Technological University Dublin

ARROW@TU Dublin

Conference Papers

School of Manufacturing and Design Engineering

2018-9

Hybrid-Cooled Data Center Server Layout Optimization for Air-Side Heat Recovery

Assel Sakanova
Trinity College Dublin, Ireland, sakanova@tcd.ie

SAJAD ALIMOHAMMADI
sajad.alimohammadi@tudublin.ie

Jaakko McEvody
Trinity College Dublin, Ireland, jmcevoy@tcd.ie

Tim Persoons
Trinity College Dublin, Department of Mechanical and Manufacturing Engineering, Dublin, Ireland, tim.persoons@tcd.ie

Follow this and additional works at: https://arrow.tudublin.ie/engschmanconn

Part of the Mechanical Engineering Commons

Recommended Citation

Hybrid-Cooled Data Center Server Layout Optimization for Air-Side Heat Recovery

Assel Sakanova[1], Sajad Alimohammadi[1,2], Jaakko McEvoy[1], Tim Persoons[1]

[1]Department of Mechanical & Manufacturing Engineering, Parsons Building, Trinity College, Dublin 2, Ireland
[2]School of Mechanical & Design Engineering, Dublin Institute of Technology, Ireland

* Corresponding Author: sakanova@tcd.ie, +353 1 896 1383

Abstract

The rapid increase in energy demand for data center requires improved cooling techniques. This study investigates, numerically and experimentally, the energy efficiency optimization based on server level air flows and also identifies the potential for waste heat recovery from the air stream for a hybrid air/liquid cooled server. Multi-objective genetic algorithm and entropy generation minimization are chosen as tools to address the multiple objectives involved in the problem and examine the cooling performance and waste heat recovery potential. Selected experimental measurements have successfully validated the CFD simulation model.

1   Background

Data center electricity consumption has been continually increasing in recent years due to the rapid development of the semiconductor industry and persistent growth in information technology demands. The forecasts are for a further continuation of the power consumption in the upcoming years. This increasing trend requires a profound rethinking of current data center thermal management strategies to ensure a more sustainable use of energy.

Most of the academic and industry studies have investigated data center thermal management at the room or rack level. The typical current cooling strategy relies on air cooling techniques with a limited cooling ability, with a view toward liquid cooling as a more efficient thermal management technique, and which also holds more promise for waste heat recuperation.

IBM Corporation has played an important role in this research field. Iyengar and Schmidt [1] developed a model which predicted the energy consumption and heat transfer characteristics in a data center. Ellsworth et al. [2] gave an overview of the water cooling unit and rack manifold. They also highlighted the techniques to improve cooling performance and enhance energy efficiency.

Among academic studies, Kumar and Joshi [3] experimentally investigated the effect of tile air flow rate on the server air distributions located at different places in the rack. Khalifa and Demetriou [4] developed a simplified model in order to analytically determine the energy-efficient design of air cooled data centers.

The internal flow inside a server considering temperature and airflow characteristics has not received much attention, although a detailed analysis addressing the thermal challenges at each level starting from the chip to the entire system could lead to an overall efficiency improvement. Some of the server studies are highlighted as follow. Han and Joshi [5] numerically developed a proper orthogonal decomposition (POD) reduced-order model of a server CPU and heat sink which provided faster simulation time with an acceptable accuracy.

One underexplored approach to conserving energy is the optimization of the internal layout of server components. Entropy generation minimization (EGM) is a technique to examine the optimization parameters and evaluate the cooling performance. The target of this study is server layout optimization with multiple design variables and global solutions based on EGM. The approach will use multi-objective genetic algorithm (MOGA) optimization [6], which is underutilized in server optimization yet is a proven approach capable of finding trade-off solutions to multiple objective problems. The paper aims to determine numerically and experimentally the validity of this approach for the layout optimization of a generic hybrid air/liquid cooled data center server, based on a single server level analysis along with potential considerations for air-side heat recovery.

2   Methodology

The baseline generic server under investigation in this study is inspired on the Intel Server Board S2600TP with two Intel Xeon E5-2600 v3 processors and sixteen DDR3 dual in-line memory modules (DIMMs), with a CPU thermal design power (TDP) of 145W [7] and DIMM TDP of 6W [8] (see Fig. 1). A computational fluid dynamics (CFD) model is developed for both 2D and 3D investigations of the internal server flow.

The experimental set up consists of a power supply, server board, pressure and flow rate sensors, and a data acquisition system as shown in Fig. 2. The pressure drop across the channel is monitored for a flow rate range from 8.8 CFM to 72.4 CFM which corresponds to a fan rotational speed range of 2,500-20,500 RPM. The results are compared with 2D simulations as well as the fan manufacturer's specification

sheets [9] as shown in Fig. 3. The 2D and 3D CFD simulations and manufacturer's specification sheet results show a reasonable agreement with our own experimental data with a maximum deviation of 18%, 22% and 23% respectively.

Figure 1: Intel server board S2600TP and sample CFD simulated temperature distribution (air-cooled situation)
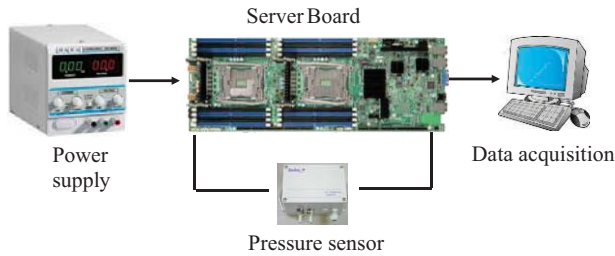

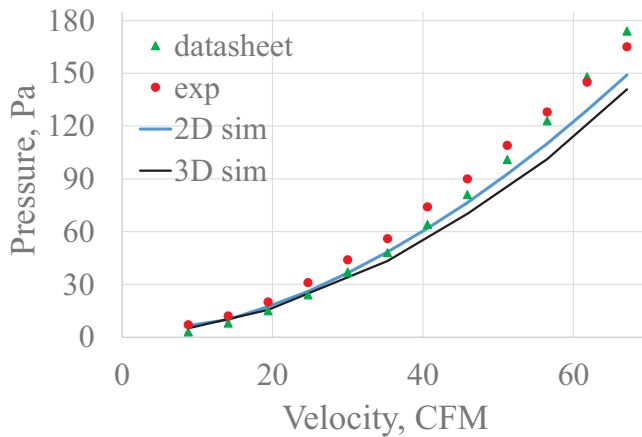*Figure 2: Experimental setup (fans, flow conditioning and flow rate sensor not shown)*

Figure 3: Experimental validation results (rotational speed range of 2,500-20,500 RPM)

2.1 Model description

In the layout optimization scenario, the CPUs are assumed to be liquid cooled and therefore thermally excluded from the simulation. The aim here is to optimize the residual air cooling of the DIMMs only. The CPU heatsinks are removed, and the presence of the liquid-cooled cold plates is assumed not to significantly affect the air flow.

The optimization process consists of the variation of three parameters, such as top upstream DIMM angle A, cross-stream distance between DIMMs C and mass flow rate as shown on Fig. 4. The downstream DIMM angle is fixed at 90^0. At the inlet to the server, air is blocked from entering into the central region (containing the CPUs) by a deflection wall (see Fig. 4). The open inlet area varies with angle A since the deflection wall ends at a distance of 7 mm from the leading edge of the bottom front DIMM, as illustrated in Fig. 4. Dashed red lines indicates the DIMM direction when $A>90^0$.
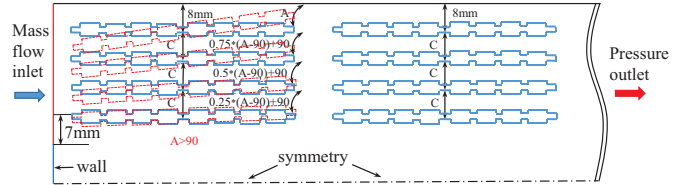
The parameter range is summarized in Table 1.

Figure 4: Numerical optimization parameter definition

Table 1: Design parameters and thermal constraint

Parameter	Range
Upstream DIMM angle A	85-95°
DIMM spacing C	10-15mm
Mass flow rate	0.15-0.18kg/s
Maximum DIMM temperature	363K

The optimization includes the following constraints and goals, aimed at maximizing heat recovery potential and coefficient of performance while safeguarding reliability: 1) limit maximum DIMMs surface temperature; 2) minimize pumping power; 3) maximize air temperature uniformity at the outlet; 4) maximize the average outlet temperature.

EGM is an established method of optimization, where the objective function is the sum of the entropy generation rate due to heat transfer $\dot{S}_{\Delta T,int} + \dot{S}_{\Delta T,ext}$ and the entropy generation rate due to fluid friction $\dot{S}_{\Delta p}$. To evaluate the potential of heat recovery from the air stream, the term $\dot{S}_{\Delta T,ext}$ was introduced, representing entropy generation rate in an external air/liquid heat exchanger at the outlet of the server:

$$\dot{S}_{\Delta T,ext} \cong \sum_i \frac{q_i(T_{out,i} - T_{HEX})}{T_{HEX}^2}$$

(1)

The outlet is divided into N=30 parallel plates, where in each segment i=1..N part of the air stream transfers its heat q_i at temperature $T_{out,i}$ to the heat exchanger which is at the constant temperature T_{HEX}=291K. $T_{out,i}$ is the local air temperature at the outlet at location y_i.

3 Results

3.1 NLPQL and MOGA comparison

Both the non-linear programming by quadratic Lagrangian

THERMINIC 2018
((24th INTERNATIONAL WORKSHOP
on Thermal Investigations of ICs and Systems))
September 2018, Stockholm / SE
www.therminic2018.eu

(NLPQL) and MOGA methods were employed to optimize the server parameters with the *single* objective of minimizing $\dot{S}_{tot}$ together with the constraint of keeping the DIMM temperature below 363K. NLPQL is a gradient-based method which might not find the global optimum solution without a good starting point. As such, the Screening Direct optimization is performed first and then the results are used as the starting point in NLPQL [10]. The best candidate solutions from both methods are highlighted in Table 2. The solutions of both optimization approaches are in reasonable agreement. This verifies the potential of the MOGA method, which will be used in the remainder of the paper for *multiple* objective optimization.

Table 2: Candidate solution comparison for single objective optimization

Parameters and outputs	Candidate solutions	
	NLPQL	**MOGA**
Flow rate, kg/s	0.179	0.169
Spacing C, mm	10.14	10.13
S_{tot}, kg·m²/(s³·K)	0.216	0.228
T_{max}, K	362.1	361.6

3.2 Sensitivity analysis

The sensitivity analysis reflects the parameter influence on the optimization targets. The higher the parameter sensitivity coefficient, the greater the impact it has on the output. Hence, the sensitivity analysis helps to determine the dominant parameters. From Fig. 5, DIMM angle A does not significantly affect the objectives except for $\dot{S}_{\Delta p}$. The positive impact means that $\dot{S}_{\Delta p}$ increases for increasing A. Flow rate has a considerable negative sensitivity coefficient on all objectives apart from $\dot{S}_{\Delta p}$. The higher the flow rate, the smaller $\dot{S}_{\Delta T,int}$, ΔT_{out} and $\dot{S}_{\Delta T,ext}$ which corresponds to our optimization targets. Meanwhile it increases $\dot{S}_{\Delta p}$ and decreases $T_{out,av}$ which is against the optimization targets. The spacing C has a negative impact on $\dot{S}_{\Delta p}$, ΔT_{out} and a positive impact on $T_{out,av}$ which is in line with the optimization targets. This happens while $\dot{S}_{\Delta T,int}$ increases with C and has an adverse effect on optimization targets.
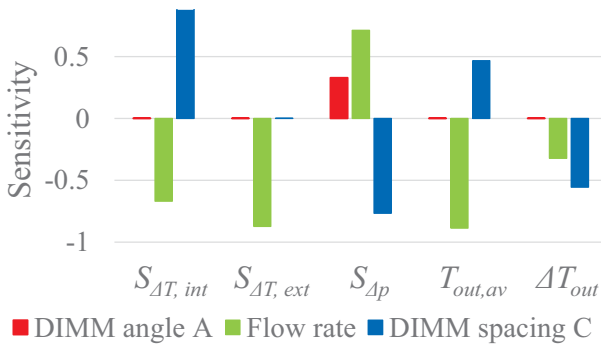

Figure 5: Design parameter sensitivity analysis

3.3 Optimization effect on EGM and $T_{out,av}$

Figure 6 illustrates the obtained Pareto-optimal front for the optimization goals. The Pareto-optimal front is a set of non-dominated solutions where all solutions are considered equally good and none of the objective functions can be improved without worsening the other one. However, by including a subjective preference, i.e., using a single goal on any of the optimization goals, the red dots would be considered the best compromise based on EGM minimization (Fig.6a,b) and $T_{out,av}$ maximization (Fig. 6c) targets.
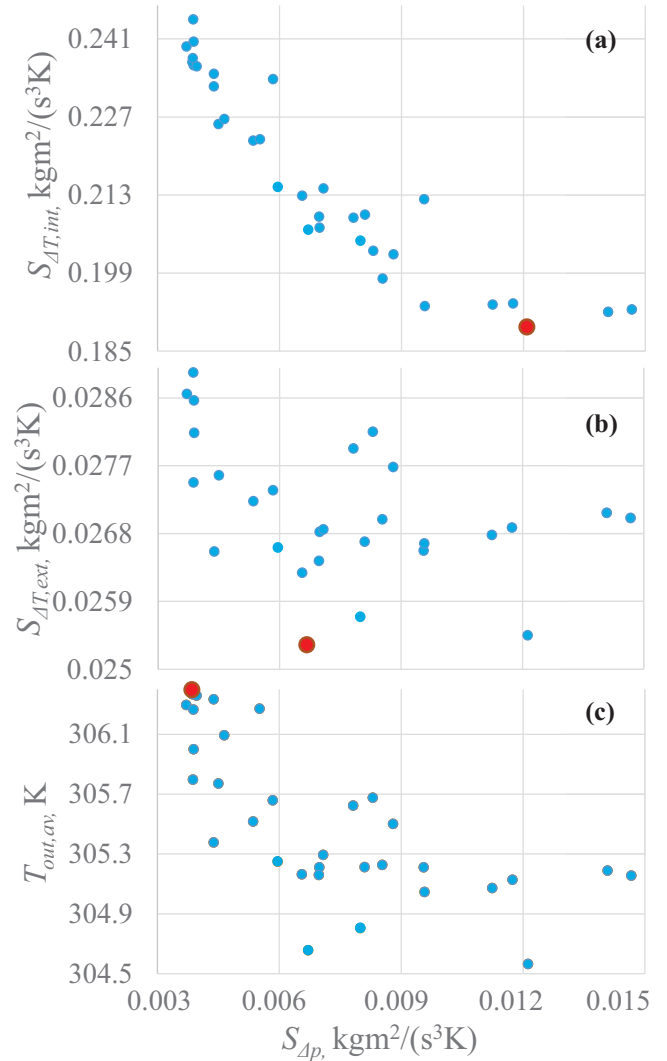

Figure 6: Pareto-optimal front

3.4 Optimization effect on T_{out} uniformity

Figure 7 shows the average outlet temperature $T_{out,av}$ (dashed line) and T_{out} uniformity (solid line) of the baseline and optimized server layout selected based on the highest possible T_{out} uniformity. Only a slight increase in $T_{out,av}$ is observed in the optimized server layout, while the T_{out} uniformity is more affected.

4   Conclusions

In the present study, an internal layout optimization for a hybrid air/liquid cooled data center server is performed using the MOGA approach. Entropy generation minimization (EGM) and average outlet temperature $T_{out,av}$ maximization are specified as the optimization targets, and the effect of DIMM angle and cross-stream distance between DIMMs are investigated. The following conclusions are made during the analysis:
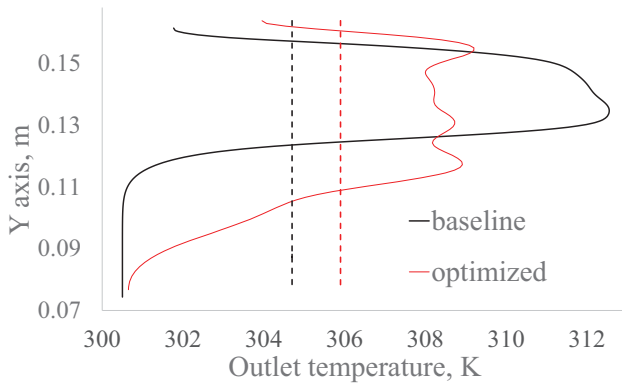

Figure 7: Outlet temperature profiles showing $T_{out,av}$ and T_{out} uniformity for the baseline and optimized server

- The DIMM angle A is the parameter with the least influence on optimization goals.
- Increasing the flow rate is beneficial for $\dot{S}_{\Delta T,int}$, $\dot{S}_{\Delta T,ext}$ minimization and ΔT_{out} uniformity. While $\dot{S}_{\Delta p}$ maximization and therefore increase in pump power as well as $T_{out,av}$ minimization are considered as the penalty for that improvement.
- An increase of DIMM spacing C minimizes $\dot{S}_{\Delta p}$, maximizes $T_{out,av}$ and decreases the ΔT_{out}, while $\dot{S}_{\Delta T,int}$ increases.
- The MOGA approach shows great potential, making it a suitable tool for hybrid-cooled data server layout optimization design.

Acknowledgements

Literature

[1]   M. Iyengar and R. Schmidt, "Analytical Modeling for Thermodynamic Characterization of Data Center Cooling Systems," Journal of Electronic Packaging, vol. 131, no. 2, pp. 021009-021009-9, 2009.

[2]   J. M. J. Ellsworth, G. F. Goth, R. J. Zoodsma, A. Arvelo, L. A. Campbell, and W. J. Anderl, "An Overview of the IBM Power 775 Supercomputer Water Cooling System," Journal of Electronic Packaging, vol. 134, no. 2, pp. 020906-020906-9, 2012.

[3]   P. Kumar and Y. Joshi, "Experimental investigations on the effect of perforated tile air jet velocity on server air distribution in a high density data center," in 2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2010, pp. 1-7.

[4]   H. E. Khalifa and D. W. Demetriou, "Energy Optimization of Air-Cooled Data Centers," Journal of Thermal Science and Engineering Applications, vol. 2, no. 4, pp. 041005-041005-13, 2011.

[5]   H. Xuefei and Y. Joshi, "Energy reduction in server cooling via real time thermal control," in 2012 28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2012, pp. 20-27.

[6]   Y. O. Abdelsalam, S. Alimohammadi, Q. Pelletier, and T. Persoons, "A multi-objective genetic algorithm optimisation of plate-fin heatsinks," in 2017 23rd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC), 2017, pp. 1-6.

[7]   Intel server board S2600TP Product Family and Intel Compute Module HNS2600TP Product Family, 2016 Available at:
https://www02.cp-static.com/objects/pdf/f/f88/1349633133_1_server-barebone-intel-compute-module-hns2600tp24sr-hns2600tp24sr.pdf

[8]   M. Iyengar et al., "Server liquid cooling with chiller-less data center design to enable significant energy savings," in 2012 28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2012, pp. 212-223.

[9]   Nidec, Fan Specification Sheets, 2014 Available at:
http://www.nidec.com/en-NA/product/fan/category/F010/G030/P2000104/

[10]   Ansys 14.5, Using Adaptive Single-Objective Optimization, Available at:
http://www.cadfamily.com/download-pdf/ANSYS-DesignXplorer/tutr_dx_ASO.pdf