Environmentally opportunistic computing: A distributed waste heat reutilization approach to energy-efficient buildings and data centers

J. Zachary Woodruff<sup>a</sup>, Paul Brenner<sup>b</sup>, Aimee P.C. Buccellato<sup>c</sup>, David B. Go<sup>a,d,s</sup>

<sup>a</sup> Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, United States
<sup>b</sup> Center for Research Computing and Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, United States
<sup>c</sup> School of Architecture, University of Notre Dame, Notre Dame, IN 46556, United States
<sup>d</sup> Department of Chemical Engineering and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, United States

**A R T I C L E   I N F O**

Article history:
Received 22 April 2013
Received in revised form 20 September 2013
Accepted 29 September 2013

Keywords:
Distributed computing
Building efficiency
Waste heat
Data center
Modeling framework

**A B S T R A C T**

With building energy consumption rising in industrial nations, new approaches for energy efficiency are required. Similarly, the data centers that house information and communications technology continue to consume significant amounts of energy, especially for cooling the equipment, which in turn produces vast amounts of waste heat. A new strategy to overcome these challenges is called environmentally opportunistic computing (EOC), which conceptualizes the data center as a series of distributed heat providers (nodes) for other-purposed buildings that use the waste heat from the data center nodes to offset their own heating costs. In this paper, a general framework for evaluating the deployment of EOC is developed and select model cases are analyzed. The results show that by redefining a centralized data center as distributed nodes across multiple buildings, the overall energy consumption of an organization decreases significantly. The advantages of applying EOC to buildings that require constant water heat as opposed to seasonal space heat are explained, and the method of distributing the computational load among data center nodes is evaluated.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

It is well known that energy consumption by commercial and residential buildings continues to rise worldwide in developed nations, with most of the energy going to heating, ventilation, air conditioning (HVAC), and water heating [1]. For example, a 2011 report by the United States (U.S.) Department of Energy showed that building energy consumption accounts for approximately 40% of the energy consumption in the U.S. today, or nearly 11.61 trillion kW-h/yr [2]. Further, nearly all of this energy is generated by non-renewable energy resources (e.g., petroleum, natural gas), thus presenting a challenge to develop both more energy-efficient buildings and buildings that integrate renewable energy sources. For this reason, a number of studies have explored ways to reduce energy consumption in buildings, such as optimizing the control strategy for the building management system [3] and incorporating novel construction practices and materials [4]. Alternatively, a variety of renewable energy concepts, utilizing, for example, solar energy [5,6] or wind energy [7], have begun to emerge at the single building scale. However, given the magnitude of the problem, more aggressive, scalable solutions and alternative routes to energy efficiency need to be developed.

Congruently, data centers, which house the information and communications technology (ICT) that supports our economic, government, and social infrastructure, consume a significant amount of electricity. Recent estimates place data center consumption at nearly 198.8 billion kW-h/yr and rising or ~1.3% of all electricity use worldwide (and ~2% of electricity use in the U.S.) [8]. Further, this usage is widely spread among many data centers and organizations. For example, one of the largest web-based presences, Google Inc., was estimated to account for less than 1% of all data center electricity consumption worldwide. One main concern is that not all of this electricity goes to operate the ICT equipment; a significant amount goes to facility operation, especially cooling the data center. Because nearly all of the ICT electricity consumption is manifested as heat, and overheating directly impacts reliability and performance, data centers cool and condition their ICT equipment continuously in order to meet customer demand for consistent availability and uptime. Recent studies have suggested that on average ~40% of the electricity consumed in a data center powers equipment required to maintain operating conditions within the facility, and nearly all of this is for thermal management.

One perspective on sustainable development is to address both of these challenges symbiotically by using the heat generated by data centers as space and/or water heat for commercial, residential,
or even industrial buildings and facilities. By harvesting the heat produced by data centers for other–purposed buildings, the potential exists to reduce the overall (combined building and data center) energy consumption of an organization. The idea of harvesting the heat produced by a data center has begun to receive more attention as the significant energy costs of data centers continue to rise [9]. For example, adsorption systems that harvest the waste heat to drive the primary cooling system with no additional power input have been suggested by multiple groups [10,11]. Alternatively, a novel use of relatively hot water (~60 °C) to liquid cool a data center rack demonstrated the feasibility of using liquid cooling to generate waste heat that can be directly used for building heating [12]. Practically, a number of groups worldwide have begun to implement waste heat harvesting strategies [13] to the extent that an energy reuse effectiveness (ERE) metric has been proposed to evaluate data centers that reuse waste heat [14]. However, these all use a single centralized data center to service a single other–purposed building immediately adjacent to the data center. A perhaps more effective waste heat utilization philosophy is to decentralize large data centers into smaller data center nodes that are directly integrated into the buildings they serve. This philosophy, which we call environmentally opportunistic computing (EOC) [15], takes the concept of distributed computing and reprioritizes it as distributed heating, where the data centers are treated not only as entities that meet the needs of their computing end–users but also the needs of the buildings they serve as heat sources. Further, if integrated with the buildings’ existing HVAC systems, the data center nodes can potentially benefit from cooling provided by the building.

In practice, EOC would consist of distributed “containerized” data center nodes attached to or integrated with other–purposed buildings such as office buildings, apartment complexes, hotels, or university/municipal buildings and facilities. Fig. 1 shows a vision for EOC as a series of nodes implemented across a municipality, community, university, or industrial campus. Buildings throughout the organization would be outfitted with EOC nodes that either provide space or water heat depending on the needs and function of the building. Computational jobs would then be migrated from node to node based on the computational requirements of the job, the availability of servers in the node, and the waste heat required by the integrated building. This, in effect, creates a market place where both the buildings and the end-users act as both consumers and providers – end-users providing heat to the buildings and the buildings providing computational services to the end-users. Further this vision can be extended across multiple communities where local utility availability and cost could also play essential roles in the EOC marketplace. This approach is similar to the concept of the locally integrated energy sector where waste heat and renewable energy sources are integrated and shared across a community to reduce the overall carbon footprint [16], but includes the additional complexity of consumers dictating the production of heat based on their computational demand.

How a single node is integrated into a building would depend on the specific needs of the building (does it require space or water heat), the structure and function of the building, the local climate, as well as numerous other factors that would need to be considered on a case–by–case basis. EOC is built around the concept that each EOC node operates with free cooling to keep the ICT equipment functional and reliable, using either unconditioned ambient air (or return air from the building) for air cooling or the building’s existing plumbing for liquid cooling. The four basic EOC node types are then: (1) air cooling to space heat, (2) air cooling to water heat (requiring a heat exchanger), (3) water cooling to space heat (requiring a heat exchanger), and (4) water cooling to water heat (with heat exchanger optional). From this perspective, the energy savings from EOC comes from multiple sources. The other–purposed buildings’ energy usage would be reduced by the free heat from the EOC node, and the cost to cool and condition a large, centralized data center would be removed. While there are hurdles to broad EOC adoption [17], such as security concerns, distributed server administration, and coordinating building and EOC node control systems, EOC is a compelling approach to manage energy resources as energy–hungry computing technologies become even more integrated into society.

To demonstrate the concept of EOC, an EOC node has been developed and integrated with a local greenhouse in a collaboration between the University of Notre Dame and the City of South Bend, Indiana, U.S.A. [15]. As shown in Fig. 1b, the EOC node uses free ambient air cooling and exhausts its waste heat directly into the greenhouse as space heat. With three racks of servers connected directly to Notre Dame’s research network and actively running computational jobs, the EOC node has been shown to deliver ~15–40 kW of waste heat to offset the space heating needs of the greenhouse during cooler months. During warm months, the waste heat is not utilized and exhausted directly to ambient.

While this prototype demonstrates a practical implementation of EOC, it does not reveal the benefits of more realistic and broad deployment of EOC. In this work, we take a higher–level perspective to analyze the deployment of EOC for various building sectors – commercial office buildings and apartmet buildings or hotels – to understand the benefits of scale. We establish metrics to not only understand performance but also to evaluate the deployment in order to guide future design and implementation decisions.
2. Model for analysis of EOC deployment

2.1. Definition of the problem and system

There are different perspectives on how to evaluate the performance of an EOC deployment. If we consider that the energy used for ICT equipment will be spent regardless of whether the waste heat is utilized, then one perspective is that any utilization of the waste heat is a successful deployment of EOC. However, in order to evaluate an effective deployment of EOC, how much the EOC deployment actually reduces the overall energy consumption of the organization needs to be evaluated. Further, on an individual node/building basis, an effective deployment should utilize as much of the waste heat being produced by the node (that is, the supply of waste heat should not exceed the heating demand of the building) and the waste heat should appreciably reduce the power consumption for heating the building. In other words, the waste heat used by the building should perfectly offset the heating requirement for the building. If a significant amount of waste heat is rejected or the available waste heat is small compared to the building’s needs, then alternative node/building integration should be considered. Note that here the effectiveness of a deployment accounts for how the reutilized waste heat is used to offset other heating costs, and is thus more inclusive than the existing ERE metric, which is used to evaluate the impact on the data center alone [14].

In this analysis, we consider the energy consumption of buildings, data centers, and nodes in various model EOC deployments. We do not include factors such as capital costs to build these integrated node/buildings, to retrofit existing buildings, or to modify data centers for reduced ICT loads. In the following, we define the problem and system at hand and overview the key parameters and definitions that describe EOC deployment. A full mathematical description of each of the variables in the problem is included in the Appendix.

For our model system, outlined in Fig. 2, we consider an organization that has a single large, centralized data center that produces a total amount of computational power \( P_{\text{org,ICT}} \) for the organization and then a number of buildings \( N \) that each require some amount of space and water heat \( q_{b,\text{req}} \) to function properly. When EOC is deployed, we consider that the total amount of computational power for the organization remains constant, but that it is distributed to individual nodes \( P_{n,\text{ICT}} \) integrated with the various buildings; if there are an insufficient number of nodes to handle the total computational load, then the remainder is computed at the centralized data center. Thus, the ICT power in each node \( P_{n,\text{ICT}} \) corresponds to some amount of waste heat \( q_w \) delivered to the building. In theory, all the node ICT power manifests itself as usable waste heat, but in practice, there are heat transfer losses such that \( q_w < P_{n,\text{ICT}} \). Ideally the usable waste heat delivered to the building perfectly offsets the heating need for the building, \( q_w = q_{b,\text{req}} \).

Fig. 1. (a) Schematic of EOC deployed across an organization (municipality) with EOC nodes integrated into a variety of buildings and compute jobs migrated between the nodes. (b) Schematic and photo (inset) of EOC node integrated in a greenhouse via a collaboration between the University of Notre Dame and the City of South Bend, IN. This is an air-to-space configuration where the EOC node uses free ambient air cooling and delivers the waste heat as space heat for the greenhouse.

Fig. 2. Schematic of a model organization alone and with EOC deployment.
2.2. EOC effectiveness metrics

There are two essential metrics that characterize the performance of EOC for our model organization—how an integrated node reduces the power consumption of a single building and how the EOC deployment reduces the overall power consumption of the organization.

The building power consumption effectiveness metric $\varepsilon_{bldg}$ assesses how much of the building’s heating power consumption is reduced by the presence of an EOC node or

$$
\varepsilon_{bldg} = \frac{\text{change in building power consumption with EOC}}{\text{building power consumption without EOC}}. \tag{1}
$$

The building power consumption without EOC ($P_b$) is simply the power consumption of the building’s native heating system delivering the required heating $q_{b,req}$. When there is an EOC node, the useable waste heat delivered to the building $q_w$ offsets some or all of $q_{b,req}$. However, there are additional energy expenditures accompanied with the delivery of this waste heat. If, for example, the waste heat is less than the needs of the building ($q_w < q_{b,req}$), the building will still expend heating energy to make up the difference. Additionally, in many cases the waste heat might not come at a suitable temperature for direct use by the building—for instance, the temperature of the waste air heat will typically be $\sim 90 - 100 \degree F$ and the building will need to expend additional energy to not only enhance the heat so it can be utilized but also to deliver it around the building via pumps or fans that would not normally be required. We group all this building energy expenditure associated with incorporating an EOC node into the term $P_{b,EOC}$, which is defined explicitly in the Appendix, noting that $P_{b,EOC}$ can vary in time depending on how much useable waste heat is delivered to the building under specific circumstances.

The building’s power consumption effectiveness can thus be defined as

$$
\varepsilon_{bldg} = \frac{P_b - P_{b,EOC}}{P_b}. \tag{2}
$$

Under this definition, $\varepsilon_{bldg}$ is not constrained to be between 0 and 1 by any physical limits. Ideally, $\varepsilon_{bldg} = 1$ implying that EOC relieves all heating power consumption by the building such that the building expends no energy when EOC is used ($P_{b,EOC} = 0$). However, even if the waste heat meets the energy needs of the building ($q_w = q_{b,req}$), there will still be some energy costs associated with utilizing and delivering the waste heat to the building. Thus, realistically, $\varepsilon_{bldg} < 1$ under even the best circumstances. If for some reason, these additional energy costs are unreasonably large, then it is even possible for $\varepsilon_{bldg} < 0$, implying a deployment of EOC that actually increases the energy consumption of the building. Finally, if $\varepsilon_{bldg} = 0$, then there is no net improvement or detriment to the building.

The organization power consumption effectiveness metric $\varepsilon_{org}$ assesses how much of the organization’s power consumption is reduced by extracting some servers from a data center to be placed in an EOC node or

$$
\varepsilon_{org} = \frac{\text{change in organization power consumption} - P_{org,EICT}}{\text{organization power consumption} - P_{org,EICT}} \tag{3}
$$

In this definition, we have subtracted the total ICT power consumption of the organization ($P_{org,EICT}$) from the total energy consumption so that its magnitude does not arbitrarily impact $\varepsilon_{org}$. We assume this to be an inherent energy expenditure that is always present, i.e., the organization always requires a specific amount of computing power. Since $P_{org,EICT}$ is a fixed constant, we can divide ICT power consumption into that loaded into the data center ($P_{dc,EICT}$) and that loaded into the EOC nodes ($P_{n,EICT}$) such that:

$$
P_{dc,EICT} = P_{org,EICT} \quad \text{when EOC is not implemented};
$$

$$
P_{dc,EICT} = P_{org,EICT} - \sum P_{n,EICT} \quad \text{when EOC is implemented}.
$$

The organization’s power consumption without EOC is simply the summation of the power consumption of all the buildings in the organization plus that of the centralized data center, $P_{org} = \sum P_b + P_{dc}$. The energy consumption of the data center $P_{dc}$ is the sum of the ICT equipment in the data center and the power required for the facility including thermal management (air conditioning) as well as lighting and other utilities.

When an organization includes EOC nodes, the power consumption of each building decreases ($P_{b,EOC}$) as well as that of the data center ($P_{dc,EOC}$). However, additional power is now needed to run each individual node $P_{n,EOC}$, including both the ICT equipment in the node and any additional node facility requirements. Thus, the organization’s power consumption with EOC is simply $P_{org,EOC} = \sum P_{b,EOC} + P_{dc,EOC} + \sum P_{n,EOC}$, where the summations are over each of the building/nodes pairs in the entire organization. If a building in the organization does not have a node, then $q_w = 0$ and $P_{n,EOC} = 0$ and so forth for that building.

The organization’s power consumption effectiveness can thus be defined as

$$
\varepsilon_{org} = \frac{P_{org} - P_{org,EICT}}{(P_{org,EICT} - P_{org,EICT})}. \tag{4}
$$

Like $\varepsilon_{bldg}$, $\varepsilon_{org}$ is not constrained to be between 0 and 1 by any physical limits. In an ideal case, there are no cooling requirements for either the nodes or data center when EOC is implemented because their reduced sizes now enable efficient free cooling and the facility power requirements for each approach 0. Additionally, if each building/node pair is ideally matched, such that the waste heat perfectly offsets the building heating need and there is no cost to deliver the waste heat, then $\varepsilon_{org} = 1$. However, there will always be some facility costs for both the data center and nodes as well as some cost to deliver the waste heat, so that this is never realistically achievable. Again, if these supplementary energy costs are unreasonably large then it is possible for $\varepsilon_{org} < 0$ corresponding to a deployment of EOC that actually increases the net power consumption of the organization.

2.3. Assumptions and cases considered

In analyzing a building or organization, both metrics can be considered in both a time dependent and time average manner, as the operation and performance of EOC will necessarily vary over time and season. Thus, we consider two types of buildings in the organization that are illustrative of typical heating expenditures that could be offset by EOC. The first is a commercial office building where the majority of the heating demand is for space heating during colder seasons, and the water heat requirements are fairly constant but small by comparison. The second is an apartment building where there is a need for space heat, but also a significant and fairly constant need for water heat for bathing, cooking, and laundry year-round that is much greater than a typical office building. Fig. 3 illustrates representative space and water heat usage (normalized to the maximum consumption of the apartment) for these two building types over a typical year based on data in [18] and [19]. The data has been normalized since the exact values have been interpreted from the literature, and we use these as characteristic consumptions that can be scaled with our model building size. Clearly, type, size, function, construction, and location will influence the exact behavior of these buildings, but we consider these generic cases as a baseline for this study.
To offset this building usage, there are two approaches to distributed waste heat generation. One is that the node could deliver a constant amount of waste heat to the building, which is the equivalent to the servers in the node operating constantly, i.e., \( P_{\text{ICT}} \neq f(t) \). With the constraint that the total ICT load for the organization is fixed, the ICT load in the data center is also constant under these conditions. We label this static loading, and in this case there can be times when the waste heat perfectly offsets the building needs, only offsets a portion of the heating need, or the waste heat produced exceeds the needs of the building. In this case, some of the waste heat is rejected, as shown conceptually in Fig. 3a for the apartment building. It is non-optimal to reject this waste heat, so the alternative is that the node's ICT load varies in order to match the building's heating needs perfectly. Under this dynamic loading, whenever the building's heating needs increase then the ICT load in the node increases and vice versa, as shown in Fig. 3b for the office building. When less waste heat is needed, \( P_{\text{ICT}} \) decreases and the computing jobs are migrated back to the data center (increasing \( P_{\text{dc,ICT}} \)) or other nodes. Thus both \( P_{\text{ICT}} \) and \( P_{\text{dc,ICT}} \) vary in time with the constraint that the total ICT load for the organization is constant.

In this work, both of these loading conditions are considered.

The facility power consumption of the data center and that of a node are also key parameters. As noted above, if the data center is appreciably reduced in size due to redistribution to nodes, then the facility operating costs in principle decrease. This can viewed from different perspectives. If ICT hardware is removed from a data center and placed in nodes, the cooling power expenditure would initially decrease because the air conditioning units would not have to remove as much heat. But since the cooling equipment would have been sized for a full data center, if too much ICT hardware is removed then the air conditioning units would begin operating inefficiently because they would be oversized for the heat load. Thus the facility power consumption is typically some non-linear function of the ICT load, \( P_{\text{dc,ICT}} \). An alternative perspective is that not only is the ICT equipment removed, but the data center itself is reduced in physical size with new air conditioning units properly sized for the new ICT load. In this case, the facility power consumption will decrease monotonically as \( P_{\text{dc,ICT}} \) decreases. Further, both of these scenarios are also predicated on which loading approach is taken – static or dynamic. If static loading is used in the nodes, then the physical size of data center can be scaled down to account for its reduced ICT load. However, if dynamic loading is used in the nodes, then the data center cannot be scaled because it has to be sized to manage the worst-case scenario, or conditions when buildings do not need any waste heat and all of the ICT is redirected to the data center.

The power consumption of a data center can be characterized by the power usage effectiveness (PUE) [20], which is the ratio of the total power required to operate the data center to the power used to operate the ICT hardware. The industry average is PUE ~ 1.8, implying that operational power is ~80% of the ICT power [21]. In this work, when static loading was assumed, we scaled data center facility power consumption linearly based on the remaining ICT load in the data center and assumed that \( \text{PUE}_{\text{dc}} = 1.8 \), regardless of the data center’s size. For dynamic loading, we assume the PUE fluctuates as the load in the data center fluctuates. Because the exact form of the non-linear relationship between facility and ICT power consumption is not known, we assumed the relationship scaled linearly so that the PUE was a constant at \( \text{PUE}_{\text{dc}} = 1.8 \). Thus while we made the same assumption of \( \text{PUE}_{\text{dc}} = 1.8 \) for both loading conditions, they represent two physically different scenarios. For each individual EOC node, we assumed that the node would benefit from free cooling – either unconditioned ambient air or liquid cooling via integration into the building’s plumbing system – driving the node facility power consumption down, and we estimated a constant value of \( \text{PUE}_{\text{n}} = 1.3 \). However, if a node is oversized for a building, then free cooling is no longer as effective and additional facility power is required to cool and remove the excess heat in the node, adding to the facility power consumption of the node \( P_{\text{P,E0C,facility}} \), causing the PUE of the node to increase. We assumed this was the case whenever the node ICT load exceeded the heating needs of the building during static loading and increased the node facility power consumption to include removing this excess heat.

### 3. Results and analysis

For this analysis, both an organization with a single EOC node/building and an organization with multiple EOC node/buildings are considered. While there is a large range of possible parameters for both the relative size and energy consumptions of the buildings, the problem was simplified by assuming the maximum building heating power consumption during the year was 10% of the total ICT load of the organization. A reasonable approximation for a mid-rise apartment is 300 kW [19], such that our model data center was 3 MW. To make it possible to compare the energy savings between office and apartment buildings, the total yearly energy consumption was assumed to be equal. This caused the peak power consumption of the office building to be greater than that.
for the apartment building, but with a lower annual water heating requirement as shown in Fig. 3. We also assumed that a given node could be generically designed to produce space and water heat, rather than specifying a particular waste heat mode for each node.

3.1. Single building deployment

First consider an organization consisting of a data center and a single building (sized to 1/10 of the data center) in order to understand the impact of the two loading conditions as shown in Fig. 3. Fig. 4 shows the building and organization effectiveness for the static loading condition when the node is sized to offset half of the peak energy consumption of the building, i.e., \( P_{n,ICT} = 0.5 \cdot \max(q_{b,req}) \), where \( \max(q_{b,req}) \) is defined as the peak energy consumption that occurs during the winter. We observe an interesting trend that highlights the competing interests at stake. The building’s effectiveness is maximized from the spring through autumn months, as the waste heat from the node completely offsets the needs of the building. (Note that \( \epsilon_{bldg} < 1 \) due to assumed inefficiencies in the node/building system – see Appendix.) However, \( \epsilon_{bldg} \) decreases dramatically in the winter months as the node only offsets part of the building’s increasing heating requirements. Inverse to this relationship, we see that the organization effectiveness \( \epsilon_{org} \) peaks in the winter months as the reutilized waste heat offsets the entire organization’s energy use but decreases in the summer months as the building rejects the waste heat provided by the node. Additionally, because only a single node sized to less than 10% of the data center is considered, we see that the organization effectiveness is relatively small as most of the ICT power still resides in the data center. The office building is slightly better than the apartment building in the winter months due to its greater need of consistent space heat, but the apartment building has superior performance in the summer because it still requires significant, consistent water heat (Fig. 3). Trivially, the time-averaged building effectiveness is maximized when the node is optimally sized so that its waste heat always offsets any heating needs of the building or \( P_{n,ICT} = \max(q_{b,req}) \). More interestingly, the time-averaged organization effectiveness is also greatest at \( P_{n,ICT} = \max(q_{b,req}) \), even though for the majority of the year the heat is being rejected to the ambient.

For dynamic loading, the building heating needs are always perfectly offset by the EOC node such that \( \epsilon_{bldg} \) is constant and at its maximum value. The time variation of \( \epsilon_{org} \) follows a similar trend as that for static loading, but the magnitude is not the same, as shown in Fig. 5 for an apartment building. During the summer months, when the building does not require space heat, the ICT load is migrated back to the data center, whereas for static loading the waste heat produced by the node and not used for water heat is simply rejected to the environment. Because of the underlying assumption that the PUE of the node is better than that of the data center (\( \text{PUE}_n < \text{PUE}_{dc} \)) due to free cooling, it actually is more efficient for the organization to run the ICT load in the node rather than the data center. That is, there is an inherent benefit to removing power from the data center and distributing it to the nodes where the PUE is more favorable. Clearly, if \( \text{PUE}_{dc} \) can itself be reduced by taking advantage of free cooling, albeit at a much larger scale, this benefit becomes negligible. This also manifests itself in the time-averaged organization effectiveness shown in Table 1, where the static loading is marginally better than the dynamic loading. We also note that the apartment building is marginally better than the office building, and this is because of the greater contribution of year-round water heat to the building’s heating needs, such that less node waste heat is rejected in the summer months. The absolute values of \( \epsilon_{org} \), ranging from 0.025 to 0.17 (representing 2.5–17% energy savings), and the time average of \( \epsilon_{org} \sim 0.1 \) suggest that the deployment of a single node containing 10% of the data center’s ICT load could improve the organization’s energy consumption by ~10%, as expected. However, we note that these absolute values are a strong function of the parameters we choose for this model system, and inefficiencies in any of the system components could alter this value.

The main benefit of dynamic loading is when there are many different building types in the organization that require heating at different times – such as if the buildings are at different locales with different climates. In this situation, the computing loads can be distributed to match the heat requirements of each individual building, limiting the amount of waste heat that is rejected by the buildings. Because this analysis does not examine buildings with yearly heating requirements that offset each other this is an area to explore in future studies.

3.2. Multiple building deployment

When considering multiple buildings, there are numerous configurations and combinations of buildings and nodes that could be considered. But to illustrate the effect here, we assume that a given organization either has only office or only apartment
buildings. Consider first an organization with two buildings and one data center and the following two cases: (a) one large node for one of the two buildings and (b) two smaller nodes, one for each building. In case (a), a single large node could cause a significant portion of the waste heat to be rejected, whereas in case (b) the two smaller nodes could distribute the waste heat more effectively and allow more of it to be utilized. Fig. 6 shows the time-averaged organization effectiveness $\bar{\varepsilon}_{\text{org}}$ as a function of the total ICT power in all the nodes, in the single node in case (a) or in the sum of the two nodes in case (b). We normalize this total ICT power to the heat requirement of a single building, or $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}})$. For case (a), the single node is optimally sized when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 1$ and is twice the optimal size when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 2$. For case (b), each node is optimally sized when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 2$ and only half the optimal size when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 1$.

We note a few interesting features on this plot. Consider first when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 1$. The optimization effectiveness for the two nodes sized to half optimal is $\sim 50\%$ greater than that of a single node that is optimally sized. This advantage in distributing the nodes is because two buildings reject less waste heat with undersized nodes than one building with an optimally sized node, i.e., the waste heat is better utilized in moderate spring and autumn seasons by two buildings than one. Now consider when $\sum P_{n,\text{ICT}}/\max(q_{b,\text{req}}) = 2$, and there is a $\sim 0.09$ (100% gain) increase in $\bar{\varepsilon}_{\text{org}}$ by distributing to two nodes. Again, this is because the two buildings are rejecting less waste than the single building. For example, in the case of the single node, the node is so oversized that the majority of the waste heat it produces is rejected, even in the cold winter months. In fact, we observe that for a single node, $\bar{\varepsilon}_{\text{org}}$ essentially saturates at the optimal sizing and any oversizing results in continuously rejected waste heat. Thus by taking the additional node ICT power and distributing it to multiple buildings there is a net overall gain for the organization.

We can extend this distributed analysis to an organization consisting of many buildings, equally sized so that each building’s heating needs are again 10% of the total ICT load of the organization. As an example, we consider an organization consisting of 20 buildings and one data center, so that in theory the total ICT load could offset the heating needs of half of the buildings. Fig. 7 shows $\bar{\varepsilon}_{\text{org}}$ as a function of the number of buildings in the organization that have EOC nodes. If 10 buildings or less have nodes, each node can be optimally sized to the building, i.e., $P_{n,\text{ICT}} = \max(q_{b,\text{req}})$. However, if more than 10 buildings have nodes, each node will be undersized, $P_{n,\text{ICT}} < \max(q_{b,\text{req}})$, assuming the ICT load is distributed equally among the nodes. Fig. 7 illustrates that the trend observed in Fig. 6 holds as the number of buildings is scaled — the effectiveness increases when the organization’s ICT load is distributed widely, even if the nodes are undersized. Of note, however, is the slope of the trend. When there are fewer than 10 nodes and they can be optimally sized, the inclusion of each additional node increases $\bar{\varepsilon}_{\text{org}}$ by $\sim 0.03$, with the improvements asymptotically approaching a plateau. Thus, while it is beneficial to break up a data center into nodes distributed as heat sources across an organization, the return on investment is most efficient when the number of nodes is chosen so that they are each optimally sized.

4. Discussion

This modeling framework and analysis shows that there is potential benefit to an organization by implementing EOC across a series of buildings. However, implementation of EOC at a building or even community-scale will still involve its integration with technologies that enable the effective use and/or distribution of waste heat and/or the storage of waste heat. Recent implementations of these integrated technologies suggest some potential approaches. For example, a data center in Canada has taken the simple approach of retrofitting air cooling to provide space heat, where the data center node resides on a bottom floor and the occupied office area is on the floor above [22]. Space heat is then delivered to both the occupied office and an adjacent warehouse. Implementation required a second thermostat in the office area to control pneumatic baffles, complementing the thermostat for the traditional furnace-based heat. An alternative strategy used in Finland uses liquid-based cooling to cool a data center and meanwhile generate warm water for a district heating network, a system of water-heated pipes that then are used to provide space heat to domestic homes [23]. A similar concept has also been employed by the U.S. Department of Energy’s National Renewable Energy Laboratory, using warm water-based cooling and then piping the heated water for use as space heat in offices and research laboratories [24].

While these approaches take advantage of traditional HVAC and plumbing to reutilize the waste heat, intelligent design can also take advantage of energy storage in the building. Though ideally computational load would be redirected to a given EOC node to offset a building’s current need for waste heat, our analysis shows that in many instances (and likely in many practical applications),
some of the waste heat will be rejected. However, if the waste heat can be stored and reused later, this could also improve overall building and organization efficiency. Different energy storage strategies exist for buildings, often taking advantage of latent heat properties such as incorporating phase change materials in the structure [25] or high heat capacity reservoirs (e.g., water tanks, aquifers) [26]. For example, the U.S. Department of Energy Research Support Facility currently uses a labyrinth of concrete (increasing the thermal mass) as a thermal storage medium for waste heat produced by a data center in the facility’s basement [27], and similar concepts using latent or sensible heat properties could be employed for EOC nodes integrated into buildings. We note, however, that while all these examples highlight potential integration approaches, they would have to be adapted for the scale of the EOC nodes and buildings under consideration for a given organization.

5. Conclusions

This work has presented an approach to increasing energy efficiency by using distributed waste heat reutilization, where the waste heat is provided by the computational hardware that underpins our society. By distributing ICT equipment into building-integrated nodes rather than a centralized data center, waste heat can be harvested locally to offset a building’s energy needs. Further, integrating smaller data center nodes with buildings also allows for free cooling. The combination of these features leads to a net benefit, reducing the overall energy consumption of an organization. In this work, a framework for evaluating the benefit of reconfiguring an organization to utilize EOC-based distributed waste heat reutilization has been presented, and analysis on simple, model systems shows that this is an effective strategy for increasing the energy efficiency of an organization. Further, the concept of distributed waste heat reutilization can be expanded to explore other applications of heat sources and sinks, such as industrial processes across a manufacturing campus as distributed heat sources. Future work will explore optimizing this strategy by realistically accounting for integration inefficiencies as well as other sources of energy gain and loss, such as embodied energy.

Acknowledgements

The authors acknowledge the donation of over 100 servers by the eBay Corporation and the partnership of the City of South Bend and South Bend Botanical Society for the prototype EOC node. J.Z.W. acknowledges support from the National Science Foundation Research Experiences for Undergraduates.

Appendix.

A.1. Building effectiveness metric

The building power consumption effectiveness metric $\varepsilon_{\text{bldg}}$ assesses how much of the building’s heating power consumption is reduced by the presence of an EOC node or

$$\varepsilon_{\text{bldg}} = \frac{\text{change in building power consumption with EOC}}{\text{building power consumption without EOC}}.$$  (A.1)

The building power consumption without EOC is simply the heating power consumption of the building’s native heating system. If the building requires a certain amount of heat (space or water) to operate properly ($q_{\text{req}}$), this is simply

$$P_b = \frac{q_{\text{req}}}{\varepsilon_{\text{bldg}}}.$$  (A.3a)

The parameter $\varepsilon_{\text{bldg}}$ generically represents the overall thermal efficiency of the building’s native heating system (including thermal losses due to the envelope of the building). This reflects the fact that the building consumes more power for heat than is actually required to keep the building at a target temperature. In the cases considered in the main manuscript, the building efficiency was always assumed to be $\eta_{\text{bldg}} = 0.8$, which was considered a reasonable value.

The building power consumption with EOC is simply the power consumption required to supplement the waste heat provided by the EOC node. This includes two components. If $q_w < q_{\text{req}}$, then the building needs to expend its own energy to fulfill the remaining heating requirement. Additionally, the waste heat might not come at a suitable temperature for direct use by the building; for instance, the temperature of the waste air heat will typically be $\sim 90$–$100$ °F and the building will need to expend energy to not only enhance the heat so that it can be utilized ($P_{\text{utilize}}$) but also deliver it around the building via pumps or fans that would not normally be required ($P_{\text{deliver}}$). As such, the building’s power consumption with an EOC node is

$$P_{\text{b, EOC}} = \frac{q_{\text{req}} - q_w}{\eta_{\text{bldg}}} + P_{\text{utilize}} + P_{\text{deliver}}.$$  (A.3a)

Note that this assumes that the power to operate the ICT in the node ($P_{\text{b, ICT}}$) is “free” to the building as it would have been consumed regardless of whether the ICT was in the EOC node or the centralized data center.

The waste heat provided by the node may not necessarily be identical to the heat produced by the ICT equipment; i.e., $q_w \neq P_{\text{b, ICT}}$. The reason for this is that the heat transfer between the node and building (whether the node is adjacent to the building or embedded in the building) will inherently have losses. These different loss mechanisms are shown conceptually in Fig. A.1. We therefore assume that some percentage of the waste heat is lost $\eta_{\text{loss}}$ such that the waste heat delivered to the building is $q_w = P_{\text{b, ICT}}(1 - \eta_{\text{loss}})$. For these studies, we estimated this waste heat loss to be $\eta_{\text{loss}} = 0.19$ corresponding to a 10% loss from the node itself and a 10% loss during transportation to the building. If the node is directly integrated into the building, rather than adjacent to the node, then this loss factor will be different. Finally, we also assumed that $P_{\text{utilize}}$ and $P_{\text{deliver}}$ were both 10% of the ICT load in the node.

The building’s power consumption effectiveness can thus be defined as

$$\varepsilon_{\text{bldg}} = \frac{P_b - P_{\text{b, EOC}}}{P_b} = 1 - \frac{(q_{\text{req}} - q_w)/\eta_{\text{bldg}} + P_{\text{utilize}} + P_{\text{deliver}}}{q_{\text{req}}/\eta_{\text{bldg}}};$$  (A.3a)

or including the heat transfer loss

$$\varepsilon_{\text{bldg}} = 1 - \frac{(q_{\text{req}} - P_{\text{b, ICT}}(1 - \eta_{\text{loss}}))/\eta_{\text{bldg}} + P_{\text{utilize}} + P_{\text{deliver}}}{q_{\text{req}}/\eta_{\text{bldg}}};$$  (A.3a)
A.2. Organization effectiveness metric

The organization power consumption effectiveness metric $e_{org}$ assesses how much of the organization’s power consumption is reduced by extracting some servers from a data center to be placed in an EOC node or

$$e_{org} = \frac{\text{(change in organization power consumption – $P_{org, ICT}$)} \text{ with EOC}}{\text{(organization power consumption – $P_{org, ICT}$)} \text{ without EOC}}.$$

The total ICT power consumption of the organization ($P_{org, ICT}$), which is an inherent cost that is assumed to always be present,

$$e_{org} = 1 - \frac{\sum (q_{b, req} - q_w)/\eta_{bldg} + P_{w, utilize} + P_{w, deliver} + P_{dc, ICT, EOC}(PUE_{dc} - 1) + P_{n, ICT}(PUE_{dc} - 1)}{\sum (q_{b, req} - q_w)/\eta_{bldg} + P_{dc, ICT}(PUE_{dc} - 1)}$$

is subtracted from these quantities so that its magnitude does not arbitrarily impact $e_{org}$.

The organization’s power consumption without EOC is simply the summation of the power consumption of the buildings in the organization and centralized data center or $P_{org} = \sum P_b + P_{dc}$. The data center energy consumption $P_{dc}$ is determined based on the power usage effectiveness PUE, which is defined as the ratio of the total power to operate the data center to the ICT power used by the data center

$$PUE_{dc} = \frac{P_{dc}}{P_{dc, ICT}} = \frac{P_{dc, ICT} + P_{dc, facility}}{P_{dc, ICT}}$$

Here, $P_{dc, facility}$ is defined as the facility power required to operate the data center, such as the air conditioning systems. Based on Eqs. (A.1) and (A.4), the total organization power is thus

$$P_{org} = \sum P_b + P_{dc} = \sum q_{b, req}/\eta_{bldg} + PUE_{dc} - P_{dc, ICT}.$$  (A.5)

The organization power consumption with EOC is simply the power consumption of the buildings, nodes, and data center or $P_{org, EOC} = \sum q_{b, EOC} + P_{dc, ICT} + \sum P_b$. Using Eq. (A.2) for the building power consumption and an equivalent PUE definition for the nodes as in Eq. (A.4), this becomes

$$P_{org, EOC} = \sum \left( q_{b, req} - q_w/\eta_{bldg} + P_{w, utilize} + P_{w, deliver} \right) + PUE_{dc} - P_{dc, ICT}.$$  (A.6)

Eqs. (A.5) and (A.6) have the additional constraint on the ICT power of

$P_{dc, ICT} = P_{org, ICT}$ when EOC is not implemented;

$P_{dc, ICT} = P_{org, ICT} - P_{n, ICT}$ when EOC is implemented.

In some cases, the waste heat production is greater than the building need, and the building rejects some portion of the waste heat. However, the heat must still be removed from the node so that the ICT equipment is not damaged. If the free cooling is insufficient to mitigate this rejected waste heat, additional energy must be expended to cool the ICT equipment in the node increasing $P_{n, facility}$ as well as PUE$_n$. For this model, we assumed that PUE$_n = 1.3$ if no waste heat was rejected, but when waste heat was rejected we modified $P_{n, facility}$ such that the additional energy required to dispose of the rejected waste heat was 50% of the rejected heat load itself. When the node was appreciably oversized for a building, this could subsequently increase PUE$_n$ to as much as 1.6 for the conditions considered in our model studies. As noted in the main manuscript, for all cases it was assumed that PUE$_n = 1.8$.

The organization’s power consumption effectiveness can thus be defined, with appropriate algebraic manipulation, as

where the summations are over each of the building/nodes pairs in the entire organization. In terms of PUE, this can equivalently be written as

$$P_{dc, ICT, EOC} = P_{org, ICT}$$

noting that $P_{dc, ICT, EOC}$ in the numerator and $P_{dc, ICT}$ in the denominator are not equivalent.

References