Exploiting Thermal Energy Storage to Reduce Data Center Capital and Operating Expenses*

Wenli Zheng, Kai Ma, and Xiaorui Wang
The Ohio State University, Columbus, OH 43210
{*zheng.691, ma.495, wang.3596*}*@osu.edu*

Abstract

Power shaving has recently been proposed to dynamically shave the power peaks of a data center with energy storage devices (ESD), such that more servers can be safely hosted. In addition to the reduction of capital investment (cap-ex), power shaving also helps cut the electricity bills (op-ex) of a data center by reducing the high utility tariffs related to peak power. However, existing work on power shaving focuses exclusively on electrical ESDs (e.g., UPS batteries) to shave the server-side power demand. In this paper, we propose TE-Shave, a generalized power shaving framework that exploits both UPS batteries and a new knob, thermal energy storage (TES) tanks equipped in many data centers. Specifically, TE-Shave utilizes stored cold water or ice to manipulate the cooling power, which accounts for 30-40% of the total power cost of a data center. Our extensive evaluation with real-world workload traces shows that TE-Shave saves cap-ex and op-ex up to $2,668/day and $825/day, respectively, for a data center with 17,920 servers. Even for future data centers that are projected to have more efficient cooling and thus a smaller portion of cooling power, e.g., a quarter of today's level, TE-Shave still leads to 28% more savings than existing work that focuses only on the server-side power. TE-Shave is also coordinated with traditional TES solutions for further reduced op-ex.

1 Introduction

As cloud computing is gradually becoming a major paradigm in the IT industry, the increasing business demands for computing are driving data centers to rapidly increase their hosted servers. As a result, the power distribution and cooling systems in many of today's data centers have already approached their peak capacities. Since the upgrades in data center power systems can be extremely expensive (e.g., ranging in hundreds of millions of dollars) and often lag behind the required increases of hosted servers to support new business, data centers are seeking new ways to operate the existing power facilities as close as possible to their maximum capacity, such that the high capital expenses (cap-ex) of constructing new power facilities can be delayed. In the meantime, with the rapid growth of hosted servers, the operating expenses (op-ex) have also become a serious concern for data centers. For example, the energy bills of data centers in the US are estimated to be approximately $7.4 billion in 2011 [1]. Therefore, it is important for data centers to find feasible ways to better amortize the non-recurring cap-ex and reduce the recurring op-ex.

Power over-subscription has recently been identified as an important methodology for data centers to gain a high return on their investment in power facilities. The key advantage of over-subscription is placing more servers on the power infrastructure than it can support if all the servers maximize their power consumption at the same time. Since it has been shown in many studies [2, 3, 4, 5] that servers rarely peak simultaneously, over-subscription allows much more servers to be hosted than traditional provisioning that relies on the server nameplate power values, without the need of upgrading the power infrastructure. While the power peaks of the power distribution units (PDUs) in a data center can be shaved by balancing the workload across the PDUs or managing the power delivery topology between the PDUs and servers as in [6], in order to implement power over-subscription at the data center level for safely increased utilization of the power infrastructure, data centers must ensure that the total power consumption never exceeds its limitation. The power infrastructure is best utilized for a minimized cap-ex, if the power demand of a data center can be shaped approximately as a constant to eliminate any power peaks that may cause a power overload. Dynamic reduction of power peaks can also help reduce the op-ex because many utility companies charge a high tariff that is directly related to the peak power demand of a data center [7].

Existing approaches to peak power reduction primarily rely on device throttling [8, 9, 10], e.g., DVFS (dynamic voltage and frequency scaling), or workload shaping by postponing delay-insensitive workloads [11]. However, either of them may cause undesired degradation of system performance. Recently, *power shaving* has been proposed to dynamically shave the power peaks of a data center with energy storage devices (ESD). During a power peak, the extra energy can be drawn from ESDs, and during a power valley, the remaining power budget can be used to recharge ESDs, such

*This work was supported, in part, by NSF under grants CCF-1143605 and CNS-1143607 (CAREER Award) and by ONR under grant N00014-09-1-0750.

that the power drawn from the grid can be maintained approximately as a constant. However, existing work on power shaving [12, 7, 13] focuses exclusively on electrical ESDs (e.g., uninterruptible power supply (UPS) batteries) to shave the server-side power demand. Unfortunately, batteries have several major limitations. First, batteries are well known to be environmentally unfriendly and the recycling costs can be high. Second, batteries often have limited energy capacities for power shaving, because their original purpose is to temporarily handle power outages for only 5 to 10 minutes, before a diesel generator starts up. Third, power shaving needs to frequently discharge/recharge batteries, which hurts the battery lifetime and availability. Finally, the usable capacity of a battery decreases exponentially with the increase of discharge current based on Peukert's effect. Therefore, it is important to explore other energy storage methodologies for data center power shaving.

In this paper, we propose TE-Shave, a generalized framework that exploits both UPS batteries and a new knob, thermal energy storage (TES) tanks, for power shaving without performance degradation. TES tanks have been equipped in many data centers [14], and previously used to make cold water or ice at nighttime (when the power price is low) and provide additional cooling at daytime (when the power price is high) for reduced op-ex [14]. The key novelty of TE-Shave is that it discharges the TES when the power load (instead of price) is high. As a result, TE-Shave can effectively reduce both cap-ex and op-ex. Specifically, during a power peak, in addition to using UPS batteries, TE-Shave can adaptively throttle the flow rate of the chillers to reduce the cooling power. To meet the cooling requirement, TE-Shave switches to the TES for stored cold water or ice to supplement the chillers and exchange heat with the warm air returned to the computer room air conditioning (CRAC) systems. Similarly, during a power valley, TE-Shave increases the cooling power to charge the TES by making cold water or ice. By manipulating not only the server-side power but also the cooling power, TE-Shave achieves a much higher power shaving capacity than existing solutions that solely rely on electrical ESDs. Our extensive results show that TE-Shave allows a data center with 17,920 servers to save cap-ex and op-ex up to $2,668/day and $825/day, respectively. Even for future data centers projected to have more efficient cooling and thus a smaller portion of cooling power, e.g., just a quarter of today's level, TE-Shave still leads to 28% more savings than existing work that focuses only on the server-side power. Specifically, our major contributions are as follows:

- We propose to investigate a new knob, thermal energy storage (TES) tanks, for data center power shaving. We discuss the characteristics of different kinds of TES tanks and model their impacts on data center cooling.
- We design TE-Shave, a generalized framework that features different strategies to shave both cooling-side and server-side power based on the different characteristics of TES and UPS. We evaluate TE-Shave with real-world workload traces and show that it saves more cap-ex and op-ex than utilizing either of the two ESDs.
- We coordinate TE-Shave with the traditional TES solution that discharges TES tanks at daytime when the electricity price is high and recharges them at nighttime when the electricity is cheap. As a result, the data center op-ex is further reduced.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents background information about UPS batteries and TES tanks. We discuss the design of TE-Shave in Section 4, and describe the evaluation methodology in Section 5. Section 6 presents the experiment results. Section 7 is the conclusion.

2 Related Work

It is demonstrated that provisioning the servers according to the nameplate power leads to a low utilization of the data center power facilities. Fan et al. [4] have studied the potential of safe over-subscription with real Google traces. Existing solutions on power capping (e.g., [8, 9, 10]) have presented various schemes to control the data center power by DVFS to support safe over-subscription. However, those solutions may result in degradation of computing performance because they reduce the CPU frequency during power peaks when the workloads need power the most.

Recently researchers have begun to exploit the ESDs already installed in data centers for emergency use, to develop over-subscription solutions. Govindan et al. [7, 12] have presented power shaving algorithms that utilize the UPS batteries to minimize the op-ex and cap-ex of data centers. Kontorinis et al. [13] have explored the distributed UPS topology adopted by Google and developed a coordination scheme for power shaving at the cluster or PDU level. Wang et al. [15] have considered some other electrical ESDs besides batteries, such as supercapacitors (expensive), fly wheels (fast self-discharging) and compressed air energy storage (energy inefficient). All these methods can successfully cut down the costs without performance degradation, but the efficiency is still limited by the attributes of those electrical ESDs.

As an emerging technology, TES has been equipped at I/O Data Centers' Phoenix ONE data center, Digital Realty Trust's data warehouse, the National Petascale Computing Facility, and The National Oceanographic and Atmospheric Administration [14]. The TES system of Intel IT at a large regional hub data center can keep working for several hours during an outage [16]. Some data centers and other buildings [17, 18] have already been using (or suggested to use) TES tanks to store cold water or ice created at nighttime (when the power price is low) and then provide additional cooling at daytime (when the power price is high) for reduced op-ex. However, their schemes do not address power shaving or capping, and hence cannot support safe power over-subscription to save cap-ex.

3 Background

In this section we briefly introduce some background information about UPS batteries and TES tanks.

Table 1. Parameters for UPS batteries [13, 15].

Parameters	LA Battery	LFP Battery
Energy density	80 Wh/L	150 Wh/L
Energy efficiency	95%	95%
Typical price	$2/Ah	$5/Ah
Typical lifetime	4 years	10 years

3.1 UPS Batteries

Traditionally, UPS batteries are installed between the automatic transfer switch (ATS) and PDUs to sustain the power supply during an outage emergency. Such a centralized deployment leads to 10-15% energy loss on the AC-to-DC and DC-to-AC conversion (double-conversion) [7]. Google solves this problem by distributing the UPS batteries to the server level and installing them after the power supply units of servers, such that the double-conversion loss is avoided. However, the volumetric constraint can be a critical problem for the distributed UPS [15] if a large battery capacity is required to handle a long emergency. Several recent projects have already utilized UPS batteries (e.g., Lead-acid (LA) or Lithium Iron Phosphate (LFP), whose parameters are listed in Table 1) for power shaving [7, 12, 13].

3.2 TES Tanks

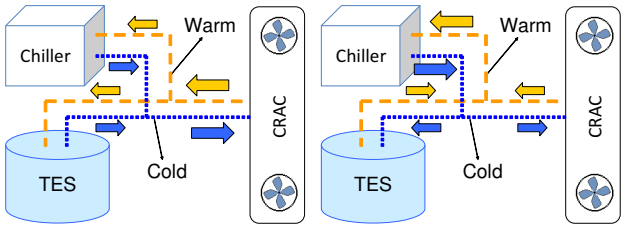
Besides the UPS battery, another type of ESD is also equipped in many data centers, which is the TES tank (e.g., Google has installed TES tanks in its $300M data center in Taiwan [14]). TES tanks are basically prepared to maintain the cooling when the chiller system fails; they usually store some cold materials, e.g., cold water or ice, which can absorb a great amount of heat dissipated by the IT equipment. The heat loss of TES tanks is 1% to 5% every day [19], while ice tanks are supposed to have a faster heat loss than water tanks because of their lower temperature.

Recently TES tanks are more actively utilized to store cheap electricity in night or unstable renewable energy. Compared with UPS batteries, TES tanks cost much less but have a longer lifetime (e.g., 20-30 years [20]), and also require less maintenance due to almost no need of replacement or special recycling technologies. When the TES is discharged, we can reduce the chiller power by raising the temperature set point of the supplied water [18, 21, 22, 23] and/or decreasing the flow rate of the water passing through the chiller [18, 23, 24]. If a cooling tower is built to cool the chiller, its power can also be adjusted along with the changing of chiller power. One limitation is that the chiller power cannot be changed too frequently, because 3-4 minutes are needed for the temperature in a data center to achieve the steady state after the configuration of the cooling system changes [25]. Therefore, TE-Shave utilizes TES tanks to shave long-last power peaks that cannot be effectively handled by only UPS batteries due to their limited capacities.

There are different types of TES tanks and some of them work as buffers for the coolant, such as the water tank. Figure 1 shows how a water tank works. When discharged, it provides part of the cold water to the CRAC, reducing the amount of cold water from the chiller and its power as well. When the data center power demand is low, the remaining power budget can be used to recharge the TES so it can be used repeatedly. Recharging the water tank requires the chiller to increase its power to produce more cold water, storing the extra part into the water tank after meeting the needs of the CRAC. The temperature difference between the cold water and the warm water can be 6 to 9 °C [26], and hence the energy density of a water tank is 25.2-37.8 kJ/kg, as the specific heat capacity of water is 4.2 kJ/(kg°C).

Table 2. Parameters for the TES tanks [19].

Parameters	Water Tank	Ice Tank
Energy density	25.2-37.8 kJ/kg	336 kJ/kg
Heat loss	1-5% per day	
Cooling device	chiller	refrigerator
COP	higher	lower
Price	$8.5-28.4/kWh	$14.2-19.9/kWh



(a) Discharge Mode (b) Recharge Mode

Figure 1. Water flows when discharging and recharging a water tank for power shaving.

Table 2 lists the parameters of water tanks as well as ice tanks. An ice tank plays a role as another chiller besides the primary chiller of the cooling system. Instead of directly buffering the cold coolant, an ice tank melts the stored ice to absorb heat from the coolant that passes through its cooling coil during discharging; during recharging, it activates its own refrigerator to make ice rather than demanding the primary chiller of more work. The cooling efficiency, which is represented by COP (coefficient of performance), of the ice tank refrigerator is 20% lower than that of the primary chiller during power peaks [19], leading to more energy consumption overall. However, the energy density of ice tank is about 336 kJ/kg, which is much larger than that of the water tank because melting ice absorbs much greater amount of heat. Therefore although an ice tank is usually more expensive than a water tank of the same size, the former can be cheaper than the latter given the same energy capacity.

4 Design of TE-Shave

In this section, we discuss how TE-Shave exploits the merits of both TES and UPS to address the limitations of each other, which significantly improves the effectiveness of power shaving. It considers the two major parts that constitute the power consumption of data centers, i.e., the server-

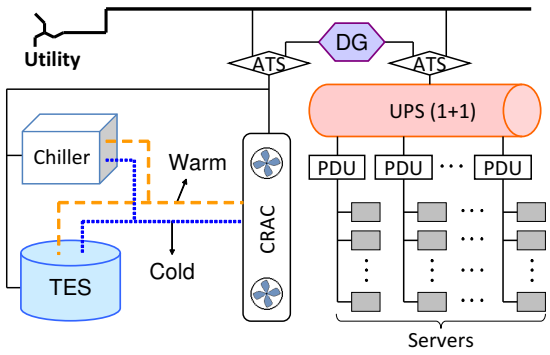


Figure 2. System diagram of the data center. UPS/TES shaves the server/cooling power.

side power (consumed by the servers and UPS) and the cooling power (consumed by the cooling system). By operating the ESDs, TE-Shave keeps the total power drawn from the grid below a power threshold (determined offline). As a result, TE-Shave can support safe power over-subscription at the data center level and flexibly shifts power between IT equipment and the cooling system. Figure 2 illustrates how the system works, where the TES can be either water or ice tanks. Though the UPS deployment in the figure is centralized, we model distributed UPS as well.

Some data centers recently reported that they reached a low power usage effectiveness (PUE), e.g., Google claims their annual average PUE is only 1.12, benefiting from their highly customized cooling systems that draw cold water directly from rivers or seas for free cooling and other special facilities [27]. However, due to various constraints such as location and climate, a PUE around 2.0 is still common to most existing commercial or governmental data centers. For example, according to the Uptime Institute [28], the average PUE for data centers worldwide is between 1.8 and 1.9 in 2012. Digital Realty Trust's survey in 2012 [29] also shows that the average reported PUEs in North America and Europe are 2.8 and 2.61, respectively. Therefore, despite that a much lower PUE has been shown to be possible in a few experimental data centers with highly customized cooling systems, it is unrealistic to assume that most existing data centers would significantly improve their PUEs in a few years, if they are not located at riverside or seaside.

4.1 On Cooling Side with TES Tanks

TE-Shave benefits from cooling power shaving because TES tanks cost little investment to provide large energy capacity, which decides the size of peak that can be shaved. Over-provisioning UPS batteries is an expensive approach to increase the energy capacity. Furthermore, power shaving needs frequent discharge/recharge of ESDs, which can shorten the battery lifetime and further raise the cost. In contrast, the equipment cost of TES tanks is much lower. For example, to have 1 kWh of electrical energy capacity to be used for 10 years, a battery whose maximum depth of discharge (DoD) is 60% costs at least $700 [13], while a TES tank costs at most $140 (water tank) / $72 (ice tank) when the average COP of the chiller is 5, or only $40 (water tank) / $20 (ice tank) when the average COP is 1.4 [19]. Note that TES tanks actually store thermal energy but we convert it to the equivalent electrical energy for this comparison:

$$ElectricalEnergy = ThermalEnergy/COP \quad (1)$$

In addition, TES tanks are also more friendly to the environment than the UPS batteries.

To decrease the cooling power to shave a peak, we throttle the water flow rate in the chillers to allow the reduction of chiller power. We assume the chiller power is continuously and precisely controllable, which may in fact depend on the specific devices and control knobs. Since the heat generated on the server side determines the amount of cold water needed by the CRAC, throttling the flow rate in the chiller may lead to lack of cold water. To supplement the cooling capacity, we discharge the thermal energy stored in the TES tanks (in the form of cold water or ice). Similar to UPS batteries, TES tanks must reserve some amount of the stored energy to satisfy the cooling requirement for 12 minutes when the servers run at full load but the chillers fail [16].

In order to decide the water flow rate of TES based on the chiller power and the heat generation, we need to model the discharge/recharge operations. The discharge models of water tank and ice tank can be established in the same way, though water tanks supply cold water for cooling while ice tanks melt stored ice to absorb heat from the coolant. The real chiller power P_{chi} is allowed to be reduced to:

$$P_{chi} = max(P^O_{chi} - (P^O_{sum} - P_T), P^{min}_{chi}) \quad (2)$$

where P^O_{chi} is the original chiller power if the TES is not activated; P^O_{sum} is the original total power demand (the sum of server-side power and cooling power); P_T is the power threshold, so $P^O_{sum} - P_T$ is the cooling power substituted by the cold water from the TES tanks; P^{min}_{chi} is the minimum power consumption of the chiller. We avoid to completely turn off the chiller (though it can save more power) because of the long time overhead to resume the chiller for operation. Equation (3) derives the thermal energy flow rate TF_{dis} (in terms of kJ/s or kW) discharged from the TES tanks:

$$TF_{dis} = H_{ser} + H_{other} - AbsorbHeat(P_{chi}) \quad (3)$$

where H_{ser} and H_{other} are the heat generation rates of the servers and the other devices (mainly the UPS in this paper), respectively, which are equal to their power consumptions; the function $AbsorbHeat$ returns the amount of heat absorbed by the cold water from the chiller when its power is P_{chi}, based on the specific cooling power model (in our evaluation, we use the cooling model described in Section 5.1). This equation means the rest of the heat dissipated by the servers and the other devices must be absorbed by the cold water from the TES tanks. We calculate the mass flow rate of water MF_{dis} discharged out of the TES tanks as:

$$MF_{dis} = TF_{dis}/(SpecHeat \times \Delta T) \quad (4)$$

where $SpecHeat$ is the specific heat capacity of water; ΔT is the temperature difference between warm and cold water.

The recharge models of water tanks and ice tanks are different. For water tanks, the chiller is operated to produce

more cold water than required by the CRAC during power valleys, and store the extra amount in the tanks. In Equation (5), P_{chi}^{max} is the maximum chiller power.

$$P_{chi} = min(P_{chi}^{O} + (P_T - P_{sum}^{O}), P_{chi}^{max}) \quad (5)$$

We compute the thermal energy flow rate recharged into the water tanks with Equation (6) and calculate the recharging mass flow rate with Equation (7):

$$TF_{rec} = AbsorbHeat(P_{chi}) - H_{ser} - H_{other} \quad (6)$$

$$MF_{rec} = TF_{rec}/(SpecHeat \times \Delta T) \quad (7)$$

When ice tanks are employed, they activate their own refrigerators to make ice during power valleys by utilizing the remaining power budget, so the chiller power is not affected. We assume a constant refrigerator power for each ice tank.

Both water tanks and ice tanks lose some stored energy because of the heat exchange with outside environment, but the heat loss rate is acceptable (1-5% per day [19]), and can be offset by the improved cooling efficiency during the power valleys when we recharge the TES tanks (the quantitative calculation is included in our experiments). The heat loss caused by moving water out of/into the TES tanks should be negligible since the extra pipes are usually short.

4.2 On Server Side with UPS Batteries

Although the efficiency of TES is promising, employing it as the only ESD for power shaving still has some shortcomings. Sometimes there can exist several transient power spikes during a short time interval (e.g., 3 minutes), but the operation period of the chillers and TES tanks might be relatively too long to shave such spikes efficiently, because it takes several minutes (e.g., 3-4 minutes [25]) for the cooling system to settle down after each adjustment. To deal with such transient spikes, TE-Shave resorts to UPS batteries due to their short reaction time (at most several milliseconds are needed for UPS to achieve the required power [15]). As a result, the reaction time of the integrated solution (TE-Shave) is also short since the UPS can shave the transient spikes. On the other hand, if a data center already equips more battery capacity than really needed during an outage emergency, TE-Shave will also make use of it to enlarge the energy capacity. We set lower bounds for the energy stored in UPS batteries because some amount must be reserved for power outage (enough to support the full load for 1-5 minutes [7, 13]). The reserved amount is also affected by the DoD limit to guarantee the battery lifetime. A larger DoD will lead to fewer discharge/recharge cycles (e.g., an LA battery with DoD = 20% can be operated for about 2,800 cycles, but only 500 cycles when DoD = 80% [13]). TE-Shave prevents battery overuse based on the DoDs to avoid hurting the lifetime.

TE-Shave can work with all kinds of UPS options. Due to space limitations, we discuss only datacenter-level centralized and server-level distributed UPS deployments (referred to as *centralized UPS* and *distributed UPS* in the rest of this paper). Although distributed UPS has better flexibility to handle frequently oscillating workloads and no double-conversion loss, every server is assigned a dedicated battery and hence the servers cannot share the battery energy with each other, while the energy stored in a centralized UPS can be shared among all servers. The volumetric constraint is another critical problem for the distributed UPS [15].

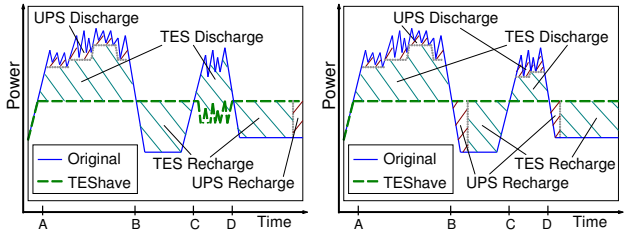
To model the usage of centralized UPS for power shaving, we consider a scenario of two UPS devices (1+1 redundancy). One of them is the main working UPS and the other is the backup, but both can be utilized for power shaving. We use configurable switches as [7] to allow the centralized UPS to share the power load out between the batteries and the power grid when a discharge is required, to avoid wasting the power budget. For the alternative distributed UPS deployment, each battery can be set on normal, discharge, or recharge mode individually as in [13], and just enough batteries will be discharged to shave a power peak. When the batteries get a chance to recharge, the recharge rate will be limited by an upper bound, since it typically takes several hours for an exhausted battery to be fully charged [30, 13].

4.3 Integration of Two Sides

TE-Shave performs power shaving on both the cooling and server sides, jointly managing TES and UPS to exploit the different advantages of the two types of ESDs for improved power shaving ability. Given a power demand curve and a power threshold, we calculate in every time slot how much energy should be discharged from the ESDs to cut down the power peaks if the power demand is higher than the threshold. Likewise, we analyze how much energy can be recharged into the ESDs by utilizing the remaining power budget if the power demand is lower than the threshold. Hence we have a long-term plan about the ESD usage after going through the demand curve. A violation of the threshold occurs if the stored energy is inadequate, or some UPS energy is still available but a new discharge/recharge cycle is prohibited by the strategy to protect the battery lifetime. We begin the offline analysis with setting the power threshold equal to the original peak power; after each run through the demand curve, we gradually lower the threshold until a certain run triggers the violation, or the power budget is always fully utilized. Such exhaustive search as done in [7, 13] guarantees that this algorithm can result in the lowest allowable threshold (i.e., shaving the most peak power).

Figure 3 describes two integration strategies, each of which makes a long-term plan about the ESD usage using the above algorithm, and derives a power threshold (shown as the dashed line labeled TE-Shave). When a peak comes (from time point A to B in Figure 3), the TES energy is discharged as the main supplement to the reduced grid power, and the transient spikes can be shaved by the UPS energy. The two strategies differ when a valley comes (from B to C): TES is recharged first in Figure 3(a) while UPS is recharged first in Figure 3(b). This difference can be influential when the valley is not large enough to fully recharge the ESDs. Later when another peak comes (from C to D), the strategy recharging TES first (Figure 3(a)) has no UPS energy to handle the transient spikes, and thus it has to discharge more

TES energy than needed for cooling to maintain the power threshold, leading to some waste of stored energy and power budget (power shaving with only TES has this problem as well). Thus we propose to recharge UPS first in power valleys since it avoids such waste.



(a) *Recharging TES first* (b) *Recharging UPS first*

Figure 3. Illustration of integration strategies.

4.4 Time-varying Electricity Price

The primary TE-Shave strategy introduced above discharges the ESDs as little as possible in the peak time and recharges them as soon as possible in the valley time. Such conservative ESD usage guarantees to shave the peak power to the lowest level, which means the most cap-ex savings. However, the traditional way of utilizing TES tanks in data centers is to discharge them at daytime (when the electricity price is high) and recharge them at nighttime (when the price is low), to save the op-ex. Therefore, to coordinate with the traditional TES management strategy, we now extend TE-Shave to exploit the time-varying electricity price for further reduced op-ex, probably at the cost of slightly higher cap-ex (than the primary TE-Shave strategy).

To be specific, there are two existing strategies that use TES tanks to exploit the price variation. The first one discharges the TES for 6 hours at daytime (12pm to 6 pm) and recharges the TES for 6 hours at nighttime (11pm to 5 am) [19]. The second strategy is similar, which discharges the TES whenever the electricity price is high and recharges the TES whenever the price is low. Both the two solutions focus only on the op-ex, while the cap-ex may not be saved because the peak power is not considered. In fact, the peak power can be even increased if the peak arrives at nighttime, or the electricity price happens to be low in the peak time.

We design TE-Shave-P, a variant of TE-Shave, for the time-varying price model. After the long-term plan of ESD usage and the power threshold are derived by TE-Shave, TE-Shave-P further adjusts them according to the electricity price. Depending on whether such adjustment is allowed to cause a higher power threshold and less cap-ex savings, TE-Shave-P can result in three solutions: *Cap-ex Preferred* forbids any change of the threshold; *Cap-ex Flexible* allows to raise the threshold 10% higher; *Total-ex Preferred* finds the best threshold that maximizes the sum of cap-ex savings and op-ex savings. Without violating the new threshold, the long-term plan is modified to discharge more energy from the TES in the time slots with higher prices. If an extra discharge leads to a lack of stored energy and thus causes a violation of threshold, the TES operations in the time slots with lower prices will be changed to recharge adequate energy. Such adjustment is repeated from the time slot with the highest price to the one with the lowest price.

5 Evaluation Methodology

In this section, we introduce our data center, ESDs and workloads used in the simulation, as well as the evaluation metrics. TE-Shave is compared with two baseline power shaving strategies: E-Shave and T-Shave. E-Shave utilizes only UPS batteries as the energy storage, similar to the *e-Buff* knob in [7] (centralized UPS) and the *ClustCtrl* policy in [13] (distributed UPS). T-Shave exploits only TES tanks and adjusts the chiller power to realize power shaving.

5.1 Simulated Data Center

We simulate a data center hosting 17,920 servers, which are divided into 16 areas. Each area is similar to the data center modeled in [31], i.e., hosting 1,120 HP Proliant DL360 G3 servers and 4 chilled-water CRAC units (each costs 10 kW to drive the fans). To simplify the air circulation modeling, we neglect the heat exchange between different areas and assume there is one chiller for each area, such that the cooling model established in [31] is applicable. Each chiller can consume at most 300 kW as the rated power, and at least 30 kW in the lowest power-consuming state.

We compute the power of each server P_{ser} according to the workload Uti (in terms of CPU utilization) based on a simplified linear model:

$$P_{ser} = P_{idle} + (P_{full} - P_{idle}) \times Uti; \quad (8)$$

where P_{idle} and P_{full} are the idle power (150W) and fully-loaded power (285W), respectively. Those parameters are directly taken from [31] to be consistent with the cooling model from the same reference. We assume the workload is allocated evenly across all the servers as in [7, 12], and hence every server consumes the same amount of power. While workload balancing is not the focus of this paper, the uniform allocation is assumed to support the cooling model *UniformWorkload* in [31], which needs to be modified if the data center adopts a different type of workload allocation.

The total cooling load is equal to the heat generation rate of the server side, shared by the chillers and the TES. When the TES is unused, we calculate the cooling power based on the server-side power and the cooling power model in [31], where the COP decreases with the increase of cooling load, and the impact of outside ambient is considered to be negligible. Since the CRAC fans, pumps and valves must keep working whether the coolant comes from the chillers or the TES, not all the cooling power can be shaved by using TES tanks. Typically, the chiller power (including the power of cooling tower) accounts for about 90% of the cooling power excluding the fan power [32]. This portion of power can be partially saved when the TES provides part of the coolant. In the simulation, we compute the instant cooling power and

chiller power in every time slot, based on the instant server-side power, the power threshold, and the usage of TES.

We have two options for the UPS deployment: centralized UPS and distributed UPS. The centralized UPS (1+1 redundancy) is equipped with 12V LA batteries, which can be fully charged in 3 hours [30] once completely drained (though not allowed), and its double-conversion loss is set to be 10% of the input power. The distributed UPS is equipped with 12V LFP batteries and each UPS supports a 2U server [13] (two servers are coupled to resemble a 2U server as in [31]). We use the maximum recharge power to minimize the recharge time (2 hours) as in [13]. The DoDs are limited at 40% for LA batteries and 60% for LFP batteries, which are the optimal values according to [13].

We also have two options for the TES: water tank and ice tank. For water tank, the temperature difference between cold and warm water is 6 °C and the trivial heat exchange between them is ignored as in [18]. For ice tank, the COP of its refrigerator is derived to be 1.12, due to its energy efficiency is 20% lower than that of the primary chiller in peak time [19]. We set the refrigerator power to be 31.6 W/gallon for an exhausted ice tank can be fully charged in 9 hours (usually 6-12 hours [33]). The equipment costs of water tank and ice tank are $28.4 and $19.9 per kWh (equivalent to $0.75 and $6.33 per gallon), respectively. The prices are selected as the maximum values in their price ranges (see Table 2) in order to compensate for the labor cost of installation.

To facilitate the design that UPS batteries can handle the transient spikes, whose durations are shorter than the operation period of the TES tanks, we configure a longer operation period for the TES tanks (15 minutes) and a shorter one for the UPS batteries (3 minutes) in the simulation.

5.2 Workload Traces

We evaluate TE-Shave and the baselines with the power demand curves created according to the power models and the workload traces. Figure 4 shows the four real-world workload traces used in our experiments (*Google*, *IBM*, *Wiki* and *HTTP*). Each trace records the data center utilization every 15 minutes, except for *HTTP* whose granularity is finer (3 minutes). *Google* [13] has 3-day data with typically one valley and one peak every day. *IBM* [9] records the utilization of 5,415 servers in a week, and we consolidate the workload and evenly distribute it onto the 1,120 servers in each area of our data center. *Wiki* (26-day) and *HTTP* (7-day) are both from Wikipedia [34], but the latter contains more transient spikes. Since the entire *IBM*, *Wiki* and *HTTP* traces are too long to be clearly displayed in the figures, we only present the three days with the highest utilization, though we test them for the whole lengths in the experiments.

5.3 Evaluation Metrics

While the total cost of a data center includes expenses in various aspects (e.g., building and employee salaries), power shaving only addresses the cost of the power infrastructure

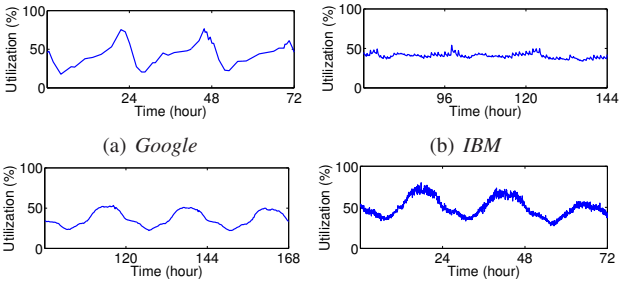


Figure 4. The four real-world workload traces.

(cap-ex) and the monthly electricity bills (op-ex) related to the servers, the cooling system and the energy storage, without affecting the other expenses. Hence our evaluation metrics are selected based on the cost model in [3] as follows:

$$OpSave = \Delta Energy * Price_E + \Delta Peak * Price_P; \quad (9)$$

$$CapSave = (\Delta Peak * Price_I - ESD)/10years; \quad (10)$$

where $\Delta Energy$ is the daily average difference between the original energy demand and the energy demand after shaving; $\Delta Peak$ is the difference between the original peak power and the shaved peak power; ESD is the extra cost of the increased ESD capacities, calculated based on their typical prices and lifetimes (Tables 1 and 2), while the TES tanks need not to be replaced because their lifetime is typically 20-30 years [20]; $Price_E$ is the energy price ($0.05/kWh [7, 13]); $Price_P$ is the peak power price ($12/kW/month [7]) charged based on the highest power peak in a month; $Price_I$ is the price of power provisioning, which is set to be $5/W here because over one third of the total cap-ex cost ($10/W to $25/W) is spent on power infrastructure [12]. The data center is assumed to be amortized over 10 years [13]. Particularly, when the time-varying electricity price model is applied, we set $Price_P$ to be zero assuming there is no extra charge for the peak power draw, and $Price_E$ varies along with time according to the price trace from NYISO [35], which records the price data in August, 2011.

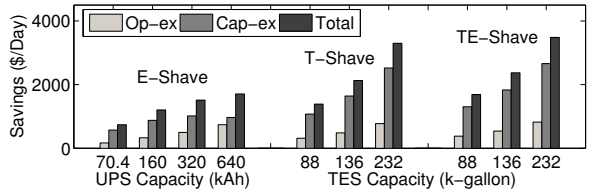
6 Results

In this section, we compare TE-Shave and the baselines based on the simulation results and discuss different ESD options. Due to the space limitation, some subsections present the results with only a subset of the four traces, as the unpresented results show similar trends. The unpresented results with ice tank are close to those with water tank when the ice tank is only 1/12 as large as the water tank.

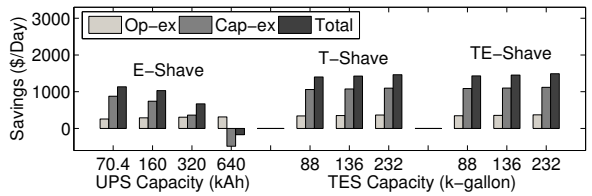
6.1 Integration of TES and UPS

We first evaluate whether TE-Shave can realize more savings than the baselines E-Shave and T-Shave. For E-Shave and TE-Shave, the default battery capacity of each centralized UPS is 70.4 kAh (not counted in the extra battery cost). It can support the fully utilized data center for 10 minutes. In order to reduce more peak power, E-Shave may need to over-provision the battery capacity up to 640 kAh (640 kAh

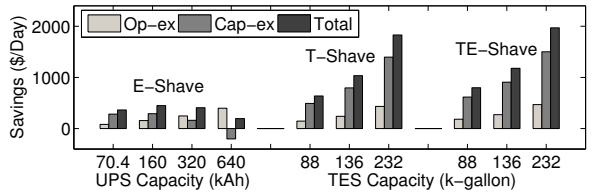
can sustain the full load for 128 minutes, so we consider it as a large capacity). The default water tank is 40 k-gallon (reserved for emergency), but T-Shave and TE-Shave need larger TES capacity such that the extra amount can be utilized for power shaving. The heat loss rate of water tank is set to be 3%/day, as the common range of TES heat loss is 1-5%/day [19]. Figure 5 shows the op-ex and cap-ex savings realized by the three strategies with different ESD capacities. The average PUE is 1.79, 1.78, 1.72 or 1.81 with the workload trace *Google*, *IBM*, *Wiki* or *HTTP*, respectively.



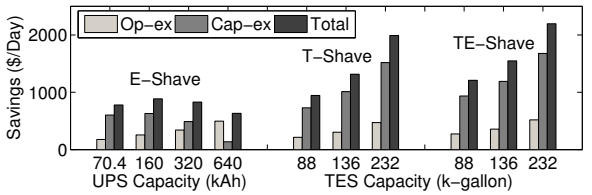
(a) *Google*



(b) *IBM*



(c) *Wiki*

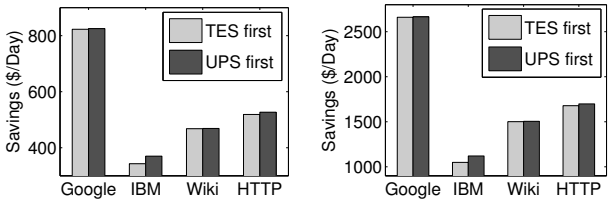


(d) *HTTP*

Figure 5. Daily average savings in 10 years. In default, UPS is 70.4 kAh and TES is 40 k-gallon. TE-Shave increases only TES size.

Greater ESD capacities enable saving more op-ex as expected because more peak power can be reduced, but the efficacy depends on the workload characteristics. For workloads with sharp peaks like *Google* (Figure 4(a)), the efficacy is greater; for workloads with flat peaks like *Wiki* (Figure 4(c)), the contribution of power shaving is much less. On the other hand, though the peak power decreases with the increase of ESD capacity, having more batteries can make a considerable side effect on the net cap-ex savings (subtracting the extra ESD cost from the savings on power provisioning). For E-Shave with *Wiki* and *HTTP*, the 320 kAh or 640 kAh UPS capacity leads to less cap-ex savings and total savings than 160 kAh, since the batteries are too expensive (even cost more than save, such as E-Shave with 640 kAh for *Wiki*). The results of TE-Shave clearly show the merits of TES tanks for their low price. For example, with *Google*, TE-Shave and E-Shave save 9.4-19.3% and 4.1-7.4% of the cap-ex, respectively, by shaving 9.5-19.6% and 4.1-18.3% of the peak power. T-Shave saves a little less than TE-Shave because TE-Shave also exploits the default UPS batteries and thus avoids wasting energy when shaving transient power spikes. Based on our simulation results, TE-Shave discharges 19.4% less energy from UPS batteries than E-Shave on average, and thus the battery lifetime of TE-Shave can be 2.45 times longer (according to the lifetime curve in [36]), given the default battery capacity. TE-Shave also leads to 39.2% smaller discharge current on average, and hence more efficiently utilizes the batteries according to Peukert's effect.

In Section 4.3, we analyze the advantage of recharging UPS before TES in the valley time. Here we use the simulation results to confirm this analysis. We run TE-Shave with a 232 k-gallon water tank to test the four workload traces. Figure 6 shows that recharging UPS first saves more in both op-ex and cap-ex. The most significant difference appears in the simulation with the *IBM* trace (Figure 4(b)), because the valleys of *IBM* are relatively shallow, i.e., not much power budget is available to recharge the ESDs, and thus the waste of power budget caused by recharging TES first becomes a more significant problem. For the other traces whose valleys are deeper, recharging UPS first makes more savings as well, though the differences are smaller.



(a) *Op-ex* (b) *Cap-ex*

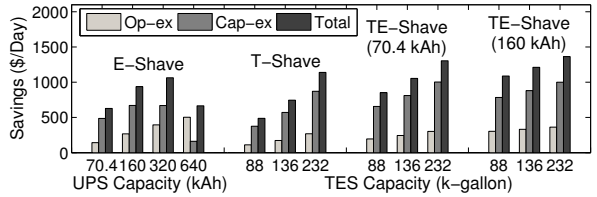
Figure 6. Impact of recharging preference.

6.2 Data Centers with Low PUE

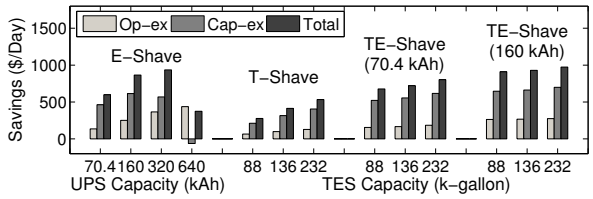
Some data centers are projected to have more efficient cooling and thus a smaller portion of cooling power in the future. Consequently, manipulating only the cooling-side power consumption with TES tanks may be insufficient for power shaving of the entire data center. We simulate such changes by modifying the chiller power to be 25% or 10% of the original value, meaning the energy efficiency of the chiller is 4 or 10 times higher, respectively. We use the *Google* trace here to study the influence of highly efficient cooling on TE-Shave and the baselines.

Figure 7 shows the op-ex and cap-ex savings. For 25% chiller power (average PUE is 1.31), a small TES capacity is unable to support T-Shave to achieve as much savings as E-Shave, but T-Shave with a large TES capacity is still more profitable than E-Shave with any size of UPS capacity. For

10% chiller power (average PUE is 1.21), the performance of T-Shave is further limited, because the ratio of cooling power over the total power consumption is only 3-4% and operating the TES tanks cannot affect the power draw adequately. TE-Shave can outperform T-Shave in both cases by utilizing the UPS energy. When the chilling efficiency is 4 or 10 times higher, TE-Shave with 160 kAh UPS capacity and 232 k-gallon TES capacity can achieve 28% or 4% more savings, respectively, than E-Shave with 320 kAh UPS capacity. This comparison indicates that even for data centers with very low PUEs, integrating TES with UPS for power shaving is still more beneficial than investing only on the UPS batteries.



(a) 25% Chiller Power



(b) 10% Chiller Power

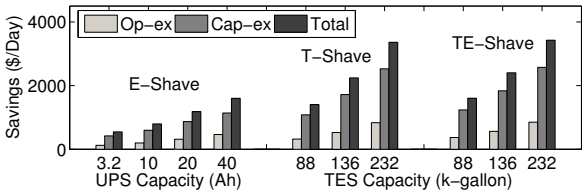
Figure 7. Savings with higher cooling efficiencies. The percentage of Chiller Power is the ratio of the chiller power with higher cooling efficiency over the original chiller power.

6.3 Different UPS Deployment

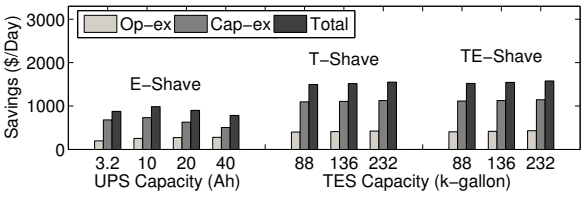
Figure 8 shows the simulation results with the *Google* and *IBM* traces when the data center employs distributed UPS with LFP batteries, and the battery capacity in each 2U server varies from 3.2 Ah (default capacity) to 40 Ah (maximum capacity that can be accommodated, according to [13]). The results are similar to those in Section 6.1 because we make the centralized UPS to appropriately allocate the power load between the batteries and the power grid to avoid wasting power budget, and thus the rest power demands supplied by the batteries are the same for the two UPS topologies, given the same total power demand and budget. Hence we think the distributed UPS deployment does not affect the effectiveness of TE-Shave and its advantage over the baselines. In addition, the battery capacity of distributed UPS is strictly constrained because of the volumetric limit, which may prevent E-Shave from saving more cost when larger battery capacity is needed.

6.4 Time-Varying Electricity Price

Although we use T-Shave as a baseline for comparison, the traditional TES solutions in data centers mainly shift the



(a) *Google*



(b) *IBM*

Figure 8. Savings with Distributed UPS.

cooling load from daytime to nighttime (*Day Night*) or from high-price periods to low-price periods (*Op-ex Preferred*), without addressing power shaving. Here we use the time-varying electricity price model to compare TE-Shave-P with the two traditional TES solutions. We assume that *Day Night* (which represents the current practice in data centers) discharges or recharges the TES at a uniform rate within each 6-hour period. For *Op-ex Preferred*, the discharge/recharge rate is proportional to the difference between the instant price and the average price.

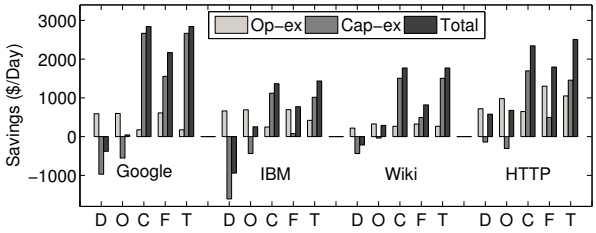


Figure 9. Savings for handling time-varying electricity price. Baselines: D (*Day Night*), O (*Op-ex Preferred*); TE-Shave-P: C (*Cap-ex Preferred*), F (*Cap-ex Flexible*), T (*Total-ex Preferred*).

Figure 9 shows the savings obtained by the five strategies when a 232 k-gallon water tank is equipped. Since *Day Night* and *Op-ex Preferred* focus only on reducing op-ex, they can make a power peak higher if the peak comes at night or when the price is low, because the TES is being recharged at those moments. Consequently, the data center needs to be provisioned based on a power demand higher than the original peak power, leading to the negative cap-ex savings. TE-Shave-P solves this problem by applying a power threshold to constrain the peak power draw. Among the three solutions of TE-Shave-P, *Cap-ex Preferred* saves the most cap-ex by keeping the threshold unchanged; *Cap-ex Flexible* saves the most op-ex by raising the threshold for 10%, which sacrifices some cap-ex savings; *Total-ex Preferred* gradually adjusts the threshold to achieve the most total savings, which can be 3 to 60 times higher than the total savings of *Op-ex Preferred*, while *Op-ex Preferred* outper-

forms *Day Night* with all the four traces. Based on our calculation, the average amount of discharged water of *Total-ex Preferred* TE-Shave-P (10.4 k-gallons/day) is less than that of *Day Night* (10.7 k-gallons/day). Therefore TE-Shave-P should not increase the cost of TES maintenance.

7 Conclusion

Existing work on power shaving focuses exclusively on electrical energy storage devices (e.g., UPS batteries) to shave the server-side power demand. In this paper, we have presented a novel power shaving framework, TE-Shave, which exploits both UPS and a new knob, thermal energy storage tanks equipped in many data centers. Specifically, TE-Shave utilizes stored cold water or ice to manipulate the cooling power, which accounts for 30-40% of the total power cost of a data center. Our extensive evaluation with real-world workload traces shows that TE-Shave saves capex and op-ex up to $2,668/day and $825/day, respectively, for a data center with 17,920 servers. Even for future data centers that are projected to have more efficient cooling and thus a smaller portion of cooling power, e.g., just a quarter of today's level, TE-Shave still leads to 28% more cost savings than existing work that focuses only on the server-side power. TE-Shave is coordinated with the traditional TES solutions to take advantage of the price differences in the time-varying electricity market for further reduced op-ex.

References

[1] United States Environmental Protection Agency, "Report to Congress on server and data center energy efficiency," 2007.

[2] P. Ranganathan, P. Leech, D. E. Irwin, and J. S. Chase, "Ensemble-level power management for dense blade servers," in *ISCA*, 2006.

[3] L. A. Barroso and U. Holzle, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Morgan and Claypool, 2009.

[4] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *ISCA*, 2007.

[5] Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," *Proceedings of SIGCOMM CCR*, 2008.

[6] S. Pelley, D. Meisner, P. Zandevakili, T. F. Wenisch, and J. Underwood, "Power routing: Dynamic power provisioning in the data center," in *ASPLOS*, 2010.

[7] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and limitations of tapping into stored energy for datacenters," in *ISCA*, 2011.

[8] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No power struggles: Coordinated multi-level power management for the data center," in *ASPLOS*, 2008.

[9] X. Wang, M. Chen, C. Lefurgy, and T. Keller, "Ship: Scalable hierarchical power control for large-scale data centers," in *PACT*, 2009.

[10] X. Wang and M. Chen, "Cluster-level feedback power control for performance optimization," in *HPCA*, 2008.

[11] H. Lim, A. Kansal, and J. Liu, "Power budgeting for virtualized data centers," in *USENIX ATC*, 2011.

[12] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar, "Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters," in *ASPLOS*, 2012.

[13] V. Kontorinisy, L. Zhangy, B. Aksanliy, J. Sampsony, H. Homayouny, E. Pettisz, M. Tullseny, and T. Rosing, "Managing distributed ups energy for effective power capping in data centers," in *ISCA*, 2012.

[14] R. Miller, "Google embraces thermal storage in taiwan," http://www.datacenterknowledge.com/archives/2012/04/03/google-embraces-thermal-storage-in-taiwan/.

[15] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy, "Energy storage in datacenters: What, where, and how much?" in *SIGMETRICS*, 2012.

[16] D. Garday and J. Housley, "Thermal storage system provides emergency data center cooling," Intel Information Technology on Computer Manufacturing Thermal Management, Tech. Rep., 2007.

[17] Y. Zhang, Y. Wang, and X. Wang, "Testore: Exploiting thermal and energy storage to cut the electricity bill for datacenter cooling," in *CNSM*, 2012.

[18] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves, "Model predictive control for the operation of building cooling systems," *IEEE Transactions on Control Systems Technology*, 2012.

[19] K. Roth, D. Westphalen, J. Dieckmann, S. Hamilton, and W. Goetzler, "Energy consumption characteristics of commercial building hvac systems volume iii: Energy savings potential," 2002.

[20] "Icebank - ice storage systems," http://www.calmac.com/benefits/general.pdf.

[21] M. Iyengar, R. Schmidt, and J. Caricari, "Reducing energy usage in data centers through control of room air conditioning units," in *IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, 2010.

[22] W. Huang, M. Allen-Ware, J. Carter, E. Elnozahy, H. Hamann, T. Keller, C. Lefurgy, J. Li, K. Rajamani, and J. Rubio, "Tapo: Thermal-aware power optimization techniques for servers and data centers," in *Green Computing Conference and Workshops*, 2011.

[23] Y. Chang and H. Tu, "An effective method for reducing power consumption - optimal chiller load distribution," in *Green Computing Conference and Workshops*, 2011.

[24] M. International, "Chiller plant design," 2002.

[25] F. Ahmad and T.Vijaykumar, "Joint optimization of idle and cooling power in data centers while maintaining response time," in *ASPLOS*, 2010.

[26] R. Beversdorf and J. Andrepont, "Thermal energy in raleigh, nc - modernizing infrastructure using an inventive district cooling solution financed via performance contract," Pepco Energy Services and The Cool Solutions Company, Tech. Rep., 2006.

[27] "Efficiency: How we do it," http://www.google.com/about/datacenters/efficiency/internal/.

[28] D. Chernicoff, "The results are in: The uptime institute 2012 data center survey," http://www.zdnet.com/blog/datacenter/.

[29] "Campos survey results 2012," http://knowledge.digitalrealtytrust.com/2012/03/campos-survey-results-2012/.

[30] "Smart-ups on-line, apc rt 1500va rack tower 120v," http://www.apc.com/products/resource/include/techspec_index.cfm?base_sku=SURTA1500RMXL2U&total_watts=200.

[31] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, "Making scheduling "cool": Temperature-aware workload placement in data centers," in *USENIX ATC*, 2005.

[32] M. Iyengar and R. Schmidt, "Energy consumption of information technology data centers," *Electronics Cooling*, December, 2002.

[33] "Calmac thermal energy storage - icebank," http://www.calmac.com/products/icebank.asp.

[34] G. Urdaneta, G. Pierre, and M. Steen, "Wikipedia workload analysis for decentralized hosting," *Elsevier Computer Networks*, 2009.

[35] "New york independent system operator," http://www.nyiso.com/.

[36] "Battery life (and death)," http://www.mpoweruk.com/life.htm.