Applied Thermal Engineering 105 (2016) 622–638

Contents lists available at ScienceDirect

Applied Thermal Engineering

journal homepage: www.elsevier.com/locate/apthermeng

Cooling of server electronics: A design review of existing technology

Ali C. Kheirabadi, Dominic Groulx *

Dept. of Mechanical Engineering, Dalhousie University, Halifax, NS, Canada

HIGHLIGHTS

- Review heat transfer/flow parameters from studies on server cooling technologies.
- Perform direct quantitative comparison between cooling technologies.
- Identify heat transfer limitations, power requirements, etc.
- Summarize industry implementations of each cooling solution from a design perspective.
- Identify the capacity of cooling solution relative to heat loads expected by 2020.

ARTICLE INFO

Article history:
Received 31 December 2015
Accepted 15 March 2016
Available online 19 March 2016

Keywords:
Server electronics
Air cooling
Liquid cooling
Heat pipes
Pool boiling
Spray cooling
Jet impingement
Heat transfer

ABSTRACT

This review quantitatively examines and compares the heat transfer characteristics of several cooling technologies with potential application in the server electronics industry. Strategies that have been examined include traditional air cooling, single and two-phase indirect liquid cooling, heat pipes, pool boiling, spray cooling, and jet impingement. The characteristics that have been examined include heat flux values, coolant temperatures, and coolant flowrates; which serve as indicators of the heat transfer limitations and power requirements of each cooling solution. A direct comparison against anticipated server heat loads has shown that some form of liquid cooling is necessary in high performance computing applications; where individual processor heat loads are expected to reach 300 W by the year 2020. While in the case of general purpose computing, where individual processor heat loads are expected to reach 190 W, air cooling remains a viable option; although other factors such as operating costs, chip reliability, and waste heat recovery may still encourage the use of liquid cooling.

1. Introduction

Cooling of server electronics is a growing concern for businesses and government organizations involved with data processing, storage, and telecommunication. The main basis for this concern is the increasing cost and complexity of thermal management in server housing facilities. As the industry continues to adopt newer generations of processors that offer improved computational performance and capabilities, the rise in generated heat becomes unavoidable.

In a typical data center, roughly 33% of the total electricity consumed is allocated to thermal management of server electronics [1]. Concurrently, total global electricity usage in data centers has increased from approximately 152 billion kW h/year in 2005 to 238 billion kW h/year in 2010; roughly 1.3% of global electricity use [2]. In the United States alone, total data center energy usage amounted to 91 billion kW h/year in 2013 and is projected to reach 140 billion kW h/year in 2020; resulting in operating electricity costs of $13 billion/year and annual carbon emissions of 150 million metric tons [3]. These trends have exacerbated issues pertaining to energy consumption, operational costs, and the overall carbon footprint of server housing facilities; thus generating major discussions on the limitations and practicality of traditional air cooling strategies. Consequently, substantial efforts have been made to research and develop alternative liquid cooling solutions in an attempt to mitigate such issues.

Several reviews have been published in recent years dealing with thermal management challenges within data centers as well as available cooling technologies. Kant [4] chiefly reviews storage, networking, and management challenges within large virtualized data centers; however detailed discussions regarding standard server rack layouts and traditional air cooling systems are also provided. Garimella et al. [2] identify major drivers necessary for progress at both business and technological levels within data centers. Implementations of liquid cooling within industry are

* Corresponding author. Tel.: +1 (902) 494 8835.
E-mail address: dominic.groulx@dal.ca (D. Groulx).

http://dx.doi.org/10.1016/j.applthermaleng.2016.03.056
1359-4311/

Nomenclature

A_{HS}	heat spreader area (cm^2)
A_w	heated surface area (cm^2)
h_{eff}	effective heat transfer coefficient (W/cm^2 K)
P_{fan}	fan power consumption (W)
P_{pump}	pumping power consumption (W)
ΔP	pressure drop (kPa)
Q	coolant flowrate (cm^3/s)
q	total heat transfer rate (W)
q''_{HS}	heat spreader heat flux (W/cm^2)
R_{HS}	heat spreader thermal resistance (K/W)
R''_{TIM}	TIM thermal resistance (K cm^2/W)
T_{bath}	coolant bath temperature (°C)
T_i	coolant inlet temperature (°C)
T_o	coolant outlet temperature (°C)
T_{sat}	coolant saturation temperature (°C)
T_w	heated surface temperature (°C)

Acronyms

CDU	coolant distribution unit
CRAC	computer room air conditioning
CRAH	computer room air handling
HP	high power
HPC	high performance computing
LHP	loop heat pipe
LP	low power
LTS	loop thermosyphon
MCHS	microchannel heat sink
PM	porous medium
RDHx	rear door heat exchanger
SCHx	side car heat exchanger

discussed with focus upon energy savings resulting from waste heat recovery. Garimella et al. [1] provide a comprehensive qualitative review of the various thermal management technologies available within industry. Solutions examined include air cooling, liquid cooling (direct and indirect, single and two-phase), and miscellaneous cooling technologies such as thermoelectrics and synthetic jets. Ebrahimi et al. [5] focus upon methods and technologies available for data center waste heat recovery. Solutions such as air cooling and indirect single and two-phase liquid cooling are discussed quantitatively in terms of heat loads and system operating temperatures. Zhang et al. [6] review various technologies and configurations available for data center free cooling; estimated costs and energy savings are presented for airside, waterside, and heat pipe free cooling systems. Fulpagare and Bhargav [7] provide a general review on recent advances in data center thermal management. Numerical and experimental studies regarding air cooled data centers are summarized with focus upon facility layouts, efficiencies, performance metrics, and lifecycle analyses. Liquid cooling solutions such as on-chip water cooling and rear door heat exchangers are also discussed briefly. Oró et al. [8] focus upon the integration of renewable energy technologies into existing data center infrastructures and their respective impacts upon data center energy efficiencies and carbon footprints. Air cooling and various liquid cooling strategies are described along with their potential for free cooling. Finally, the *Data Center Handbook* [9] provides a comprehensive overview of the various cooling technologies and configurations available to data center engineers. Emphasis is placed upon air cooling and its numerous arrangements while alternative liquid cooling strategies such as on-chip single and two-phase cooling, pool boiling, and sealed dielectric systems are also discussed in great detail. Estimated energy savings associated with these technologies are also presented.

Collectively, these reviews paint a comprehensive picture of the various technologies available for data center thermal management. There is however a lack of direct quantitative comparison between the numerous cooling solutions proposed by researchers or industry. It is the aim of this design review to provide such an assessment through a fixed set of quantitative evaluation criteria. Additionally, the heat transfer mechanisms of each cooling solution are briefly reviewed while emphasis is placed upon their benefits and drawbacks as seen in industry applications. The overall intent is to provide a useful design tool for thermal engineers involved in server cooling applications; particularly at a time when new and innovative cooling technologies are frequently emerging.

2. Evaluation criteria

In this review, cooling solutions are evaluated quantitatively through their heat transfer limitations and power requirements. The following sections summarize the motivation and significance behind these criteria.

2.1. Heat transfer limitations

The foremost characteristic of any cooling solution is its heat transfer limitation. This is defined as the maximum rate at which heat may be generated by a processor while die operating temperatures remain below acceptable limits; typically recommended to be 85 °C [5]. This definition must also assume that a practical set of operating conditions apply. For example, an air cooling system's heat transfer limitation may be increased considerably by reducing inlet air temperatures; however this approach is rendered unpractical beyond a certain point due to amplified vapor-compression cooling costs.

Heat transfer rates of proposed cooling solutions are well documented in literature and are usually quantified through either the wall heat flux metric or the heat transfer coefficient. The wall heat flux metric is simply the total heat transfer rate divided by the heated surface cross sectional area. The heat transfer coefficient instead employs the total convective heat transfer area (*i.e.* base and fin surface areas) while also taking into account the heated surface and coolant temperatures.

Although very commonly used to rate the efficacy of a given cooling solution, these metrics fail to provide sufficient information necessary for direct comparison between alternative technologies. The wall heat flux metric provides no indication of the heat spreading area required for effective heat dissipation. An air cooled heat sink for instance, will require larger base and fin areas in relation to a waterblock when cooling processors of identical power and size. The waterblock is therefore a more effective cooling solution as it requires less server area for similar heat dissipation; the chip heat flux metric fails to identify this. The heat transfer coefficient faces a similar problem. The total convective heat transfer area, which accounts for extended surface areas, does not directly characterize the server area required for a given cooling solution.

In sight of this, this review quantifies the heat transfer rate of a given cooling solution through the *heat spreader heat flux metric*; the total heat transfer rate divided by the overall base area of the heat spreader. This is not a perfect metric as it has relinquished