2019-4

Multi-objective Layout Optimization of a Generic Hybrid-cooled Data Centre Blade Server

Assel Sakanova
Trinity College Dublin, sakanova@tcd.ie

Sajad Alimohammadi
Technological University Dublin, sajad.alimohammadi@tudublin.ie

Jaakko McEvody
Trinity College Dublin

See next page for additional authors

Recommended Citation

Sakanova, A., Alimohammadi, S. McEvody, J., Battaglioi, S. & Persoons, T. (2019). Multi-objective layout optimization of a generic hybrid-cooled data centre blade server. *Applied Thermal Engineering* V156, 25 June 2019,pp. 514-523. doi:10.1016/j.applthermaleng.2019.04.071

Authors
Assel Sakanova, Sajad Alimohammadi, Jaakko McEvody, Sara Battaglioli, and Tim Persoons

Multi-objective Optimization of a Generic Hybrid-Cooled Data Centre Single Blade Server

Assel Sakanova[1,*], Sajad Alimohammadi[1,2], Jaakko McEvoy[1], Sara Battaglioli[1] and Tim Persoons[1]

[1] Department of Mechanical & Manufacturing Engineering, Parsons Building, Trinity College, Dublin 2, Ireland
[2] School of Mechanical & Design Engineering, Technological University Dublin, City Campus, Ireland
* Correspondence: email: assel001@e.ntu.edu.sg; Tel.: +353-830476842

Abstract: The rapid global increase in energy demand for data centres requires the continuous improvement of cooling solutions and techniques implemented. In a standard data centre approximately a third of the total power consumption can be attributed to the cooling infrastructure, resulting in raised power usage effectiveness (PUE) values. The main culprits of raised PUE are outdated legacy (air cooled) data centres, where only low grade waste heat is available and its capture and re-use remains a challenge. This study investigates numerically the potential for energy recuperation by a server-level internal layout optimization for a hybrid air/liquid cooled server. The study uses multi-objective genetic algorithms (MOGA) and entropy generation minimization (EGM) techniques to incorporate the multiple objectives involved in solving this problem, and examines the cooling performance and waste heat recovery potential. In order to evaluate the potential of the waste heat recovery the term $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta T},\boldsymbol{ext}}$ is introduced in this study. Effect of modified server component layout on pressure drop and maximum outlet temperature were of main interest, due to their role in fan power usage and energy recuperation potential. The base unmodified configuration CFD models were validated through experimental pressure measurements conducted on a real server blade module. The research concluded that a basic server layout optimization such as changing the

memory module angles and spacing could not only enhance the cooling efficiency but also improve the potential for higher grade waste heat recovery. Overall the decrease in EGM due to server layout optimization could be as high as 15%, while the quality of the waste heat due to temperature uniformity might reach up to 42%.

Keywords: Data centre; thermal management; server cooling; multi-objective genetic algorithm optimization; entropy generation minimization; waste heat recovery

Nomenclature

A	area (m^2)
D_h	hydraulic diameter (m)
h	heat transfer coefficient (W/m^2K)
m	mass flow rate (kg/s)
Δp	pressure drop (Pa)
P_{pump}	pumping power (W)
P	perimeter (m)
Re	Reynolds number
$\dot{S}_b$	entropy generation rate of the baseline server (W/K)
$\dot{S}_{\Delta T,int}$	entropy generation rate due to heat transfer from DIMM's surface temperature to air (W/K)
$\dot{S}_{\Delta T,ext}$	entropy generation rate due to heat transfer in an external air/liquid heat exchanger (W/K)

$\dot{S}_{\Delta T,int} + \dot{S}_{\Delta T,ext}$	entropy generation rate due to heat transfer (W/K)
$\dot{S}_{\Delta p}$	entropy generation rate due to fluid friction (W/K)
$\dot{S}_{gen}$	total entropy generation rate (W/K)
ΔT	temperature difference (K)
u	velocity (m/s)

Subscripts

av	average
b	baseline
in	inlet
max	maximum value
out	outlet
R	ratio
w	wall

32

1. Introduction

Data centre electricity consumption has been steadily increasing in recent years due technological advancements in the semiconductor industry and persistent growth in information technology demands. Industry surveys show no slowdown in the power demand for data centre facilities world-wide with an estimated tripling in the next decade [1]. This increasing trend requires improved data centre thermal management solutions to ensure their power consumption is maintained within sustainable limits. The latest summary of the thermal management techniques of data centres from the chip to the cooling system is presented by Khalaj and Halgamuge [2]. Their research investigates state-of-the-art multi-level hybrid thermal management systems, which employ both air and water as the working fluid.

The majority of the studies related to data center thermal management have been concerned with the room or rack level configurations. The existing cooling strategy includes air cooling techniques with limited cooling ability and liquid cooling as highly efficient thermal management technique.

IBM Corporation has played an important role in an industrial research on the advanced electronics cooling. Iyengar and Schmidt [3] developed a model which predicted the energy consumption and the heat transfer characteristics in a data centre. They considered the case study with the load of 5.88MW and concluded the chiller energy is the highest consumer of total cooling energy. In 2008 IBM introduced the water cooling Power 775 Supercomputing system. Ellsworth *et al.* [4] gave an overview of the water cooling unit and rack manifold. They also highlighted the techniques to improve the cooling performance and enhance the energy efficiency. David *et al.* [5] presented the experimental characterization of the chiller-less air cooled data centre. The paper discussed the thermal, hydraulic characteristics of the air-to liquid and liquid-to-liquid rack heat exchanger cooling performance. Schmidt and Cruz [6] numerically investigated the effect of the chilled air exiting the hot aisle of a raised floor data centre. The results could be used as a guidance for the data centre layout design. Schmidt [7] described the set of measurement of data centre thermal profile above the raised floor. Schmidt *et al.* [8] stated that a significant amount of energy could be saved by preventing the mixing of cold and hot air streams. To implement the technique, they introduced separate hot and cold aisles with exhaust chimneys, resulting in energy savings of up to 59%.

Among academic studies, Kumar and Joshi [9] experimentally investigated the effect of tile air flow rate on the server air distributions located at different places in the rack. They came to the conclusion that increasing the perforted tile air flow rate is not the best way to provide cooling to high density racks. Khalifa and Demetriou [10] developed a simplified analytical model in order to identify the optimum energy-efficient design for air cooled data centres. The methodology also showed the trade-off between the cooling infrastructure and performance characteristics. Karki and Patankar [11] numerically examined the flow field and pressure distribution in the under-floor plenum of the raised-floor data centre within a one-dimensional framework. They also compared the results between a one-dimensional model with that of a three-dimensional model and found a good agreement. Fouladi *et al.* [12] demonstrated a novel hybrid modelling strategy for data centre cooling system optimization. The employment of proper orthogonal decomposition (POD) airflow modelling scheme demonstrated 23% and 43% energy and exergy savings, respectively. Shah *et al.* [13] proposed the exergy-based approach for data centre thermal management and energy efficiency evaluation. The proposed model quantifies the amount of energy utilized from other components for thermal management purposes. Samadiani *et al.* [14] developed a simulation-based design approach to achieve an adaptable energy efficient data centre design. This approach helps to investigate the effect of the design parameters in terms of reliability and power consumption minimization. Alkharabsheh *et al*. [15] summarized the recent advancement in the data centre modelling and energy optimization. The paper presented the areas of the potential research for data centre thermal magamanet. Song [16] examined the use of organized fan-assisted tile systems and investigated their impact on the cooling effectiveness. The author came to the conclusion that the technique can significantly improve the cooling performance and might be an interesting research direction in the future.

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 follows. Han and Joshi [17] numerically developed the server CPU and heat sink by using POD which provided faster simulation time with acceptable accuracy. This was used together with the fan controller in order to study the energy consumption reduction in server CPU fans. Sarma and Ambali [18] numerically investigated the thermal design and pressure drop across the 2U configuration computing server. They concluded that all server components remained under the allowable limit and operated without any failures. Iyengar *et al.* [19] experimentally studied the concept of hybrid cooling systems, which incorporates air cooling flow low power components (power supplies, storage disk drives, printed circuit board) and water cooling for higher power components (microprocessors and memory cards). Hybrid cooled systems have demonstrated energy savings of up to 30%. According to Garimella *et al.* [20] the shift from air to liquid cooling will happen in the near future, which will significantly improve the energy efficiency of next generation data centre facilities.

One underexplored approach to conserving energy is the optimization of the internal layout of the server components, whether by rearranging the components themselves or using baffles to redirect air flow. Here, the former more invasive approach will be taken. Entropy generation minimization is one of the techniques to carry out the optimization and evaluate the cooling performance [21]. Server optimization with multiple design variables and global solutions is the aim of this study. Multi-objective genetic algorithm is a proven powerful approach capable of finding trade-off solutions to multiple objective problems.

Due to limited supply of fossil fuels and negative consequences of its usage, waste heat recovery as a method to conserve energy is gaining more attention nowadays [22, 23]. The main challenge in this field consists of capturing and transferring the waste heat in order to increase the recovery system efficiency. Almost 100% of all electrical power supplied to the data centre server is converted to the heat, which requires a cooling system capable of maintaining the server components below their maximum allowable temperatures. Some research is ongoing to reuse the waste heat in order to reduce data centre operational costs [24]. The author proposed the approach which is able to measure the exergy destruction and eliminate irreversibility in the thermal path. Data centre companies such as Facebook and Google with up to 300 MW of power usage are making an effort to reduce energy usage and the waste heat decrease their facilities rate of power waste [25]. Even though most of the data centres operate at 20% of the maximum load and rarely operates at full load, the total requirement for heat dissipation continues to increase [26]. This indicates that waste heat recovery has the capability to decrease the electricity consumption in data centres.

In many studies, for the sake of simplicity in the theoretical calculation, the air temperature and flow distribution at the inlet of the heat exchanger is by default assumed to be uniform [27]. However, in reality the air temperature and flow is generally non-uniform. The non-uniformity significantly degrades both the air-side thermal and hydraulic performance [28]. Therefore, temperature and airflow non-uniformity is important and cannot be neglected by the assumption of a uniform distribution.

In this study, the optimisation of a generic hybrid cooled single blade server is investigated to improve the potential for air-side heat waste recovery, taking into account the outlet air flow non-uniformity. This work extends an earlier study reported on by Sakanova et al. [29] towards a larger parameter space. Also it gives more in-depth insights into the optimization process with a detailed analysis of the effect of the design parameters. At the end the optimized server is compared with the baseline server in terms of the improved ability to recuperate waste heat.

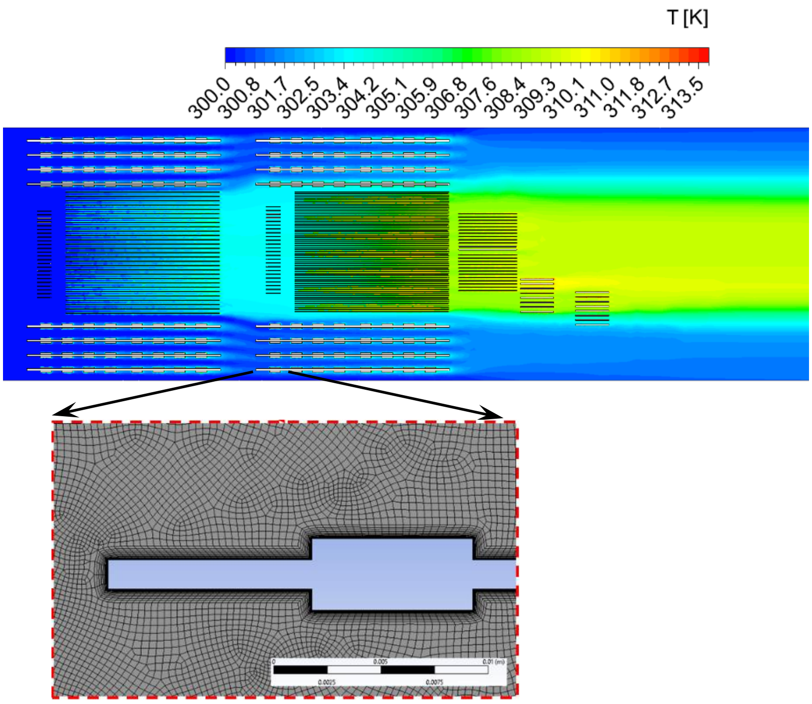
2. Flow simulation and optimization methodology

2.1. CFD model setup and grid independency

An Intel server board S2600TP with two Intel Xeon E5-2600 v3 processors and 16 DDR3 dual in-line memory modules (DIMMs) is the baseline server under investigation in this study, with a CPU thermal design power (TDP) of 145W [30] and DIMM TDP of 6W per module [19] (Fig. 1). A computational fluid dynamics (CFD) model is developed for both 2D and 3D investigations of the internal server flow. An implicit solver is used to solve governing equations. The second order upwind scheme is applied for energy and momentum discretization. The standard interpolation scheme is adopted for pressure discretization. The pressure-velocity coupling is solved by the SIMPLE algorithm. The convergence criteria are at 10^{-4} and 10^{-7} for velocity and energy, respectively. The *κ-ω SST* turbulence model is applied. According to the *κ-ω SST* model requirements, the y^+ value is kept below 1 with the number of prismatic layers of 15.

The detailed mesh independence study is performed and summarized in Table 1. The grid independence studied based on five consequentially refined mesh generated from the baseline mesh 1. The next following mesh in refined by refinement factor of 1.32-1.36. The effect of the refining mesh grid is presented in terms of pressure drop across the server. There is a negligible difference (less than 1%) in pressure drop between mesh 4 and 5. The discretization error is also estimated using a method of grid convergence index (GCI) which is performed based on four mesh grids. GCI indicated the maximum uncertainty at each mesh grid stating that the solution are within asymptotic range of convergence [31]. Mesh 4 is chosen as a good trade-off between accuracy and computational time.

(a)



(b)

Fig. 1. (a) Intel server board S2600TP and (b) sample CFD simulation temperature distribution (air-cooled situation) at 20500 RPM, heat power at CPU, RAM modules, chipset and BMC is 145W, 100.8W (6.3W per each), 7W and 1.5W accordingly and simulation grid.

Table 1. Mesh independence study results

Grid number	Element numbers (mln)	Refinement factor	Output quantity (server pressure drop) (Pa)	Grid convergence index (GSI) (%)	Asymptotic convergence check
1	1	N/A	28		
2	1.36	1.36	30.6	9.1	
3	1.8	1.32	31.8	4.04	1.039
4	2.45	1.36	36.5	1.09	
5	3.29	1.34	36.8	0.07	1.008

2.2. Experimental validation

The experimental validation aims to check the global quantities such as the pressure drop and flow rate and not the local quantities such as velocity fields. The validation is strictly hydraulic in isothermal conditions. The experimental set up consists of a power supply (1), server board, differential pressure transducer (2), flow rate sensors, and a data acquisition system (DAQ) (3) as shown in Fig. 2. The internal three fans [32] boost the room temperature air through the server board at rotational speed of 2,500-20,500 RPM which corresponds to flow rate of 8.8-72.4 cubic feet per meter (CFM). In order to measure the pressure drop across the channel, two pressure taps are installed at the inlet and outlet of the server. The flow sensor is mounted at the server outlet and fixed by means of a 3D-printed mount. The data is logged through National Instruments DAQs controlled through LabVIEW. Additionally, a 70 cm long smooth duct is attached at the server inlet to establish hydrodynamically fully developed flow at the server inlet. According to [33] the hydrodynamic entrance length in the turbulent flow becomes insignificant beyond a pipe length of 10 times the hydraulic diameter ($D_h = 4A/P$=6.5cm). At the end, the full system is properly insulated to prevent any additional air to be sucked in. The obtained experimental results are compared with 2D and 3D simulations as well as with the server fan data sheets [32], as shown in Fig. 3. 2D and 3D simulations and fan server data sheet results show a reasonable agreement with experimental data, with a maximum deviation of 22%, 23% and 18%, respectively.

Maximum deviation of 23% between both 3D and 2D simulation, and experimental results occurs at the lowest flow rate. This increased error can be contributed to a number of different factors. Firstly the differential pressure transmitter/transducer performance is expressed as a percentage of max calibrated span or full scale, which states that the error contribution of the differential pressure transmitter increases as the flow rate/differential pressure drops. For example, according to data sheet, the differential pressure transmitter uncertainty is +/-0.20% of flow rate at maximum rate; it changes to +1.27/-1.28% of flow rate at the minimum rate. At the lowest flow rate, the pressure discrepancy between simulations and experiments (23%) was found to be below 2 Pa, well within a reasonable range. Secondly, the simulations cannot capture the complex and rather rough surface of the server board which would contribute to an increased pressure drop.

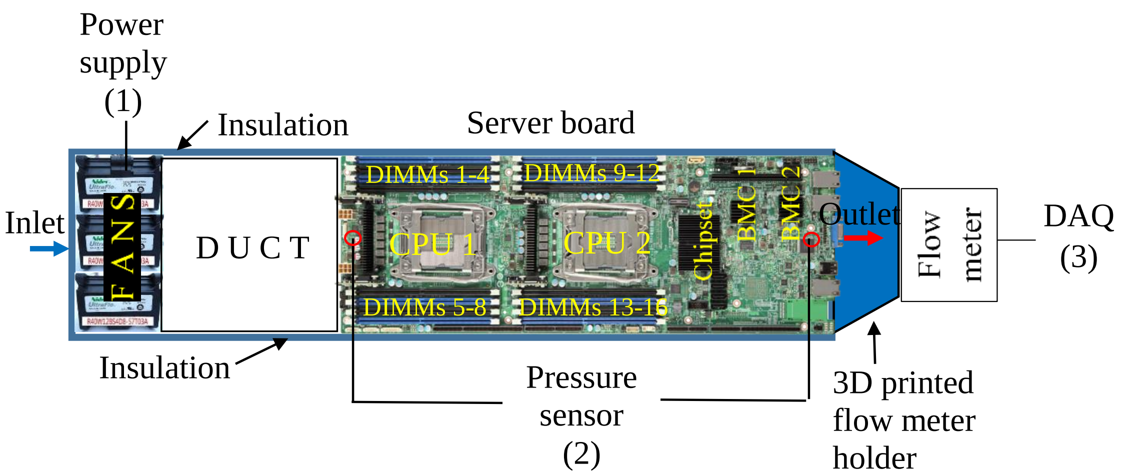


Fig. 2. *Experimental setup*

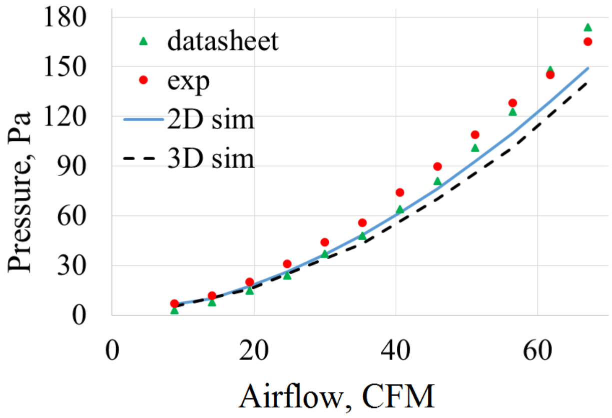


Fig. 3. *Experimental validation results (for fans speed ranges 2,500-20,500 RPM)*

2.3. Numerical optimization approach

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 air cooling of the remaining components, i.e. the DIMM modules. The CPU heatsinks are removed, and the presence of the liquid-cooled cold plates is assumed to have no significant effect on the air flow due to their significantly reduced profile.

The optimization process consists of the variation of four parameters, such as the top upstream DIMM angle A, downstream DIMM angle B, the cross-stream distance between DIMMs C and the mass flow rate (parameter D), as shown on Fig. 4. At the inlet to the server, air is blocked from entering the central region (containing the CPUs) by a deflection wall (Fig. 4). The open inlet area varies with the angle A since the deflection wall ends at a distance of 7 mm from the leading edge of the bottom upstream DIMM, as illustrated in Fig. 4. The dashed red lines indicate the DIMM direction when A > 90^0. The range of the parameters is summarized in Table 2.

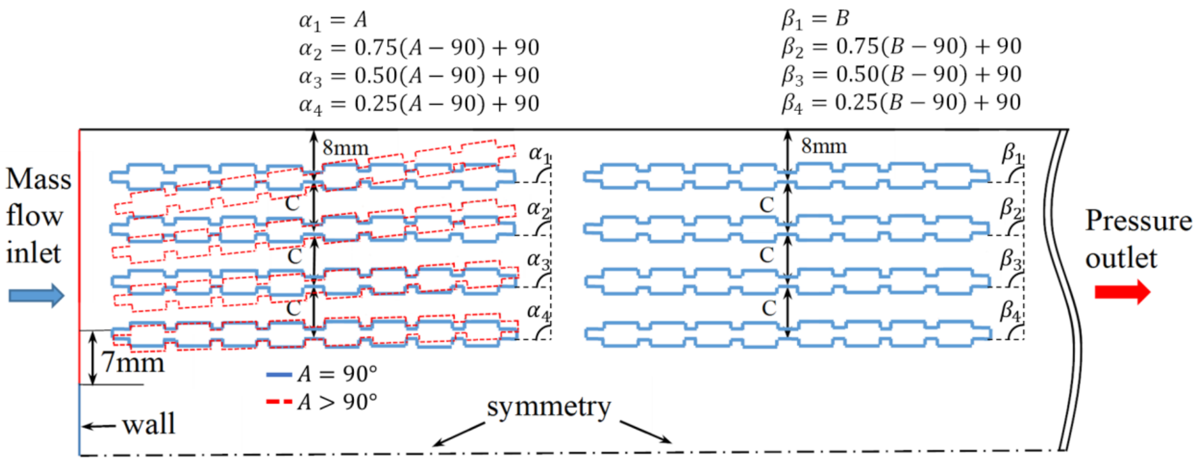


Fig. 4. *Numerical optimization parameter definition*

2.4. MOGA optimization procedure and EGM technique

Figure 5 depicts the block diagram of MOGA execution. Firstly, the initial population containing $Npop$=50 is generated. Then the values of the objective functions are obtained for the initial population and evaluated based on the optimization criteria. If the criteria is satisfied then MOGA produces a set of optimal Pareto solitons. However, if the optimization criteria are not met, then based on the reproduction, which includes the crossover and mutation, a new set of candidates is created for evaluation. The procedure is repeated until the optimum solution is found.

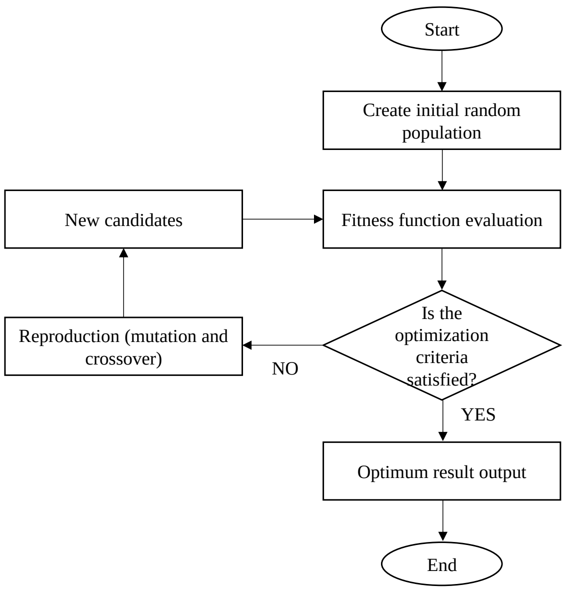


__Fig. 5.__ The flow chat of optimization procedure

Table 2 includes the design parameters range at which the MOGA optimization analysis is considered. Also the maximum allowable temperature constraint of 90°C for DIMM modules is introduced according to [34]. The referenced values of the baseline parameters are taken from the relevant Intel datasheet [30].

Table 2. Overview of the design parameters and thermal constraints

Parameter	Range	Reference (baseline server)
Upstream/downstream DIMM angle A and B	85-95°	90°
DIMM spacing C	10-15mm	10mm
Mass flow rate (parameter D)	20-24CFM (0.15-0.18kg/s)	up to 90CFM (0.675kg/s)
Maximum DIMM temperature	90°C	90°C

The optimization process in Table 3 includes the following goals, constraints and corresponding objective functions, aimed at maximizing the waste heat recovery potential and the coefficient of performance while safeguarding reliability.

Table 3. Overview of the optimization goals, objective functions and constraints

Goal	Objective function	Constraint
1. Minimise DIMMs surface temperature	minimize $\dot{S}_{\Delta T,int}$	Keep the maximum DIMM temperature within the margin of 90°C
2. Minimize the entropy generation rate in an external air/liquid heat exchanger at the outlet of the server	minimize $\dot{S}_{\Delta T,ext}$	
3. Minimize the pumping power	minimize $\dot{S}_{\Delta p}$	
4. Maximize the average outlet temperature	maximize $T_{out,av}$	

EGM is an established method of the 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 the fluid friction $\dot{S}_{\Delta p}$, as follows:

$$\dot{S}_{\Delta T,int} \cong \sum_{k=1}^{N_{DIMM}} \sum_{j=1}^{N_{chip}} \frac{q_{j,k}(T_{w,j,k} - T_{in})}{T_{in}^2} \tag{1}$$

where N_{DIMM} is the number of DIMMs (N_{DIMM}=16) and $N_{chip}$is number of chips per DIMM (N_{chip}=16). T_{in} =300K. T_w is the wall temperature.

The entropy generation rate due to fluid friction $\dot{S}_{\Delta p}$ is defined as

$$\dot{S}_{\Delta p} = \frac{\Delta p \; uA}{T_{in}} \tag{2}$$

A rear door heat exchanger is assumed in this work to be mounted on the back of the server to capture the heat of the hot air coming out from the board server and transports it to the chilled water, as shown in Fig.6. This rear door heat exchanger could be a finned coil air/liquid heat exchanger where the fins help to increase the air-side heat transfer area [35].

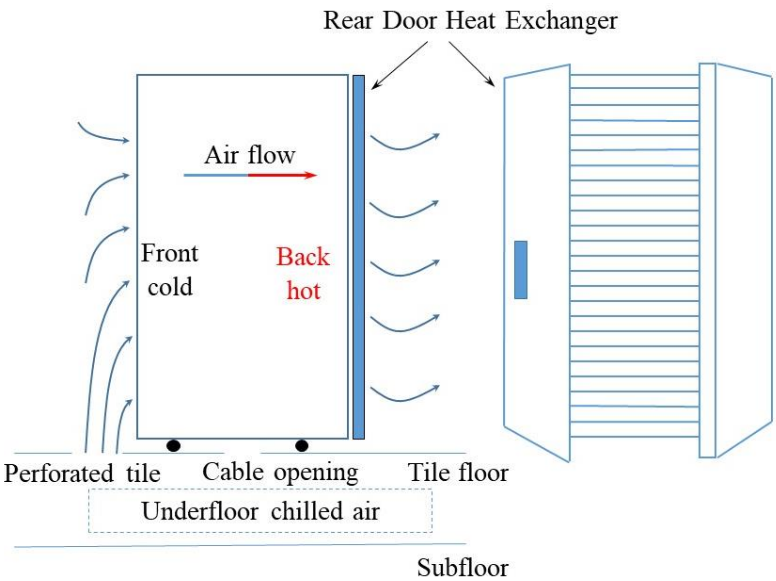


Fig. 6. *Rear door heat exchanger*

To evaluate the potential of heat recovery from the air stream, the term $\dot{S}_{\Delta T,ext}$ is introduced by this study, representing entropy generation rate in an external air/liquid heat exchanger at the outlet of the server:

$$\dot{S}_{\Delta T,ext} \cong \sum_i \frac{q_i\left(T_{out,i} - T_{HEX}\right)}{T_{HEX}^2} \tag{3}$$

The outlet location is divided into a total of N = 29 parallel plates, with the distance between plates equal to 0.003m. At each plate $i = 1 \ldots N$, a part of the air stream transfers its heat q_i at the temperature $T_{out,i}$ to the heat exchanger which is at a constant temperature $T_{HE\,X} = 291K$ [35]. $T_{out,i}$ is the local air temperature, evaluated from CFD results at the outlet location y_i. Where $q_i = hA\Delta T$ and $h = Nu \cdot k/D_h$, where Nu, Nusselt number, is for laminar flow between two parallel surfaces is defined from [36] with D_h as twice the distance between the plates. The total entropy generation rate is the sum of the above mentioned components of entropy generation:

$$\dot{S}_{gen} = \dot{S}_{\Delta T,int} + \dot{S}_{\Delta p} + \dot{S}_{\Delta T,ext} \tag{4the }$$

3. Results

3.1. NLPQL and MOGA comparison

Both non-linear programming by quadratic Lagrangian (NLPQL) and the MOGA method were employed to optimize the server parameters with the single objective of minimizing $\dot{S}_{gen}$ together with the constraint of keeping the DIMM temperature below 90°C. 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 4. The solutions from both optimization approaches are in reasonable agreement with each other. This verifies the potential of the MOGA method, which will be used in the remainder of this article for a multi-objective optimization.

Table 4. Candidate solution comparison for single objective optimization

Parameters and outputs	Candidate solutions	
	NLPQL	MOGA
Flow rate D, kg/s	0.179	0.169
Spacing C, mm	10.14	10.13
$\dot{S}gen$, W/K	0.216	0.228
$Tmax$, K	362.1	361.6

3.2. Local and global 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. There are two types of sensitivity charts: local and global. The local sensitivity calculates the outputs based on the change of inputs independently. The local parameter sensitivity is based on the difference between the minimum and maximum values obtained by varying one of the input parameter while holding the remaining parameters constant. It means that the local sensitivity depends on the input parameters which are held constant. In this study, the input parameters which are held constant are the manufacturing values of the baseline server. The global sensitivity is based on the correlation analysis of the generated sample point which are located throughout the entire space of input parameters. It means the global sensitivity does not depend on the input (manufacturing) value since all possible values of the inputs have been considered during the analysis.

3.2.1 Local sensitivity of input parameters on optimization targets

Figure 7 shows the local sensitivity trend while Fig. 8 shows the detailed analysis of the local sensitivity stating the actual values of the input/output parameters at the different points. In order to see the impact of input parameters on air flow temperature uniformity at the outlet of the server, the term ΔT_{out} is introduced and added to Fig. 7-9. From Fig. 8 DIMM angle A has a noticeable positive impact on $\dot{S}_{\Delta p}$ and less significant influence on $\dot{S}_{\Delta T,ext}$. The positive impact means that $\dot{S}_{\Delta p}$and $\dot{S}_{\Delta T,ext}$increases with A. These impacts are opposite to our optimization goals.

An increase in DIMM angle B decreases $\dot{S}_{\Delta T,int}$ more significantly with less penalty in $\dot{S}_{\Delta p}$ as compared with DIMM angle A. However as DIMM angle B increases it also contributes to the flow and temperature non-uniformity ΔT_{out}. In a heat exchanger due to heat transfer, the temperature gradients take place between cold and hot fluids and in the wall separating the fluids. In most of the heat exchanger's heat transfer and pressure drop calculations, the inlet flow and temperature distribution are considered to be uniform, however this assumption could not be accepted as the realistic one. In practice, the air flow and temperature distribution are non-uniform. This non-uniformity has an impact on thermal and hydraulic deteriorations. Temperature uniformity is simply defined as difference between min and max temperatures at the outlet of the server or the inlet to heat exchanger.

The spacing C has a negative impact on $\Delta T_{out}$and a positive impact on $T_{out,av}$which is in line with the optimization targets. This happens while $\dot{S}_{\Delta T,int}$ dramatically increases with C which has an adverse effect on optimization targets.

Mass flow rate D 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.

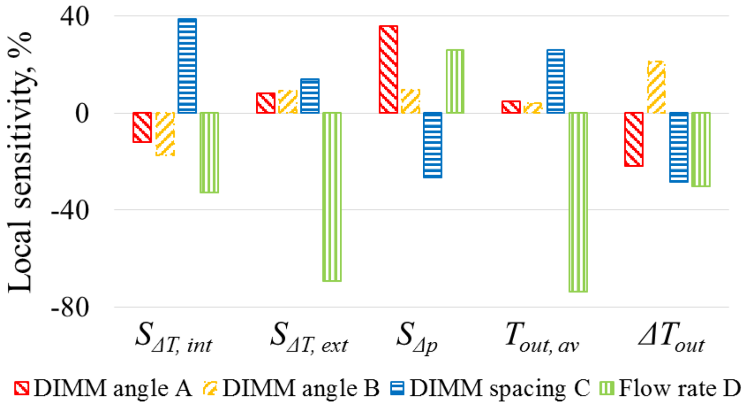
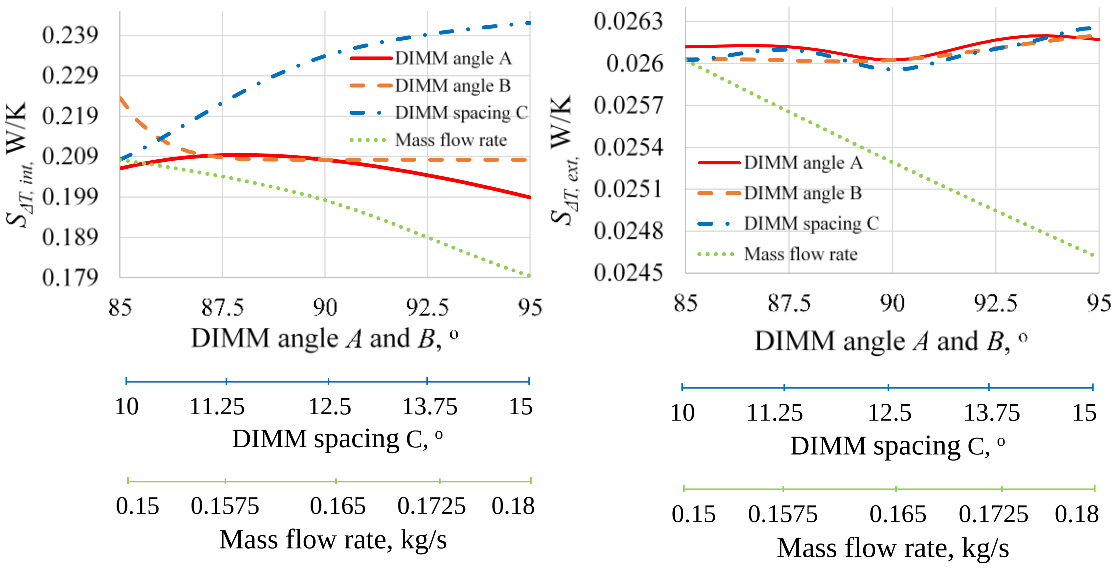
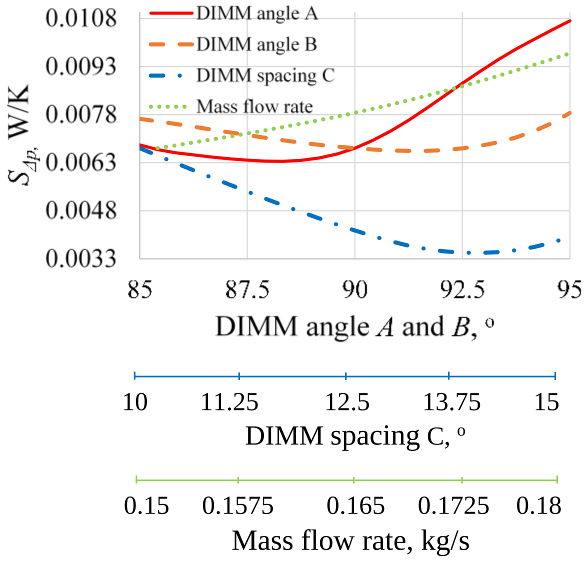
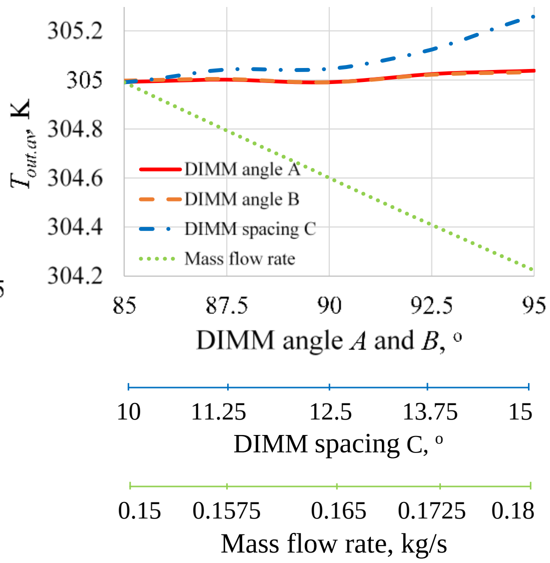


Fig. 7. *Local sensitivity of input parameters on optimization targets and ΔT_{out} uniformity*

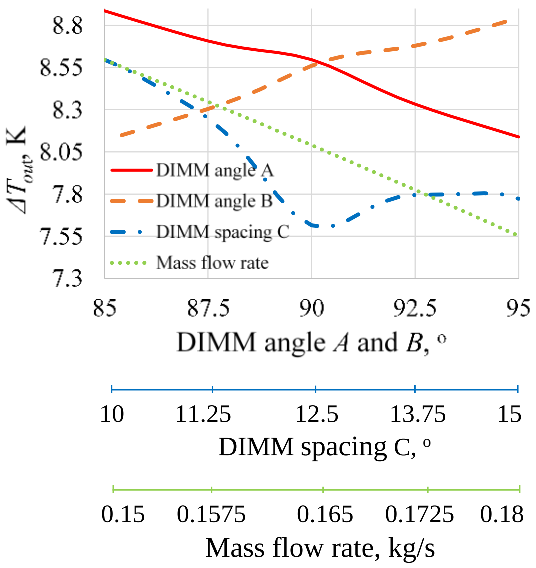


(a) (b)





(c) (d)



(e)

Fig. 8. *The detailed local sensitivity analysis of input parameters on optimization targets*

3.2.2 The global sensitivity of input parameters on optimization targets

The global sensitivity chart is slightly different from the local one (Fig. 9). However, the main trends are kept the same. The highest effect among all input variables on the optimization targets has mass flow rate D followed by DIMM spacing C and DIMM angle A. The least influenced parameter is DIMM angle B. An increase in flow rate D is in line with all optimization goals by neglecting the $T_{out,av}$ due to small range of change. However the increase in $\dot{S}_{\Delta p}$ is a penalty for the improved optimization goals. An increase in DIMM spacing C can decrease this penalty but at the price of degrading other outputs.

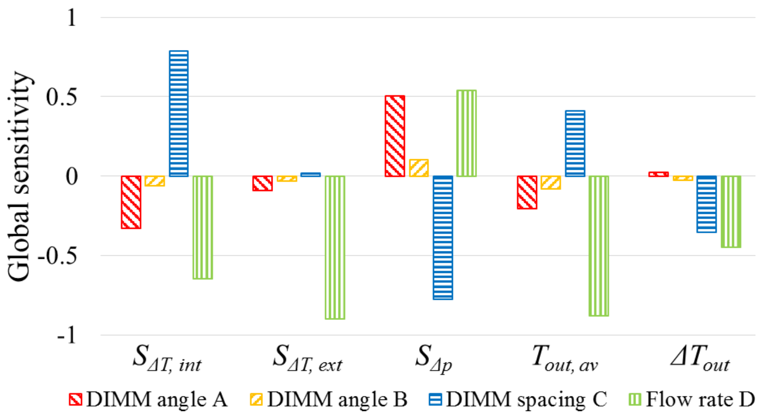
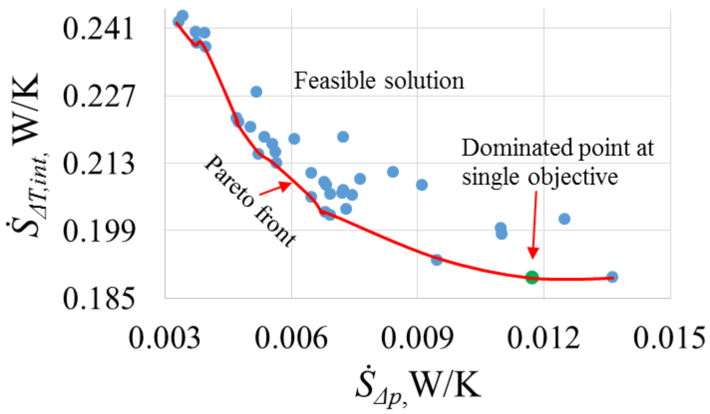


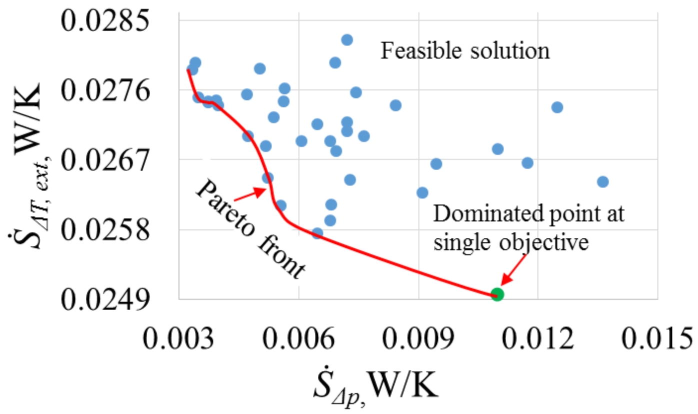
Fig. 9. *Global sensitivity of input parameters on optimization targets*

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

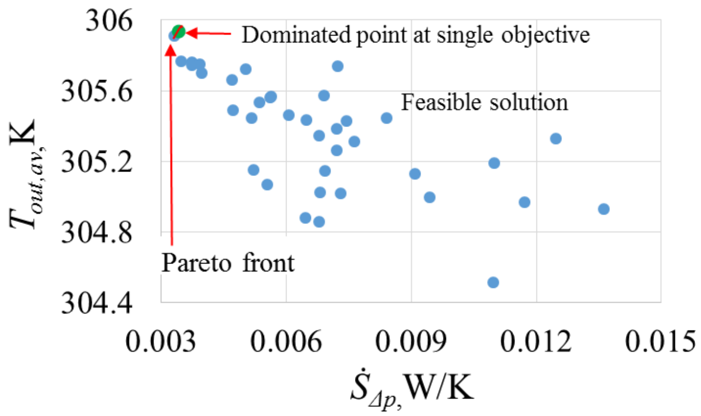
Fig. 10 illustrates the obtained Pareto-optimal front for the optimization goals. The Pareto-optimal front is a set of non-dominated solutions where all solutions are considered equally good and none of the objective functions can be improved without worsening the other one. However, by including a subjective preference, i.e., using a single goal on any of the optimization goals, the green dots would be considered the best compromise based on EGM and $T_{out,av}$ maximization targets.



(a)



(b)



(c)

Fig. 10. *Pareto-optimal front for the optimisation goals*

3.4. Optimization effect on ΔT_{out} uniformity

Fig. 11 shows the average outlet temperature $T_{out,av}$ (dashed line) and ΔT_{out} uniformity (solid line) of the baseline and optimized server layout selected based on the highest possible ΔT_{out} uniformity. Only a slight increase in $T_{out,av}$is observed in the optimized server layout, while the ΔT_{out} uniformity is more affected. In order to measure the thermodynamic effect of the improved geometry of the server with respect to the baseline, the entropy generation ratio and temperature uniformity ratio are introduced and defined as $\dot{S}_R = 1 - \dot{S}_{tot}/\dot{S}_b$, $\Delta T_{out,R} = 1 - \Delta T_{out} / \Delta T_{out,b}$. Table 5 summarizes a few examples of some improved geometry compared to the baseline. According to it, $\dot{S}_R$ may decrease by up to15% and the temperature uniformity, $\Delta T_{out,R}$ which defines the waste heat quality by up to 42%.

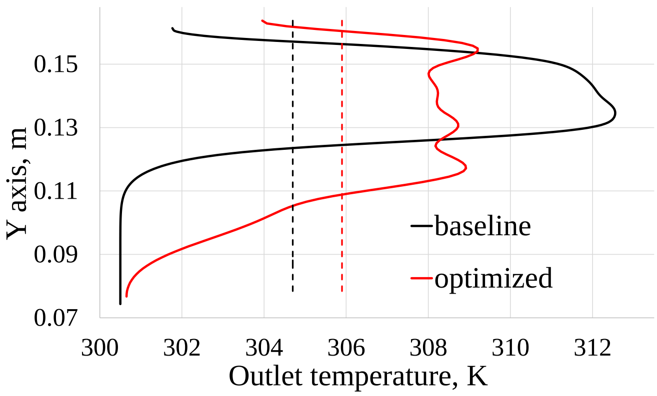


Fig. 11. Outlet temperature profiles showing $T_{out,av}$ and ΔT_{out} uniformity for the baseline and optimized server

Table 5. Comparison of the optimized server with the baseline one

Geometry	*Input parameters*	$\dot{S}_{\Delta T,int}$, *W/K*	$\dot{S}_{\Delta p}$, *W/K*	$\dot{S}_{\Delta T,ext}$, *W/K*	$\dot{S}_{gen}$, *W/K*	ΔT_{out}, *K*	$T_{out,av}$, *K*	$\dot{S}_R$, *%*	$\Delta T_{out,R}$, *%*
Baseline	*A=90^{0}C* *B=90^{0}C*	*0.22907*	*0.0058*	*0.0268*	*0.2677*	*12.0*	*304.25*	-	-

	C=10mm								
	D=0.2237kg/s								
Geometry 1	*A=92.78^{0}C*	*0.18926*	*0.01172*	*0.0266*	*0.22765*	*9.3*	*304.97*	*15*	*22.5*
	B=87.075^{0}C								
	C=10.284mm								
	D=0.1791kg/s								
Geometry 2	*A=94.78^{0}C*	*0.1996*	*0.0109*	*0.0249*	*0.2355*	*7.1*	*304.52*	*12*	*42*
	B=90.41^{0}C								
	C=12.784								
	D=0.1797kg/s								
Geometry 3	*A=94.38^{0}C*	*0.18945*	*0.0136*	*0.0264*	*0.2295*	*8.4*	*304.93*	*14*	*30*
	B=94.483^{0}C								
	C=11.066mm								
	D=0.1767kg/s								

4. Conclusions

In the present study, an internal layout optimization for a hybrid air/liquid cooled data centre server is performed using the MOGA approach. Entropy generation minimization and an average outlet temperature $\boldsymbol{T_{out,av}}$ maximization are specified as the optimization goal. The effect of DIMM angles A and B, cross-stream distance between DIMMs C and mass flow rate D are also investigated. In order to evaluate the potential of the waste heat recovery the term $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta T,ext}}$ was introduced in this study. The effect of air flow and temperature non-uniformity, $\Delta \boldsymbol{T_{out}}$ is considered and monitored. The following conclusions are made during the analysis:

- The flow rate is found to be the most dominated input parameter, followed by DIMM spacing C and DIMM angle A accordingly. The least influenced parameter is DIMM angle B;
- An increase in flow rate is in line with the optimization goals but it costs an essential increase in $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta p}}$. An increase in DIMM angle A is also noticeably contribute to the $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta p}}$;
- An increase in DIMM spacing C can curb the growth of $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta p}}$ with increasing angle A, but it markedly contributes to $\dot{\boldsymbol{S}}_{\boldsymbol{\Delta T,int}}$which is counter to the optimization goals;
- All input parameters do not significantly change the range of $\boldsymbol{T}_{\boldsymbol{out,av}}$. The difference in $\boldsymbol{T}_{\boldsymbol{out,av}}$ is within few degrees between the baseline and optimized server;
- The visible contribution on $\Delta\boldsymbol{T}_{\boldsymbol{out}}$ uniformity during the optimization are made by an increase in all parameters except DIMM angle B;
- Overall the decrease in EGM due to server layout optimization ($\dot{\boldsymbol{S}}_{\boldsymbol{R}}$**)** could be as high as 15%, while the quality of the waste heat due to temperature uniformity $\Delta\boldsymbol{T}_{\boldsymbol{out,R}}$ might reach up to 42%.

The MOGA approach shows a great potential, making it a suitable tool for hybrid-cooled data server layout optimization design. The server layout optimization design seems a promising way of improving the data centre cooling system reliability and also increasing the potentiality for waste heat recovery. The future research direction is considered as the investigation of the cooling optimization measures at the rack level.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under the SFI Strategic Partnership Programme Grant Number SFI/15/SPP/E3125. The authors also acknowledge the financial support of the Irish Research Council (IRC) under grant number GOIPD/2016/216, and acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

References

[1] Global warming: Data centres to consume three times as much energy in next decade, experts warn. https://www.independent.co.uk/environment/global-warming-data-centres-to-consume-three-times-as-much-energy-in-next-decade-experts-warn-a6830086.html, 2016 (accessed 1 December 2018)

[2] A. Habibi Khalaj and S. K. Halgamuge, A Review on efficient thermal management of air- and liquid-cooled data centers: From chip to the cooling system, Appl Energy 205 (2017) 1165-1188. https://doi.org/10.1016/j.apenergy.2017.08.037.

[3] M. Iyengar and R. Schmidt, Analytical Modeling for Thermodynamic Characterization of Data Center Cooling Systems, J. Electron. Packag. 131 (2) (2009) 021009-021009-9. https://doi.org/10.1115/1.3103952.

[4] 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, J. Electron. Packag. 134 (2) (2012) 020906-020906-9. https://doi.org/10.1115/1.4006140.

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

[6] R. Schmidt and E. Cruz, Raised floor computer data center: effect on rack inlet temperatures of chilled air exiting both the hot and cold aisles, in ITherm 2002. Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (Cat. No.02CH37258), 2002, pp. 580-594.

[7] R. Schmidt, Thermal profile of a high-density data center: Methodology to thermally characterize a data center, ASHRAE Trans 110 (2) (2004) 635-342.

[8] R. Schmidt, Energy savings through hot and cold aisle containment configurations for air cooled servers in data centers, Proceedings of the ASME 2011 Pacific Rim Technical Conference and Exposition of Packaging and Integration of Electronic and Photonic Systems (2011) 1-6. https://doi.org/10.1115/IPACK2011-52206.

[9] 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.

[10] H. E. Khalifa and D. W. Demetriou, Energy Optimization of Air-Cooled Data Centers, J. Thermal Sci. Eng. Appl. 2 (4) (2011) 041005-041005-13. https://doi.org/10.1115/1.4003427.

[11] K. C. Karki and S. V. Patankar, Airflow distribution through perforated tiles in raised-floor data centers, Build Environ. 41 (6) (2006) 734-744. https://doi.org/10.1016/j.buildenv.2005.03.005.

[12] K. Fouladi, A. P. Wemhoff, L. Silva-Llanca, K. Abbasi, and A. Ortega, Optimization of data center cooling efficiency using reduced order flow modeling within a flow network modeling approach, Appl. Therm. Eng. 124 (2017) 929-939. https://doi.org/10.1016/j.applthermaleng.2017.06.057.

[13] A. J. Shah, V. P. Carey, C. E. Bash, and C. D. Patel, Exergy Analysis of Data Center Thermal Management Systems, J. Heat Transfer 130 (2) (2008) 021401-021401-10. https://doi.org/10.1115/1.2787024.

[14] E. Samadiani, H. Amur, B. Krishnan, Y. Joshi, and K. Schwan, Coordinated Optimization of Cooling and IT Power in Data Centers, J. Electron. Packag. 132 (3) (2010) 031006-031006-14. https://doi.org/10.1115/1.4001858.

[15] S. Alkharabsheh *et al.*, A brief overview of recent developments in thermal management in data centers, J. Electron. Packag. 137 (2015) 040801-1-040801-19. https://doi.org/10.1115/1.4031326.

[16] Z. Song, Numerical investigation for performance indices and categorical designs of a fan-assisted data center cooling system, Appl. Therm. Eng. 118 (2017) 714-723. https://doi.org/10.1016/j.applthermaleng.2017.02.122.

[17] 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.

[18] A. N. S. V. S. Sarma and V. D. Ambali, Cooling solution for computing and storage server, in 2017 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2017, pp. 840-849.

[19] 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.

[20] S. V. Garimella, T. Persoons, J. A. Weibel, and V. Gektin, Electronics Thermal Management in Information and Communications Technologies: Challenges and Future Directions, IEEE Trans. Compon., Packag., Manuf. 7 (8) (2017) 1191-1205. https://doi.org/10.1109/TCPMT.2016.2603600.

[21] A. Bejan, Entropy generation minimization: The new thermodynamics of finite-size devices and finite-time processes, J. Appl. Phys. 79 (3) (1996) 1191-1218. https://doi.org/10.1063/1.362674.

[22] K. Zhang, Y. Zhang, J. Liu, and X. Niu, Recent advancements on thermal management and evaluation for data centers, Appl. Therm. Eng. 142 (2018) 215-231. https://doi.org/10.1016/j.applthermaleng.2018.07.004.

[23] Z. He, T. Ding, Y. Liu, and Z. Li, Analysis of a district heating system using waste heat in a distributed cooling data center, Appl. Therm. Eng. 141 (2018) 1131-1140. https://doi.org/10.1016/j.applthermaleng.2018.06.036.

[24] C. D. Patel, A vision of energy aware computing from chips to data centers., In: Proceedings of ISMME 2003. Tsuchiura, Japan;

[25] J. Glanz, Power, pollution and internet. https://www.huffingtonpost.com/tim-mohin/power-pollution-and-the-i_b_1966893.html, 2012 (accessed 1 December 2018)

[26] C. D. Patel, C. E. Bash, C. Belady, L. Stahl, and D. Sullivan, Computational fluid dynamics modeling of high compute density data centers to assure system inlet air specifications. , In; Proceesings of IPACK'01

[27] K. Shen, Z. Zhang, Z. Zhang, and Y. Yang, Investigation of effect on cross-flow heat exchanger with air flow non-uniformity under low Reynolds number, Adv. Mech. Eng. 9 (7) (2017) https://doi.org/10.1177/1687814017708088.

[28] T. Gao, J. Geer, and B. Sammakia, Nonuniform temperature boundary condition effects on data center cross flow heat exchanger dynamic performance, Int. J. Heat Mass Transfer 79 (2014) 1048-1058. https://doi.org/10.1016/j.ijheatmasstransfer.2014.09.011.

[29] A. Sakanova, S. Alimohammadi, J. McEvoy, and T. Persoons, Hybrid-Cooled Data Center Server Layout Optimization for Air-Side Heat Recovery, in 2018 24rd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC), 2018. https://doi.org/10.1109/THERMINIC.2018.8593309.

[30] Intel® Server Board S2600TP Product Family and Intel® Compute Module HNS2600TP Product Family. https://www02.cp-static.com/objects/pdf/f/f88/1349633133_1_server-barebone-intel-compute-module-hns2600tp24sr-hns2600tp24sr.pdf, 2016 (accessed 13 March 2018)

[31] S. Alimohammadi, D. B. Murray, and T. Persoons, Experimental validation of a computational fluid dynamics methodology for transitional flow heat transfer characteristics of a steady impinging jet, J. Heat Transfer 136 (9) (2014) https://doi.org/10.1115/1.4027840.

[32] Nidec fan R40W12BS4D8-07. https://datasheet.octopart.com/R40W12BS4D8-07-Nidec-datasheet-16750012.pdf, 2014 (accessed 13 March 2018)

[33] Yungas A.Çengel and J. M. Cimbala, Fluid mechanics : fundamentals and applications (1st ed.). (McGraw-Hill series in mechanical engineering). Boston: McGraw-Hill Higher Education 2006.

[34] 4Gb: x8, x16 Automotive DDR3L SDRAM. https://www.micron.com/resource-details/cf973726-4e7c-449f-826b-7391513641b9, 2018 (accessed 13 March 2018)

[35] IBM Rear Door Heat eXchanger. ftp://ftp.software.ibm.com/software/uk/itsolutions/optimiseit/energy-efficiency-solutions/3-product-ibm-rear-door-heat-exchanger.pdf, 2005 (accessed 13 March 2018)

[36] R. K. Shah and A. L. London, Chapter VI - Parallel Plates (Laminar Flow Forced Convection in Ducts). Academic Press, 1978, pp. 153-195.