, chiller + dry cooler		full DLC / Immersion cooling	
	Worst case scenario	Power, W	% of IT power	Power, W	% of IT power	Power, W	% of IT power	Power, W	% of IT power
Nodes total, pcs	40	40		40		40		40	
Single node power, W	567	567		549		549		507	
Total IT Cooling Power needed, W	22 694	22 694	100 %	21 978	100 %	21 978	100 %	20 278	100 %
Compute Air Cooled (chiller w/free cooling), W	22 694	22 694	100 %	4 452	20 %	4 452	20 %	627	3 %
Compute DLC cooled (chiller/dry cooler), W				17 267	79 %	17 267	79 %	19 652	97 %
DC Fan Load, W		908	4 %	178	1 %	178	1 %	25	0 %
Chiller Load + dry cooler, W		4 312	19 %	4 127	19 %	1 537	7 %	905	4 %
Water Pumps, rack coling, W				59	0 %	59	0 %	0	0 %
Water Pumps, large pipes, W		259	1 %	248	1 %	248	1 %	231	1 %
UPS Losses (On-Line), W		1 557	7 %	1 538	7 %	1 538	7 %	1 419	7 %
Power Distribution Losses, W		1 362	6 %	1 319	6 %	1 319	6 %	1 217	6 %
IT Load PUE, ratio		1.37		1.33		1.21		1.19	
Total Power Consumption	45 388/ 200 %	31 090	137 %	29 188	133 %	26 598	121 %	24 076	119 %

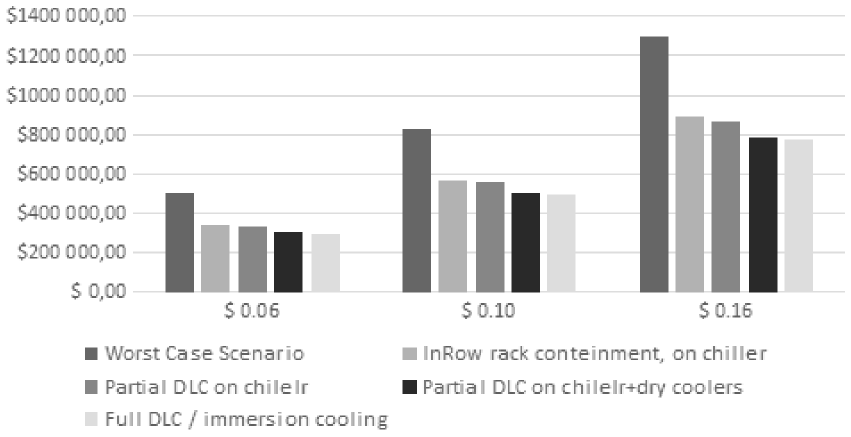


Fig. 6. 5 years energy bill for the system with different energy cost for 100 KW load at $0.06, $0.1, $0,16 price for KW/h.

The above given diagram helps to understand what could be a portion of energy bill in your system total cost of ownership. The 5 year cost is calculated with the simple formula of number of hours multiplied by PUE and power consumption of the system and different price for KW/h.

The system price changes with time. According to our experience, current high density and high performance computing systems (excluding infrastructure cost) is about $2–4 mlns per 100 KW computing system. Otherwise, 10 years ago the same amount of power was consumed by the high performance computing systems with

$200–400 K price. This happens because of the different reasons, among which the complexity and price of the CPU/GPU/Memory units are constantly increasing.

When hosting the system, inside of the datacenter we have no need to take into account half of the above mentioned power consumer categories.

While installing system in the commercial datacenter we will not take care of:

Datacenter water pumps consumption.

UPS losses.

Power distribution losses / except inside of the rack ones.

We only pay energy bill, but this will include the datacenter losses in the energy cost. We should take care of the most efficient way of system cooling. Datacenter should be able to control the energy costs on different sources: electrical power, air cooling, cold water, warm water.

Otherwise, for the overall datacenter 5–10 % power distribution losses are the standard ratio by our practice, some users managed to get energy bill on the server room entry point and used few efficient devised for energy distribution, lowering the cost of the current expense down to 1 %, with using simple PDU modules in the server room and at the room energy count we are getting roughly 1 % power distribution losses. The same scenario is met in the datacenter hosting, when we only should take care of rack PDU modules (Table 10).

Table 10

Power consumption rack level

Component	Air, InROw water, on chiller		Air+DLC CPU/Memory, chiller		CPU/Memory, chiller + dry cooler		full DLC / Immersion cooling	
	Power, W	% of IT power	Power, W	% of IT power	Power, W	% of IT power	Power, W	% of IT power
UPS Losses (On-Line), (DC), W	-1 557	-7 %	-1 538	-7 %	-1 538	-7 %	-1 419	-7 %
Power Distribution Losses (DC), W	-1 362	-6 %	-1 319	-6 %	-1 319	-6 %	-1 217	-6 %
Power Distribution Losses (rack), W	227	1 %	220	1 %	220	1 %	203	1 %
IT Load PUE, ratio	1.25		1.21		1.09		1.07	
Total Power Consumption, W	28 399	125 %	26 550	121 %	23 960	109 %	21 642	107 %

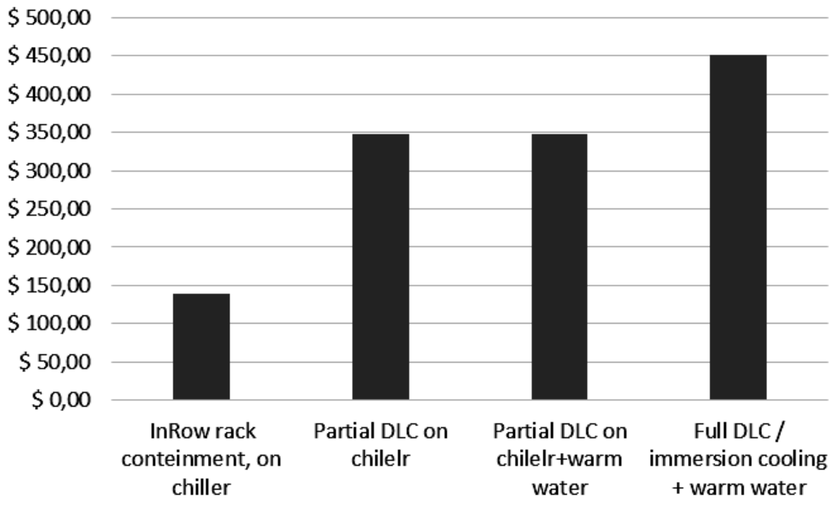


Fig. 7. Inside of the rack cooling solution cost per node

Table 10 is needed to analyze the cost of the usage of high density computing equipment installed in the commercial datacenter. This calculation will work if the datacenter is able to differentiate the power delivered by different sources: electric power, cold water, warm water.

As discussed above, we were focused on OPEX because of the clear reasons. But to give a general understanding of the CAPEX influence with the different cooling solutions, we have made calculations with the sample systems price on internal and external (to the server room) components. The prices are approximate and can differ for different modes and with different discounts.

Let’s check how much the cooling system + energy will cost when installing the high density computing system in the datacenter which has water to cool your system (Fig. 7).

According to our investigation, the current energy cost in an advanced datacenter in Ukraine, to host your rack is in average $0.2 per KW/h. Let’s use calculation from Table 10 and Fig. 6 to evaluate the CAPEX+OPEX cooling and energy costs for one rack of system type 1 (Table 11).

The below diagram gives visual representation of the Table above (Fig. 8).

Table 11

$/node cooling equipment add-on CAPEX + 5 years full energy cost (OPEX)

	$/node cooling cost	Node power, W	PUE	5 year power (KW)	5 year $0,20
InRow rack containment, on chiller	$138.89	567	1.25	31 062.12	$6 212.42
Partial DLC on chiller	$347.22	538	1.21	28 537.93	$5 707.59
Partial DLC on chiller + warm water	$347.22	549	1.09	26 232.37	$5 246.47
Full DLC / immersion cooling + warm water	$451.39	507	1.07	23 758.78	$4 751.76

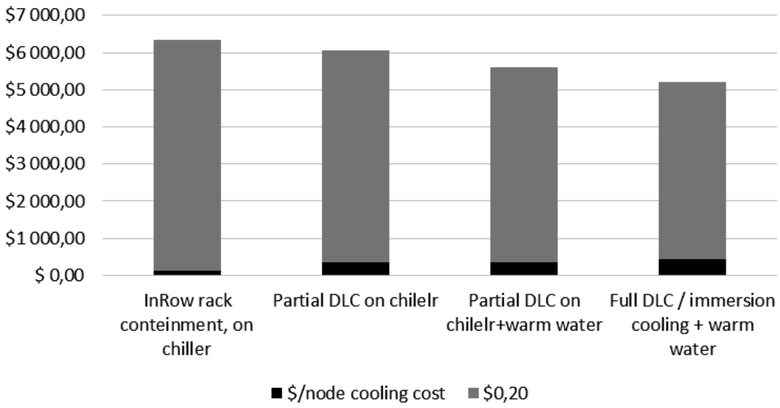


Fig. 8. $/node cooling equipment add-on CAPEX + 5 years full energy cost (OPEX)

According to the above mentioned scenario full DLC or immersion cooling looks to be the cheapest solution, when energy costs are high, which a typical issue for the datacenter hosting.

We also should take into consideration, that using DLC or immersion cooling should also be evaluated in more details, as of better heat transfer abilities and, thus, lower leakage possible currents, which leads to better results. We made quite rough calculations in the above research, but we tried to pay the attention to all major influencing factors.

There are many minor influencing factors, and we are unable to evaluate everything in the small research. And also we understand that many factors, including equipment efficiency, outside temperatures and others are not possible to predict 100 % and they can only be evaluated in the exact project. As an example, the outside temperature in different years will differ, depending on the climate change, sun activity, and other weather factors. The average year temperature in Kyiv region can differ 2–4 % [16], resulting in 1–2 % of the energy consumption results deviation according to our calculations above.

Datacenters can also benefit from using warm water cooling, as they can reuse the heat for different purposes, but that is already a different story.

VI. CONCLUSIONS

The first high performance clusters in Ukraine, based on the industry standard components, were built at the end of previous century. National Academy of Science of Ukraine invested into high performance cluster in early 2000s and as the result the system with 16 x computing nodes was built in the year 2003 in the Institute of Cybernetics of Ukraine. Back then, the efficient cooling was not a major topic on the table of supercomputing cluster builders as other topics seemed more viable and the cost of energy in Ukraine was dramatically low. As the price of electricity raised ~10 times during last 15 years and its costs heavily influence high performance computing projects in Ukraine, the importance of highly efficient cooling becomes obvious here as it has been already for many years in Europe and other countries, where the energy costs are high.

The research conducted in the article should help users, planning high performance computing system or any other high density computing system purchase/construction to make right decision, efficiently utilizing funds and what is even more important making their projects green and consuming less energy for the sake of humanity and their own country. The given above model gives an opportunity for the researchers to calculate their own power and cooling needs and financials for the power usage during 5 years of equipment use.

We conducted the research focusing on the discovery of the best technics and cooling methods comparison while building high density computing system. The exact data will vary depending on the exact compute and cooling equipment, but the general concepts will not change.

For the future work we are planning to investigate immersion cooling in details, as the most advanced and possible to implement more precise cooling and to build a prototype of the system, using standard immersion cooling liquids. The development of the new type of liquid for immersion cooling could also be one of the ways of our research development.

We believe in immersion cooling as it can be the most: Effective, Efficient, and Safe.

REFERENCES

[1] Golovinsky A. L., Malenko A. L., Modern methods of cooling supercomputer centers, Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, 2014

[2] HPE ProLiant Gen10 Server Extended Ambient Temperature Guidelines. Hewlett Packard Enterprise. 2017, 2018.

[3] HPE Power Advisor Tool. Hewlett Packard Enterprise. https://paonline56.itcs.hpe.com/ .

[4] Data Center Liquid Cooling – Global Market Outlook (2017–2023). StatisticsMRC, January 2018. ID code: SMRC16949.

[5] ASHRAE TC9.9, Thermal Guidelines for Data Processing Environments, Fourth Edition, ASHRAE, Atlanta Georgia, 2015.

[6] Salimifard, Parichehr. (2014). Comparison of part load model prediction of commercial office subsystem energy consumption with sub-metered data. 10.13140/2.1.1033.2963.

[7] High Temperature Electronics. F. Patrick McCluskey, Thomas Podlesak, Richard Grzybowski. CRC Press, 1997.

[8] Vysokie toki utechki. NVIDIA GeForce GTX 480 (GF100)", https://overclockers.ru/blog/RussOver/show/14701/Vysokie_toki_utechki_NVIDIA_GeForce_GTX_480_GF100

[9] MaxxAir IF18 3000-CFM 18-Inch Blade Heavy-Duty Exhaust Fan Specification. http://www.maxxair.com/products/maxxfan/maxxfan.php#

[10] James R. Welty; Charles E. Wicks; Robert E. Wilson; Gregory L. Rorrer (2007). Fundamentals of Momentum, Heat and Mass transfer (5th edition). John Wiley and Sons. ISBN 978-0470128688.

[11] Heat Transfer Coefficient. https://en.wikipedia.org/wiki/Heat_transfer_coefficient

[12] "Average overall heat transmission coefficients for some common fluids and surface combinations as Water to Air, Water to Water, Air to Air, Steam to Water and more". https://www.engineeringtoolbox.com/overall-heat-transfer-coefficients-d_284.html

[13] Thermodynamics. An engineering approach (7th ed.). Boles, Yunus A. Çengel, Michael A. (2011). New York: McGraw-Hill. p. 608.

[14] 35 kW Active Rectifier with Integrated Power Modules. Peter Wiedemuth, HÜTTINGER Elektronik GmbH + Co. KG, Germany. Serge Bontemps, Microsemi PPG power module Products, France. Johann Miniböck, m-pec, Austria and ETH Zurich, power electronic systems laboratory, Switzerland.

[15] https://rp5.ru/Weather_in_the_world

[16] https://birdinflight.com/ru/infografica/12042016-climate-kiev-100-years.html

Anatoliy Melnyk is a Head of Computer Engineering Department at Lviv Polytechnic National University since 1994. He graduated from Lviv Polytechnic Institute with the Engineer Degree in Computer Engineering in 1978. In 1985 he obtained his Ph.D in Computer Systems from Moscow Power Engineering Institute. In 1992 he received his D.Sc. degree from the Institute of Modeling Problems in Power Engineering of the National Academy of Science of Ukraine. He was recognized for his outstanding contributions to high-performance computer systems design as a Fellow Scientific Researcher in 1988. He became a Professor of Computer Engineering in 1996. Since 1982 to 1994 he has been a Head of Department of Signal Processing Systems at Lviv Radio Engineering Research Institute. Since 1994 to 2008 he has been Scientific Director of the Institute of Measurement and Computer Technique at Lviv Polytechnic National University. Since 1999 to 2009 he has been Dean of the Department of Computer and Information Technologies at the Institute of Business and Perspective Technologies, Lviv, Ukraine. He has served since 2000 as President and CEO of Intron ltd. He has also been a visiting professor at Kielce University of Technology, University of Information Technology and Management, Rzeszow, University of Bielsko-Biala. Currently, he is a visiting professor at the Department of Artificial Intelligence of John Paul II Catholic University of Lublin.

Volodymyr Saviak, Hewlett-Packard Enterprise AI, HPC & POD Manager in CEE countries

Interests: Artificial Intelligence, High Performance Computing, Performance Optimized Datacenters, Computer Architectures

More than 25 years designs IT solutions on the market of the Central Eastern Europe countries, focused on High Performance Computing, Big Data and Performance Optimized Datacenters. Volodymyr helps organizations understand benefits and implement technical computing solutions and gets most of their research. For the development of the "First Ukrainian Supercomputer Center" and Ukrainian HPC community, he was granted with a top level Ukraine's Parliament Award for Significant Achievements.