Renewable and Sustainable Energy Reviews 31 (2014) 622–638

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities

Khosrow Ebrahimi, Gerard F. Jones, Amy S. Fleischer*

Department of Mechanical Engineering, Villanova University, Villanova, 800 Lancaster Avenue, PA 19085, USA

ARTICLE INFO

Article history:
Received 22 May 2013
Received in revised form
22 November 2013
Accepted 20 December 2013
Available online 21 January 2014

Keywords:
Data center
Waste heat recovery
Waste energy reuse
Absorption refrigeration
Organic Rankine cycle
Thermoelectric

ABSTRACT

The depletion of the world's limited reservoirs of fossil fuels, the worldwide impact of global warming and the high cost of energy are among the primary issues driving a renewed interest in the capture and reuse of waste energy. A major source of waste energy is being created by data centers through the increasing demand for cloud based connectivity and performance. In fact, recent figures show that data centers are responsible for more than 2% of the US total electricity usage. Almost half of this power is used for cooling the electronics, creating a significant stream of waste heat. The difficulty associated with recovering and reusing this stream of waste heat is that the heat is of low quality. In this paper, the most promising methods and technologies for recovering data center low-grade waste heat in an effective and economically reasonable way are identified and discussed.

A number of currently available and developmental low-grade waste heat recovery techniques including district/plant/water heating, absorption cooling, direct power generation (piezoelectric and thermoelectric), indirect power generation (steam and organic Rankine cycle), biomass co-location, and desalination/clean water are reviewed along with their operational requirements in order to assess the suitability and effectiveness of each technology for data center applications. Based on a comparison between data centers' operational thermodynamic conditions and the operational requirements of the discussed waste heat recovery techniques, absorption cooling and organic Rankine cycle are found to be among the most promising technologies for data center waste heat reuse.

Contents

* Corresponding author. Tel.: +1 610 519 4996.
E-mail address: amy.fleischer@villanova.edu (A.S. Fleischer).

1364-0321/$ - see front matter
http://dx.doi.org/10.1016/j.rser.2013.12.007

1. Introduction

During the past two decades, the increasing demand for data processing, data storage systems and digital telecommunications coupled with the simultaneous advances in computer and electronic technology have resulted in a dramatic growth rate in the data center industry. In United States, during the years 1998–2010, the number of data centers increased from 432 to 2094 [1] and the overall power usage of data centers experienced a doubling between the years 2000 and 2006. Recent energy statistics indicate that the data center industry is responsible for 1.3% of the world and 2% of the United States electricity consumption [1]. In fact, between 25% and 35% of the worldwide power consumption of data centers (30 Billion Watts) is consumed by US data centers [1].

The growth has been not only in the number of data centers, but also in the size, floor area and/or computing density of datacenters [2]. The data center industry operates on many different scales, with the construction of huge datacenters with floor areas as large as ~9000 m^2 containing thousands of server racks and several megawatts power usage [2], yet also with the design of compact data centers with more and more computing power packed into smaller and smaller spaces. The construction and annual operating costs for a typical data center are ~$15,000/$m^2$ and $1500/$m^2$, respectively [2].

Considering the ongoing growth of the data center market coupled with the development of higher power density server components, it is expected that the share of electricity consumption by data centers will continue to increase for the foreseeable future. The predictions for annual increases in data center power demand are as high as 15–20% [3].

Despite increasing efforts to integrate renewable energy sources into the overall US power grid, fossil fuels are still the major energy source for electricity production. Considering the limited supply of fossil fuels and the consequent air pollution and global warming of using them to produce electricity, not to mention the increasing cost of electricity in many markets, there is a growing effort to capture and reuse waste heat in all types of energy conversion systems. In many cases, this effort will need to include substantial work into improving capture and transport of the waste heat, and into increasing the efficiency of the recovery system.

Large data center operators, including familiar companies such as Google (with ~300 million watts of power usage) and Facebook (with ~60 million watts of power usage) are taking action to increase the fuel efficiency and decrease their wasted power [1]. In fact, most data centers rarely operate at their maximum rated load. The majority of servers operate at or below 20% of their maximum capacity most of the time yet even when the system is idle, 60–100% of the maximum power is still drawn from the grid [4,5]. Almost all the electrical power supplied to the server is dissipated into heat, necessitating the use of large scale cooling systems to keep the server rack temperatures in a safe operational range. There are significant research efforts ongoing to develop dynamic need-based resource allocation as one way to reduce energy dissipated by data centers [6,7]. In combination with recovery and reuse of the waste heat, these efforts can lead to more efficient and cost-effective operation of data centers.

Many data centers feature servers with power densities in excess of 100 W/cm^2 and even up to as high as 200 W/cm^2, which means that a rack with a 0.65 m^2 footprint has heat dissipation requirements as high as 30 kW, or roughly 30 times higher than the amount of energy dissipated by a typical rack with the same footprint in 1990 [8]. This increased heat dissipation demand [9,10], increases the costs associated with powering and cooling datacenters. Thus, the recovery and reuse of waste heat energy has the potential to significantly reduce data center operational costs.

The main barrier to the implementation of waste heat recovery and reuse systems into operational data centers in that in contrast to many industrial waste heat recovery systems, the heat, although plentiful, is of low quality. The capture temperature is limited by the temperature limits of the electronics, which remain below 85 °C in most cases. This low quality temperature makes it quite challenging to be reuse the heat through conventional thermodynamic cycles and processes.

This study provides a much needed comprehensive review of all commonly available waste heat reuse techniques with a specific focus on their particular application to data center operating conditions. Each technique will be examined through the lens of a data center operator with consideration of both retrofit to legacy air cooled data centers and to integration into newer water cooled data centers. Through a comparison between the technology's operational requirements with the specifications of waste heat sources and streams in different type of data centers, the most promising solutions for waste heat reuse are identified.

This review paper is presented in four sections. In the first section, the significance and necessity of research on waste heat recovery in data centers was explained through a review of the past, current, and projected future growth trend of data center industry and its impact on energy consumption and environment. In the second section, the energy generating components of data centers are introduced and their physical arrangement is described. A comprehensive literature review is given in which thermal loads and temperature limits in different components are categorized and tabulated depending upon the data center cooling type. This section also presents a detailed review of heat sources and streams in different data center designs which provides the reader with a broad and detailed background on data center thermodynamic operating conditions. In the third section, eight potential; low-grade waste heat recovery technologies are reviewed along with their operational requirements. The benefits and challenges associated with the implementation of each technology for data center waste heat recovery are discussed and listed. In the final section, the suitability of each technology for different data center designs is discussed and the most promising options for data center applications are identified for further investigation.

2. Overview of data centers

2.1. Physical organization

A data center is a space allocated to house most of the ICT (Information and Communication Technology) modular assets such as servers, switches, and storage facilities and to control the environmental conditions (temperature, humidity, and dust) to ensure that the ICT systems operate reliably and in a safe and efficient manner [2,11]. Depending upon the scale of the business, a data center might include a single rack of equipment or a few or