Energy Storage in Datacenters: What, Where, and How much?

Di Wang, Chuangang Ren, Anand Sivasubramaniam,
Bhuvan Urgaonkar, and Hosam Fathy*
Dept. of Computer Science and Engineering, *Dept. of Mechanical and Nuclear Engineering
The Pennsylvania State University
{diw5108, cyr5126, anand, bhuvan}@cse.psu.edu, hkf2@engr.psu.edu

ABSTRACT

Energy storage - in the form of UPS units - in a datacenter has been primarily used to fail-over to diesel generators upon power outages. There has been recent interest in using these Energy Storage Devices (ESDs) for demand-response (DR) to either shift peak demand away from high tariff periods, or to shave demand allowing aggressive under-provisioning of the power infrastructure. All such prior work has only considered a single/specific type of ESD (typically re-chargeable lead-acid batteries), and has only employed them at a single level of the power delivery network. Continuing technological advances have provided us a plethora of competitive ESD options ranging from ultra-capacitors, to different kinds of batteries, flywheels and even compressed air-based storage. These ESDs offer very different trade-offs between their power and energy costs, densities, lifetimes, and energy efficiency, among other factors, suggesting that employing hybrid combinations of these may allow more effective DR than with a single technology. Furthermore, ESDs can be placed at different, and possibly multiple, levels of the power delivery hierarchy with different associated trade-offs. To our knowledge, no prior work has studied the extensive design space involving multiple ESD technology provisioning and placement options. This paper intends to fill this critical void, by presenting a theoretical framework for capturing important characteristics of different ESD technologies, the trade-offs of placing them at different levels of the power hierarchy, and quantifying the resulting cost-benefit trade-offs as a function of workload properties.

Categories and Subject Descriptors: C.0 [Computer Systems Organization]: General

Keywords: Datacenters, Power Provisioning, Energy Storage, Cost Reduction.

1. INTRODUCTION

Datacenter power consumption is raising serious cost, environmental footprint, and scalability concerns. A large datacenter may spend millions of dollars in annual operational expenditures (op-ex) paying its electricity bills. An even larger capital expenditure (cap-ex) goes into provisioning the power delivery network to accommodate the peak power draw, even if this draw never/rarely occurs (see Figure 1). In fact, the peak power draw also impacts op-ex, because utility tariffs typically dis-incentivize high peak power draws (especially when they come simultaneously from numerous customers) with time-of-day pricing, peak-based pricing, etc. Demand-Response (DR) mechanisms are commonly deployed for modulating the power draws in electrical grids to address these concerns, and such mechanisms are also being adopted in the datacenter context. Datacenter DR primarily uses IT-based knobs - consolidation, scheduling, migration, and power state modulation - to shape power draw. One common mechanism - energy storage - that is used in grids (and other domains) [2, 45] to hoard power when available and draw from it when needed (to address supply-demand mismatch) has only recently drawn attention for datacenter DR [18, 20, 43].

Energy storage in the datacenter has been primarily used as a back-up mechanism to temporarily handle power outages until diesel generators are brought online. This typically takes a few seconds (at most a couple of minutes), and most datacenters provision UPS devices with lead-acid batteries of possibly 2X-3X this capacity. Many datacenters also have multiple (e.g., N+1 redundancy) UPS units for enhancing availability. Recently some proposals have been made to leverage either these devices, or add additional storage capacities, for DR within a datacenter [18, 20, 43].

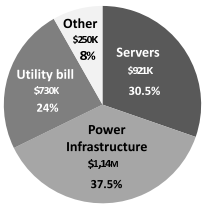

Figure 1: Amortized *Monthly* Costs for Physical Infrastructure of a 10MW Datacenter. Source: [3, 21, 32].

Concurrently, newer high-density datacenters are being built with "distributed" ESDs, e.g., server-level UPS units in Google [17], rack and power distribution unit (PDU) level units in Facebook [12] and Microsoft [31]. The primary motivation for such decentralization is to reduce energy losses: centralized UPS units cause a double-conversion wherein they convert from AC → DC → AC which is distributed to the servers and then converted back to DC by server-level power supplies. Moving the UPS units close to the server can help avoid losses due to such double-conversion (another alternative is to build DC-based datacenters as in [42]). In

SIGMETRICS'12, June 11–15, 2012, London, England, UK.

all these cases, the ESD - whether at a server or in a rack - is still used only for handling power outages.

While recent research has shown the benefits of lead-acid battery based UPS for datacenter DR - whether it be in shaving peaks to under-provision the power infrastructure [18, 20] or for shifting demand away from high tariff periods [43], all such prior work has only considered a single type of ESD, placed at a single location in the datacenter power hierarchy. Provisioning and placement of energy storage is a complex problem, with (i) a plethora of ESD options, (ii) each option with its idiosyncracies in cost, density, wear, and operational characteristics, (iii) a multitude of choices when placing them at different - possibly multiple - levels of the power hierarchy, and (iv) controlling their usage based on workload characteristics. This paper presents a formal framework for modeling and optimizing these design choices. Specifically, we address the following questions in the provisioning and operation of ESDs:

- Which ESDs should we employ? With a plethora of technologies available today, we need to model the trade-offs in cost, density, wear/health, efficiency and other operational characteristics when provisioning.
- Why should we be restricted to any one ESD technology? Given diversity in load demands temporally and spatially (across regions of the datacenter), should we consider multiple options given that different technologies are suited for different kinds of workloads?
- Where should these ESDs be placed? Why should we restrict them to any one level of the datacenter power hierarchy (e.g., central or server-level)? Does a multi-level energy storage hierarchy make sense given the statistical multiplexing effects (less burstiness) as we move up the hierarchy? An important consideration is the amount of space/volume that can be devoted for such ESDs, since datacenter real-estate is expensive.

While this paper provides specific answers to these questions (e.g., (i) single technologies suffice when there is little temporal/spatial heterogeneity, though the choice of technology would depend on the workload (ultra-capacitors and flywheels for tall and narrow workload peaks, but compressed air and/or batteries for wide peaks); (ii) regions of diversity in workload characteristics where hybrid ESD options are attractive; (iii) the benefits of multi-level ESD solutions, and the impact of volumetric constraints on these solutions), the more important contribution is a modeling and optimization framework that can take a range of input workload specifications, operating conditions, and power-related costs to derive ESD provisioning and dispatch solutions towards reducing cap-ex and op-ex. Using a combination of synthetic workloads which exercise the intricacies of ESD operation, and loads reported in 4 real datacenters, we demonstrate how this framework can be used to answer the above questions. We believe our framework can be very helpful when designing and operating a datacenter power infrastructure that leverages ESDs for DR.

2. BACKGROUND

Power Hierarchy: Figure 2 shows a typical datacenter power hierarchy. At the highest layer, a switchgear scales down the voltage of utility power, which then passes through centralized UPS units (with redundancy for availability). The UPS units also get another feed from backup diesel generators. Power then goes through Power Distribution Units (PDUs) which route it to different racks. At the next level, each rack may have regulated power distributors/outlets, with power then going down finally to the power supply units within each server (lowest level). In all, there are 4 levels shown in this particular hierarchy, though for simplicity we generally use 3 levels (datacenter, rack, and server) in this paper.

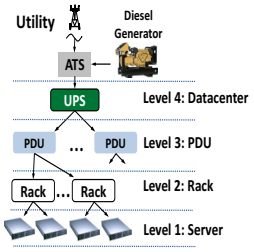

Figure 2: Typical Datacenter Power Hierarchy.

ESDs to Lower Power Costs: The cap-ex of provisioning the power infrastructure is reported to cost between $10-20 per watt [3, 21]. For safety reasons, it needs to be provisioned for the peak draw that may ever happen (even if it is a rare event). Shaving peak draws can, therefore, reduce cap-ex costs. Furthermore, shaving at the bottom/lower layers of the hierarchy benefits a larger portion of the hierarchy compared to shaving only at the top level - shaving at a level allows all the equipment in that level and above to be provisioned for a smaller peak. The op-ex that goes into paying the monthly power bills can have different tariff structures depending on the utility. Since peak draws at the utility scale are bound by the capacity of the grid and the generation capacity of power plants, utility companies disincentivize high peak draw with either (i) time-of-day pricing [5] (high prices occur when everyone is expected to have high draws), and/or (ii) peak slab based pricing [11] (where the maximum draw is separately tracked and priced for each watt of this draw). Thus, reducing the peak draw can also reduce op-ex. ESDs can hoard energy during a "valley" (period of low power draw), which can then be used to supplement the utility when a peak occurs - in effect, hiding a portion of the peak from higher up in the hierarchy (and the utility) - to provide both cap-ex and op-ex savings.

Centralized vs. Distributed and Hierarchical ESDs: ESDs, while continuing their role in handling outages, can be introduced at each level of the hierarchy. Most current datacenters use centralized UPS units, i.e., only at the datacenter level, which are typically lead-acid based with enough capacity to sustain the datacenter needs for a few minutes. Newer high-density datacenters are starting to avoid centralized UPS, and are going for de-centralized options, e.g., rack-level in Facebook [12], Microsoft [31], or server-level as in Google [17]. Whether centralized or distributed, all these are still single-level ESD placement options. We could have ESDs (possibly different technologies) simultaneously present at different layers hierarchically. We note the following qualitative comparisons between centralized vs. distributed/hierarchical placement options:

- Distributed ESD placement, particularly deep down (e.g., server level), allows cap-ex savings in a larger portion of the hierarchy.
- There is an analogy with issues in shared vs. private caches in chip-multiprocessors. While a shared (centralized ESD) cache allows better resource utilization when there is imbalance across different users, resource contention can become a problem leading to potential unfairness. Private (distributed ESDs) caches are isolated from each other from this perspective. For example, two server-level ESDs that can each shave a single peak, could do so

for their respective servers. It is possible that when replacing them with a single ESD at the higher level (with sum of their capacities), this ESD ends up shaving two peaks of 1 server and none of the other (becoming unfair).
- As noted earlier, a completely distributed server-level ESD option can help avoid double-conversion related energy losses compared to centralized placement.
- A centralized ESD solution may not have a serious volume/real estate constraint. In fact, many facilities place such ESDs outside of the datacenter itself (e.g., separate room of shelves with lead-acid batteries, flywheels, basements of buildings for compressed air, etc.), since datacenter floor/rack space is precious (running to thousands of dollars per square foot). A distributed/hierarchical solution would need real-estate within the datacenter.

Consequently, it would be interesting to find distributed and/or hierarchical ESD solutions, provided we can stipulate volume constraints for the ESDs at each level.

Availability Issues: Prior work [18, 19, 27] has shown that ESDs can be used for DR without affecting their ability to handle power outages, as long as a few residual minutes of capacity is always retained. We could either adopt this option, or have separate sets of ESDs (one set for handling outages and the other for DR). This relieves us of concerns of our ESD-based DR hurting datacenter availability.

3. ENERGY STORAGE DEVICES (ESD)

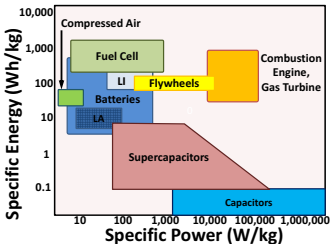

Figure 3: Ragone Plot.

Significant advances have been made in energy storage technologies, both in extending the practicality of their applications (by boosting densities, increasing lifetimes, lowering cost, etc.), as well as in bringing new technologies to the market. A broad spectrum of energy storage options exists, as is typically depicted using Ragone plots [36], such as the one shown in Figure 3. A Ragone plot compares different ESD technologies in terms of their power densities and energy densities, where "density" can be either volumetric or mass-based. As can be seen in Figure 3, there are sharp differences between various ESDs. For instance, compressed air-based energy storage (CAES), has a relatively high specific energy but a low specific power, implying that it is better suited for holding a large amount of energy as long as this energy does not need to be discharged very fast. On the other hand, capacitors have high power densities, albeit they are unable to sustain a large power draw for an extended period of time. In addition to this broad characterization, there are numerous other factors that can impact ESD suitability for DR, which we discuss below. In the interest of space and clarity, we use the following representative ESD technologies (which are also gradually finding their way into the datacenter [29]) from the Ragone design space in our discussions:

- *Ultra/Super-capacitors (UC):* These improve on conventional capacitors using double-layer electrochemistry between the electrodes, allowing a thousand-fold increase in energy stored. While the double-layer restricts it to low voltage operation, these capacitors can be connected in series for higher voltage needs. There has been nearly two orders of magnitude improvement in their cost per Farad over the past decade, and we are already seeing some commercial offerings of these products [28] for the datacenter market as UPS devices. We use this ESD as representative of the lower right end of the Ragone plot.
- *Compressed Air Energy Storage (CAES):* Air can be compressed (with compressors which consume energy) in confined spaces, and this pressurized air can subsequently be used to drive turbines for electricity generation. Since compression and decompression generate and absorb heat respectively, heat exchange mechanisms are needed to ensure proper operation. More importantly, CAES [9] may require significant real-estate to create such confined spaces, and the cost/availability of such real-estate needs to be taken into consideration (e.g., building basements, tanks in parking lots, etc.). We use CAES as representative of the upper left end of the Ragone plot.
- *Flywheels (FW):* The momentum of a rotating wheel/cylinder is gaining acceptance [47] as a UPS device for the datacenter, to temporarily handle the load until diesel generators kick in. Even though they are not intended for extended operation (somewhat inferior to even batteries from this perspective), they can provide the high power needs in the brief interlude between the outage and generator start-up.
- *Batteries (LA or LI):* These (particularly lead-acid) are the most commonly used storage devices in datacenters, where the electrochemical reactions (of the appropriate chemistry within), is used to store and generate electricity. There are several kinds of batteries [10] - lead-acid, lead-carbon, lithium-ion, vanadium flow, sodium-sulphur, etc., and we will consider lead-acid (LA) and lithium-ion (LI), which are more prevalent and representive of two ends of battery spectrum on the Ragone plot, and with very different costs, in our evaluations.

While some consider fuel cells, diesel generators, and other captive sources as ESDs, in this work we do not consider these options and limit ourselves to these 5 technologies.

ESD	LA	LI	UC	FW	CAES
Energy Cost C_k^{eng} ($/kWh)	200	525	10000	5000	50
Power Cost C_k^{pow} ($/kW)	125	175	100	250	600
Energy Density v_k^{eng} (Wh/L)	80	150	30	80	6
Power Density v_k^{pow} (W/L)	128	450	3000	1600	0.5
Discharge:Charge Rate γ_k	10	5	1	1	4
Life Cycle $Lcyc_k$ (# discharges x 1000)	2	5	1000	200	15
Max. DoD DoD_k^{max} (%)	80	80	100	100	100
Float Life T_k^{max} (years)	4	8	12	12	12
Energy Efficiency η_k (%)	75%	85%	95%	95%	68%
Self-discharge μ_k per Day	0.3%	0.1%	20%	100%	low
Ramp Time T_k^{ramp} (sec)	0.001	0.001	0.001	0.001	600

Table 1: ESD parameter values [9, 10, 38, 39].

We now explain the different factors that need to be taken into consideration when incorporating such ESDs into the datacenter. In Table 1, we quantify relevant parameters for the 5 technologies that we evaluate.

Cost (Energy and Power): When optimizing electricity costs with ESDs, we need to account for the costs of the ESDs themselves. The cost of ESD depends on 2 factors -

the total energy that is to be stored/discharged, and the rate (power) at which the energy is to be charged/discharged. This is somewhat indicated by where the device falls on the Ragone plot. Rows 1 and 2 of Table1 show these two components of the cost for the 5 ESDs under consideration. As can be expected, from the energy point of view, CAES is the least expensive (50 $/kWh) with ultra-capacitors at the other end of the spectrum at 200X this cost. On the other hand, with respect to power, ultra-capacitors are the most attractive option with CAES being 6X more expensive. Batteries offer a good compromise between these extremes.

Density (Energy and Power): Beyond costs, it is also important to consider the densities of these technologies required to provide a certain energy and power demand. Density determines the "volume" (real-estate) that needs to be provisioned in the datacenter to sustain the demands. Since datacenter real-estate is very precious - whether it be rack space or floor space, volume constraints may need to be imposed when provisioning ESDs. ESDs which may be attractive based on energy or power costs may not necessarily be suitable because of space constraints. For instance, CAES - the most cost-attractive option - is the worst from the density viewpoint even if we are only trying to cater to energy demands (and willing to tolerate the slow discharge rate offered by CAES). In fact, as we will find, trying to provide CAES for each server (or even a rack), is prohibitive in real-estate demands. At best, we can consider CAES at a datacenter scale, where basements, sealed tanks in parking lots, etc., may be options.

Discharge/Charge Rate Ratio: Since ESDs alternate between charging and discharging, it is important that there be sufficient charging time to hoard the required capacity before the next discharge. We can capture this by the discharge/charge ratio, which is larger than 1 (i.e., it takes longer to charge than discharge) for many ESDs. Ultra-capacitors and flywheels may come close to this ideal behavior, while batteries (particularly LA) are not as attractive.

Replacement Costs (Charge/Discharge cycles and Lifetime): We need to consider the costs of ESD replacement, since the datacenter infrastructure may itself have a much longer lifetime (e.g., 12 years as suggested in [21]). The lifetime of an ESD, especially batteries, depends (among other factors) on the number of charge-discharge cycles and the Depth-of-Discharge (DoD) of each discharge [9]. In addition, the internal chemistry itself has certain properties such as lead-out, which can also impact the lifetime orthogonal to usage. Row 8 of Table 1 gives the average lifetime (in years) of these ESDs based on typical usage. Batteries typically need replacement while our other technologies can possibly match the expected datacenter lifetime. It is not that batteries stop working abruptly - rather, their capacity for holding charge degrades over time, and replacement is done when it drops below 80% of the original capacity.

Energy Efficiency: The energy used to charge an ESD is higher than what can be drawn out subsequently, implying losses. Ultra-capacitors and flywheels are very energy efficient, while batteries can incur losses of 15-25% based on their chemistries. The efficiency of CAES is even worse.

Self-Discharge Losses: ESDs can lose charge even when they are not being discharged, with the loss proportional to the time since the last charge. Fly-wheels can be poor from this perspective, and so are ultra-capacitors. Consequently, it may be desirable that such devices be charged just before a discharge, rather than hoarding the charge for a long time.

Ramp Rate: While power density is one factor influencing the rate at which energy can be drawn, the ramp rate is another consideration in some ESDs. One can view the ramp rate as a start up latency to change the power output (analogous to how combustion engines of automobiles can accelerate from zero to a given speed in a certain amount of time, and then sustain it at that speed). In most ESDs we consider the ramp rate is very high (i.e., it takes at most a few milliseconds to start supplying requisite power draw), except in CAES where it can take several minutes. Hence, CAES cannot instantaneously start sustaining any desired draw, requiring either (i) anticipating the draw and taking pro-active measures, or (ii) using some other ESD until CAES becomes ready to sustain the draw.

4. EFFICACY OF DIFFERENT ESDS

To achieve effective selection and placement of ESDs in the datacenter, it is important to understand the efficacy of each ESD technology in shaping a given power demand time-series, and use this to understand the trade-offs across these technologies. Towards this, we intentionally keep the power demand representation simplistic at this stage (more extensive representations and real loads are considered later): it is simply an ON-OFF series where the ON periods correspond to a high demand value ("peaks") while the OFF periods correspond to a low demand value ("valleys"). We denote the mean amplitude ("height") and duration ("width") of the peaks and valleys of this time-series as (h_{peak},w_{peak}) and (h_{valley},w_{valley}), respectively. Note that both the capex and op-ex costs grow with the tallest peak of the series as discussed in Section 2. Hence to understand the efficacy of an ESD, we focus on finding its size/cost that is needed to "shave" a certain specified portion of *all* the peaks (rather than any one peak) in the time-series. We develop a simple model to compare their cost-efficacy in shaping different kinds of power demands.

4.1 Model for a single ESD

We denote the portion (amplitude) of the peak that is to be shaved as h_{shave}. It is important that h_{shave} be assigned a "realizable" value, e.g., clearly we must have $h_{shave} < h_{peak} - h_{valley}$, since an ESD capacity is finite and it needs to be charged at some point (in a valley). We denote the "frequency" of peak occurrences as $f_{peak} = \frac{1}{w_{peak}+w_{valley}}$. We use $k \in \{1, 2, ..., K\}$, to denote an ESD technology (so $K = 5$ in our evaluations). We can now translate the ESD properties in Table 1 into the following constraints to calculate the cost of ESD k (C_k^{total}) that must be provisioned to shave h_{shave} from the demand.

First, the power and energy densities of technology k impose these lower bounds on the required ESD capacity (cost):

$$\frac{C_k^{total}}{C_k^{pow}} \geq h_{shave}; \quad \frac{C_k^{total}}{C_k^{eng}} \times DoD_k^{max} \geq h_{shave} \times w_{peak}.$$

Next, the device must possess enough energy at the beginning of a peak to shave h_{shave} of it. It may have to acquire all of this energy by re-charging during the preceding valley, implying the following dependence on the discharge/charge rate ratio (γ_k). (Note that this re-charging will increase the power drawn during the valley making it higher than h_{valley}, but this increased value would still be less than $h_{peak} - h_{shave}$

by our "realizability" assumption above):

$$\frac{C_k^{total}}{C_k^{pow}} \times w_{valley} \times \frac{1}{\gamma_k} \geq h_{shave} \times w_{peak}.$$

Finally, the cost of ESD replacement should also be factored. This is governed by its expected float-life (T_k^{max}), as well as (i) the wear caused by repeated discharges and (ii) the extent/depth of each these discharges. Except for batteries, the other 3 ESDs have a large enough T_k^{max} to be less affected. To quantify the effect of wear on the lifetime, we can de-rate the expected number of lifetime charge-discharge cycles ($Lcyc_k$) which is calculated pessimistically at a Depth-of-Discharge of DoD_k^{max}, by the actual depth to which it is discharged DoD_k^{actual}, to get the expected lifetime of the ESD k as

$$Life_k = min\left(\frac{1}{f_{peak}} \times Lcyc_k \times \frac{DoD_k^{max}}{DoD_k^{actual}}, T_k^{max}\right).$$

This expected lifetime can be used to find the amortized cost (yearly) as per each of the above three requirements (energy needs, power needs, re-charging rates), and the cost of the employed ESD k is given by the maximum of these three requirements:

$$max\left(\frac{h_{shave}C_k^{pow}}{Life_k}, \frac{h_{shave}w_{peak}C_k^{eng}}{Life_k \times DoD_k^{max}}, \frac{h_{shave}w_{peak}\gamma_k C_k^{pow}}{w_{valley}Life_k}\right).$$

4.2 ESD Suitability for Different Demands

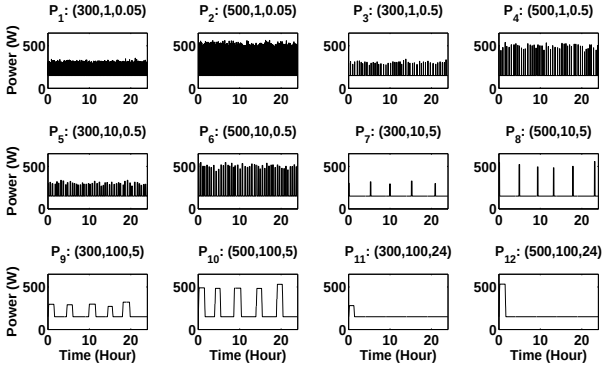

Figure 4: Power demand building blocks with mean values of (h_{peak} watt, w_{peak} min, $\frac{1}{f_{peak}}$ hour)

4.2.1 Model Evaluation Methodology

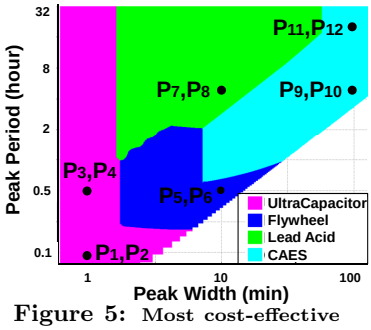

Figure 5: Most cost-effective ESD for different Peak types.

Since at this stage our goal is to evaluate the match between ESD properties and workload characteristics, we begin with synthetic power demands (using our ON-OFF series described above) for which we vary the parameters over a wide range. We use normal distributions for each of these parameters with specified mean and variance. We pick a set of 12 different combinations of means (2 each for peak height and frequency, 3 for peak width, and 1 for valley height) that we use as our synthetic workload "building blocks". Figures 4 (P_1-P_{12}) shows the 12 resulting power demands. For each of these 12 power demands, we choose an h_{shave} that is realizable for that demand. Using the model described in previous subsection, we compute the (peak width, peak frequency) region over which each of the 5 ESDs under consideration is the most cost-effective (since h_{shave} is set to be the same for all ESDs, the resulting cap-ex and op-ex savings are identical, and hence we need only compare the ESD costs). We show these regions in Figure 5, together with where our 12 workloads fall.

4.2.2 Key Findings from our Evaluation

No Single Technology Always Best: For each storage technology, there is a portion of the workload region, where it is cost-superior to other technologies. For example, ultra-capacitor is best when we have extremely narrow peaks (tens of seconds to a minute) as in P_1-P_4. Among the 5 technologies, ultra-capacitors offer the cheapest cost/power draw ($/kW). They can also re-charge fast enough within the high frequency of peak occurrence in these demands. Although they are the most expensive in terms of cost/unit energy ($/kWh), this does not become prohibitive since the peaks to be shaved for these demands require only small amounts of energy. At the other end, CAES is an attractive option for demands P_9-P_{12}, which are more "energy-demanding" with wide peaks (and ultra-capacitors are high cost options for these demands). CAES, on the other hand, requires very high capacities (and costs) to handle the narrow high power peaks where ultra-capacitors are attractive. There are regions in the middle where batteries (P_7-P_8) and flywheels (P_5-P_6) are the better options. *Therefore, a datacenter may need to consider these workload idiosyncrasies and variances over its lifetime in provisioning ESDs rather than always employing a fixed technology.*

Hybrid ESDs May Be Desirable: Certain technologies appear complementary to each other in terms of their pros and cons. The most stark contrast is between ultra-capacitors and CAES: while ultra-capacitors are the most cost-effective for P_1-P_4, CAES turns out to be the most prohibitive; the reverse holds for P_9-P_{12}. Such complementary behavior is also seen, to different degrees, for other pairs of technologies as well. *Therefore, when a datacenter may house different kinds of workloads - either across its different spatial regions or temporally over its lifetime - it may be worthwhile to consider hybrid ESD options.*

Multi-level ESD May Be Desirable: Within a datacenter, the nature of the power demand seen at different levels of its power hierarchy can be different, e.g., higher averages and smaller variances because of statistical multiplexing effects as we move up. When we compare a power demand with smaller average but higher variance (e.g, P_1) against those with higher average but smaller variance (e.g., P_{10}), we find different technologies being the most cost-effective. Whereas P_1 is best shaped using ultra-capacitors, CAES is the best choice for P_{10}. *Therefore, it may be desirable to employ appropriate (possibly different) ESD technologies at multiple - server, rack, and datacenter - levels of the power hierarchy. Furthermore, pushing ESDs deeper down the hierarchy can allow higher cap-ex savings.*

5. A FRAMEWORK FOR PROVISIONING AND CONTROL OF ESDS

Motivated by these insights, we develop an optimization

framework for provisioning ESDs within a datacenter. In particular, we design it to allow for multiple kinds of ESD technologies as well as the possibility of having them at multiple locations within the datacenter. Such placement should also take into consideration the volumetric/real-estate constraints that may restrict the usage/capacity of the ESDs at these locations (servers and racks) in the datacenter. Provisioning is closely tied with the associated control problem: how should these ESDs be used (i.e., charged/discharged) for a given datacenter configuration, utility tariffs, and power needs of various servers? Consequently, we design our framework to jointly address the provisioning and control problems. The goal is to determine an ESD based solution that maximizes the amortized net cost savings (i.e., cost savings in power-related cap-ex and op-ex minus cost of procuring and operating the ESDs).

5.1 Inputs

Workload (Power Demand): There is considerable prior work on characterizing and predicting server workloads (e.g., [1]) and properties such as time-of-day effects, etc., have been observed. For this work, we are concerned with a time-series of power draws at different levels of the datacenter. Since we already have a lot of ground to cover, we assume that prior work on load prediction can be leveraged, and combined with power modeling work (e.g., [4]), to derive a reasonable, accurate time-series of power draws at a server granularity over the given optimization horizon (e.g., a day). Further, our work is intended to provide guidelines when building (ESD solutions for) datacenters, at which point some estimate of load characteristics is in any case assumed for right-sizing of IT and power equipment. A more detailed treatment of these issues can be considered in future work. Specifically, for server i, we assume its power demand time-series given by $P_{1,i,t}, t \in \{1, ..., T\}$, where $T \times \delta$ represents our optimization horizon. We will consider different such time-series - both synthetic and real workloads - in our evaluations.

Power Infrastructure (Cap-Ex): We assume L levels (e.g., datacenter, rack, server) in the power hierarchy and use the variable $l \in \{1, ..., L\}$ to denote a particular level, with $l = L$ corresponding to the highest level and $l = 1$ corresponding to the server-level. Within a level l, we denote the number of power supply equipment (e.g., transformers, switchgear, and centralized UPS at datacenter level, and so on until individual power supplies at the server level) by n_l. The equipment also includes any ESDs we may need to provision at that level, and we begin with homogeneous equipment at a level for simplicity, though this can be generalized (particularly when we have different parts of the datacenter running different workloads as in some of our experiments). Prior work has pointed to cap-ex costs ranging between \$10-20 per watt of power provisioning (i.e., for our set of $\sum_l n_l$ equipment). Though the costs are not explicitly stated for power provisioning at each level of the hierarchy, it is typically more effective to start under-provisioning from deep down the hierarchy (i.e., $l = 1$) since it would allow larger portion of the hierarchy to benefit from cap-ex savings. We consider a conservative cap-ex saving of \$10/watt resulting from peak shaving as a starting point, and go as low as \$1/watt in our experiments. This is the saving that would be obtained by reducing a watt from the maximum draw P_L^{max}. The power draw under l at any time is given by the sum of the power draws of all servers under this sub-hierarchy, and the maximum of this sum over our optimization horizon is denoted as P_l^{max}.

Utility Tariffs (Op-Ex): Utilities base their tariffs on the actual energy consumption (say a \$/kWh), and the need to sustain the maximum power draw across all their customers within the constraints of their existing capacity. To address the latter concern, utilities dis-incentivize high power draws (especially simultaneously from multiple customers) by two mechanisms: (i) vary a (say as $a(t)$) based on the time-of-day [5]; and/or (ii) track the peak draw (typically averaged over 15 minute windows) and impose a cost of b \$/watt (e.g., as in [11]). Our framework is generic enough to accommodate either, and we simply use mechanism (ii) in our discussions/evaluations. Consequently, we need to track P_L^{max} (i.e. the maximum power draw at any time at the datacenter scale), and associate a b \$/watt cost to it.

5.2 Optimization Problem Formulation

Decision Variables: Given our goal to jointly address provisioning and subsequent control, we choose decision variables that capture the operational aspects of ESDs as well as the decisions about their sizing and placement. In the subsequent discussions, the subscripts l and i of the variables indicate the level in the datacenter hierarchy, and the index of associated equipment instance at that level, respectively. E.g., at the leaf level, $l = 1$ and i can take values from 1 to the number of servers; at rack level, $l = 2$ and i can take values from 1 to the number of racks and so on. In general, the tuple (l,i) denotes the root of the sub-hierarchy governing a certain set of servers. A server can source its power only from the ESDs that are in the path from itself to the root. Subscript k is used to denote the ESD technology.

First, let $S_{k,l,i}$ denote the "size" - the energy capacity of an ESD of type k placed at (l,i). Second, for each such device, to capture its "usage", we use variables $D_{k,l,i,t}$ and $R_{k,l,i,t}$ to represent the discharge and re-charge rate, respectively, during time slot t. To carry over the residual ESD energy capacity from one time slot to the next, we use $E_{k,l,i,t}$, which is the energy left in this device at the beginning of time slot t. Finally, $P_{l,i}^{realize}$ denotes the realized peak as a result of our shaving in sub-hierarchy (l,i).

Objective: We can now express various components of our overall objective function. All of these have been normalized/amortized to the horizon of our time series. The expected cap-ex savings in power infrastructure due to under-provisioning is given as $CapExSavings = \sum_{l=1}^{L} \sum_{i=1}^{n_l} \alpha_{l,i} \times (P_{l,i}^{max} - P_{l,i}^{realize})$, where $\alpha_{l,i}$ is the savings for each watt of under-provisioning at level l. The expected op-ex savings can be expressed as $OpExSavings = (\sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \sum_{t=1}^{T} a \times (D_{k,l,i,t} - \frac{R_{k,l,i,t}}{\eta_k})) + b \times (P_{L,1}^{max} - P_{L,1}^{realize})$, where a and b are the unit costs for energy and peak power draw in the utility tariff explained in previous subsection. Finally, the additional cost of ESDs themselves is given by $EStoreCost = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{i=1}^{n_l} (S_{k,l,i} \times C_{k,l,i})$. Here $C_{k,l,i}$ is the normalized cost of ESD k per unit energy adjusted to its actual lifetime, which depends on how the device is used (e.g., the same battery would last longer if it undergoes shallow discharges), and hence is itself unknown. Rather than dealing with a non-linear program in which $C_{k,l,i}$ is treated as an unknown, we keep our program linear and run it successively with the value of $C_{k,l,i}$ yielded by one run fed into the next run till convergence is achieved.

Finally, putting these components together, we have our objective as:
Maximize $(CapExSaving + OpExSaving - EStoreCost)$.
Constraints: We assume that all ESDs are fully charged at the beginning of the time-series, and need to leave them in the same state at the end of the time-series. In any time slot, an ESD may only hold energy between a lower threshold allowed by its recommended DoD and its maximum capacity. To capture these we have:

$$E_{k,l,i,1} = E_{k,l,i,T+1} = S_{k,l,i}, \forall k,l,i, (1a)$$

$$(1 - DoD_k^{max}) \times S_{k,l,i} \leq E_{k,l,i,t} \leq S_{k,l,i}, \forall k,l,i,t, (1b)$$

For each ESD, the amount of energy that can be discharged or stored is bounded by the product of its provisioned size and corresponding discharge ($r_k^{discharge}$) and recharge ($r_k^{recharge}$) rates:

$$0 \leq D_{k,l,i,t} \leq S_{k,l,i} \times r_k^{discharge}, \forall k,l,i,t, (2a)$$

$$0 \leq R_{k,l,i,t} \leq S_{k,l,i} \times r_k^{recharge}, \forall k,l,i,t, (2b)$$

We account for the conversion losses (energy efficiency η_k) of an ESD, during the charging process as

$$P_{l,i,t} = \sum_{j=1}^{n_{l-1,i}} \left(P_{l-1,j,t} + \sum_{k=1}^{K} \frac{R_{k,l-1,j,t}}{\eta_k} - \sum_{k=1}^{K} D_{k,l-1,j,t} \right), \forall l \geq 2, i,j,t, (3a)$$

Furthermore, when charging we should still ensure that the net power draw (including the power for all the equipment under l) is bound by $P_{l,i}^{realize}$ (which is in turn less than $P_{l,i}^{max}$). This gives us:

$$0 \leq P_{l,i,t} + \sum_{k=1}^{K} \frac{R_{k,l,i,t}}{\eta_k} - \sum_{k=1}^{K} D_{k,l,i,t} \leq P_{l,i}^{realize}, \forall l,i,t, (3b)$$

$$0 \leq P_{l,i}^{realize} \leq P_{l,i}^{max}, \forall l,i, (3c)$$

To account for energy losses due to self-discharge (μ_k) characteristics, we have

$$E_{k,l,i,t} = E_{k,l,i,t-1} + R_{k,l,i,t-1}\delta - D_{k,l,i,t-1}\delta - E_{k,l,i,t-1}\mu_k, \forall k,l,i,t \geq 2, (4)$$

The ramp-up properties of ESDs pose restrictions on how fast the rate of discharge can itself increase over time (recall the analogy of car acceleration). We capture this as:

$$\frac{D_{k,l,i,t+1} - D_{k,l,i,t}}{\delta} \leq \frac{S_{k,l,i} \times r_{k,l,i,t}^{discharge}}{T_k^{ramp}}, \forall k,l,i,t, (5)$$

Our final constraints restricts the volume ($V_{l,i}^{max}$) within which the ESDs must be accommodated at various levels:

$$\sum_{k=1}^{K} \left(\frac{S_{k,l,i}}{v_k^{eng}} \right) \leq V_{l,i}^{max}, \forall l,i, (6a)$$

and

$$\sum_{k=1}^{K} \left(\frac{S_{k,l,i} \times r_k^{discharge}}{v_k^{pow}} \right) \leq V_{l,i}^{max}, \forall l,i, (6b)$$

Discussion: When solving for a large datacenter with thousands of servers, this strategy can become intractable. We exploit the homogeneity of load across servers to appropriately scale down the number of decision variables, together with associated parameters. Finally, it should be noted that our framework automatically allows one ESD to charge by discharging another ESD instead of only using utility power.

6. EXPERIMENTS AND EVALUATION

6.1 Experimental Setup and Methodology

Max. Power	4 MW
# racks	256
# servers	8192
V_2^{max}	20% of rack vol.
V_1^{max}	10% of server vol.
Cap-ex (α)	\$10,\$1/Watt
Op-ex energy (a)	\$0.05/kWh
Op-ex peak (b)	\$12/kW/Month

Table 2: Parameter Values.

Configuration and Parameters: Our evaluations use a 4 MW datacenter of 8192 servers (placed in racks of 32 servers/rack), each with a 500W power supply, organized in a hierarchy as in Figure 2. We choose three levels for placing ESDs (L=3): top of the hierarchy (l=3, as in most datacenters today), rack-level (l=2, corresponding to some reported datacenters of Facebook [12] and Microsoft [31]), and server-level (l=1, as in some reported datacenters of Google [17]). Again note that these reported datacenters still place their ESDs at only one level, and not as a multi-level hierarchy which we allow.

We present results for $K = 5$ ESD technologies discussed earlier (however, LI based solutions do not explicitly appear in the results since whenever LI comes up on top of LA, the CAES option does even better). As discussed in Section 2, we use an op-ex cost model (Table 2) that is representative of that charged by Duke Electric [11], which has a monthly peak component charge of \$12/kW/month in addition to the energy usage charge. We consider cap-ex cost of \$10/W for the datacenter power infrastructure which is at the lower end of the \$10-20/W range reported in literature [3, 21]. Since it is difficult to quantify the cap-ex benefits for peak reduction only at a lower level of the hierarchy, we also consider sensitivity experiments where we push cap-ex costs as low as \$1/W in our technical report [48]. Note that even a cent of difference in additional cap-ex savings (i.e., $\alpha_{l+1} - \alpha_l > \epsilon$) when we push ESDs deeper down the hierarchy, will automatically enable our formulation to place ESDs deeper down the hierarchy as allowed by volume constraints. The cap-ex (12 year datacenter lifetime), op-ex (typically charged monthly), and ESD costs (appropriate actual lifetimes) are amortized for the horizon of the time series.
ESD Placement/Configurations: For comparison, we use 3 existing baseline strategies which are all "non-hybrid & single-level" placement styles (i.e., only one ESD technology x used at only one level in a given solution): $B_{dc,x}$, $B_{rack,x}$, $B_{server,x}$. Though these represent some datacenter ESD placement styles in use today, note that x=lead-acid batteries in most of them, and their capacities are chosen just for power outages. Our solutions will consider other options for x together with higher capacities. We compare these baselines with: (i) "hybrid & single-level," where we allow multiple technologies to be provisioned at a fixed level - centralized ($HybSin_{dc}$), rack-level ($HybSin_{rack}$), or server-level ($HybSin_{server}$), and (ii) "hybrid & multi-level," where we allow multiple technologies to be provisioned at possibly all of these levels ($HybMul$).
Workloads (Power Demands): To stress and understand the impact of workload parameters, we take the individual synthetic power demands P_1 ... P_{12} described earlier in section 4 and combine them in interesting ways (pair-wise) to bring out the impact of homogeneity and heterogeneity of workload behavior on ESD provisioning (section 6.2). Each resulting server-level power demand time-series spans a day, with each point in the series corresponding to the power

needs over a one minute duration. Broadly, we refer to the peaks with mean width upto a minute as "narrow," 1-10 minutes as "medium," and higher (10 minutes to 2 hours) as "wide." In addition, Section 6.3 studies these issues with power demands of real-world datacenters and clusters (for Google [13], TCS [44], MSN [8] and stream media clusters [23]), reported in prior studies.

6.2 Synthetic Workload Experiments

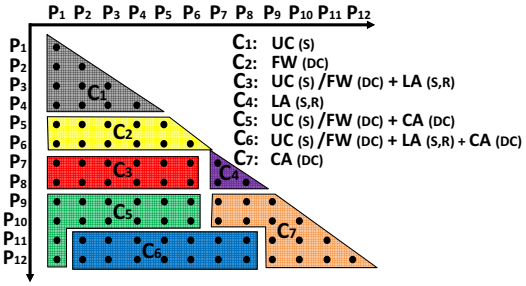

Figure 6: Results for the 78 synthetic workload combinations grouped in Clusters of "Best Configuration". The element (P_i, P_j) shows the ESD configuration when the per-server power demand is $P_{i,j}$.

While in Section 4 we had considered power demands that were homogeneous in time and in space (i.e., across servers) to get an overall idea of which single technology is better suited for a workload characteristic, we now use pair-wise combinations of the P_1 ... P_{12} time-series to study whether (i) hybrid combinations make sense, and (ii) whether multi-level ESDs are promising. This results in a 78-combination design-space of experiments to capture a large range of peak widths, heights and frequencies. This combination can create temporal heterogeneity. For example, Figure 7 shows one such power demand ($P_{2,9}$): it has two "phases" with significantly different properties. We will later see that this kind of workload is representative of some real world behavior with multi time-scale variations, e.g., time-of-day effects of load change at a macro-scale, with finer scale variations at each of these loads (as in the MSN workload). Dealing with such "temporal" heterogeneity, may mandate different solutions (hybrid or multi-level) compared to spatial heterogeneity, where different regions of the datacenter/racks may have different behaviors.

Results at a Glance: Figure 6 presents a simple way of viewing the results from our design space of experiments, by showing the best (or comparable to the best) ESD configurations for various $P_{i,j}$. For ease of discussion, we group these into "iso-configuration" clusters. The exact sizing decisions for various ESD technologies may vary across elements of a cluster. With a few exceptions, we find that our power demands fall into seven different clusters, and we present the ESD configuration for each cluster. For example, C_6 is best served by an ESD configuration which uses a combination of (i) ultra-capacitors at server level or fly-wheel at datacenter, plus (ii) lead-acid batteries at server and/or rack levels, plus (iii) CAES at datacenter level. We find such a representation easier to parse, than go through every data point in the results. The savings from such ESD provisioning are typically between 10-40% of total datacenter power-related costs, after factoring in the cost of ESD provisioning itself. Below we discuss the provisioning choices made by our framework for each cluster, and give detailed cost savings results for the $P_{2,9}$ workload.

Temporal Homogeneity (Peak Widths are all Narrow or Medium or Wide): Recall that our analysis in Section 4 had suggested that ultra-capacitors are the most cost-effective in dealing with narrow peaks (P_1-P_4 in Figure 4) due to their low power cost, superior lifetime, and excellent charge/discharge ratio. With the additional volume constraints, self-discharge, and energy efficiency that our optimization framework captures, ultra-capacitors continue to serve as the most cost-effective ESD technology for such power demands, and suffice by themselves without requiring any other ESD technology. The capacity can be met right at the server-level, within its volume constraints (thereby providing higher cap-ex savings in the hierarchy). This set of results is depicted by cluster C_1 in Figure 6.

At the other end of the spectrum, we find that a solely CAES-based ESD design is best for demands whose peaks are wide (P_9-P_{12}), again in agreement with our model. However, the volume constraints mandate that its capacity be provisioned at the datacenter level, rather than rack/server levels. This set of results is depicted by cluster C_7.

However, in the middle of the spectrum, lead-acid battery at the server-level and/or rack-level is the most cost-effective ESD in dealing with less frequent medium width peaks (cluster C_4). However, when the frequency of such peaks increases, the lifetime deterioration of lead-acid batteries comes into play, making flywheel at the datacenter-level as the better option (cluster C_2).

Note that homogeneous peak widths of pair-wise collation of P_is refers to the region close to the diagonal in Figure 6. In these regions, creating power demand mixes is not significantly changing the individual demand properties, keeping these results in agreement with those in Section 4.

Temporal Heterogeneity (Mix of Narrow, Medium and Wide Peaks):

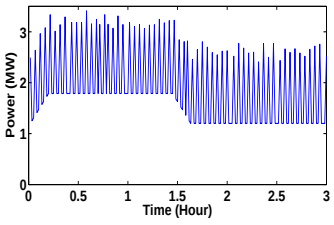

Figure 7: $P_{2,9}$ demand.

Clusters C_3, C_5 and C_6 bring out the need for hybrid, and possibly multi-level, ESD solutions. E.g., C_3 suggests both lead-acid batteries and ultra-capacitors at the server-level, or lead-acid batteries at server-level and flywheels at the datacenter-level; C_5 suggests CAES at the datacenter-level and ultra-capacitors at the server-level or both CAES and flywheel at the datacenter-level. In addition to the peak characteristics, the cap-ex savings when pushing ESDs deeper down the hierarchy as well as the associated volumetric constraints lead to these more extensive options rather than a single-level single-technology solution. To understand these issues in detail, we take $P_{2,9}$ (shown pictorially in Figure 7) and examine savings and ESD costs for different options in Table 3. We have the following observations and key take-aways from these results:

	B_ (savings, cost)	**HybSin_** (savings, cost)	**HybMul** (savings, cost)
Datacenter	LA (2.6k,0.5k)	UC+FW+CAES (3.0k, 0.2k)	FW+CAES
Rack	LA (2.3k,0.3k)	UC+LA (3.0k,0.2k)	-
Server	LA (1.7k,0.1k)	UC+LA (3.0k,0.2k)	UC (3k,0.2k)

Table 3: $P_{2,9}$: (Savings($/day), ESD costs($/day)). Total cost without ESD is $10K/day.

- Even the baselines $B\_$ (i.e., current datacenters endowed with sufficient lead-acid battery capacities) can offer substantial cost savings (16-25% even after factoring in the storage cost). In this case, the savings are better with a centralized ESD that does not have volume constraints as opposed to restricted capacity sizes deeper in the hierarchy, since we do not have explicit cap-ex savings values when shaving peaks at each level of the power hierarchy. This lack of additional cap-ex information about savings when we go deeper down the hierarchy (which is conservatively set to ϵ) is also the reason why the cost savings with $HybMul$ are not different from those of $HybSin_{dc}$ (though in reality the savings are likely to be higher for $HybMul$).
- For datacenters with centralized ESD, CAES (in combination with flywheels or ultra-capacitors if there are tall and narrow peaks to compensate for the slow ramp rate of CAES) appears to be a better option if space is not a problem. For instance, CAES-based hybrid datacenter level solutions provide 15% better savings than just a lead-acid centralized solution.
- The improvements that our `HybSin_` techniques offer over their corresponding baselines improve as we go down the hierarchy ($HybSin_{dc}$ offers 15% more cost savings that B_{dc} while $HybSin_{server}$ offers over 75% more cost savings than B_{server}). This results from the more stringent volume constraints at lower levels, where hybrid solutions that include ESDs with higher power and/or energy density offer larger gains. This may suggest that recent datacenters with distributed ESDs, like those at Google, Facebook and Microsoft, may benefit further from a move to hybrid ESDs.

Spatial Heterogeneity: Until now, we have only discussed spatially homogeneous power demands (i.e., all servers across the datacenter experience the same power demands). However, our techniques can address heterogeneous demands, and in fact hybrid and multi-level ESDs make even more sense in such environments. Many datacenters host different applications, with diverse power demands, e.g. a search engine may have applications which directly cater to Internet requests along with crawlers running in the background. These may not necessarily run on the same servers, and may even fall in different power sub-hierarchies in the datacenter. In the interest of space, we summarize one result to illustrate this point. When half the servers run P_1 and the other half run P_{12}, and we compare our $HybSin_{server}$ which uses ultra-capacitors and lead-acid batteries in the two sets of servers respectively, it provides over 50% more cost savings than both homogeneous ESD baselines $B_{server,UC}$ and $B_{server,LA}$.

6.3 Real Workload Experiments

In this section, we construct power demands to mimic behavior reported previously in four different real-world datacenters/clusters.

6.3.1 TCS

Figure 8(a) presents the power demand based on measurements reported from a TCS cluster [44] (that runs a range of many standard enterprise class applications) scaled to our assumed 4 MW datacenter. This power demand has an extremely small peak (the peak to average ratio is about 1.1), which lasts roughly 2 hours each day. We find that the net

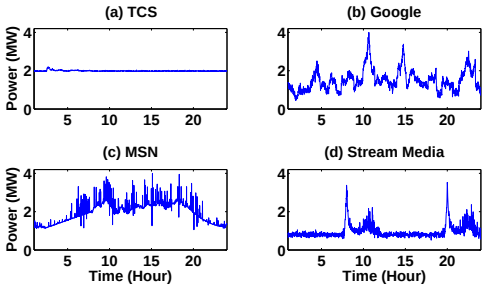

Figure 8: Real-world power profiles

savings offered by any ESD configuration is small (just 2%, though still off-setting the cost of the storage provisioning).

6.3.2 Google

Figure 8(b) presents the power demand from a Google cluster [13] scaled to our assumed datacenter. This power demand shows several peaks during the day with high variances in the power demand, suggesting that ESDs can help. Furthermore, given that server-level LA ESDs are reportedly employed in their datacenters, we would like to examine whether we can improve upon them.

	B_ (savings, cost)	**HybSin_** (savings, cost)	**HybMul** (savings, cost)
Datacenter	CAES (4.9k,0.4k)	FW+LA+CAES (5.2k,0.3k)	FW+CAES
Rack	LA (4.7k,0.3k)	UC+LA (4.8k,0.3k)	-
Server	LA (3.9k,0.2k)	UC+LA (4.7k,0.5k)	LA (5.2k,0.3k)

Table 4: Google Workload: (Savings($/day), ESD costs($/day). Total cost without ESD is $12K/day.

Table 4 presents results for the ESD configurations under consideration. First, we see that if we are to use just a single technology, single-level solution, a datacenter level provisioning (using CAES) does provide considerably higher savings (nearly 25%) than using only lead-acid batteries at each server. The volume constraint does play an important role in limiting the benefits of ESDs in the lower levels of the hierarchy. However, considerations such as double-conversion, and more significant cap-ex benefits (than what we have conservatively used) across the layers, may be reasons for going with a server-level option as in Google. Our framework suggests that even if we are restricted to a server-level placement, a hybrid option of ultra-capacitors together with lead-acid batteries can mitigate the volume constraints to bridge this 25% cost savings gap. If we can remove restrictions even further and explore multi-level hybrid options, our framework suggests flywheel and CAES at the datacenter level, together with lead-acid batteries at the server level to provide as much as 30% benefits over just server-level batteries. As can be seen in Figure 8(b), this workload shows burstiness at different time scales, allowing a richer set of ESD options to shape this power profile.

6.3.3 MSN

Figure 8(c) shows the load in a MSN facility [8], which has been translated to a power demand for our 4 MW datacenter. Table 5 shows the cost savings with the different ESD configurations. To a large extent, the results/savings are similar to those in the Google workload - capacity limitations and conservative cap-ex cost assumptions limit the extent of savings with just server-level lead-acid batteries, and CAES/flywheels can better augment the power needs centrally. The main difference is that this workload has

very diverse sets of peaks - some last as long as 8 hours per day, and some which are at most a few minutes. The short and bursty tall power spikes favor a ultra-capacitor based solution at the server level, whether it be in $HybSin_{server}$ or $HybMul$. Overall we find $HybMul$ giving around 30% and 20% better savings than server level alone, or rack-level alone placement of lead-acid batteries.

	B_ (savings, cost)	**HybSin_** (savings, cost)	**HybMul** (savings, cost)
Datacenter	LA (4.0k,0.5k)	UC+FW+CAES (4.4k,0.3k)	FW+CAES
Rack	LA (3.8k,0.3k)	UC+LA (4.3k,0.3k)	-
Server	LA (3.4k,0.1k)	UC+LA (4.2k,0.2k)	UC (4.4k,0.3k)

Table 5: MSN Workload: (Savings($/day), ESD costs($/day). Total cost without ESD is $15K/day.

Until now, we have presented only overall summary results. In order to give more detailed facets of ESD operation, we zoom in on a small time window of the MSN power profile and show the operation of different ESDs for $HybMul$ in Figure 9. For instance, the horizontal straight line shows the draw from the "utility", and we can see that despite the "demand" varying over time (which is also shown in this figure), the draw from the utility remains constant. To bridge this gap, the lines at the bottom show the charge (negative values of power in the y-axis), and discharge (positive values in y-axis) of the ESDs. We can see that CAES takes a bulk of the gap for significant portions of the time. However, when there is a sudden spike, the ultra-capacitor meets the difference. In order to charge this ultra-capacitor, we see that its curve goes through a negative spike just before it serves the required surge. Note that this negative spike is served by the energy sourced from the CAES, rather than pose an additional load on "utility".

Next, we focus on $HydSin_{rack}$ to take a closer look at the impact of device lifetime and charge/discharge ratio on provisioning efficacy. As in the Google environment, we arrive at a hybrid solution that combines ultra-capacitors with lead-acid batteries. Of the three main ways in which ultra-capacitors are better than lead-acid batteries - lifetime, charge/discharge ratio, and power density - we would like to understand which precise ones allow ultra-capacitors to complement lead-acid batteries so well. To do this, we relaxed the constraints related to each of these one at a time, and compared those results with the solution in Table 5. Though not explicitly shown here, we note that all three aspects have a role to play in why this particular hybrid solution was suggested. We show sensitivity results of these aspects in our technical report [48].

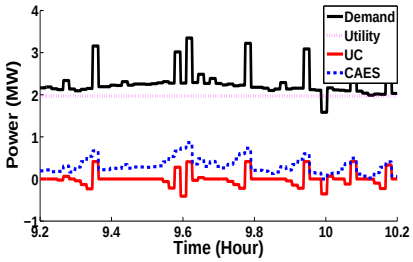

Figure 9: ESD charge/discharge control for the MSN power demand.

6.3.4 *Streaming Media*

Temporal heterogeneity can also occur at shorter (than MSN) time-scales, as in a server running a streaming media server where there is a power spike once every hour or few hours when new clients login to start watching a new show. We see this in the Media server load [23] shown in Figure 8(d), which has again been scaled for our 4 MW datacenter. Table 6 shows the correspond savings with ESD options.

	B_ (savings, cost)	**HybSin_** (savings, cost)	**HybMul** (savings, cost)
Datacenter	LA (5.6k,0.2k)	FW+LA+CAES (5.7k, 0.2k)	FW+CAES
Rack	LA (5.6k,0.2k)	LA (5.6k,0.2k)	LA
Server	LA (4.0k,0.1k)	UC+LA (5.4k,0.1k)	LA (5.7k,0.2k)

Table 6: Streaming media: (Savings($/day), ESD costs($/day)). Total cost without ESD is $10K/day.

As before, if considering a single level and single technology provisioning, the savings are better at the higher levels because of the restrictions we have imposed. However, even the server level hybrid provisioning (comprising lead-acid batteries and ultra-capacitors) does as well as any centralized provisioning, since the peaks are not as wide as in the MSN workload - benefits of ultra-capacitors can out-weigh any volume constraints of lead-acid batteries. Adding higher level storage capacities does not buy much more (just around 6% improvement).

6.4 Key Insights from Evaluation

1. ESDs help reduce power-related cap-ex and op-ex by up to 50%, even accounting for their provisioning costs.
2. With a single-level restriction, real-estate constraints limit the savings that we can get with server/rack level distributed solutions. In such cases, centralized ESD placement is a better option (e.g. 25% better in the Google workload).
3. Allowing hybrid ESD technologies even at the server level, improves the savings by upto 35% (e.g. in MSN workload) compared to a single ESD option.
4. Overall, a multi-level multi-ESD solution provides the best savings, giving improvements between 10-30% with respect to the best single-level single-ESD solution.

7. RELATED WORK

While there has been considerable work on datacenter energy and peak power management (summarized below), energy storage has been discussed mainly from the viewpoint of handling utility failures and/or intermittent power availability (e.g. with renewables [22, 24, 40, 41]). Exploiting energy storage for DR has been studied in battery-powered embedded, mobile, sensor domains [6, 35]), with such DR in datacenters only recently emerging [18, 20, 43].

Reducing Datacenter Energy Consumption: Much of the early work on datacenter power focused on reducing energy consumption by: (i) exploiting server-level performance knobs (e.g., DVFS-capable CPUs [15]), (ii) scheduling, placement/consolidation, and migration of computation across servers [7, 14, 26], (iii) reducing energy losses within the overall power infrastructure [30], and (iv) improving the energy consumed by the cooling infrastructure [37].

Cost-aware Provisioning and DR for Datacenters: Recent work [13, 16, 33] has proposed ideas to reduce provisioning costs or improve power infrastructure utilization with underprovisioning and statistically multiplexing it among workloads of complementary power needs. Such work relies upon reactive power control mechanisms to ensure safety

and limit performance overheads during such emergencies [37]. DR techniques have been explored to reduce the utility bill by adapting power consumption to the vagaries of electricity cost and availability. These include: (i) across datacenters - dynamic workload redistribution to datacenter sites with cheaper prices or power availability [34, 46] and (ii) within a datacenter - complementing the utility draw with energy storage during periods when energy is expensive or unavailable [43], workload scheduling/postponement to match electricity price [25], etc.

8. CONCLUSIONS AND FUTURE WORK

For the first time, this paper has investigated the novel problem of energy storage provisioning - what, where and how much - in the datacenter for Demand Response (DR) purposes. With a plethora of energy storage options, intricacies in their characteristics, pros and cons of placing them in different layers of the power hierarchy, and their suitability to diverse workload characteristics, there are numerous design choices when provisioning and dispatching power from these energy storage devices. We have presented a detailed treatise on energy storage technologies/characteristics impacting their operation in the datacenter. We have presented a simple model to gauge the suitability of a given technology to different workload power profiles, and used this to identify the regions where each becomes cost-effective. We have then formalized a systematic framework for capturing the different intricacies of ESD operations, and developed a generalized optimization platform for ESD placement and control in the datacenter. This platform can be invaluable in datacenter design, capturing a whole spectrum of costs, constraints and workload demands.

Using a wide spectrum of synthetic workloads that stress different aspects of these ESDs, we have shown (i) homogeneous ESD technologies suffice when there is not much heterogeneity in the workload, though the region of operation will decide which ESD should be deployed (e.g., narrow, tall and frequent peaks suited for ultra-capacitors/flywheels vs. broad and infrequent peaks better suited for compressed air and possibly batteries); (ii) even when placing ESDs in a single layer of the power hierarchy, considerations such as how much of the power hierarchy to optimize, volume constraints deciding storage capacity, and statistical multiplexing effects of the workload, influence where (server, rack or datacenter levels) the ESDs should be placed; (iii) even at a single layer of the power hierarchy, hybrid ESD solutions employing multiple technologies can offset the limitations of these constraints to provide substantial benefits (e.g., 25% improvement in cost savings in MSN at the server level); (iv) even if a hybrid ESD option is not employed at each level, a multi-layer hybrid solution can provide as much, if not better, savings across the spectrum of workloads that we have studied. The hybrid and multi-level ESD solutions are even more beneficial when the temporal (over different time scales) and spatial (across regions of the datacenter) heterogeneity in the workload increases.

We are building a small scale prototype with actual ESDs to study these issues experimentally. Further, we are looking to improve control/dispatch algorithms of the ESDs available in the datacenter based on continuously evolving properties. Our contributions are analogous to those in capacity planning of IT load for provisioning computing equipment in the datacenter, with the difference being that we are focusing on power provisioning. Such planning usually involves awareness of the workload, to help in right-sizing of the equipment. We believe this paper provides the valuable tools toward right-sizing the power infrastructure (when a datacenter is built) given workload characteristics.

9. ACKNOWLEDGMENTS

This work was supported, in part, by NSF grants CNS-0720456, CNS-0615097, CAREER award 0953541, and research awards from Google and HP.

10. REFERENCES

[1] M. F. Arlitt and C. L. Williamson. Internet Web Servers: Workload Characterization and Performance Implications. *IEEE Trans. Netw.*, 5(5):631–645, 1997.

[2] A. Bar-Noy, M. P. Johnson, and O. Liu. Peak Shaving Through Resource Buffering. In *Workshop On Approximation and Online Algorithms*, 2008.

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

[4] D. Brooks, V. Tiwari, and M. Martonosi. Wattch: a framework for architectural-level power analysis and optimizations. In *Proceedings of ISCA*, 2000.

[5] California ISO Open Access Same-time Information System Hourly Average Energy Prices, Nov. 2011. `http://oasishis.caiso.com/`.

[6] Q. Cao, D. Kassa, N. Pham, Y. Sarwar, and T. Abdelzaher. Virtual Battery: An Energy Reserve Abstraction for Embedded Sensor Networks. In *Proceedings of RTSS*, 2008.

[7] J. Chase, D. Anderson, P. Thakur, and A. Vahdat. Managing Energy and Server Resources in Hosting Centers. In *Proceedings of SOSP*, 2001.

[8] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao. Energy-aware Server Provisioning and Load Dispatching for Connection-intensive Internet Services. In *Proceedings of NSDI*, 2008.

[9] H. Chen, T. N. Cong, W. Yang, C. Tan, Y. Li, and Y. Ding. Progress in Electrical Energy Storage System: A Critical Review. *Progress in Natural Science*, 19(3), 2009.

[10] K. C. Divya and J. Stergaard. Battery Energy Storage Technology for Power Systems - An Overview. *Electric Power Systems Research*, 79(4), 2009.

[11] Duke Utility Bill Tariff, Oct. 2011. `http://www.duke-energy.com/pdfs/scscheduleopt.pdf`.

[12] Facebook Open Compute Project, Nov. 2011. `opencompute.org`.

[13] X. Fan, W.-D. Weber, and L. A. Barroso. Power Provisioning for a Warehouse-Sized Computer. In *Proceedings of ISCA*, 2007.

[14] A. Gandhi, V. Gupta, M. Harchol-Balter, and M. Kozuch. Optimality Analysis of Energy-Performance Trade-off for Server Farm Management. In *Proceedings of SIGMETRICS*, 2010.

[15] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy. Optimal Power Allocation in Server Farms. In *Proceedings of SIGMETRICS*, 2009.

[16] D. Gmach, J. Rolia, C. Bash, Y. Chen, T. Christian, A. Shah, R. K. Sharma, and Z. Wang. Capacity

Planning and Power Management to Exploit Sustainable Energy. In *Proceedings of CNSM*, 2010.

[17] Google Server-level UPS for Improved Efficiency, Apr. 2009. `http://news.cnet.com/8301-1001_3-10209580-92.html`.

[18] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar. Benefits and Limitations of Tapping into Stored Energy For Datacenters. In *Proceedings of ISCA*, 2011.

[19] S. Govindan, D. Wang, L. Chen, A. Sivasubramaniam, and B.Urgaonkar. Towards Realizing a Low Cost and Highly Available Datacenter Power Infrastructure. In *Workshop on HotPower*, 2011.

[20] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar. Leveraging Stored Energy for Handling Power Emergencies in Aggressively Provisioned Datacenters. In *Proceedings of ASPLOS*, 2012.

[21] J. Hamilton. Internet-scale Service Infrastructure Efficiency, ISCA 2009, Keynote. `http://perspectives.mvdirona.com/`.

[22] K. Le, R. Bianchini, T. D. Nguyen, O. Bilgir, and M. Martonosi. Capping the Brown Energy Consumption of Internet Services at Low Cost. In *Proceedings of IGCC*, 2010.

[23] B. Li, G. Y. Keung, S. Xie, F. Liu, Y. Sun, and H. Yin. An Empirical Study of Flash Crowd Dynamics in a P2P-Based Live Video Streaming System. In *Proceedings of GLOBECOM*, 2008.

[24] C. Li, A. Qouneh, and T. Li. Characterizing and Analyzing Renewable Energy Driven Data Centers. In *Proceedings of SIGMETRICS*, 2011.

[25] J. Li, Z. Li, K. Ren, X. Liu, and H. Su. Towards Optimal Electric Demand Management for Internet Data Centers. *IEEE Trans. on Smart Grid*, 2011.

[26] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew. Greening Geographical Load Balancing. In *Proceedings of SIGMETRICS*, 2011.

[27] M. Marwah, P. Maciel, A. Shah, R. Sharma, T. Christian, V. Almeida, C. Araújo, E. Souza, G. Callou, B. Silva, S. Galdino, and J. Pires. Quantifying the Sustainability Impact of Data Center Availability. *SIGMETRICS Performance Evaluation Review*, 37(4), 2010.

[28] Maxwell Ultracapacitor Uninterruptible Power Supply Solutions, Oct. 2011. `http://www.maxwell.com/products/ultracapacitors/industries/`.

[29] S. McCluer and J. F. Christin. APC White Paper: Comparing Data Center Batteries, Flywheels, and Ultracapacitors, Oct. 2011. `http://www.apcmedia.com/salestools/DBOY-77FNCT_R2_EN.pdf`.

[30] D. Meisner, C. M. Sadler, L. A. Barroso, W. Weber, and T. F. Wenisch. Power Management of Online Data-intensive Services. In *Proceedings of ISCA*, 2011.

[31] Microsoft Reveals its Specialty Servers, Racks, Apr. 2011. `http://www.datacenterknowledge.com/archives/`.

[32] M. K. Patterson, D. G. Costello, P. F. Grimm, and M. Loeffler. Data Center TCO; A Comparison of High-density and Low-density Spaces. In *Proceedings of THERMES*, 2007.

[33] S. Pelley, D. Meisner, P. Zandevakili, T. F. Wenisch, and J. Underwood. Power Routing: Dynamic Power Provisioning in the Data Center. In *Proceedings of ASPLOS*, 2010.

[34] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the Electric Bill for Internet-Scale Systems. In *Proceedings of SIGCOMM*, 2009.

[35] M. Ra, J. Paek, A. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely. Energy-delay Tradeoffs in Smartphone Applications. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*, 2010.

[36] D. V. Ragone. *Review of Battery Systems for Electrically Powered Vehicles*. Society of Automotive Engineers, 1968.

[37] P. Ranganathan, P. Leech, D. Irwin, and J. Chase. Ensemble-level Power Management for Dense Blade Servers. In *Proceedings of ISCA*, 2006.

[38] D. Rastler. Electricity Energy Storage Technology Options. Technical Report 1020676, Electric Power Research Institue, 2010.

[39] S. M. Schoenung and W. V. Hassenzahl. Long- vs. Short-Term Energy Storage Technologies Analysis. A Life-Cycle Cost Study. A Study for the DOE Energy Storage Systems Program. Technical Report SAND2003-2783, Sandia National Laboratories, 2003.

[40] N. Sharma, S. Barker, D. Irwin, and P. Shenoy. Blink: Managing Server Clusters on Intermittent Power. In *Proceedings of ASPLOS*, 2011.

[41] C. Stewart and K. Shen. Some Joules Are More Precious Than Others: Managing Renewable Energy in the Datacenter. In *Workshop on HotPower*, 2009.

[42] M. Ton, B. Fortenbery, and W. Tschudi. DC Power for Improved Data Center Efficiency, Mar. 2008. `http://hightech.lbl.gov/dcpowering/about.html`.

[43] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam. Optimal Power Cost Management Using Stored Energy in Data Centers. In *Proceedings of SIGMETRICS*, 2011.

[44] A. Vasan, A. Sivasubramaniam, V. Shimpi, T. Sivabalan, and R. Subbiah. Worth Their Watts? - An Empirical Study of Datacenter Servers. In *Proceedings of HPCA*, 2010.

[45] P. Ven, N. Hegde, L. Massoulie, and T. Salonidis. Optimal Control of Residential Energy Storage Under Price Fluctuations. In *Proceedings of the International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies*, 2011.

[46] A. Verma, P. De, V. Mann, T. Nayak, A. Purohit, D. Gargi, and K. Ravi. BrownMap : Enforcing Power Budget in Shared Data Centers. In *Proceedings of the USENIX International Middleware Conference*, 2010.

[47] VYCON: Flywheel Based UPS System in Data Centers, Nov. 2011. `http://www.vyconenergy.com/pq/ups.htm`.

[48] D. Wang, C. Ren, A. Sivasubramaniam, B.Urgaonkar, and H. Fathy. Energy Storage in Datacenters: What, Where, and How much? Technical Report CSE-11-016, The Pennsylvania State University, November 2011.