Energy 77 (2014) 265–270

Contents lists available at ScienceDirect

Energy

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

Heat pipe based thermal management systems for energy-efficient data centres

Hussam Jouhara [a, *], Richard Meskimmon [b]

[a] Institute of Energy Futures, RCUK Centre for Sustainable Energy Use in Food Chains, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
[b] S & P Coil Products Ltd, SPC House, Evington Valley Road, Leicester, LE5 5LU, UK

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form
26 August 2014
Accepted 28 August 2014
Available online 23 September 2014

Keywords:
Heat pipes
Data centres
Free cooling
Heat exchangers
Energy efficiency

ABSTRACT

This paper investigates the potential applications for heat-pipe based heat exchangers in enhancing the efficiency of data centres' cooling. The paper starts by assessing current industry practise and highlighting the challenges facing the data-storage industry; illustrating the legislative, technical and commercial constraints that are now, or will be prevalent in the industry as the sector continues to grow to cater for the ever increasing appetite for public sector, commercial and consumer remote data storage. The concept of free cooling and its potential application in data-centres is then introduced and analysed. A theoretical model is then constructed based on the established, proven performance characteristics of heat-pipe technologies and the weather data for a typical region in the UK. A case study has been conducted thereon and the results indicate potential energy savings of up to 75% are achievable when utilising heat pipe based free cooling systems.

1. Introduction

Climate change has established itself as the greatest driving force for innovation in the industry for the first time in over two centuries [1–4]. Therefore, vigorous research activities to reduce the energy consumption of energy-intensive establishments such as data centres are being carried out by researchers and engineers over the last 30 or so years.

Data-centres are facilities used to house IT systems and associated hardware, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (i.e. HVAC – heating, ventilation & air conditioning), fire suppression and security devices. Continuous growth in this sector, particularly in storage, network and computer capacities have resulted in increasing power density and heat dissipation in data-centres, thereby increasing the cooling needs and costs. The UK government is already making efforts to control this increase and promote energy efficiency through the CRC (Carbon Reduction Commitment) legislation, as part of the Climate Change Act 2008 [5]. This legislation's main purpose is a mandatory carbon reduction and energy efficiency scheme aimed at changing energy use behaviours and whilst it is not specifically aimed at data-centres by its definition data-centres will be significantly impacted.

The Green Grid, which a non-profit organisation, identifies data-centre power and cooling as two of the biggest issues facing IT organisations today [6]; however it is not expected that data-centres will trade off resiliency or performance for energy efficiency due to their critical function and the increasing reliance on the internet. These centres require lots of flexibility, and will, for example, want to fit high density servers in any location around the world. Data-centres have been classified; the simplest being a Tier 1 data-centre and the most stringent being a Tier 4 data-centre, which would be designed to host mission critical computer systems. Further details are available in the TIA-942: Data-centre Standards Overview [7].

Current and planned legislation is also driving the need to find commercially viable alternatives for reducing or recovering the cost of cooling critical data storage infrastructure. Data storage service providers are faced with the challenge of developing storage technologies that generate less heat or finding the means to cool them that has lower net energy consumption.

In October 2008, the EU launched its Code of Conduct on Data-centres Energy Efficiency [8], which focuses on two areas. There were:

* Corresponding author. Tel.: +44 (0) 1895 267805; fax: +44 (0) 1895 269777.
E-mail address: hussam.jouhara@brunel.ac.uk (H. Jouhara).

http://dx.doi.org/10.1016/j.energy.2014.08.085
0360-5442/

- IT Load – this relates to the consumption efficiency of the IT equipment in the data-centres,
- The Facilities Load – this relates to the mechanical and electrical systems that support the IT electrical load such as cooling systems.

This Code of Conduct states that electricity consumed in data-centres, including enterprise servers, ICT (Information and Communication Technology) equipment, cooling equipment and power equipment, is expected to continue to contribute substantially to the electricity consumed in the EU (European Union). Western European electricity consumption for the sector of 104 TWh per year can be estimated for the year 2020. These projected energy consumption rises pose a problem for EU energy and environmental policies, it is therefore important that the energy efficiency of data-centres is maximised to ensure the carbon emissions and other impacts such as strain on infrastructure associated with increases in energy consumption are mitigated. The Code has already identified an increasing willingness of manufacturers and vendors to compete on the basis of energy efficiency in data-centres. It appears businesses are becoming increasingly aware of their environmental imperatives and the need to reduce energy consumption. However awareness does not necessarily lead to good decision making. The successful data-centre operators of the future will be aware of the financial, environmental and infrastructure benefits to be gained from improving the energy efficiency of their facilities through optimisation of power distribution, cooling infrastructure, IT equipment and IT output.

Based on the above, the thermal management of data-centres continues to be an expensive process, justified thus far because of the critical nature of such facilities. Increasingly high energy costs and the need for more efficient and sustainable cooling solutions will continue to be a top priority for engineers and planners. Taking the Silicon Valley in California as an example, the air conditioning of its data-centres is responsible for about 40% of its total energy consumption [9]. Other energy data from around the world report similar findings where these centres were found to consume many time higher than department stores, commercial buildings, and other high-energy consuming installations. Therefore, the utilisation of free cooling opportunities to reduce the energy cost and greenhouse gases emissions is a must-have if the climate conditions are suitable [10].

1.1. Free cooling

ASHRAE's Thermal Guidelines for Data Processing Environments [11] recommends a temperature range of 18–27 °C for general operational areas. In order to successfully adopt a free cooling strategy the outside air will need to be at least 5 °C cooler than that being maintained inside which imposes a limit on the applicability of these systems to temperate climates. Free cooling involves the replacement of 'warm' internal air with 'cool' outside air and is employed whenever the outside air is suitable for direct introduction into the data centre. Examples where free cooling is not appropriate are situations whereby the humidity of the outside air is not suitable or the contaminant level of outside air is high or there is a risk of smoke being drawn into the operating environment. In these instances indirect free cooling is appropriate whereby there is no air exchange just heat transfer via a heat exchanger.

Rather than conditioning the temperature of the entire data centre, contemporary techniques concentrate on the direct cooling of the server racks using hot or cold aisle containment. Such techniques involve the cooled supply air being delivered directly to the servers and the hot return air from the servers (typically around 36 °C) cooled back to the supply temperature. These elevated return temperatures open up the possibility to extend the range of outside conditions which are appropriate for free cooling.

The utilisation of spontaneous free cooling systems is proven to be feasible in cold to moderate-temperature climates where the ambient temperature is below the design temperature of the data-centre for most of the year. However, delayed free cooling is also possible with the increasing use of PCMs (phase change materials) to absorb the heat energy by melting the PCM material (latent heat), when the outside temperature values are not suitable for direct free cooling, and to release this heat energy when the outside ambient temperature drops to a low-enough value to secure the release of the absorbed latent heat and solidify the PCM within the heat exchanging system [12–15]. This storage capability, when coupled with a suitable heat exchanging system, can provide a free cooling solution even in desert-like climates, by utilising the wide day–night temperature difference.

In some cases, when free cooling solutions are considered, the pumping power that is needed to sustain the heat exchange process can prove costly and it can, in extreme cases, jeopardise the feasibility of the free cooling solution due to high running costs. This is the case when low temperature differences exist for most of the year between the data-centre return air and the ambient. Under such conditions, high flow rates of the hot and cold streams are required to secure the necessary heat transfer rate [16,17].

Taking the running cost of the heat exchangers into account, free cooling periods of data-centres can then be extended if this cost is lowered or even eliminated if passive cooling solutions, such as the heat pipe technology, are designed and adopted.

Based on the above and when the climate conditions allow for free cooling solutions, heat pipes, especially the gravity-assisted designs, are more economical than conventional system due to their passive operation, efficiency, reliability, and the relatively low cost of manufacturing when compared with conventional air cooling solutions that are used in data-centres. In addition, this technology provides an additional level of contingency, which helps in extending the intervals between regular maintenance schedules [18–22].

1.2. Heat pipes

Heat pipes have been utilised in small-scale electronics applications for decades for cooling critical microprocessor components [23,24]. The wicked heat pipe, which was developed in the 1960's by NASA for their aerospace and satellite needs has subsequently been universally adopted by the electronics industry [23]. These small-scale heat pipes, which work in any orientation or gravitational conditions to cater for the mobile nature of their usage, are mass-produced globally and have become a highly commoditised manufacturing industry in their own right. The wicks in these heat pipes are required to enable the operation of the heat pipe at any orientation.

Gravity, instead of the wick, can be utilised in heat pipes to enable the return of the working fluid condensate back to the evaporator. These heat pipes are called wickless heat pipes. Wickless heat pipes, also known as gravity-assisted heat pipes, are hermetically sealed tubes containing a working fluid in both the liquid and vapour phases. They utilise the highly efficient thermal transport process of evaporation and condensation to transport heat from one end to the other where the heat can be dissipated through a heat sink. The amount of heat that can be transported by these devices is normally several orders of magnitude greater than pure conduction through a solid metal [19,25–32]. As this heat pipe configuration relies on gravity to return the condensate working