ALICE-INT-2004-014
CMS-IC-EN-0001
LHCb-2004-035 DAQ
19 April 2004

A water-cooling solution for PC-racks of the LHC experiments.

CERN Technical Note

Prepared By: Philippe Vannerem and Nuno Elias

Abstract

With ever increasing power consumption and heat dissipation of todays CPUs, cooling of rack-mounted PCs is an issue for the future online farms of the LHC experiments. In order to investigate the viability of a water-cooling solution, a prototype PC-farm rack has been equipped with a commercially available retrofitted heat exchanger. The project has been carried out as a collaboration of the four LHC experiments and the PH-ESS group[1]. This note reports on the results of a series of cooling and power measurements of the prototype rack with configurations of 30 to 48 PCs. The cooling performance of the rack-cooler is found to be adequate; it extracts the heat dissipated by the CPUs efficiently into the cooling water. Hence, the closed PC rack transfers almost no heat into the room. The measurements and the failure tests show that the rack-cooler concept is a viable solution for the future PC farms of the LHC experiments.

Document Status Sheet

Table 1 Document Status Sheet

1. Document Title: A water-cooling solution for PC-racks of the LHC experiments.			
2. Document Reference Number: [Document Reference Number]			
3. Issue	**4. Revision**	**5. Date**	**6. Reason for change**
Draft	1	6 Feb 2004	First draft
Draft	2	2 Mar 2004	Comments from Ph. Gavillet, Contributions from N. Elias
Draft	3	15 Apr 2004	Comments from A. Gaddi.
Final	1	19 Apr 2004	Minor spelling and rephrasing corrections.

[1] project collaborators:
Alice: A.Augustinus, S.Philippin.
Atlas: N.Elias, O.Jonsson, J.Godlewski, B.Martin, F.Wickens.
CMS: A.Gaddi, F.Glege, A.Racz.
LHCb: L.Brarda, B.Chadaj, G.Decreuse, D.Gasser, Ph.Gavillet, R.Lindner, D.Ruffinoni, Ph.Vannerem.
PH-ESS: P.Maley, V.Pittin, Ch.Parkman.
TS-CV: M.Santos.

Table of Contents

1. Introduction

1.1. PC racks for the LHC experiments

Each of the four LHC experiments will have an on-line PC farm for data-taking purposes. The PC farm will be a cluster of about 1000-2000 commodity PCs. The PCs are rack-mounted units that are cooled with internal fans, which provide a horizontal front-to-back airflow in the PC. High-density packing of PCs on shelves or in racks requires a solution to cool the air coming out of the PCs. Assuming e.g. 300W for a future PC, a rack with 30-40 PCs would require 9-12kW of cooling power per rack.

Industrial data centers mostly use traditional air conditioning units to cool the room containing PC racks. Best practice experience as of today however, allows for a maximum heat load in an air-conditioned room of about 1.0-1.2 kW / m2. This limit translates to a very large required surface for the data center needed to host the farm of an LHC experiment, if it were to be cooled with airco units.

Another solution is to use water-cooling for the PC farm, which allows for much higher density heat loads. A water-cooling solution is particularly attractive for the PC farms of the LHC experiments because existing water-cooling infrastructure for electronic racks at the experimental sites could be easily used for cooling farm racks.

This note reports about the results of an experimental setup in which horizontally rack-mounted PCs are cooled with a commercially available, vertically retrofitted heat exchanger, called rack-cooler hereafter. The concept of this cooling solution is shown in Figure 1. The aim of the rack-cooler tests are the following:

- validate the concept of using a heat exchanger to cool horizontal front-to-back air flows in rack-mounted PCs; measure the cooling performance of the rack-cooler, and eventual heat losses to the room.

- measure the CPU temperature and power consumption of individual PCs and power consumption of a complete rack of PCs under data processing conditions similar to the future conditions in the experiment.

- test cooling behaviour in different failure scenarios similar to the ones expected in the electronic rooms.

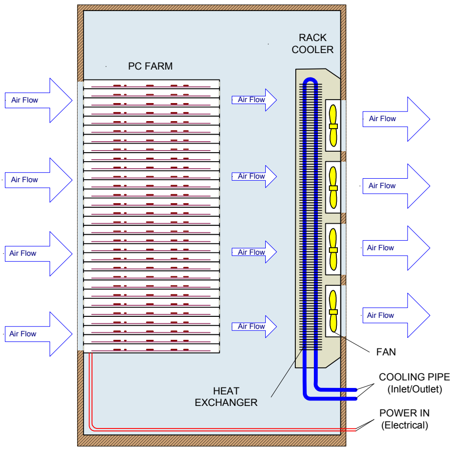


Figure 1: concept of cooling a PC-rack with a heat exchanger.

1.2. Thermodynamic parameters

In a simplified approach, the rack can be represented as a volume, through which a constant stream of air flows. The air is heated by the power dissipated in the PCs, P_{in}. This heat is removed again by the cooling water inside the heat exchanger (rack-cooler), corresponding to an outgoing power P_{out}. The thermodynamics of this system is described in detail in Appendix A.

The tested rack-cooler from Liebert has a cooling capacity of 8 kW (see section 2.1). This capacity might not be enough for a future PC rack, so here a calculation for a 10 kW rack-cooler is presented. Using the equations in the Appendix and using the desired operating conditions for PC racks in the future LHC farm rooms (water and air temperaturess), the required water flow, air flow and exchange surface of the optimal rack-cooler can be estimated. The results are shown in Table 2. For the calculations, the following assumptions have been made:

- the cold water station requires **15ºC/19ºC** for the inlet/outlet water temperatures, thus **ΔT_w=4ºC.**
- the temperature of the room should be maintained at **21ºC**, and the rack should return the air at the same temperature.
- inside the rack, the PCs heat the air up to **33ºC**.

Heat to remove	10 kW
Room Temperature	21 ºC
Inside the tubes	
• Fluid • Inlet Temperature • Outlet Temperature • Temperature difference **ΔT_w** • Flow	Water 15 ºC 19 ºC 4 ºC 2150 Kg/h
Outside the tubes	
• Fluid • Inlet Temperature • Outlet Temperature • Flow	Air 33 ºC 21 ºC 2980 Kg/h
Heat exchanger parameters	
• ΔT_{ln} • U x A	9.4 ºC 1060 W/ºC
For a realistic heat transfer coefficient U = 40 W/m^{2}K the required heat exchange surface A has to be:	
• A	26.5 m^2

Table 2: parameters for a 10 kW rackcooler

2. Test setup

2.1. Rack and rack-cooler

A rack previously used to host electronics for the LEP experiments is used in the setup. The rack has a standard width of 19inch, it is 56U high and 900mm deep. The rack with the mounted PCs is shown in Figure 2. The rack prototype was built up in the testlab of the PH-ESS group, where one could profit from available measuring infrastructure also used for cooling tests of electronic racks [1].

The rack-cooling unit tested is the commercially available RackCooler from Liebert [2]. The module is ~1.22 m long and weighs ~27 kg. It consists of a heat exchanger and four panel-mounted fans. Chilled water circulates through the module. The fans extract the hot air from the rack, run it through the heat exchanger and return it to the room at about room temperature. The design capacity of the RackCooler is 8 kW.

The rack-cooler has been mounted on the CERN rack with custom heavy-duty steel hinges. The hinges allow the unit to be moved out of the rack and to the side without being rotated and without twisting the water inlet and outlet flexible pipes. The movement to the side is required to assure access to the backside of the PCs for service purposes. Once the rack-cooler installed, the rack was made airtight in the back with panels covering the complete back plane. The panels have round holes just at the positions of the fans of the rack-cooler in order to allow air extraction.

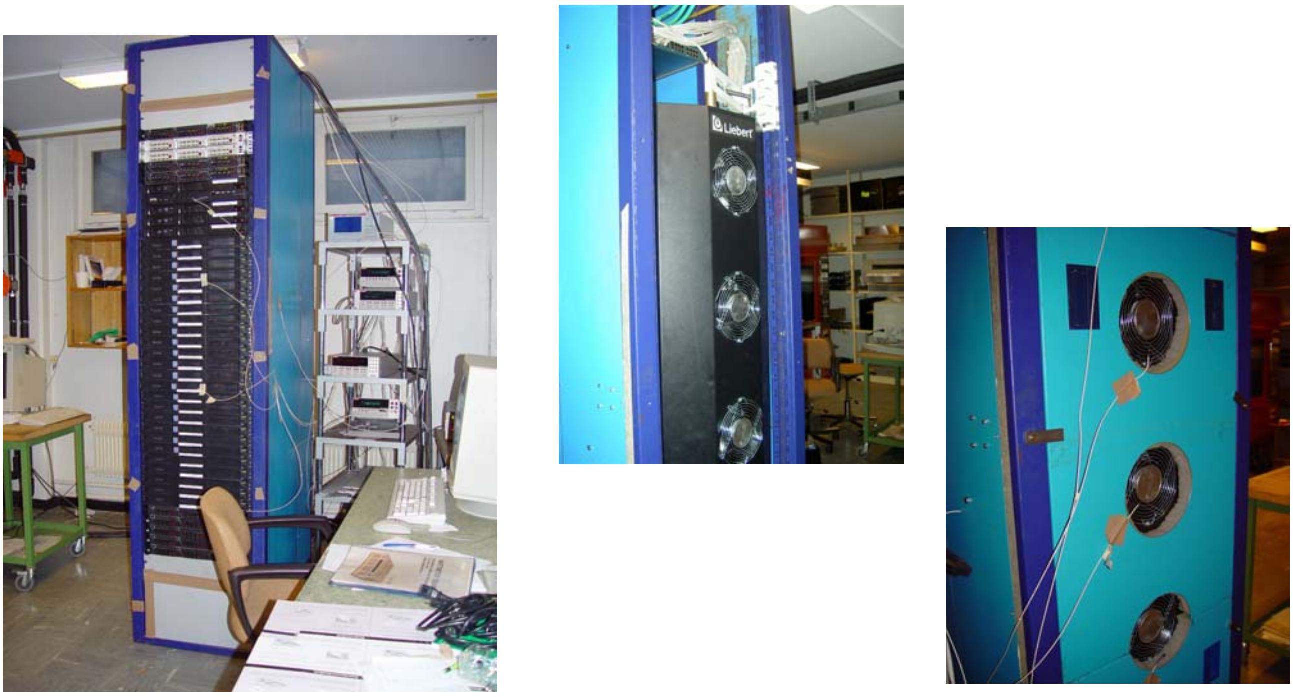


Figure 2: Pictures of the PC-rack with rack-cooler used in the test set-up.

2.2. Electrical distribution and cabling

The rack was equipped with a repartition box powered in 3-phase mode. The box distributes standard 220V to 4 CERN power distribution units equipped with differential circuit brakers that provide each 10 standard 220V outlets. The powering of PC-racks, including the startup procedure for a complete farm, is an issue on its own. First measurements to study the problems involved have been undertaken on this prototype PC rack by the TS-EL group and are documented in another note [3].

At the back, the space between the sliders of the PCs and the side panel of the rack was about ~15cm as recommended by best practice [4], and sufficient to pass the power cables and the ethernet cables, as can be seen in Figure 3. An 48-port 100Mbit ethernet switch from 3Com was installed at the top of the rack. Two ethernet connections per PC were foreseen, as in the future farm: one for the DAQ dataflow and one for the connection to the controls server.

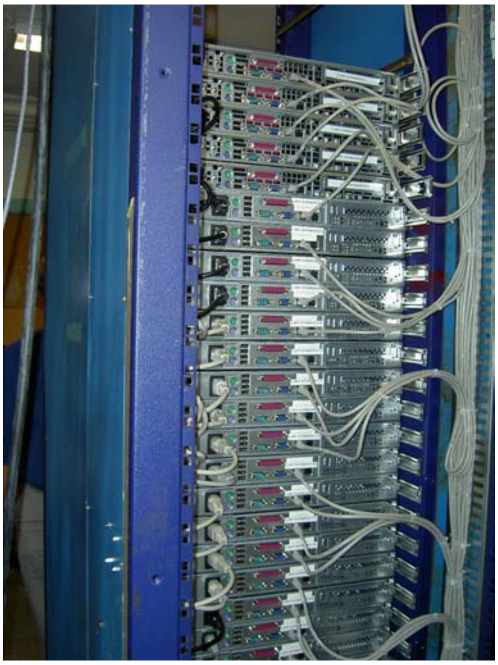


Figure 3: cabling of the PCs in the rack.

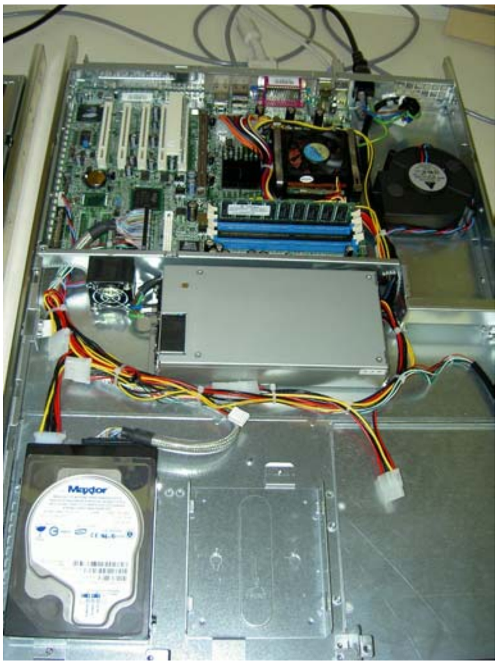
2.3. Mounting of the 1U PCs

The rack was equipped with custom rails to host 1U PCs. PCs from 2 different vendors, Melrow (NL) and Elonex (UK), were installed. These PCs were found to be cheapest available PCs complying to the recommendations of the CPU manufacturer for installation in rack-mounted boxes. PCs with the standard CPU speed for an office desktop at the time of purchase was chosen; on purpose not the highest speed available. Each PC has a 2.4 GHz Intel Pentium IV processor, two Intel Xeon processors in case of the Elonex dual CPU PC. The mono CPU machines have 512MB of RAM, a 40 GB harddisk and 2 network interface cards; they have a 250W PFC power supply. The pizza boxes contain commodity motherboards. The PCs have perforated front panels in order to serve as air inlet. Compared to common desktop PCs, the vendors have added an additional blower in the box which circulates air over the CPU and sends the hot air out through the perforated backside panel of the PC (Figure 1). It was observed that the physical dimensions (width, length) of the pizza box vary with the vendor and with the type of PC (mono CPU, dual CPU). Also the internal layout of the components (PWS, blower,…) may vary from one manufacturer to the other. It was found out during the test that the majority of the Melrow PCs had an unsuited heat sink mounted on the CPU, of which the fins were orientated perpendicularly to the air flow from the additional blower, whereas they should be parallel. This resulted in a higher operating temperature for the CPUs compared to the ones with suited heat sinks.

The number of mounted PCs increased during the tests, the various configurations are listed in Table 3. With each configuration, the rack was made as airtight as possible; empty positions were covered with panels.

Configuration	Melrow mono CPU	Elonex mono CPU	Elonex dual CPU
30 PCs	25	5	0
40 PCs	25	5	10
48 PCs	25	5	18

Table 3: The various configurations for the cooling tests.



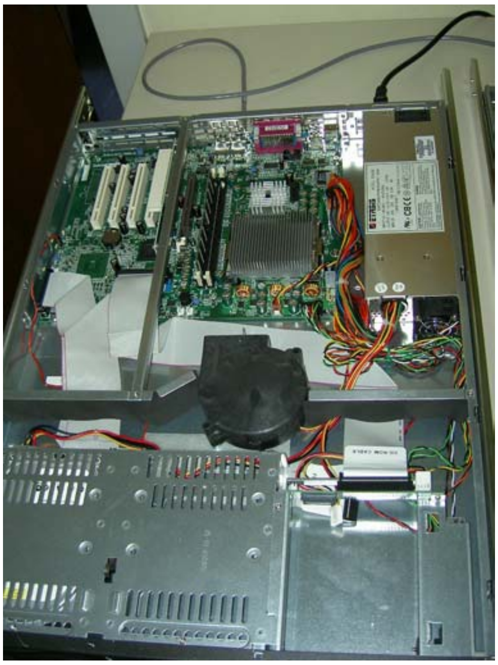


Figure 4: a look inside 1U mono CPU PCs from Melrow (left) and Elonex (right).

2.4. Measuring equipment

A PC measurement station running PVSS [5] has been collecting readings from several meters. A schematic view of the setup is shown in Figure 5. The power consumption of the total rack was measured with a Siemens Wattmeter. A flowmeter connected to a Keithley multimeter was used to monitor the water flow in the cooling pipes. Air temperatures at various positions in the rack and water temperatures were measured with PT100 probes connected to a Keithley multi-channel scanner that in turn was connected to a Keithley multimeter. The server PC had an National Instruments GPIB interface card and was running a for this purpose developed C++ program that uses the NI-GPIB libraries to collect the meter readings with an interval of 10s. Subsequently the program exported the readings to PVSS using DIM [6]. The collected data is archived by a standard PVSS archive manager. An overnight backup to a large diskserver was foreseen.

The temperatures of the CPUs is monitored with the standard linux lm_sensors package. Each PC in the rack runs a small Pcmon program which exports the temperature and the load of the CPU to the server PC with DIM at fixed 10s intervals.

Apart from collecting values, PVSS was also used to start/stop loading PCs with Monte Carlo processing jobs. The jobs were chosen so to simulate the future online event filtering. The LHCb event reconstruction package Brunel was used, processing digitized Monte Carlo data produced by the Boole package stored on the local disks of each farm PC.

A few interactive PVSS GUI panels provided a status display, a logbook function, trends of all measured values and a data export panel to extract data from the PVSS archive to a flat file in order to be analyzed with other programs.

The measurement of the dissipated power depends critically on the accuracy of the water temperature measurements at inlet T(in) and outlet T(out), because the typical increase of the cooling water temperature is only a few degrees. The PT100 probes were mounted on "fingers" that position them in the middle of the tube.

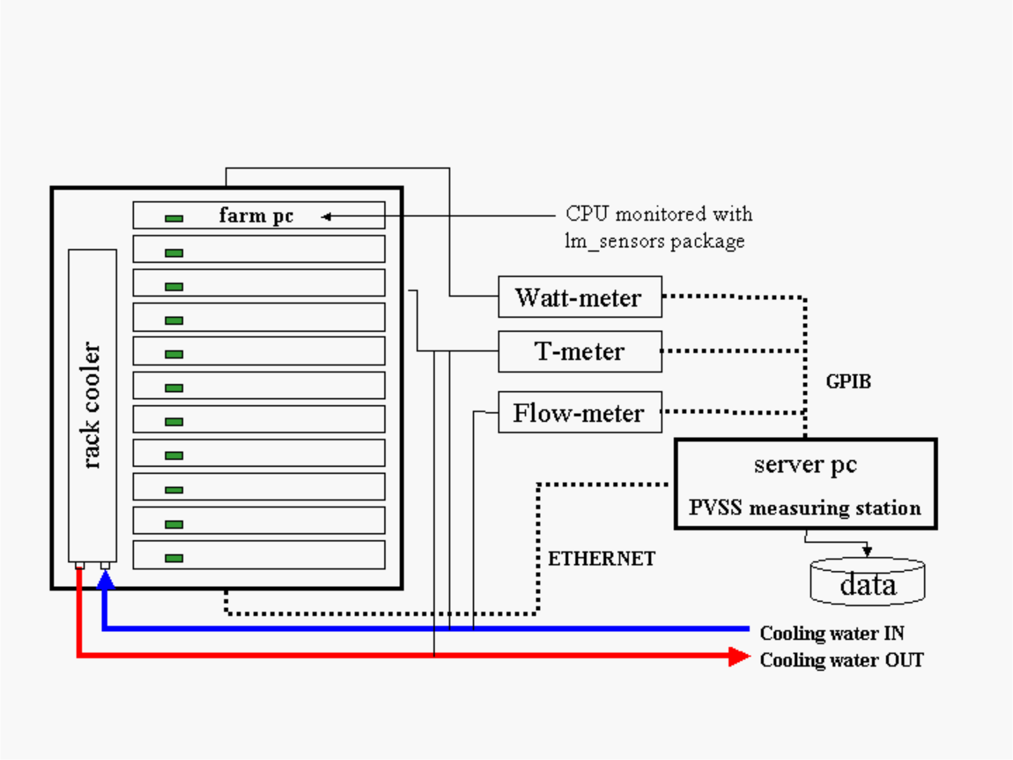


Figure 5: setup of the PC rack measurement station

The probes have been calibrated by the institute of metrology “G. Colonetti” in Turin. A quadratic fit to the (resistance, temperature) pairs was performed. The standard deviation of the residuals of the fit is 0.06 ºC, this is taken to be the accuracy for the water temperature measurements. A simple linear fit gives the same standard deviation. Hence this is consistent with the quadratic fit, and the quadratic term is negligible for these probes.

The water flow is measured with a flowmeter which passes a signal to a Keithley multimeter. Its accuracy is believed to be better than 1%.

3. Measurements and results

3.1. PC measurements

3.1.1. PC power consumption

The power consumption of individual PCs (PC configuration: see Section 2.3) has been measured with an Infratec wattmeter for three different states. The results are shown in Table 4. In the state “off”, the power chord of the PC is plugged in, but the PC is not running. In this mode the PC powers the network interface cards and the LEDs that indicate whether the PC is network connected, hence the consumption is not zero. The state “idle” is when the PCs have been booted, but have no active tasks other than the ones from the operating system, the CPU load in this state is very close to 0%. In the state “100% loaded” each CPU is running the LHCb Monte Carlo reconstruction job from the local disk, which takes the CPU load up to 100%. In the “100% loaded” state, dual CPU machines are running two reconstruction jobs. It can be seen from the table that when going from the idle state to 100% load the increase in power consumption is about 37W per CPU, independent of the manufacturer.

Models	Power consumption		
	Off	Idle	100% Loaded
Elonex mono CPU	(5.0 ± 0.5) W	(55.0 ± 0.5) W	(92.0 ± 0.5) W
	40 VA	100 VA	121 VA
Melrow mono CPU	(5.0 ± 0.5) W	(59.0 ± 0.5) W	(96.0 ± 3.0) W
	-	67 VA	103 VA
Elonex dual CPU	(9.0 ± 0.5) W	(89.0 ± 2.0) W	(167.0 ± 5.0) W
	24 VA	99 VA	180 VA

Table 4: Power consumption of individual rack-mounted PCs.

3.1.2. CPU operating temperatures

During the measurements the CPU temperatures were monitored on all the rack-mounted PCs using the linux package lm_sensors. The operating temperatures of the CPUs of the PCs, both when idle (load 0%)

open rack in air-cooled room - T(room)=23ºC		
	Load 0%	Load 100%
Elonex mono	32 ± 1	45 ± 2
Melrow mono	40 ± 3	52 ± 3
Elonex dual	30 ± 2	39 ± 2
closed rack with rack-cooler - T(room)=22ºC		
Elonex mono	29 ± 1	42 ± 1
Melrow mono	40 ± 3	52 ± 3
Elonex dual	28 ± 1	38 ± 1

Table 5: CPU operating temperatures in a rack. The average temperatures of the PCs of the same type are shown, the error quoted is the standard deviation.

and when processing (load 100%) can now be compared with and without rack-cooler. The results are shown in Table 5. For the PCs from Elonex the operating temperatures are observed to be slightly lower in a closed rack equipped with a rack cooler. This can be explained by the larger front-to-back airflow over the hot CPUs with the additional rack-cooler fans as measured in the next section. In case of the Melrow PCs there is no difference in operating temperatures observed; this is explained by the different internal layout, where the additional front-to-back airflow induced by the rack-cooler fans does not directly pass over the CPUs (see Figure 4). It can be concluded that the operating conditions for a PC in a closed rack equipped with a rack-cooler are good, and that the heat produced in the PCs is well evacuated. While this result might look trivial, it must be noted that in the open rack configuration the heat goes into the room, while in a closed rack with rack-cooler the heat is extracted into the cooling water, as will be seen in the next sections.

3.2. Air flow measurements

The air flows through PCs, rack-cooler and through a fully installed rack in the 40 PCs configuration have been measured with an anemometer. The flow measurements in Table 6 have been obtained by multiplying the measured airspeeds with the surfaces through which the flows go. The airflow intake is measured at the front of the PCs, the air flow out is measured at the back of the rack-cooler fans. The accuracy of the air speed measurement is first of all limited by the meter, one digit change gives an accuracy of about 10%. Because of inhomogenities in the flow and because of effects at the borders of the surfaces where the speed was measured, the absolute scale of the measurements is not believed to be more accurate than 15-20%. The systematic uncertainty is the same for all measurements.

Rack OPEN	
air flow intake by PCs	0.38 m^3/s
air flow out through rack-cooler	0.62 m^3/s

Rack CLOSED + rack-cooler OFF	
air flow intake by PCs	0.28 m^3/s
air flow out through rack-cooler	0.15 m^3/s
Rack CLOSED + rack-cooler ON	
air flow intake by PCs	0.46 m^3/s
air flow out through rack-cooler	0.55 m^3/s

Table 6: Air flow measurements for a rack configuration with 40 PCs.

From these measurements, the following observations can be made:

- the nominal airflow of the rack-cooler is reduced when it is installed in a closed rack, because the air has to be taken in through perforated front surfaces.

- with the rack-cooler in a closed rack, the generated air flow is about 20% stronger than the air flow generated by the PC fans in an open rack or room.

- in a closed rack without the rack-cooler fans running, the air flow generated by the PC fans alone is about 25% less than for PCs in an open rack or room. Hence, in a rack-cooler fan failure situation the airflow through the PC is reduced, but still significant.

- the rack was not completely airtight, there are some holes and gaps between the PCs and at the pass-through of cables, where additional air goes in. It will not be possible to eliminate this effect in a later large-scale farm, but attention will have to be given to large gaps and missing closing panels.

3.3. Dissipated power measurements

A simplified scheme for the total energy balance for the rack is shown in Figure 6. The electrical power supplied to the rack is converted into heat by the PC power supplies, the CPUs and the rest of the electronics in the PCs. Part of this heat will be taken out by the water circulating through the rack-cooler. It is a critical measurement to determine what fraction of dissipated power is evacuated from the rack by the water-cooling system and what fraction is lost to the environment. The power lost to the environment is constant when the temperature difference between the water inlet and outlet stabilises, hence we can make the energy balance of the rack:

$$P_{lost} = P_{in} - P_{out}$$

Heat can be lost to the environment first of all by returning air at a temperature above the intake temperature ($P_{air} > 0$), secondly by natural convection - hot side panels of the rack exposed to room temperature ($P_{side} > 0$) - and thirdly by black-body radiation. The last effect is however negligible for our tests. P_{side} could in principle be reduced by proper isolation of the side panels, but this was not undertaken as it is not foreseen in the future. But in a future farm barrack, PC racks will be side to side, thus reducing heat flow through the side panels if both racks are at the same temperature. For the power lost to the environment we can write:

$$P_{lost} = P_{air} + P_{side}$$

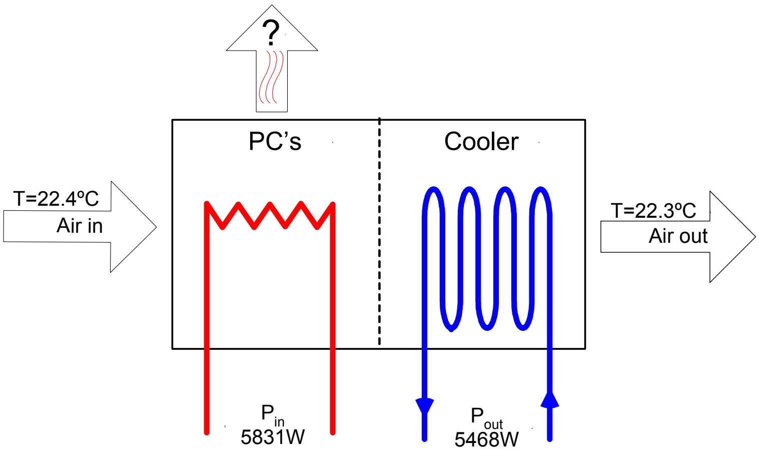


Figure 6: simple model for energy balance of a PC rack, with measurements for the 48 PCs.

The power P_{out} dissipated in the cooling water can be calculated from the measured water mass flow and from the increase of the cooling water temperature in the following way:

$$P_{out} = \dot{m}_{water} \cdot \Delta T \cdot C_p$$

where C_p is the specific heat capacity of water, taken to be 4.186 J.Kg^{-1}.K^{-1} for our tests. P_{out} has been measured for the various PC configurations of the rack. Figure 7 shows P_{out} as a function of P_{in}. The measurements are summarized in Table 7. The different power consumptions for a given configuration are obtained by varying the CPU load of the PCs in the rack from idle (0%) over half- (50%) to fully loaded (100%). In case of the 30 PCs configuration the measurements of ΔT are averages over 5 minutes, in case of 40 and 48 PCs averages over an hour are taken. The error bars are calculated from the estimated accuracies on the measurements of temperatures and flow as explained in Section 2.4; the errors are estimated to be about 4%.

The comparison with optimal cooling in the figure shows that within the error bars, up to 48 PCs can be efficiently cooled by the rack-cooler. The difference of cooling performance for the highest point of the 30 PC configuration and the lowest point of the 40 PC configuration is explained by a higher cooling water temperature for the 30 PCs. The series for the different configurations cannot be compared directly, because a different number of PCs leads to a different front surface through which cooling air is taken in, resulting in a slightly different airflow. The magnitude of the airflow is an important parameter for the cooling system as such. For the configuration with 48PCs loaded at 100%, the total consumed power P_{in} is 5.83 kW, while the power extracted by the rack-cooler, P_{out}, is 5.46 kW; this corresponds to an estimated P_{loss} of about (0.4 ± 0.2) kW or about (6 ± 4) % of P_{in}. The difference in temperature of the air in front and at the back of the rack is very small, giving an estimated P_{air} = -35W[1]. Thus the heat loss in this configuration is mainly because of a heat flow through the side panels with P_{side} ~ 0.4 kW

An important systematic check of the measurements consisted of swapping the water inlet and outlet pipes (see Figure 8). Because the heat transfer through the heat exchanger does not depend on the direction of the

[1] Taking in consideration $\dot{m}_{air} \cdot C_{p,air} = 630$ W/°C using the measured flow of air.

water flow, the difference between the ΔT 's measured in both positions (called the bias on ΔT hereafter) should be negligible.

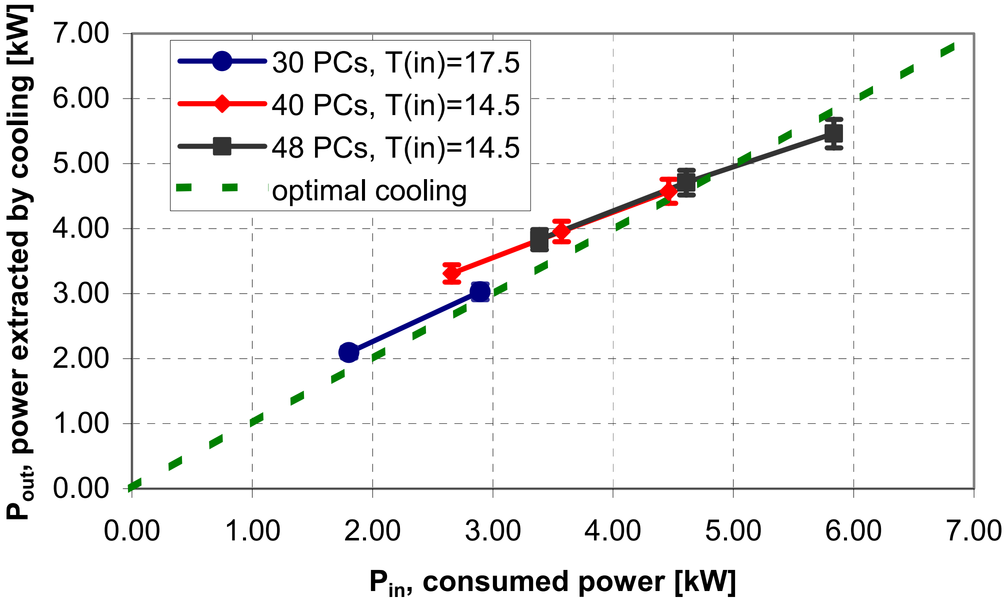


Figure 7: measured power extracted by the rack-cooler as a function of the total power consumption of the PC rack.

Configuration	CPU load	P_{in} [kW]	T_{in} [ºC]	T_{out} [ºC]	P_{out} [kW]
30 PCs	0 %	1.8	16.4	18.1	2.3
30 PCs	100 %	2.9	16.5	18.7	3.0
40 PCs	0 %	2.7	14.5	17.0	3.3
40 PCs	50 %	3.6	14.4	17.4	4.0
40 PCs	100 %	4.5	14.6	18.0	4.6
48 PCs	0 %	3.4	14.6	17.4	3.8
48 PCs	50 %	4.6	14.6	18.1	4.7
48 PCs	100 %	5.8	14.7	18.8	5.5

Table 7: measurements of P_{in} and P_{out} for different PC rack configurations and CPU loads.

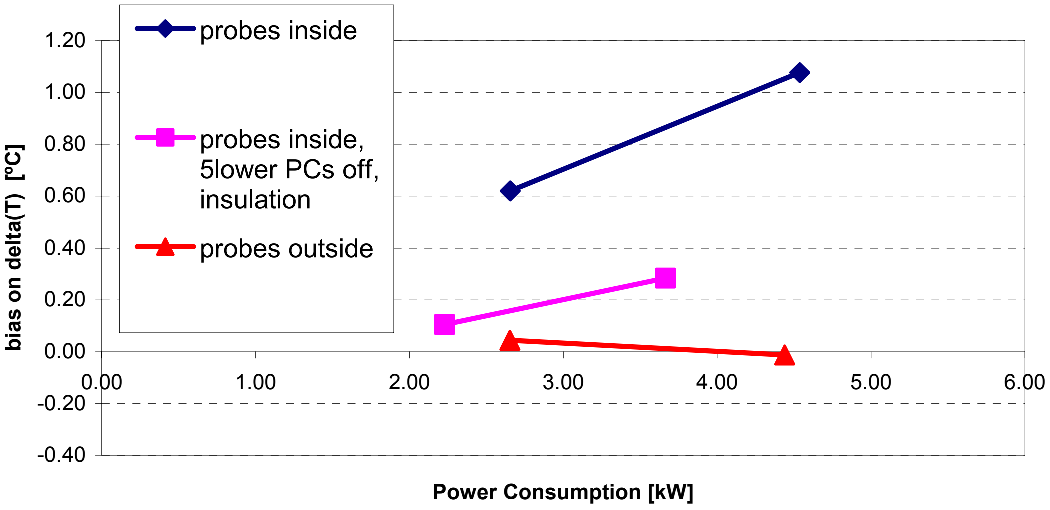


Figure 8: bias on the ΔT measurement for different probe positions.

An initial bias was discovered at the beginning of the measurements. The problem was traced back to the position of the water temperature probes; inside the rack and exposed to hot air coming from the back of the PCs. After relocating the probes outside the rack this bias was reduced to almost zero as can be seen in Figure 8. This systematic check gives confidence that the accuracy of the water temperature measurement is truly the accuracy of the probes itself.

The dependence of the cooling performance on the temperature of the incoming cooling water has been checked by varying the cooling water temperature. The measurements are shown in Figure 9. A slight dependence is observed. This is obviously expected because an increase in the cooling water temperature leads to a smaller ΔT between the hot air coming from the PCs and the surface of the heat exchanger, thus diminishing the heat transfer and decreasing the cooling performance.

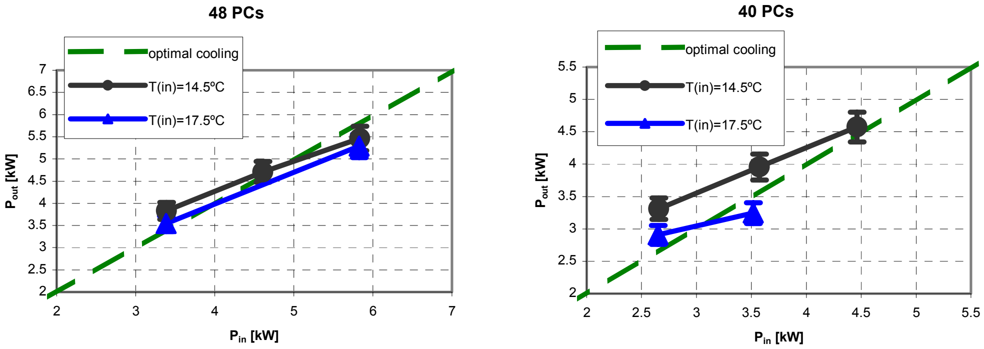


Figure 9: The dependence of the cooling performance on the cooling water temperature. All measurements are for the configuration with 48 PCs.

3.4. Air temperature measurements

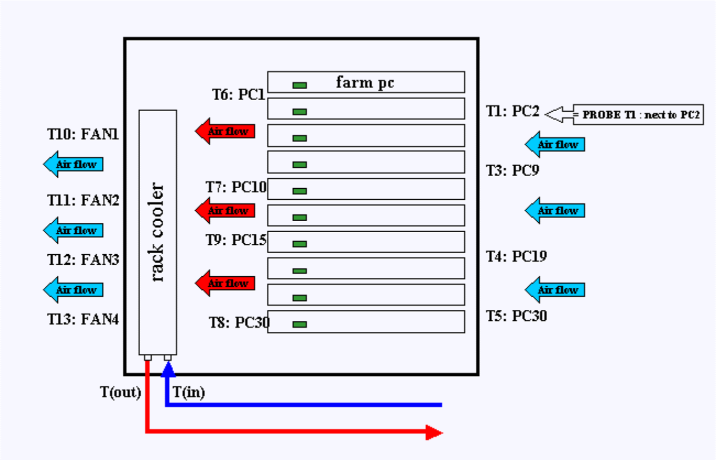


Figure 10: The position of the PT100 probes T1…T13 to measure the air temperature for the rack configuration with 30 PCs.

In order to measure the air temperature PT100 probes have been placed at various positions in and around the rack as shown in Figure 10. For this purpose, uncalibrated PT100 probes were used. A test was made in which all probes were placed together to measure the same room temperature, the standard deviation of the values of 18 different probes was 0.3 ºC , this is taken to be the accuracy for the air temperature measurements. In the first configuration with 30 PCs, 4 probes were put in the front of the rack, attached to the front panels of the PCs exactly where air is taken in, 4 probes were put in the middle of the closed rack, at the air outlets of the same PCs, and 4 probes were put after the rack-cooler fans, where the air comes out of the rack. In later configurations with 40 PCs some probes were added in the middle of the rack, after the air outlets of the additional dual CPU PCs.

The spatial distribution of the air temperature probes allows to map the cooling of the air. A complete map with 1-hour averages from 12 probes is shown in Table 6. The following effects can be observed:

- The air temperature is almost uniform at the front of the PCs, only small differences between the probes at the front are observed.

- In the middle of the rack, behind the air outlets of the PCs, the temperature of the air has increased significantly. The increase depends critically on the PC load; the air is about 5ºC hotter for a fully loaded PC than for an idle PC. The differences between the individual probes are bigger in the middle. This might be due to the exact position of the probe with respect to the hot air outlet of the PC where it was attached. Another probable explanation is that in the 40 PC configuration, 5 additional dual CPU PCs were added in positions above the 30 single CPU PCs, and 5 were added below; hence the most upper and the most lower probe in the middle of the rack were closest to the dual CPU PCs, which have longer chassis boxes than the other PCs.

- At the back of the rack, the air is coming out at a temperature which is close to the temperature at the intake, both for idling PCs and for 100% loaded PCs. Therefore, there is efficient cooling of the air coming out of up to 48 PCs. This observation is independent from the measurement of the increase of the cooling water temperature, and is consistent with the results from these measurements. Hence there is only a heat loss through the side panels of the rack.

PC Load 0% - 40 PCs		
Front	Middle	Back
21.1 ± 0.3	26.3 ± 0.3	20.7 ± 0.3
21.1 ± 0.3	27.4 ± 0.3	19.1 ± 0.3
21.1 ± 0.3	25.9 ± 0.3	20.2 ± 0.3
20.8 ± 0.3	26.0 ±0.3	19.6 ± 0.3
Front Average	Middle Average	Back Average
21.0 ± 0.1	26.5 ± 0.7	19.9 ± 0.7
PC Load 100% - 40 PCs		
Front	Middle	Back
21.8 ± 0.3	33.2 ± 0.3	22.4 ± 0.3
22.0 ± 0.3	30.4 ± 0.3	20.7 ± 0.3
21.9 ± 0.3	29.6 ± 0.3	21.8 ± 0.3
21.5 ± 0.3	31.7 ± 0.3	21.2 ± 0.3
Front Average	Middle Average	Back Average
21.8 ± 0.2	31.2 ± 1.6	21.5 ± 0.8

PC Load 0% - 48 PCs		
Front	Middle	Back
21.3 ± 0.3	27.3 ± 0.3	20.0 ± 0.3
21.5 ± 0.3	28.5 ± 0.3	20.1 ± 0.3
21.4 ± 0.3	28.2 ± 0.3	20.8 ± 0.3
21.1 ± 0.3	27.3 ±0.3	20.0 ± 0.3
Front Average	Middle Average	Back Average
21.3 ± 0.2	27.8 ± 0.7	20.2 ± 0.4
PC Load 100% -48 PCs		
Front	Middle	Back
22.3 ± 0.3	34.9 ± 0.3	22.5 ± 0.3
22.6 ± 0.3	33.5 ± 0.3	22.2 ± 0.3
22.6 ± 0.3	33.3 ± 0.3	22.6 ± 0.3
22.1 ± 0.3	34.1 ± 0.3	22.1 ± 0.3
Front Average	Middle Average	Back Average
22.4 ± 0.3	34.0 ± 0.7	22.4 ± 0.2

Table 8: Air temperature measurements for a PC-rack with rack-cooler, for different configurations. Twelve probes are positioned as shown in Figure 10. For individual probes the values are from 1-hour measurements, the errors are the estimated probe accuracy. The averages are shown, the errors are the standard deviations of the 4 probes.

An important result from the air temperature measurements is that the horizontal airflow through the PCs is to a good degree uniform in temperature. There is no apparent temperature gradient from higher to lower positions in the rack. This is clearly an advantage over systems with vertical air flow such as electronic racks where the air is taken in from a raised floor, resulting in a vertical gradient of temperature.

It should be remarked that in the configuration with 48 PCs the 9 PCs in the highest positions and the 9PCs in the lowest positions of the rack were not directly facing the surface of the heat exchanger, but nevertheless cooled as well as the central PCs. The strong airflow due to the additional rack-cooler fans effectively evacuates the hot air from the PCs independently of their position in the rack.

Because of homogeneous temperatures, one can average over the four probes at front, middle and back of the rack and present the average temperature evolution from front to back of the rack. This is shown in Figure 11. Starting from the room temperature as the reference value, it can be seen that independent from the load and the number of PCs, the heated air is cooled back to a temperature that goes slightly below room temperature.

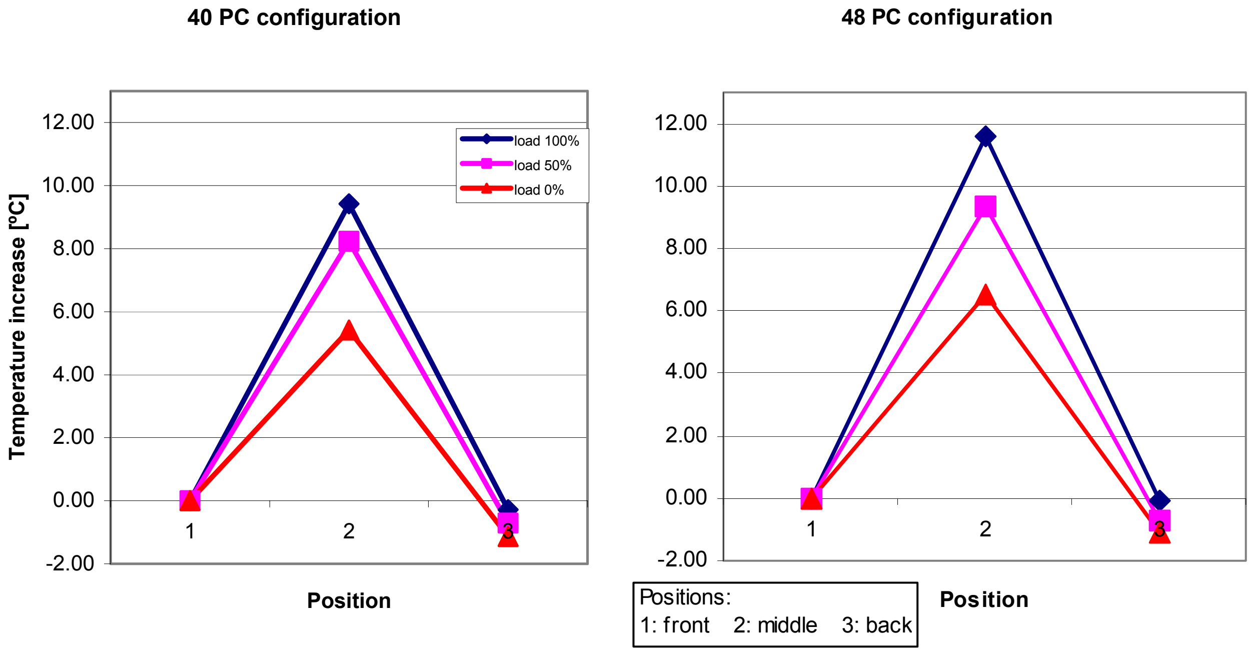


Figure 11: Air temperature evolution for a rack with rack-cooler with 40 PCs and 48 PCs.

3.5. Rack-cooling dynamics

During our tests the initial parameters were changed in order to analyse the behaviour of the rack-cooler for different measurement conditions. The power dissipated in the PC's, was one of the main parameters to be changed (from 1.8kW until 5.8kW), other critical parameter was the water inlet temperature with several values being adopted (14.5ºC, 16.5ºC and 17.5ºC).

The performance of the rack cooler (as a heat-exchanger) depends on the temperatures of both fluids, area of heat transfer, and heat transfer coefficient, which means that the rack-cooler performance varies for each of these conditions, and that for conditions different from the design point will represent positive or negative heat loss to the room, meaning an increase or decrease of the temperature of the room. But over time the room temperature will converge to a stable operating point in the following way.

In the case when the heat loss is positive, (consequently increasing the room temperature) the cooler does not have enough cooling capacity, and the air exits the rack at a higher temperature than the room. After some time, the 'hot' air leaving the rack mixes with the environment air, and the room temperature increases. At that point the PC's will be receiving air at an higher temperature than in the initial conditions, thus for the same dissipated power on the PC's a higher temperature will be reached inside the rack. A higher temperature in the middle of the rack means a higher temperature difference between the two fluids, thus promoting a better heat exchange on the rack cooler, and consequently removing more heat than initially. After some time, the room temperature will stabilize at a higher temperature and the difference $\mathbf{P_{IN} - P_{OUT}}$ will decrease.

In the opposite case, when the heat loss is negative, the room temperature decreases and also the power removed on the rack-cooler will be reduced.

It is not easy to quantify what operating point will be reached in each of these situations. But in both cases it is possible to note a balancing effect from the rack-cooler behaviour, the difference $|\ \mathbf{P_{IN} - P_{OUT}}|$ is decreased by a consequent adjustment of the room temperature. For small deviations from the design parameters, a stable operating point at a room temperature close to the room temperature foreseen in the design will be reached by the above dynamical mechanism.

3.6. Failure tests

After having verified that the rack-cooling performance is as expected, the overall behaviour of the PC rack with rack-cooler was investigated in failure situations that are expected at the future farm site. The motivation to conduct failure tests such as fan failures and cooling water stop is to find out what would be additional requirements, if any, for the Detector Control System (DCS) and the Detector Safety System (DSS) for an electronics room with a few tens of PC racks.

3.6.1. Fan failures

In order to simulate a fan failure under high load conditions, the fans of the rack-cooler have been stopped while the system load of the total rack was 100%. Meanwhile, the cooling water went on circulating through the heat exchanger. The time evolution of the air temperatures in the rack is shown in Figure 12. Because the heated air is not any longer extracted from the rack in an efficient way, the PC boxes and the complete rack structure starts heating up, as can be seen from the probe attached to the side of the rack. Under these conditions it is critical to check the CPU temperatures of the PCs.

The following observations can be made:

- though the temperature in the middle of the rack goes up quickly, it reaches a plateau after more than an hour. This is probably due to the airflow induced by the PC fans only. As can be seen from the evolution of the water temperature, the ΔT of the water is reduced, but the remaining airflow results in an average extracted power of about 3.0 kW for a given consumed power of 5.8 kW.

- the small remaining airflow through the PCs is returned to the room at a temperature lower than the temperature at the intake. This is explained by the reduced speed of the air and the interaction with the cooling water in the heat exchanger. This again indicates that in this situation the heat is escaping from the rack through the side panels.

- The temperature of the side panel of the rack goes up significantly; the air flow in the rack is not enough to remove all the heat produced by the CPUs, additional heat flows through the side panels. This leads to a rise of the room temperature as can be seen from the rising T_front.

- Taking into account the increase in room temperature, it can be seen that the working temperature of the CPUs reaches a plateau that is about 8-10 ºC above the normal operating temperature in a cooled rack after about 1 hour. These values are still below the critical maximum operating temperature as specified by the CPU manufacturer.

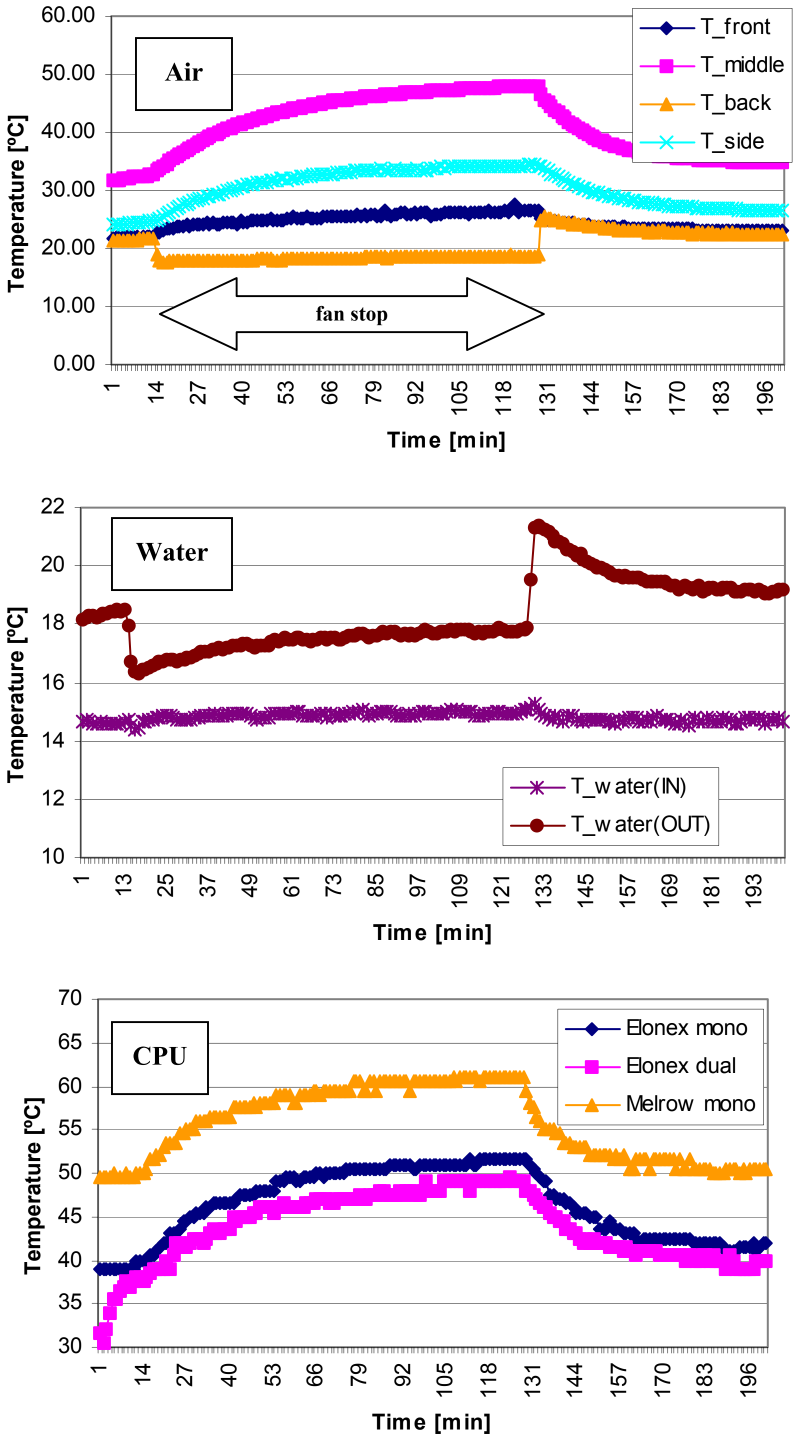


Figure 12: evolution of air, water and CPU temperatures after a fan failure.

3.6.2. Cooling water stops

The second failure situation that was simulated in the tests is a stop of cooling water flow, while the fans keep working. The time evolution of the air temperatures in the rack and of the CPU temperatures is shown in Figure 13. The following observations can be made:

- as soon as the cooling water stops, the hot air from the PCs is returned to the room at a temperature that is only slightly below the hot air temperature. At the same time the temperature of the sidepanel starts rising. So the whole rack starts warming up including the heat exchanger, as can be seen from the temperature evolution of the cooling water inside. This is a slow process because it takes a lot of energy to warm up rack+PCs+heat exchanger, at a rate of about 6ºC per hour in test setup.
- The temperature of the water in the heat exchanger starts to rise slowly to converge with room temperature.
- The CPU temperatures rise more slowly than in the case of the fan failure test, only 6-9 ºC above normal operating temperature in the rack after 2 hours of running without cooling water. They stay far below critical operating temperatures as specified by the manufacturers. The slow rise is explained by active extraction of the hot air into the room, the PCs are still cooled, but now by air that is steadily increasing in temperature.

3.6.3. PC rack and farm monitoring recommendations

From the failure tests it follows that:

- Rack-cooler fan failures for a full rack could be handled by DCS, provided that a fan failure signal is available. Single rack-cooler fan failure is even less critical and can be handled as in the previous situation.
- A situation where all the rack-cooler fans of all the racks would fail simultaneously while the PCs are still running is very improbable because both take their power from the same source.
- a cooling water stop would lead to a quick increase in room temperature. This situation is easily detected by the monitoring of the cooling water flow by DCS, and eventually by the DSS temperature probes in the farm room.
- Overheating of individual CPUs and eventual switch offs at certain tresholds can be handled by DCS, by integrating CPU temperature monitoring as in the test setup. Additionally, modern CPUs have an auto switch-off feature in the PC itself above a certain treshold.
- in critical situations where PCs would suddenly start burning, the air flow induced by the rack-cooler fans would extract eventual smoke quickly to the room where the smoke sniffers of the CSAM system, complement to DSS, would detect this eventually and cut the power of the complete room.

It is believed that all critical failure and safety situations are covered by the above descriptions. It is found that the planned DCS and DSS implementations cover these situations. On the other hand an additional fan failure signal going from the rack-cooler directly to the DCS system would allow a faster switch-off of the faulty rack, hence provide a better protection of the equipment.

In order to prevent complete destruction of the farm equipment in case of fire, an additional fire protection system for IT fabric environments might be installed. This is subject of another study [7].

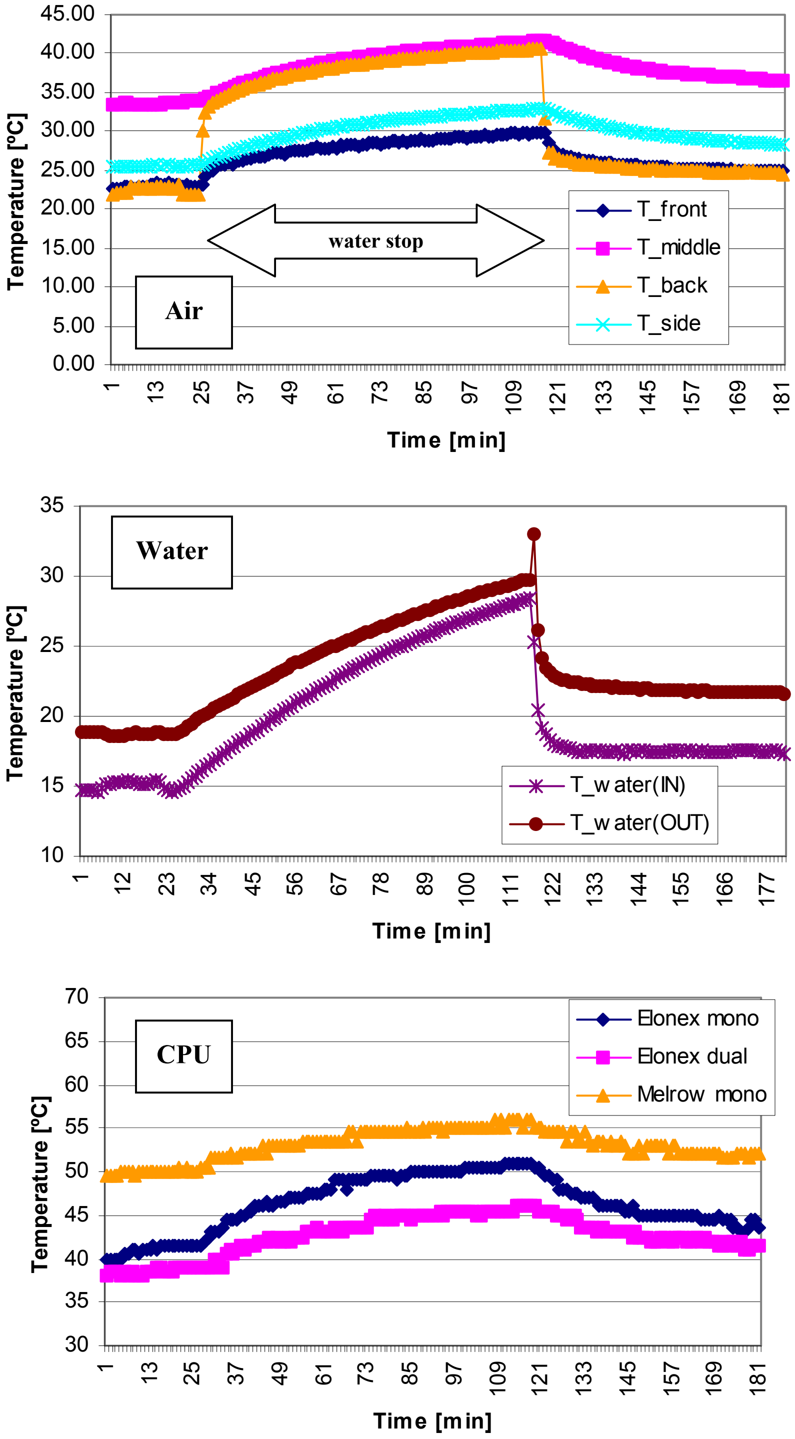


Figure 13: evolution of air, water and CPU temperatures after a cooling water stop.

4. Conclusions

A range of tests and measurements on a 19-inch rack housing rack-mounted PCs and cooled by a retrofitted heat exchanger have been presented. CPU temperature measurements and airflow measurements show that the operating conditions for PCs are very good in an airtight rack. The front-to-back airflow through the PCs is slightly higher than without a rack-cooler due the extra rack-cooler fans. Temperature mapping measurements around and inside the rack show that the airflow that is almost homogeneous in temperature, providing an effective cooling of the PCs independently from their positions in the rack.

The power consumption of standard Intel 2.4 GHz CPUs resulted in a total power consumption of 5.8 kW for a rack configuration with 30 mono-processor PCs and 18 dual-processor PCs, 48 rack-mounted PCs in total, when the CPUs are 100% loaded with Monte Carlo processing jobs. The heat produced in the rack can be effectively cooled away with the tested retrofitted rack-cooler. For this configuration a power loss of (6±4)% of the consumed power has been measured, where the heat loss is mainly due to heat flows through the side panels, which are exposed to ambient air of the room.

Both situations with a failure of the rack-cooler fans or with a cooling water stop were not immediately critical (>2hours) for the operation of the PC-rack. The ambient temperature starts to rise in both situations, but the complete rack with PCs has to be warmed up by the hot air from the PCs, which is a slow process. The CPU temperatures stayed below the critical operating temperatures, and no immediate action had to be taken. All failure scenarios for PC racks at the experimental site can be dealt with by the current DCS and DSS plans if one adds a fan failure signal to the Detector Control System.

The above results leads to the conclusion that a retrofitted heat exchanger proves to be a viable solution for the cooling of the future large-scale PC farms of the LHC experiments.

Acknowledgements

We would like to thank the PH-ESS group for their hosting of the test setup and for their support of the project.

5. Appendix: Rack-cooling thermodynamics.

Rack-cooling using the heat exchange principle

In a simplified approach, the rack can be represented as a volume, through which a constant stream of air flows. Inside the volume we have an input of heat P_{in}, representing the power dissipated in the PC's, and a heat output P_{out} representing the heat removed via cooling water inside the heat exchanger (rack cooler).

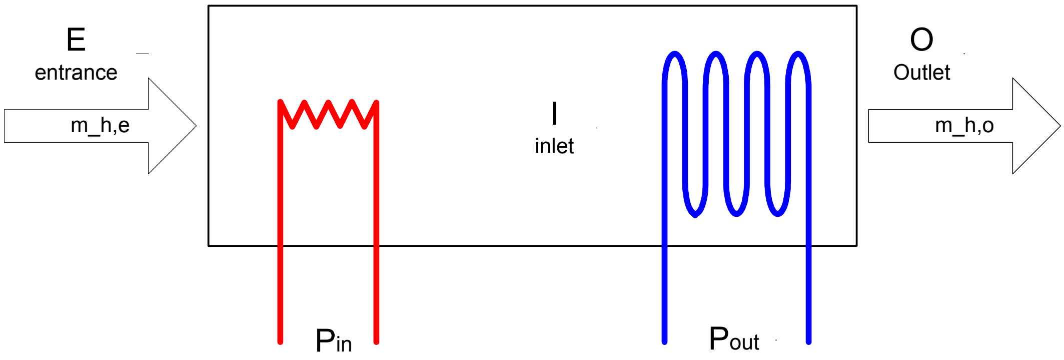


Figure 12: simple model for a PC rack.

The rack cooler is an air-water compact heat exchanger, working in cross flow. The cold fluid (water) circulates inside a tube. Outside the tube a stream of air crosses an aligned or staggered arrangement of tubes.

In order to increase the heat transfer coefficient (**U**), and the exchange surface (**A**), extended surfaces (fins) are introduced in the outer surface of the cooling pipes.

The amount of heat exchanged between the 2 fluids is:

$$P_{out} = U \times A \times \Delta T_{\ln} \qquad \textbf{Eq. 1}$$

Where:

$$\Delta T_{\ln} = \frac{\Delta T_1 - \Delta T_2}{\ln(\Delta T_1 - \Delta T_2)} \qquad \textbf{Eq. 2}$$

Figure 13 illustrates the temperature distribution inside the heat exchanger, and the used symbols.

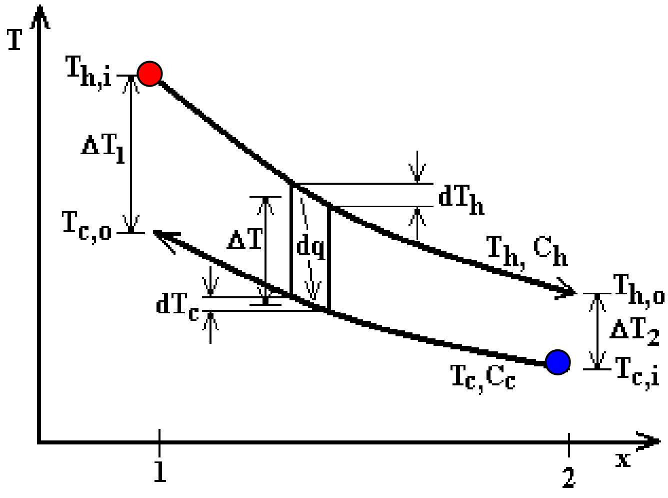


Figure 13: Temperature distributions in a counter-flow heat exchanger

Thermodynamical equations

A constant volumetric airflow is assumed here. Taking in consideration that the average temperature between 'i' and 'm', is the same as between 'm' and 'o', we can assume that the air mass flow is constant along the rack[2]

$$\dot{m}_h = \rho \times \dot{v}_h = \dot{m}\_h,e = \dot{m}\_h,o \qquad \textbf{Eq. 3}$$

We can assume also a constant mass flow, **m$_h$.**

The electrical power dissipated in the PC's due to joule effect, can be expressed according with the following equation:

$$P_{in} = \dot{m}_h \times Cp_{air} \times (T_{h,i} - T_{h,e}) = \dot{m}_h \times Cp_{air} \times \Delta T_{h,e\_i} = \dot{C}_h \times \Delta T_{h,e\_i} \qquad \textbf{Eq. 4}$$

$$\boldsymbol{\Delta T_{h,e\_i} \propto P_{in}}$$

The power taken out by the rack cooler (air side) is:

$$P_{out} = \dot{m}_h \times Cp_{air} \times (T_{h,o} - T_{h,i}) = \dot{m}_h \times Cp_{air} \times \Delta T_{a,i\_o} = \dot{C}_h \times \Delta T_{h,i\_o} \qquad \textbf{Eq. 5}$$

$$\boldsymbol{\Delta Ta_{i\_o} \propto P_{out}}$$

The power taken out by the rack cooler (water side) is:

$$P_{out} = \dot{m}_c \times Cp_{water} \times (T_{c,i} - T_{c,o}) = \dot{m}_c \times Cp_{water} \times \Delta T_{water} = \dot{C}_c \times \Delta T_c \qquad \textbf{Eq. 6}$$

[2] Assuming an air temperature of around 20ºC at inlet and 30ºC in the middle, $\rho_{[25ºC]}$=1.1717kg/m^3. The value of ρ at 20ºC is 1.1919kg/m^3, and at 30ºC is 1.1522kg/m^3

$$\Delta T_c \propto P_{out}$$

Re-arranging **eq.5** and **6]**

$$\frac{\dot{C}_h}{\dot{C}_c} = \frac{\Delta T_c}{\Delta T_{h,i\_o}} \qquad \textbf{Eq. 7}$$

The energy balance for our model can be written as:

$$\Delta P = P_{in} - P_{out} = \dot{m}_h \times Cp_{air} \times \Delta T_{a,e\_o} \qquad \textbf{Eq. 8}$$

A rack-cooler designed to remove all heat dissipated in the PC's would mean:

$$\Delta P = 0 \text{ and thus, } Ta_I = Ta_O$$

6. References

[1] G. Thomas, “Rack cooling project – final report”, CERN report EPESS-2003-014, 08 August 2002.

[2] Liebert. http://www.liebert.com

[3] K. Kahle, “Measurements of rack-mounted PCs for LHC”, CERN EDMS document nº 442180, January 2004.

[4] Server System Infrastructure initiative. http://www.ssiforum.org/docs/server_rack_spec_v1_1.pdf

[5] PVSS II, ETM AG, http://www.pvss.com.

[6] C. Gaspar et al., “DIM, a Portable, Light Weight Package for Information Publishing, Data Transfer and Inter-process Communication”, proceedings of CHEP 2000, Padova.

[7] A. Gaddi, private communication, http://www.hi-fog.com.