Multi-tier energy buffering management for IDCs with heterogeneous energy storage devices

Zahra Abbasi, Madhurima Pore, Ayan Banerjee, and Sandeep K. S. Gupta
Impact Lab, School of Computing, Informatics and Decision Systems Engineering, ASU, Tempe, Arizona
http://http://impact.asu.edu/
{zahra.abbasi,madhurima.pore,ayan.banerjee,sandeep.gupta}@asu.edu

***Abstract*—Energy buffering, has been proposed to store renewable energy and low cost electricity in Energy Storage Devices (ESDs) and use it judiciously to reduce electricity bill in Internet data centers. Recent research have considered long term variation in electricity price, renewable power and workload and have shown the efficiency of energy buffering in reducing electricity bill. However, these aspects of data centers exhibit both long and short term variation. Further, there is inherent heterogeneity in ESD physical characteristics (e.g., charging and discharging rates). We hypothesize that a multi-tier energy buffering management can leverage the heterogeneity in ESD characteristics and better optimize utilization of renewable energy and low-cost power in presence of both short and long term variabilities in a data center. This paper proposes an analytical study of multi-tier workload and energy buffering management technique that frames each tier as an optimization problem and solves them in an online and proactive way using Receding Horizon Control (RHC). Our study shows that multi-tier energy buffering management increases the utilization of the renewables by upto two times compared to one-tier management.**

***Keywords*—Data center, renewable energy, energy storage.**

I. Introduction

Usage of renewable energy is increasingly becoming an important factor for Internet data centers (IDCs) to ensure their long-term environmentally responsible operation. Further, there is an increasing push to reduce the electricity cost of data centers. To achieve these goals data centers providers, mainly small and modern data centers, have started partially/completely powering their data centers by the on-site renewable energy, primarily using solar and wind energy [1], [2]. Further, researchers have suggested energy buffering using energy storage devices (ESD) such as batteries and ultra-capacitors [3]–[9]. Current data centers mainly use ESDs as power backup [4], [7]. These ESDs can also be used to store excess renewable energy and low-cost electricity during workload valleys, which can be used to meet high power demands [4]–[8]. However, energy buffering management should be aware of: (a) the intermittent nature of available energy from renewables, (b) variability in workload, (c) variability in the electricity pricing, and (d) heterogeneity in ESDs in terms of energy storage efficiency, discharge rate and storage capacity. In the light of these variabilities, this paper hypothesizes that an integrated multi-tier management of renewables, ESDs, and data center workload will reduce the electricity cost and increase the renewable energy utilization.

Multi-tier energy buffering and workload management is expected to reap benefits from the inherent variabilities of several aspects in the data centers. These aspects typically include (a) long and short term variabilities in available energy from renewables, and workload, (b) temporal variation in electricity price [10], and (c) heterogeneity in ESDs, which differ in their lifetime, maximum charging/discharging rate, and energy storage efficiency. A fundamental problem in energy buffering management is matching the variable power demand with low-cost grid and intermittent green energy. Recently researchers have studied this problem in a single tier setting, leveraging mostly long term variabilities. The idea is to (i) adapt the power consumption to the input workload using power management technique (e.g., c-state, p-state in servers), and (ii) bank green energy in ESDs and utilize it when power demand or electricity price is high [3]–[6]. These approaches ignore the tradeoff between the cost efficiency of ESDs and close matching of power demand with low-cost grid and green energy supply. Close matching of power demand and supply needs frequent charging/discharging of ESDs. On the other hand maximum utilization of renewable necessitates high energy density ESDs. The technological limitations on high energy density (energy per unit of mass or volume) ESDs such as Lithium Ion impose a cap on the "frequency" of charge/discharge, restricting the close matching of the supply and demand. ESDs such as ultra capaitors that allow frequent charging/discharging, have low energy density. Such tradeoff necessitates heterogeneous deployment of ESDs in data centers.

Energy buffering management is typically performed in a time-stepped system where the time is discretized into intervals, namely epochs. The time interval is compatible with the charging/discharging rate of batteries (e.g., hours). The inputs to a single tier management, are statistical summary (i.e., average) of power demand and supply (i.e., workload, available renewable power and electricity pricing) over epochs. In other words, the high resolution detail of power demand and supply are hidden from the management. The management scheme, assumes those information are evenly spaced within epochs. Due to such smoothness assumption, the high resolution details that occur at a time scale less than the epochs are ignored in the decision making of the manager, resulting in non-optimal decisions on energy buffering. A multi-tier energy buffering management which operates at long and short intervals can explore the data in higher resolution and optimize use of renewable and energy cost further. In this paper we analytically study the benefit of a multi-tier compared to a single tier energy buffering management, considering the variability of the available renewable energy and the power demand across epochs. Given the previous results, which suggest a Gaussian distribution for the variability of the available wind energy within hours, we analytically calculate the expected renewable energy wastage due to a single tier energy buffering management. The results suggest that when the average available

This research has been funded in parts by NSF CNS grants #0834797, #0855277, and #1218505, and by Intel Corp.

renewable energy for an epoch is less than the average power demand, the single tier management wastes up to half of the available renewable energy depending on the power demand magnitude and variability.

Motivated by the above requirements, this paper proposes **a multi-tier coordination management of power demand and supply**. We assume a data center which deploys heterogeneous ESDs to enable the combination of short and long term energy buffering management. We formulate each tier as a cost optimization problem, which can be carried out to proactively schedule energy storage and power state of the servers. To this end, we design multi-tier workload and renewable energy prediction techniques which are input to the cost optimization processes. In specific, the contributions of the paper are:

- Theoretical study to illustrate the benefit of multi-tier over single tier energy buffering management in increasing the renewable energy utilization.
- Formulating two tier energy buffering and workload management (T1 and T2 for tier one and tier two, respectively) as linear programming problems, and solving them using Receding Horizon Control.
- Demonstrating the efficiency of the proposed scheme using realistic renewable energy and workload traces. Results show combined T1 and T2 almost doubles the renewable energy utilization compared to T1 only (Fig. 13 in §V-E).

In the rest of the paper we first present our system model, and formulate the multi-tier management schemes (§II). Then we theoretically analyze the proposed schemes in §III. §IV discusses the solution of the problems, and §V shows simulation analysis. §VI and §VII discusses the related work and conclusion, respectively.

II. System model

We envision an Internet data center that obtains its required power from mix of grid, solar, wind and energy storage devices (ESDs). Also, we assume an incoming workload that consists of short requests (or transactions). The workload is assumed to be processed by any server in the data center. Further, we assume the data center has N homogeneous servers and deploys multiple heterogeneous types of ESDs which are managed in a multi-tier management scheme. For simplicity, we assume two types of ESDs (high and low energy density ESDs), in general, however, our methodology can be extended to more than two levels. ESDs can supply their charging power from both grid and renewable power, which can be used to power every server when needed (in general ESDs can be deployed in server level, rack level or data center level [7], and our analysis is based on data center level deployment) .

We divide time into epochs of length τ and each epoch, in turn, consists of K slots of length τ'. For example, an epoch can correspond to few hours in a day and a slot can correspond to minutes. We develop T1 and T2, the tier one and the tier two workload and energy buffering managers, operating at epochs and slots respectively (see Fig. 1). Leveraging long-term variation of the available renewable energy, electricity cost and power demand, T1 decides on the energy buffering management of the high energy density batteries (e.g., lead acid) which can sustain large energy for a long duration. Similarly, Leveraging short term variation of the available renewable energy, electricity cost and power demand, T2 decides on the number of active servers as well as energy buffering management of ultra capacitors. The objective of the optimization is to minimize the electricity price by leveraging ESDs to efficiently utilize low cost electricity and renewable. In the following we first give a motivation example. Then we give the modeling of the power demand, supply and workload. Next, we formulate the tier one (T1) and tier two (T2) optimizations as linear programming problems. We summarize the notations in Table. I.

TABLE I. Symbols and definitions.

Sym.	Definition	Sym.	Definition
t	epoch index	p^{AC}	power draw from grid
T	T1 entire time horizon	p^{total}	total power cons.
W	T1 prediction window	r^{total}	average green power
k	slot index	r	renewable harvesting
K	# of slots in an epoch	p	power of each server
τ	epoch length (sec)	p_0	threshold peak power
τ'	slot length (sec)	b	variable for battery cost
p^{ESD}	charging/discharging	α	energy cost
$p^{max,discharge}$	max discharging rate	β	cost per peak power
$p^{max,charge}$	max charging rate	y	# of active servers
B^{size}	ESD capacity	μ	service rate
γ	charging/discharging cost	λ	workload arrival rate
η	ESD self-discharge rate	d^{ref}	reference delay

*a) **Motivating example**:* Fig. 1 shows three cases where the two tier management utilize more renewable power than one tier management (only T1 or only T2):

Case 1 (the need for T1): Fig. 1(a) shows a case where the excess renewable power should be stored for several epochs (long time) i.e., three epochs in the example. While ESDs such as ultra-capacitors cannot sustain energy for long time due to their high self-discharge, batteries can sustain energy for long time which can be managed in long term due to their limited charging rate.

Case 2 (the need for T2): Fig. 1(b) shows a case with given slot and epoch level variation of power demand and renewable power (each epoch is assumed to consist of three slots). Since the average renewable power over the epoch t and $t+1$ is less than that of the power demand, there is no excess renewable power in average such that T1 cannot charge battery from renewable power in these epochs. However, the available renewable power in the third slot of epoch t and in the first slot of epoch $t+1$ is higher than that of the power demand. T2 can utilize ultra capacitors to store the excess renewable power observed over these slots for using in epoch $t+1$, second slot. In other words, the short term variation of the power demand and green power that are neglected by T1, may be negatively correlated as depicted by the figure. This necessitates utilizing an energy storage and a short term management scheme to match between the two.

Case 3 (the need for T2): Fig. 1(c) shows an example where T1 charges battery from excess renewable in epoch t for using in epoch $t+1$. However due to limited charging rate of the battery, all of the excess renewable power available in epoch t, third slot cannot be harvested by the battery. A combined T1 and T2, however, can utilize the ultra capacitor to harvest the spikes of the renewable power in that slot for the future use.

From the example, it can be concluded that a combined T1 and T2, can increase the renewable utilization. Similarly, it

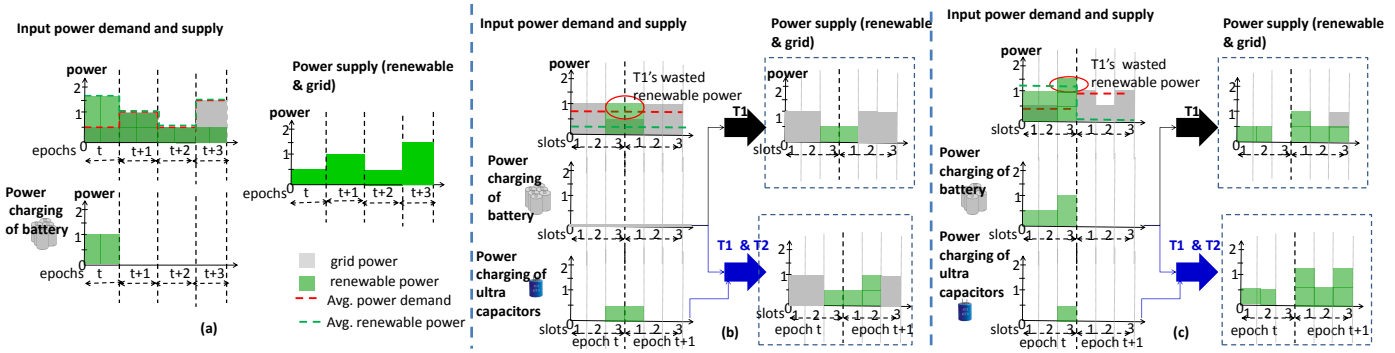


Fig. 1. Example: Renewable utilization for two tier versus one tier energy buffering management, (a) T1 to harvest renewable power long term variation by charging battery at epoch t and discharging at epoch $t+3$, (b) T2 to harvest short term variation of the renewable power by charging ultra-capacitors in third and first slots of epoch t, and $t+1$, respectively for using at epoch $t+1$, (c) T1 to harvest the average excess renewable power over epoch t using battery, and T2 to harvest the renewable power spike in epoch t, third slot using ultra-capacitor.

can be shown that a multi-tier management can decrease the cost incurred in using grid power.

b) ***Performance and workload modeling****:* The power provisioning of the servers (i.e., power state transition to active and inactive state) should be performed without compromising performance requirement of the requests. We consider service delay as the performance metric. To meet the performance requirement, the average delay of requests should not go above a *reference delay*, d^{ref}, where the value of the reference delay depends on the type of the application. The average delay of servers can be modeled using standard queuing theory results. Here without loss of generality we model each data center as a M/M/n (if a data center happens to be modeled by other queuing models such a G/G/n model, only the form of performance equations will vary yet the nature of the problem will remain the same). In M/M/n queuing model, given that all workload requests are delayed at the queue, the average service delay is expressed as: $\frac{1}{\mu} + \frac{1}{n\mu-\lambda}$ where μ denotes the service rate, and λ denotes the workload arrival rate to each data center. Other form of delay such as the 99th percentile of delay can be accounted instead of the mean delay, which require the approximation of the queue model or peak workload.

c) ***Power consumption modeling****:* For the sake of simplicity we assume servers that have only two states: active and inactive. Further, we assume servers consume zero power in the inactive state. We also assume each active server receives the same workload rate in steady-state. This is usually achieved through load balancing. So the total power consumption of the data center can be obtained by multiplying the total number of active servers (denoted by y) and power consumption value for each server (denoted by p). To keep the optimization framework linear, we set p as the power consumption of servers when they are utilized at their peak utilization.

d) ***ESD modeling****:* Energy storage are associated with several physical limitations. The cost of an energy storage depends on the total energy that is to be stored/discharged, and its charging/discharging rate. We denote the energy storage capacity as B^{size} measured in Joule. An energy storage has limits on the maximum discharge and charge rate denoted by $p^{max,discharge}$, and $p^{max,charge}$, respectively. Typically charging rate is much less than discharging rate, e.g., charging rate is 5-10 times less than discharging for lead-acid batteries. As suggested by the previous studies [4], [8], the existing UPSes in data centers can be utilized for energy management of the tier one. However, the energy level at the battery should always be sufficient to guarantee the desired availability, denoted by B_{min}. Further, the lifetime of an energy storage depends on the type of the device, the way in which it is used and several environmental factors such as temperature. Since a cost is incurred to replace the device, it is important to model its lifetime. We disregard the environmental factors and simplify the lifetime as follows. We consider that an energy storage device is associated with a cycle-life i.e., the number of charge /discharge cycles that can be accomplished during the lifetime of the device. Cycle-life is an estimate and depends upon an assumption of an average depth of discharge. We calculate the cost of a single duty cycling of an energy device, by dividing its cost (capital cost) by the number of life cycles[1]. For example the cost of a charging or discharging of an energy storage that costs \$500 and has 1000 life cycles is \$0.5. We denote such a cost as γ and incur it per charging/discharging cycle. We also consider that the ESDs are associated with self-discharge rate, denoted by η. To model energy storage, we denote the energy storage level at time t by B_t with initial value B_0 and the charge/discharge at time t by p^{ESD}, where positive or negative values mean charge or discharge, respectively. Our model and solutions account for two heterogeneous ESD in data center. To distinguish between the two we use subscripts $\{1,2\}$ for ESD parameters, e.g., η_1, and η_2 denote the self-discharge of the level one and two ESD, respectively.

e) ***Renewable Energy Modeling****:* We have used wind and solar energy as sources of renewable energy located on-site in a data center. We used the models given in [11] to generate solar power traces from irradiance and temperature and wind power traces from wind speed.

f) ***Electricity Cost modeling****:* Electricity prices usually vary over time and location. The variation is due to several factors including power generation, and more importantly supply-demand variation, and the market. The supply-demand matching is important, as any mismatch between the two can induce a high cost. The reason is that power providers may need to add or remove the generation plants or load both of which are costly. In order to minimize the high power draws by the consumers, electricity prices vary with the time of the day. Along with this, some utility providers also penalize the excess

[1]The total number of life cycles is calculated considering depth of discharge that can be achieved for the constraint of B_{min} and the maximum battery capacity B^{max}.

$$
\begin{array}{ll}
\textbf{minimize} & \sum_{t=1}^{T}(p_t^{AC}\tau\alpha_t + b_{1,t}\gamma_1) + \max_{1\le t\le T}(p_{i,t}^{AC} - p_0)^{+}\beta_i \\
\text{[Power constraint] , } \forall \text{ epoch } t: & r_t + p_t^{AC} = p_t^{total} + p_{1,t}^{ESD}, \\
\text{[Battery equations] , } \forall \text{ epoch } t: & B_{1,t+1} = min(\eta_1(B_{1,t} + p_{1,t}^{ESD}\tau), B_1^{max}), B_{1,0} = B_{min}, B_{1,t} \ge B_{min}, \\
\text{[Battery cost constraint]:} & -b_{1,t}p_1^{max,discharge} \le p_{1,t}^{ESD}, \\
 & 0 \le b_{i,t} \le 1, r_t \le r_t^{total}, \\
\text{[Max. charging/discharging]:} & p_{1,t}^{ESD} \le p_1^{max,charge}, -p_{1,t}^{ESD} \le p_1^{max,discharge}, \\
\text{[Service constraint] , } \forall \text{ epoch } t: & p_t^{total} = py_t, \text{and } y_t \le N, \\
\text{[Queuing stability constraint] , } \forall \text{ epoch } t: & y_t\mu_t > \lambda_t, \\
\text{[Performance constraint] , } \forall \text{ epoch } t: & d^{ref} \le \frac{1}{\mu} + \frac{1}{y_t\mu - \lambda_t}.
\end{array} \tag{1}
$$

power draw by imposing additional fee if the peak power draw exceeds the stipulated power in a certain time window e.g., 15 minutes [5]. To this end, we consider an electricity pricing model to account for both electricity cost per average energy consumption, α_t, and β per excess peak power draw from stipulated power (denoted by p_0).

A. T1 Problem formulation

The objective of T1 is to manage energy buffering of the level one battery to leverage the long term (epochs) variation of electricity pricing, workload and renewable in order to minimize the operational energy cost. To optimally decide on battery duty cycle, T1 should perform a cost optimization over a (possibly very long) time interval of interest T, as depicted in Eq. 1, to minimize electricity cost and energy storage cost.

As demonstrated in Eq. 1, the decision variables of T1 optimization cost framework are the average number of servers, y_t, power draw from grid, p_t^{AC}, renewable, r_t, the energy level and power draw rate for level one ESD, $B_{1,t}$, and $p_{1,t}^{ESD}$. All these variables are real, except, number of servers which we approximate it as a real variable. The integer approximation of variable y provides a small approximation ratio [12], and its effect in the objective function becomes negligible for large number of servers. Note that except, $B_{1,t}$ and $p_{1,t}^{ESD}$, all other decision variable are recalculated in the tier two optimization (T2). Three terms in the T1 objective are the cost for actual energy consumption, battery cost, and peak power cost due to exceeding from p_0, stipulated peak power. Note that we assume the peak power is calculated over T (time horizon of T1). Further, we set the renewable energy cost to zero to enable their maximum utilization. Furthermore, ESD life-cycle cost is formalized as $\gamma_1 b_{1,t}$, where γ_1 denotes the life-cycle cost per maximum discharging rate, and $b_{1,t}$ ($0 \le b_{1,t} \le 1$) is a variable to incur battery life-cycle cost proportion to the ratio of discharging rate to the maximum discharging rate. The T1 optimization is subject to the following constraints:

- Power constraint which asserts the power demand and supply balance.
- Battery constraints: (i) battery equations which asserts energy level of ESDs over time affected by its charging/discharging, p^{ESD}, and its self-discharge ratio, η, (ii) battery cost which specifies a set of equations to incur cost per discharging using a linear approximation equation, and (iii) maximum battery charging and discharging rate.
- Service constraint which specifies the total power demand and assert that the number of active servers should not exceed the total number of servers.
- Queuing stability constraint which asserts the M/M/n stability condition.
- Performance constraint which asserts the number of active servers should be chosen such that service delay is less than the workload' delay requirement (d^{ref}).

B. T2 problem formulation

The tier two optimization process accounts for the short term energy management. The optimization process is performed at small time intervals (e.g., minutes) which is called slots. The inputs are solutions to the tier one i.e., $B_{1,t}$, p_t^{ESD}, as well as the transient profile of workload and renewable. Workload and renewable transient profile is assumed to be known over the entire epoch. The optimization adjusts the power state of servers to the input workload and charging/discharging of the level two ESDs. We use k to denote a slot index, and K to denote the total number of slots in an epoch. The optimization formulation in tier two is similar to that of tier one, except that it uses some of the inputs from tier one.

Solution to the T2 problem as demonstrated in Eq. 2 specifies how many active servers are required for each slot i.e., $y_{k,t}$, and what portion of power is drawn from grid $p_{k,t}^{AC}$, renewable $r_{k,t}$, and the level two ESD $p_{2,k,t}^{ESD}$. Similar to T1, T2 cost objective accounts for electricity cost, ESD level 2 cost, and peak power cost. The peak power cost is included to enforce T2 to smooth the power consumption over every epoch. Similar to T1, T2 optimization is performed under constraints of energy storage, service and performance (see §II-B). Particularly, the power demand and supply balance of T2 considers the possible charging/discharging of the level one ESD which is decided by T1.

III. Theoretical study

This section motivates two-tier energy buffering management by providing analytical basis for its benefits under practical assumptions on the availability of the wind energy and data center power consumption. Consider a data center that utilizes on-site wind power and grid to power its servers. Further assume the grid power is offered with fixed price over time (i.e., $\alpha_t = \alpha, \forall t$, and $\beta = 0$). Furthermore, assume ESD at tier one has unlimited capacity (i.e., $B_1^{max} \simeq \infty$), is associated with infinite number of life-cycles (i.e., $\gamma_1 \simeq 0$) and has no self-discharge (i.e., $\eta = 1$). Let $\bar{p}_t$ denote the expected power demand, and $\bar{r}_t$ denote the expected available renewable power over epoch t. T1 decides on charging and discharging of battery based on $\bar{p}_t$ and $\bar{r}_t$ values to maximally utilize renewable. However, the available renewable power varies within epochs. Particularly the previous studies shows

minimize

$$\sum_{k=1}^{K}(p_{k,t}^{AC}\tau'\alpha_t + b_{2,k,t}\gamma_2) + \max_{1\le k\le T}(p_k^{AC} - p_0)^+\beta$$

[Power constraints], $\forall$ slot k : $r_{k,t} + p_{k,t}^{AC} = p_{k,t}^{total} + p_{1,t}^{ESD} + p_{2,k,t}^{ESD}$,

[Battery equations], $\forall$ slot k : $B_{2,k+1,t} = min(\eta_2(B_{2,k,t} + p_{2,k,t}^{ESD}\tau'), B_2^{max})$,

[Battery cost]: $-b_{2,k,t}p_2^{max,charge} \le p_{2,k,t}^{ESD}, r_{k,t} \le r_{k,t}^{total}$,

$0 \le b_{i,k,t} \le 1$, (2)

[Max. charging/discharging]: $p_{2,k,t}^{ESD} \le p_2^{max,discharge}, -p_{2,k,t}^{ESD} \le p_2^{max,discharge}$,

[Service constraints]: $p_{k,t}^{total} = py_{k,t}, and, y_{k,t} \le N$,

[Queuing stability constraints]: $y_{k,t}\mu_{k,t} \ge \lambda_{k,t}$,

[Performance constraints]: $d^{ref} \le \frac{1}{\mu} + \frac{1}{y_t\mu - \lambda_{k,t}}$.

that the wind generation can be modeled as a non-stationary Gaussian random process within epochs, i.e., the renewable power generation amount in kth slot of epoch t is given by [13]:$r_{k,t} \sim \mathbb{N}(\bar{r}_t, s_r^2)$, where $\bar{r}_t$ is the mean and s_r^2 is the variance. Power demand depends on the input workload as well as servers' idle and peak power. To quantify the variation of power demand within epochs we consider two cases: (i) power demand is almost constant within epochs which is true when number of active servers do not change within epochs and servers' idle power is very large i.e., $p_{k,t} = \bar{p}$, $\forall k = 1 \dots K$, and (ii) power demand varies as a Gaussian stochastic process during epochs (note that Web workload follows heavy tail stochastic process, the resulting power consumption of the workload, however, may or may not follow such a heavy-tail stochastic processes) i.e., $p_{k,t} \sim \mathbb{N}(\bar{p}_t, s_p^2)$. The two following lemma and corollary illustrate how combined T1 and T2 potentially increases the renewable utilization compared to T1 only. The lemmata are given for Optimal T1 where (i) the input data for the entire time horizon T is given, (ii) T1 is assumed to be optimally solved, and (iii) T is very large, e.g., a year.

Lemma 1: Given Gaussian distribution for the variation of the renewable power within epochs with mean $\bar{r}_t$ and variance s_r^2, and a constant power demand of $\bar{p}_t$, then Optimal T1, single tier energy buffering management for epochs of length τ, utilizing battery with characteristics of $B^{max} \simeq \infty$, $\eta = 1$, $\gamma = 0$ and $p^{max,charge}$, on average at least wastes w_t^{avg} of the available renewable energy during epoch t, where $w_t^{avg} = \min(\tau[\frac{s_r}{\sqrt{2\pi}}e^{\frac{-(\bar{p}_t-\bar{r}_t)^2}{2s_r^2}} + \frac{\bar{r}_t}{2}(1 - \text{erf}(\bar{p}_t - \bar{r}_t))], \tau[\frac{s_r}{\sqrt{2\pi}}e^{\frac{-(\bar{p}_t+p^{max,charge}-\bar{r}_t)^2}{2s_r^2}} + \frac{\bar{r}_t}{2}(1 - \text{erf}(\bar{p}_t + p^{max,charge} - \bar{r}_t))]$.

Proof: Due to given assumptions i.e., constant electricity cost, zero battery charging/discharging cycle cost and infinite battery capacity, and that T1 is expected to run for very large T (i.e., $T \to \infty$), where there is always power demand and lack of renewable power over time, we have that T1 stores all of the excess renewable energy over every epoch. We consider two cases.

Case one: assume $\bar{r}_t \le \bar{p}_t$, then there is no excess renewable energy (on average) to be stored at tier one battery. (Optimal) T1 decides on not charging the battery from the available renewable (this has been illustrated in Fig. 1). Let's define a random variable to denote the difference between the available renewable energy and the power demand for every slot within epochs as follows: $r_{t,k}^{diff} = (r_{t,k} - p_{t,k})$. Then given a T1 battery management, the expected available renewable energy wastage for a slot equals to $\mathbb{E}(r_{t,k}^{diff} > 0)$:

$$w_{k,t}^{avg,case1} = \mathbb{E}(r_{t,k}^{diff} \ge 0) = \mathbb{E}(r_{t,k} \ge \bar{p}_t)$$
$$= \int_{\bar{p}}^{+\infty} \frac{1}{s_r\sqrt{2\pi}}e^{-\frac{(x-\bar{r})^2}{2s_r^2}}\,x\,dx = \frac{s_r}{\sqrt{2\pi}}e^{\frac{-(\bar{p}_t-\bar{r}_t)^2}{2s_r^2}} + \frac{\bar{r}_t}{2}(1 - \text{erf}(\bar{p}_t - \bar{r}_t)) \quad (3)$$

Considering that there are K slots of length τ' in an epoch where $K\tau' = \tau$, and that the expected value of sum of independent random variables equals to the sum of the expected values for each variable, we have that $w_t^{avg,case1} = \tau w_{k,t}^{avg,case1}$.

Case two: assume $\bar{r}_t > \bar{p}_t$, then T1 decides on charging of the battery. However, due to limitation on the maximum charging rate of the battery, the renewable energy will be wasted if $r_{k,t} > p^{max,charge} + \bar{p}_t$. Thus the expected renewable energy wastage for every slot equals:

$$w_{k,t}^{avg,case2} = \mathbb{E}(r_{t,k}^{diff} > p^{max,charge} = r_{t,k} \ge \bar{p}_t + p^{max,charge})$$
$$= \int_{\bar{p}+p^{max,charge}}^{+\infty} \frac{1}{s_r\sqrt{2\pi}}e^{-\frac{(x-\bar{r})^2}{2s_r^2}}\,x\,dx = \tau[\frac{s_r}{\sqrt{2\pi}}e^{\frac{-(\bar{p}_t+p^{max,charge}-\bar{r}_t)^2}{2s_r^2}}]. \quad (4)$$

Considering that $w_t^{avg} = min(w_t^{avg,case1}, w_t^{avg,case2})$, the lemma follows. ■

To quantify the result of the above lemma suppose $\bar{r}_t = \bar{p}_t$, then the expected renewable energy wastage for epoch t, using only T1 battery management becomes: $w_t^{avg} = \tau(\frac{s_r}{\sqrt{2\pi}} + \frac{\bar{r}}{2})$, which means that at least half of the available renewable energy during epoch t is wasted.

Corollary 1: Given Gaussian distribution for the variation of both the available renewable power and the power demand within epochs with mean $\bar{r}_t$, $\bar{p}_t$, and variance s_r^2 and s_p^2, then T1, single tier energy buffering management management for epochs of length τ, utilizing battery with characteristics of $B^{max} \simeq \infty$,$\eta = 1$, $\gamma = 0$ and $p^{max,charge}$, on average wastes w_t^{avg} of the available renewable energy during epoch t, where $w_t^{avg} = min(\tau[\sqrt{\frac{s_r^2+s_p^2}{2\pi}}e^{\frac{-(\bar{p}_t-\bar{r}_t)^2}{2(s_r^2+s_p^2)}} + \frac{\bar{r}_t-\bar{p}_t}{2}(1 - \text{erf}(\bar{p}_t - \bar{r}_t))], \tau[\sqrt{\frac{s_r^2+s_p^2}{2\pi}}e^{\frac{-(p^{max,charge}-(\bar{r}_t-\bar{p}_t))^2}{2(s_r^2+s_p^2)}} + \frac{\bar{r}_t-\bar{p}_t}{2}(1 - \text{erf}(p^{max,charge} - (\bar{r}_t) - \bar{p}_t))]$.

The above corollary follows from Lemma 1, given that the difference of two Gaussian random variables with mean μ_1 and μ_2, and the variance σ_1^2 and σ_2^2 is a Gaussian variable with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Consequently, $r_{t,k}^{diff} \sim \mathbb{N}(\bar{r}_t - \bar{p}_t, \sqrt{s_p^2 + s_r^2})$.

To quantify the result of Corollary 1 suppose $\bar{r}_t = \bar{p}_t$, then the expected renewable energy wastage due to only T1 battery

management becomes: $w_t^{avg} = \tau \sqrt{\frac{s_r^2+s_p^2}{2\pi}}$. The above analysis is performed for wind energy and the assumption that the electricity price is constant within epochs. In reality, on-site renewable energy in data centers is dominated by both wind and solar energy, where solar energy also exhibits short term variation (see §V-B). Further electricity price may vary in intervals of less than epochs (e.g., 15 minutes [5]) which further motivates multi-tier energy buffering management. The above analytical results suggests that depending on the variability of the available renewable and power demand within epochs, a single tier energy buffering management (i.e., T1 only) cannot optimally harvest the renewable power. A combined T1 and T2, however, can explore the variation of the power demand and supply over and within epochs, resulting in less renewable energy wastage.

IV. Solutions of T1 and T2

Both T1 and T2 are developed as Linear Programming problems, where an LP solver can optimally solve them. However, the optimal solution can be achieved if all information (e.g., renewable power, workload, and electricity price) are known in advance. In reality it is almost impossible to get those information for a large time interval (at least a month, since the peak power is usually calculated monthly). For that we evaluate the efficiency of the schemes when using a prediction technique for workload and renewable and well known Rolling Horizon Control (RHC) technique. Consider a window of length $W \leq T$. RHC obtains T1 solutions at time t by solving the cost optimization of T1 over the window $(t, t+W)$, given the T1 solution at time $t-1$.

In the solution method, T1 on every epoch predicts the long term workload and renewable power variation, solves the optimization problem (Eq. 1) to determine the duty cycling of the level one ESD and reports it to T2. On every slot, T2 predicts the short term workload and renewable power over the window of length K, solves T2 cost optimization (Eq. 2) to determine the number of active servers, and the duty cycling of level two energy storage. Note that peak power is usually calculated over a long time periods (e.g., a month). Neither T1, nor T2, can smooth the peak power over a month (the prediction window of T1 is much less than a month), but they smooth the peak power over their window, which indirectly helps to smooth the peak power over a month.

V. Simulation Study

We simulate a data center with 500 homogeneous servers where each server consumes 400 W at its peak and can handle up-to 200 request per second. The data center is assumed to have Lead Acid (LA) battery and Ultra Capacitor (UC), in centralized way (all servers can access to both ESDs) and with characteristics shown in Table. II [5]. We set each decision epoch to an hour and half an hour, and a slot to a minute.

TABLE II. Specification of ESD parameters.

Parameters	**LA**	**UC**
Cost per discharge ($)	0.25	0.0115
Cycle life of one cell ($\times 10^3$ cycles)	1.2	1000
Discharge rate (KW)	5.4	42.5
Discharge-to-charge ratio	10	1
Depth-of-Discharge (%)	80	100
Self Discharge per day(%)	0.1	20

We use one month traces of renewable, and workload that are described in the following sections. We use the historical electricity prices of California for average power draw from the grid (i.e., α) which varies every 30 minutes. Further, we assume that $p_0 = 60\,kW$ a value that is 60% more than average power consumption of our simulated data center. We use a excess charge of \$12/kW for the peak power draw (i.e., β) of every hour within a window of one month. Also the prediction window of T1 (i.e., W) is set to 24 hours. The simulation environment is developed in MATLAB and GNU Linear Programming Kit (GLPK) is used to solve T1 (Eq. 1) and T2 (Eq. 2).

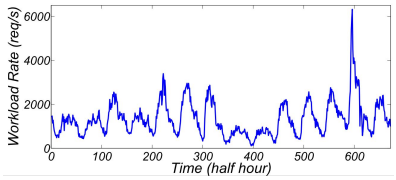
A. Workload prediction

We generate our workload from NASA workload Internet data center trace (July and August, 1995) [14] with sampled request rate (req/s) of every 1 minute and 30 minutes. We scale this workload to the capacity of our simulated data center. The workload intensity is such that on an average around 100 active servers are required. We predict the workload for the two tiers of management. The workload exhibits daily and weekly seasonality due to the usage patterns of clients that shows more activity during the peak hours of the day. Fig. 2 shows this pattern in an hourly scale. Hence, the workload with time interval for T1 can be consistently predicted using the seasonal prediction model, SARIMA [15], [16] with seasonality introduced based on the data of the previous day at same time as well as the change in the workload in few previous epochs. The initial realistic trace history is taken as the July workload trace. We then use it to predict data for the month of August. In T2, however, there are fine variations in the workload that cannot be captured in a seasonal model as depicted in Fig. 2, hence we use moving average prediction technique. The prediction data of T1 for the epoch length of 30 minutes shows a prediction error of 19.6% for forecast window of 1 and increases with the size of forecast window to up to 40.2%. For T2, with forecast window of 1, error is 27.2% and increases up to 48% with increase in the size of forecast window. For epochs of length of 1 hour we have prediction error of 21.8% that increases up to 29.67% for 24 hour forecast window.

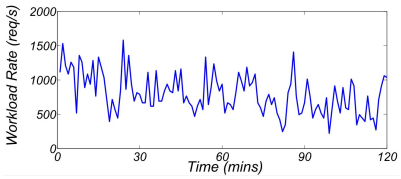
B. Renewable power traces and prediction

To capture the availability of wind and solar energy, we use traces of [17]. We use wind speed and the rated power to calculate the wind power, and Global Horizontal Irradiance (GHI) and the ambient temperature to calculate the solar power using models described in [11]. The traces are for intervals of one minutes and a duration of two month (July and August 2012). Fig. 3 and Fig. 5 demonstrate the intermittent availability of the wind and solar power sources over minutes. These observations agree with the assumption of the paper about short term variation of both wind and solar power. Similarly Fig. 3 and Fig. 5 shows every half-hour variation of wind and solar power during one day.

We vary the number of solar cells and wind turbines so at maximum rated power output of each we get two renewable power rate that supplies 10% and 60% of peak power draw of the data center. The daily pattern is observed for the solar energy with maximum energy observed in the afternoon. We characterize this pattern using Seasonal ARIMA technique to predict solar power for a day ahead to be used in T1.



(a)



(b)

Fig. 2. Workload trace from NASA 1995 trace [14] **(a)** Hourly variation of workload, and **(b)** Minute variation of workload.

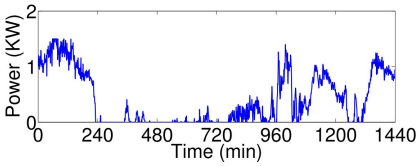


Fig. 3. Variation of wind power every minute on Aug 02, 2012.

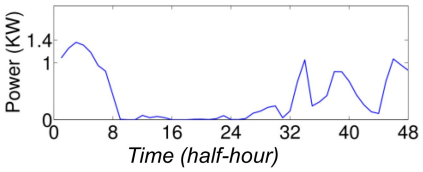


Fig. 4. Variation of average wind power over 30 minutes on Aug 02, 2012.

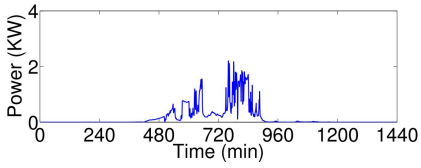


Fig. 5. Variation of solar power every minute on Aug 02, 2012.

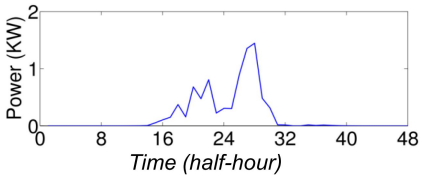


Fig. 6. Variation of average solar power over 30 minutes on Aug 02, 2012.

The variation in the wind energy is not seasonal hence we use exponential smoothening and moving average in order to predict it. Using Seasonal ARIMA [18], the Solar Power for half hour epoch (tier 1) and forecast window of one epoch shows prediction error of 17.56% which increases to 21.65% for the prediction window of a day (48 epochs). Similarly the wind power prediction for half hour epoch and forecast window of 1 shows error of 27.41% and increases to 34.76% for the prediction window of a day (48 epochs). We also use moving average technique to predict solar and wind power over an epoch to be used by T2. The prediction error in this case is higher than prediction error for T1.

C. Experiments performed

We perform experiments to evaluate the efficiency of T1 and T2 versus different capacity of ESDs, prediction error, electricity pricing models, epoch length and the available renewable power. We choose ESD capacity values that are comparable to the energy requirement of the simulated data center when working at its peak for a specific duration (e.g., ESD of size 1.2 MJ is enough for the data center to work for one second at its peak). In the experiments, combined T1 and T2 are performed when ESD capacity of both levels have non-zero values. Whereas, single tier energy buffering management, T1, is performed when ESD level one has non-zero capacity value, and ESD level two has zero capacity value. Finally, zero capacity of ESDs at both levels means that we do not perform energy buffering management at none of the tiers. Note that for the fair comparison of T1 and T2's cost saving, we always perform workload management at slots. In other words, the reported values of cost and power for all of the figures are the solutions of (Eq. 2), with the specified ESD capacity at each level. Our simulation environment is developed using MATLAB 2009. We use GNU Linear Programming Kit (GLPK) to solve T1 (Eq. 1) and T2 (Eq. 2).

D. Cost saving with low renewable power source

The experiments of this sections are performed when renewable sources contribute only 10% of the data center total power. Due to such low renewable power, the cost saving difference of combined T1 and T2, mainly comes from leveraging the electricity price variation within epochs. In order to investigate the maximum cost saving of the two tier management scheme, we run the simulation when T1 and T2 access the actual data of workload and renewable over their decision window (24 hour for T1, and 30/60 minutes for T2 depending on the epoch length). We also run the simulation for two electricity pricing models (i) $\beta = 0$, no cost per peak power draw, and (ii) β =\$12/kW, and various capacity of ESDs. Since we are interested in the costs savings of two tier energy buffering management compared to one tier management, we set the capacity of the tier one ESD to a very large value, (9 GJ, enough energy for an hour work of data center at its peak utilization). This value is chosen to ensure that the cost saving of T2 is due to variation in demand and supply and not due to low capacity of the level one ESD. Fig. 7 shows that the maximum cost saving of one tier management for zero value of β is 19% that increases to 25% using combined T1 and T2. The cost saving of combined T1 and T2 is higher for non zero β, as shown in Fig. 8, since ESDs are utilized to decrease the peak power draw from grid. It can be seen in both figures that the cost of ESDs does not increase with increase in the size of the level two ESD. The reason is that the level one ESD is the most contributor of the cost, since the level two ESD has very large number of cycles.

As shown in Fig. 9, the cost saving of both T1 and T2 are reduced for epoch length of one hour (compare Fig. 8 and Fig. 9). The reason is that T1 cannot leverage electricity variation in the intervals of half an hour since it is not aware of that information. T2 can compensate the performance of

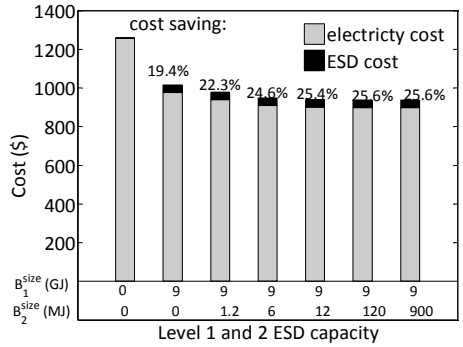


Fig. 7. Cost saving of combined T1 and T2 versus T1 for half-hour epochs, no cost per peak power draw ($\beta = 0$), and zero prediction error.

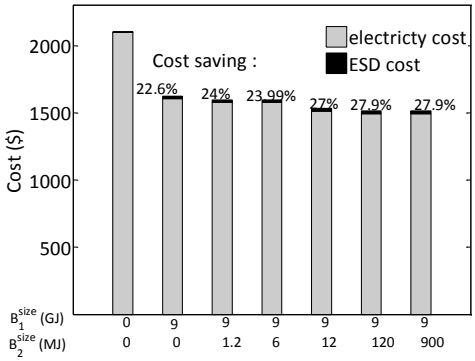


Fig. 8. Cost saving of combined T1 and T2 versus T1 for half-hour epochs, non-zero cost per peak power draw (β =\$12/kW), and zero prediction error.

T1 and incur a high cost saving by deploying relatively large ESD (i.e., 120 MJ, enough energy for 12 minutes work of data center at its peak utilization). T2 needs large ESD to be able to store energy within an hour.

As shown in Fig. 10, when T1 and T2 decide based on prediction data, their cost saving decreases due to prediction error (compare Fig. 8 and Fig. 10). However, still the combination of T1 and T2 incur up to 4% more cost saving compared to using T1 only. Note that we do not make any claims regarding the accuracy of our prediction techniques. Any prediction technique with higher accuracy can be used to improve the cost saving of T1 and T2.

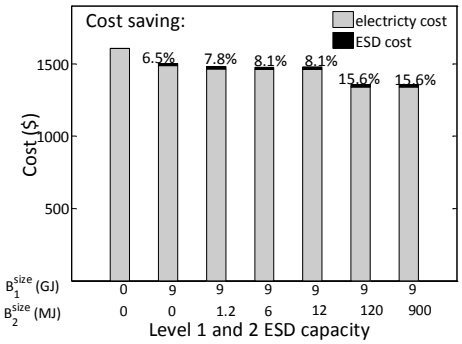


Fig. 9. Cost saving of combined T1 and T2 versus T1 for one hour epochs, non-zero per peak power draw (β =\$12/kW), and zero prediction error.

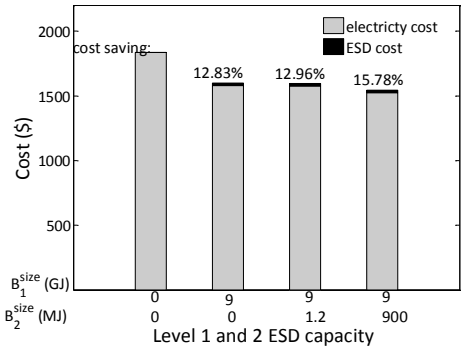


Fig. 10. Cost saving of combined T1 and T2 versus T1 for half-hour epochs, non-zero per peak power draw (β =\$12/kW) and using prediction techniques.

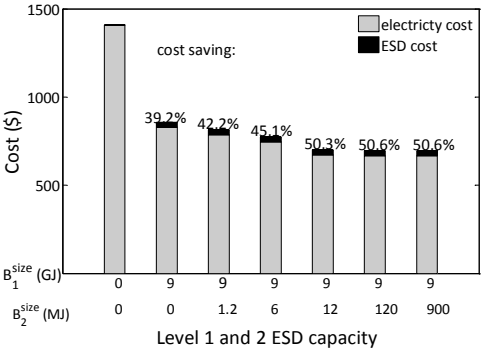


Fig. 11. Cost saving of combined T1 and T2 versus T1 for one hour epochs, non-zero per peak power draw (β =\$12/kW), zero prediction error, and high renewable generation rate.

E. Cost saving with high renewable power source

In all of the previous experiments, renewable power contributes less than 10% of the data center total required power, where renewable utilization is almost equal for with and without energy buffering management (i.e., T1 and T2). To demonstrate the efficiency of T1 and T2 in utilizing renewable, we set the renewable power generation parameters such that the renewable power sources contribute around 60% of the data center total power at its peak. Fig. 11 shows that in this case energy cost decreases by 39% when using T1 with perfect

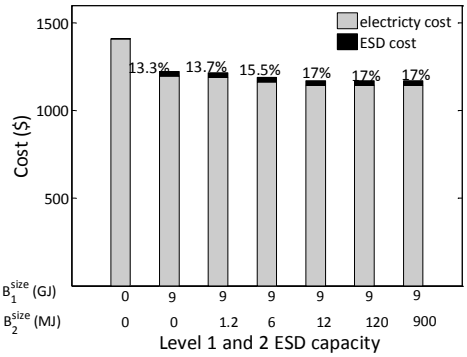
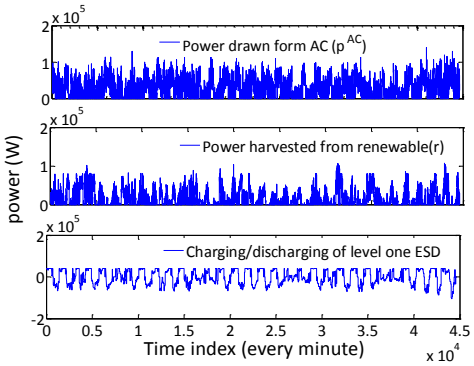
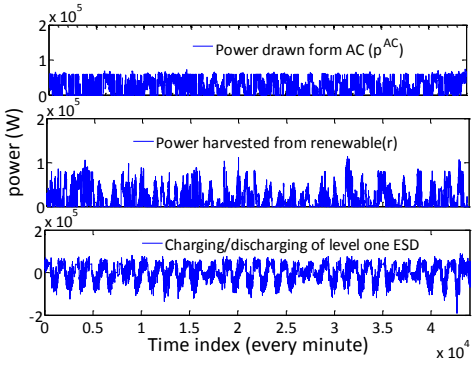


Fig. 12. Cost saving of combined T1 and T2 versus T1 for one hour epochs, non-zero per peak power draw (β =\$12/kW), using prediction technique, and high renewable generation rate.



(a)



(b)

Fig. 14. Power draw model from AC, renewable, and ESDs**(a)** T1 using $B_1^{size} = 9\,GJ$, $B_2^{size} = 0$, and **(b)** combined T1 and T2 for $B_1^{size} = 9\,GJ$, $B_2^{size} = 12\,MJ$.

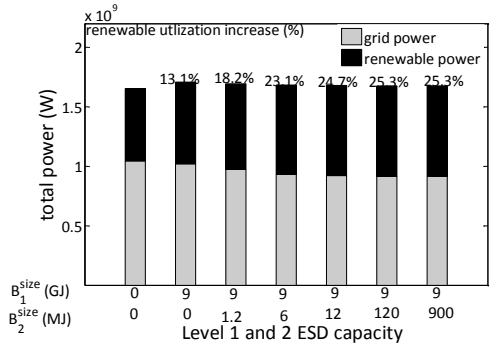


Fig. 13. Power draw model of combined T1 and T2 versus T1 for half-hour epochs, non-zero per peak power draw (β =\$12/kW), zero prediction error, and high renewable generation rate.

prediction (zero prediction error). The cost saving increases upto 50% when using combined T1 and T2. Note that we assume on-site renewable with zero operational cost. In other words, the extra cost saving of Fig. 11 compared to Fig. 8 comes from increasing the renewable utilization. This has been shown in Fig. 13 and Fig. 14. Fig. 13 shows the extra renewable that a combined T1 and T2 can harvest compared to T1. Fig. 14 shows the power draw model of T1 and the combined T1 and T2, where the level one ESD capacity at T1 for the both cases are set to 9 GJ, and the level two ESD capacity for the case of combined T1 and T2 is 900 MJ. This figure shows that the extra renewable utilization and cost saving of combined T1 and T2 comes from close matching of renewable to the power demand and from smoothing power drawn from AC at short term using frequent charging and discharging of the level two ESD of type ultra capacitor.

Fig. 12 shows that T1 and T2 when using predicted data still increase the renewable utilization and saves cost (compare Fig. 12 and Fig. 10) albeit lower than the case where data is accurately available to them (compare Fig. 12 and Fig. 11).

VI. Related Work

A. Battery management to reduce electricity cost

ESDs (in the form of UPSes) in data centers have been primarily used to supply power to the computing units during the grid utility outage for the duration it takes the backup sources to start up. Recent work explore the use of ESDs to reduce electricity operational costs at data centers [3]–[9], [19]. However, the focus of the existing studies are on developing single-tier energy buffering management schemes. Particularly, Urgaonkar et al. develop an on-line control algorithm using Lyponov optimization to exploit UPS devices to reduce cost in data centers [3]. Govindan et al. perform a comprehensive study on the feasibility of utilizing UPS to store low-cost energy, and design a Markovian based solution to schedule batteries [4]. Palasamudram et al. perform a trace-based simulation using Akamai CDN workload traces to investigate the energy cost saving that can be achieved by using batteries to shave the peak power draw from the grid [6]. Govindan et al., propose to leverage existing UPSes to temporarily augment the utility supply during emergencies (i.e., peak power) [8]. Finally, [7] presents an energy buffering management policy for distributed per-server UPSes to smoothen power draw from grid. The above works show the efficiency of ESD to reduce electricity price and handle power emergencies in data centers and consequently a motivation for our work. .

The most related work to ours is [5], where the authors investigate how data centers can leverage the existing huge set of heterogeneous ESDs. Wang et al. in this work study the physical characteristics of different types of ESDs, and their cost-benefit for utilizing in data centers. The authors also develop an offline optimization framework to decide on how heterogeneous set of ESDs can be placed in different levels of data centers power hierarchy (i.e., data center, rack, and server levels) in order to minimize the data center operational energy cost. However, the management scheme is mainly developed for the data center design, not for dynamic workload management, since it is based on the assumption that both short and long term variation of power demand for a long time horizon is given in advance. More importantly, the scheme is not aware of the on-site renewable power and only utilizes ESDs to manage short and long term variation of the power demand in order to reduce the electricity cost. However, our proposed two-tier scheme (combined T2 and T2) performs online workload and energy buffering management by being aware of the workload, the available renewable power and the electricity pricing variation.

B. Power management to maximize renewable utilization

Some related work propose green scheduling algorithms to maximally utilize renewable [9], [11], [19]–[22]. The idea

is to (i) adjust the power consumption to the available green power using power management techniques, e.g., server power state transitions [9], [11], [21], and (ii) postpone delay-tolerant jobs as much as possible to match the power demand and green power [19], [20], [23]. Our scheme also performs server power state transitioning in order to match the power consumption to the input workload at time granularity of minutes. The power efficiency of such a power management technique is previously proven [24]–[28]. Further, our scheme maximizes renewable utilization for Internet data centers not only by reducing the power demand, but also by close matching of the available renewable power to the power demand.

VII. Conclusions

This paper developed a two tier energy buffering and workload management frameworks (T1 and T2) for a data center with two types of ESDs (high energy density batteries at tier one and high power density ESD at tier two). The simulation study using realistic traces of workload, renewable, and electricity pricing, highlighted that the combined T1 and T2 significantly decreases electricity cost and increases available renewable energy utilization compared to one tier (only T1), specially when they use a prediction technique with high prediction accuracy. The reason is that on one hand, the high energy density batteries which can store large amount of energy, appropriate to leverage long-term variation of the power demand and green/low-cost power supply, cannot be quickly charged/discharged to leverage short term variation of the power demand and the green/low-cost power supply. On the other hand, high power density ESDs with high charging/discharging rate, appropriate to leverage short-term variation of the power demand and green/low-cost power supply, cannot store large amount of energy for a long duration. Therefore, a two tier energy buffering management, using two types of ESDs can achieve higher renewable energy utilization and lower electricity cost compared to single tier.

References

[1] R. Vokoun, "Renewable energy in today's data centers," White Paper, Datacenter Knowledge, May 2012.

[2] [Online]. Available: http://www.ecobusinesslinks.com/green_webhosts/

[3] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramanian, "Optimal power cost management using stored energy in data centers," in *ACM SIGMETRICS*, 2011, pp. 221–232.

[4] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and limitations of tapping into stored energy for datacenters," in *Proc. The 38th International Symposium on Computer Architecture (ISCA)*, San Jose, CA, USA, June 2011.

[5] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy, "Energy storage in datacenters: what, where, and how much?" in *12th ACM SIGMETRICS*, 2012, pp. 187–198.

[6] D. S. Palasamudram, R. K. Sitaraman, B. Urgaonkar, and R. Urgaonkar, "Using batteries to reduce the power costs of internet-scale distributed networks," in *Proc. of the Third ACM Symposium on Cloud Computing*, 2012, pp. 11:1–11:14.

[7] V. Kontorinis, L. E. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. Pettis, D. M. Tullsen, and T. S. Rosing, "Managing distributed UPS energy for effective power capping in data centers," in *Proc. of the 39th International Symposium on Computer Architecture (ISCA)*. IEEE Press, 2012, pp. 488–499.

[8] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar, "Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters," in *Proc. of the Seventeenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*. ACM, 2012, pp. 75–86.

[9] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and cooling aware workload management for sustainable data centers," in *Proc. of the 12th ACM SIGMETRICS*, 2012, pp. 175–186.

[10] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: optimization of distributed Internet data centers in a multi-electricity-market environment," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[11] Y. Zhang, Y. Wang, and X. Wang, "Greenware: greening cloud-scale data centers to maximize the use of renewable energy," *Middleware 2011*, pp. 143–164, 2011.

[12] Z. Abbasi, T. Mukherjee, G. Varsamopoulos, and S. K. S. Gupta, "DAHM: A green and dynamic web application hosting manager across geographically distributed data centers," *ACM Journal on emerging technology (JETC)*, vol. 8, no. 4, pp. 34:1–34:22, November 2012.

[13] Y. Xu and J. Chen, "Characterizing nonstationary wind speed using empirical mode decomposition," *Journal of structural Engineering*, vol. 130, no. 6, pp. 912–920, 2004.

[14] [Online]. Available: http://ita.ee.lbl.gov/html/traces.html.

[15] Z. Abbasi, M. Pore, and S. Gupta, "Impact of workload and renewable prediction on the value of geographical workload management." in *Second International Workshop on Energy Efficient Data Centers (E2DC), held as a part of ACM eEnergy*, May 2013.

[16] M. Al-Ghamdi, A. P. Chester, and S. A. Jarvis, "Predictive and dynamic resource allocation for enterprise applications," in *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, july 2010, pp. 2776 –2783.

[17] [Online]. Available: http://www.nrel.gov/midc/

[18] D. Montgomery, L. Johnson, and J. Gardiner, *Forecasting and time series analysis*. McGraw-Hill New York etc., 1990.

[19] Í. Goiri, R. Beauchea, K. Le, T. D. Nguyen, M. E. Haque, J. Guitart, J. Torres, and R. Bianchini, "GreenSlot: scheduling energy consumption in green datacenters," in *International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, 2011, pp. 20:1–20:11.

[20] Í. Goiri, K. Le, T. D. Nguyen, J. Guitart, J. Torres, and R. Bianchini, "GreenHadoop: leveraging green energy in data-processing frameworks," in *Proc. of 7th ACM European Conference on Computer Systems*, 2012, pp. 57–70.

[21] N. Sharma, S. Barker, D. Irwin, and P. Shenoy, "Blink: managing server clusters on intermittent power," *ACM SIGPLAN Notices*, vol. 46, no. 3, pp. 185–198, 2011.

[22] C. Li, A. Qouneh, and T. Li, "Characterizing and analyzing renewable energy driven data centers," *SIGMETRICS Perform. Eval. Rev.*, vol. 39, no. 1, pp. 323–324, Jun. 2011.

[23] B. Aksanli, J. Venkatesh, L. Zhang, and T. Rosing, "Utilizing green energy prediction to schedule mixed batch and service jobs in data centers," in *Proc. of the 4th Workshop on Power-Aware Computing and Systems*. ACM, 2011, pp. 53–57.

[24] D. Meisner, B. T. Gold, and T. F. Wenisch, "PowerNap: eliminating server idle power," *SIGPLAN Notices*, vol. 44, pp. 205–216, March 2009.

[25] B. Guenter, N. Jain, and C. Williams, "Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning," in *Proc. IEEE INFOCOM, Shanghai, China*, 2011, pp. 702–710.

[26] A. Krioukov, P. Mohan, S. Alspaugh, L. Keys, D. Culler, and R. Katz, "NapSAC: design and implementation of a power-proportional web cluster," in *Proc. of the first SIGCOMM workshop on Green networking*. ACM, 2010, pp. 15–22.

[27] D. Meisner, C. M. Sadler, L. A. Barroso, W.-D. Weber, and T. F. Wenisch, "Power management of online data-intensive services." in *Proc. of 38th IEEE International Symposium on Computer Architecture (ISCA)*. ACM, 2011, pp. 319–330.

[28] Z. Abbasi, G. Varsamopoulos, and S. K. S. Gupta, "TACOMA: Server and workload management in internet data centers considering cooling-computing power trade-off and energy proportionality," *ACM Trans. Archit. Code Optim.*, vol. 9, no. 2, pp. 11:1–11:37, Jun. 2012.