Should we Dual-Purpose Energy Storage in Datacenters for Power Backup and Demand Response?

Iyswarya Narayanan, Di Wang, Abdullah-Al Mamun, Anand Sivasubramaniam and Hosam K. Fathy
The Pennsylvania State University

Abstract

Prior work has shown the benefits of Energy Storage Devices (ESDs), such as batteries, to smoothen/flatten power draws in Datacenters, for reducing demand during peak tariffs (for op-ex savings) and under-provisioning the power infrastructure (for cap-ex savings). Until now, all prior studies for such smoothening, referred to as Demand Response, have considered re-purposing existing UPS unit batteries for demand response. It is not clear if such dual usage - handling power outages and demand response - is the most effective option since the needs (energy and/or power), mandates (best effort vs. hard stipulations), costs, availability and health degradation considerations could be very different. In this paper, we study the design space of choices for provisioning ESDs for these dual purposes - separate ESDs for each purpose, common pool of ESDs for both purposes, and soft-reservations in this pool with possible re-purposing dynamically based on demand. Our evaluations show that: (i) provisioning lead-acid batteries for a peak "power" load needed to handle power outages already comes with sufficient energy capacity that is more than adequate to automatically supply the energy needs for demand response; (ii) this makes it economically attractive to use the same UPS batteries, originally intended for Power Outages, for Demand Response as well, despite any consequent health degradation (due to repeated discharges); (iii) the ability to handle the needs during a power outage is not compromised despite the dual-purposing of these UPS batteries; and (iv) the non-orthogonality of the power and energy capacities of these batteries (i.e. provisioning for the high power needs during an outage automatically comes with a lot of energy capacity) suggests the possibility of having different Energy Storage Technologies for the two purposes and we show that a heterogeneous/hybrid option using Ultra-capacitors or Flywheels for Power Backup and batteries for Demand Response is a more cost-effective option.

1 Introduction

Energy storage has drawn a lot of recent attention in the datacenter context. While conventionally Energy Storage Devices (mainly Lead Acid Batteries) have been used in datacenters for temporarily handling power outages (in Un-interrupted Power Supplies) before transitioning the load to Diesel Generators, the more recent interest has been in leveraging energy storage for Demand Response (DR). By re-shaping the power demand, primarily shaving/capping peaks, there are considerable opportunities for reducing operational (deferring higher loads to non-peak tariffs) [4, 12] and/or capital (under-provisioning the power distribution infrastructure) [5, 6, 12] costs. However, all such prior explorations have used the same set of batteries (in UPS units), for both purposes - Power Backup (PB) and Demand Response (DR). It is not clear whether such a common (shared) pool of batteries is the best provisioning and usage option across these dual purposes, especially given the diversity in needs, mandates, usage and functionality across these two. Consequently, this paper explores the following important question: *should we dual-purpose the energy storage devices (ESDs) across both PB and DR, or should we have separate ESDs for each?* The answer, in a nutshell, is "yes, if we are confined to lead-acid batteries" and "no, if we could deploy hybrid ESD technologies". Figure 1 pictorially captures the requirements

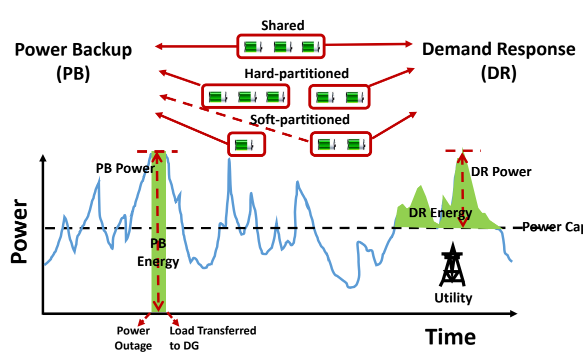

Figure 1: Battery Provisioning and Usage for PB and DR.

of ESDs for the 2 purposes - Power Backup (PB) and Demand Response (DR). These two can be quite different in the requirements and usage as described below:

- **Power needs:** PB needs to take over the entire load during outage (the entire amplitude of the y-axis in Figure 1), as opposed to handling partial load (complementing utility power) in DR. Thus, PB is more power demanding than DR.
- **Energy needs:** ESDs for PB only need to sustain power during the transition time to Diesel Generators (typically duration of few seconds/minutes [1] even when the outages are long). On the other hand, in DR, ESDs may need to supplement utility for tens of minutes or even hours [4, 6, 5, 12, 11]. Thus, DR can be more energy demanding than PB.
- **Usage Frequency:** Unlike PB which is called upon only for occasional power outages (a few times per year) [9, 8], DR could be more frequent (multiple times a day) [4, 6, 5, 12, 11]. This can cause more battery wear, thereby further impacting costs.
- **Mandates:** Batteries for PB need to provide high availability, while DR may be employed more on a best-effort basis for cost savings. Using PB batteries for DR can leave them in a semi-charged state without enough capacity to handle a power outage.

With these differences, a detailed study of the pros and cons when dual-purposing UPS batteries is warranted. Such a study should include several considerations including the Total Cost of Ownership (TCO), the wear-and-tear of the batteries due to repeated charge-discharge cycles in DR, and the ability to seamlessly handle a power outage for PB. This paper undertakes such a study, employing real world power demand traces at a rack level and examines different ways (in terms of both provisioned capacities and how they are used) of purposing the batteries for PB and DR. We find that the power demands of PB are high, and sizing the batteries for such high power demands automatically gives a sufficiently large energy capacity. This energy capacity can handle the energy needs of DR as well. Consequently, a common pool of batteries (sized based on PB power demand) that serves both PB and DR is a more cost-effective option. However, if we move away from batteries, we can leverage the more orthogonal power-energy behavior (i.e. one is less correlated with the other) of other ESDs for better cost savings. For instance, meeting the power demands of PB and DR with power-efficient ESD technologies such as Flywheels and Ultracapacitors while using batteries just for DR, gives around 50% cost savings compared to today's approach of just using batteries.

2 Provisioning for Dual Purposes

2.1 Choices

When provisioning (lead acid batteries which are the most commonly used ESDs in datacenters) for these two different purposes, we can consider the following options:

- **Shared usage:** Provision only one set of batteries, and use them for both purposes as the occasion demands. However, the capacity/number of batteries in this shared pool can be provisioned based on PB or DR needs, which we will study.
- **Hard-partitioned usage:** Provision two sets of batteries, one each for PB and DR. i.e. one set is not used for the other purpose. For each set, we will study how much power and energy capacity to provision.
- **Soft-partitioned usage:** Like the hard-partitioned case, we could have two separate sets of batteries. We provision the DR set according to the power and energy needs for a given power cap, and under-provision the PB set. But, if needed we allow DR batteries to step-in during power outages for PB. We do not consider the reverse case, where we under-provision for DR but provision in full for PB. This would be less cost-effective as PB batteries are already over-provisioned in energy as we show later.

For all these choices, we would consider tighter provisioning, and the consequent trade-offs in availability, battery lifetime/wear and cost.

2.2 Evaluations

We evaluate the different provisioning options at a rack level that has a peak power requirement of 10 KW.

Workload: We use real datacenter power demand of a production facility from [11] over a 6 month period with time resolution as fine as 20 seconds. Figure 2 shows a one week sample of the power demand. We use the term, dynamic power range, to denote $(P_{max} - P_{min})$, where P_{max} and P_{min} represent maximum and minimum power demand, respectively. We define power cap (C_f) as a fraction, f, of the dynamic power range. The power cap value is then given by $C_f = (1-f) \times (P_{max} - P_{min})$. We evaluate DR with different power caps (f = 30%, 35% and 40%), that captures the ability to flatten the demand curve using the batteries.

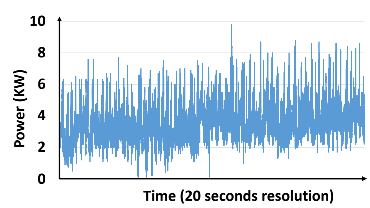

Figure 2: 1 Week of Power Demand for 10KW Rack.

Battery Modeling, Capacities and Quantity: We model OPzS Lead Acid batteries using the Kinetic Battery Model (KiBam) [7]. In this model, the battery charge is distributed over two wells: the available-charge well and the bound-charge well. The charge in the available-well can be readily used by the load, where-as the bound charge has to diffuse to the available well before being used. This diffusion is dependent on the difference in heights between these two wells. The rated capacity effect and the capacity recovery effect are captured by this model. Moreover, a rainflow counting method [3] is used to capture the battery damage (lifetime reduction) caused by DR usage. The parameter values of a unit battery that we use are shown in Table 1. Multiple such unit batteries are connected in series and/or parallel to provide power for a rack.

Energy Capacity	300 Wh
Power Capacity	300W
C-rate	1
Nominal Voltage	12V
Cutoff Voltage	10.2V

Table 1: Unit Batt. Parameters

Let N_{PB} and N_{DR} denote the number of batteries used for power backup and demand-response respectively. For a 10 KW peak power rack, we will need 33 units of the 0.3KW battery. So, our $N_{PB} = 33$ for the case where we provision only for power backup. *Note that this battery has an energy capacity of 0.3KWhr, implying that we have a highly over-provisioned energy capacity of 10 KWhr to handle any power outage that would at most take a few minutes of the battery runtime before the Diesel Generators take over.* When we separately provision for demand-response, as in hard-partitioned and soft-partitioned cases, we use the battery model to estimate the number of batteries, N_{DR}. It is the lowest number of batteries that can sustain voltage above the cut-off for the given power trace any time during DR.

Metrics for evaluation:

(i) Probability of Handling Power Outages: Dual purposing batteries can impact their ability to handle power outages, which is quantified using this metric. UPS batteries are only used to sustain power for the transition to Diesel Generators upon an outage. This transition could take a few seconds or at most a few minutes [1]. We consider an upper bound of up to 5 minutes for the transition time during which the batteries have to supply the entire power for the rack. We take published power outage durations from [9, 8] and histogram them into bins (of minute resolution), with the last bin counting all outages of 5 minutes and above. For each of these bins, we then randomly (1000 samples) pick a time in the power trace. We simulate the battery charge-discharge characteristics from the beginning and up to that point in time for any Demand Response usage. We then completely source the entire power from the battery from then on, until the outage duration corresponding to that bin. If the voltage does not fall below the cut-off voltage by then, the battery has successfully handled the outage. Using these random samples, we then calculate the corresponding probability of handling the outage by the battery. A value of 1.0 implies that the batteries are successfully able to handle all outages, while lower values indicate that DR usage is impacting their ability to handle the outages.

(ii) DR Damage index: Cycles of charge-discharge during DR can damage the battery, and this problem worsens with deeper discharges. We use Rainflow counting to compute the damage index referred to as D_{DR}. Range i is defined as the difference between the maximum and minimum depth of discharge for a given cycle. This range is divided into a number of bins. Rainflow counting histograms the number of cycles of the given range i. DR Damage index for the simulation period (6 months) is computed as in [7] as, $D_{DR} = \sum_i \frac{n_i}{N_i}$, where n_i is the number of cycles of range i that the battery has been through, and N_i is the number of cycles to fail if the battery operates at range i. We compute this metric for unit batteries that have a float life of T_{max} years (4 years for Lead-acid batteries). It essentially captures the proportion of battery life consumed due to Demand-Response. If this value exceeds 1, the battery has to be replaced earlier than the float life. If it is less than 1, the battery needs replacement only at the end of its float life. For instance, a value of 1.25 indicates that the battery was replaced before it reaches its float life, and after replacement 25% of its life-time has been consumed. So, for a float life of 4 yrs, and datacenter lifetime of 12 yrs, 3.75 (1.25*12/4) lifetime worth of batteries are needed.

(iii) Total Cost of Ownership (TCO): The TCO of batteries is computed based on the unit battery power and energy cost, the total number of batteries provisioned, and replaced over the lifetime of a datacenter as shown in the following equations. A battery is replaced at the end of its float life, or sooner because of demand-response usage. Let l be the lifetime of the datacenter, and N_{DR} be the number of batteries meeting DR power/energy needs. The total number of batteries needed for demand-response over the datacenter's life time is $N'_{DR} = Max(1, D_{DR}) * \frac{l}{T_{max}} * N_{DR}$. This captures the number of batteries replaced based on which of the two (float life or wear out) happens first. For batteries that are provisioned for outages, the total number of batteries for l years is given by $N'_{PB} = \frac{l}{T_{max}} * N_{PB}$, where N_{PB} represents the number of batteries meeting outage power/energy needs and the damage due to PB is ignored (since they are rare).

$$TCO = Max(BattPowerCost, BattEnergyCost) * N'$$
$$N' = N'_{PB} + N'_{DR}$$

The cost and float life values that we use for lead acid batteries are shown in Table 3 and we assume a datacenter lifetime of 12 years.

Configuration	Power Cap (f)	#Batt. for PB (N_{PB})	#Batt. for DR (N_{DR})	DR Damage (D_{DR}) for 4 yrs	TCO($) l=12 yrs	Prob. of Handling Outage 1 min	3 min	5 min
Baseline	N/A	33		N/A	59400	1	1	1
Shared	30%	33		0.202	59400	1	1	1
	35%	33		0.592	59400	1	1	1
	40%	33		1.819	108084	0.998	0.997	0.994
	30%	9		0.257	16200	0.198	0.15	0.128
	35%	12		0.657	21600	0.339	0.297	0.273
	40%	18		1.905	61722	0.823	0.769	0.721
Hard-partitioned	30%	33	9	0.257	75600	1	1	1
	35%	33	12	0.657	81000	1	1	1
	40%	33	18	1.905	121122	1	1	1
Soft-partitioned	30%	24	9	0.257	59400	1	1	1
	35%	21	12	0.657	59400	0.999	0.999	0.995
	40%	15	18	1.905	88722	0.988	0.961	0.947

Table 2: Comparison for Lead Acid Battery Configurations. NOTE that the TCO is only capturing battery related costs and does NOT include any savings, whether in cap-ex or op-ex due to peak capping achieved by Demand Response.

Results: Table 2 compares different configurations, together with the number of batteries employed, in terms of the TCO and ability to handle a power outage. The baseline represents today's configuration, where batteries are only provisioned for PB (33 unit batteries are needed for the peak power of the rack).

For shared usage, we consider two sets of configurations. The first set is based on capacity for power outages, i.e., for the peak power demand (33 unit batteries). The second set is based on capacity for demand response with different power caps, i.e., only for the portions above the power cap). Even though the second set of configurations significantly reduce the TCO, probabilities for handling power outages are unacceptably low mainly due to not enough power capacity during these outages. On the other hand, the energy capacity (33 KWhr) that we get for free in the first set, due to being sized for the peak power, is more than adequate for the energy needs during outages (needed only for the transition time). Even at 40% power cap, we have a probability close to 1.0 of handling outages.

Even though hard-partitioned configurations ensure battery capacities needed to always handle power outages, the separate sets of batteries introduce significantly additional costs compared to the baseline.

The Soft-partitioned configurations allow the PB needs to be fulfilled by tapping into DR batteries, when the former runs out of capacity. This allows some under-provisioning of the former, with a consequent attractive TCO at higher power capping levels. This comes at only a slight loss in the ability to handle a outage, with the probability dropping slightly below one. These configurations, thus, seem to be a better alternative than the hard-partitioned ones.

While the batteries serve a lifetime of T_{max} years upto 35% power cap, going to a 40% cap leads to frequent replacement, negating even the benefits of provisioning fewer batteries across the different configurations. This makes it imperative to consider the battery wear effect of DR when provisioning. Finally, note that though most of the configurations have equal or higher TCOs compared to Baseline, Cap-ex and Op-ex savings due to demand response may offset these additional cost as shown in prior work [5, 12, 4, 6].

3 Different ESD Technologies for These 2 Purposes?

The previous results showed that a shared set of batteries, that is provisioned to handle the power needs during an outage (for PB) is able to easily handle the energy needs of DR as well. Consequently, the shared configurations where the same batteries serve both PB and DR purposes, appear to be the best option if batteries were the only ESD technology option. However, there is a plethora of ESD options, with different trade-offs in terms of delivering energy and power efficiencies that are typically captured by a Ragone plot [12]. Also, some of these, especially ultra-capacitors [2] and flywheels [10], are also making their way into datacenters. Given the diverse needs of PB and DR, should we consider different ESD technologies for the two? Especially given the high power needs of PB, should we consider power efficient options such as Ultracapacitors (UC) and Flywheels (FW) for PB, while continuing to use (lead-acid) batteries for DR?

We consider such heterogeneous/hybrid technology configurations, called "Hybrid-Battery+FW" and "Hybrid-Battery+UC". In these configurations, we size

ESD	Batt.	Flywheel	UltraCap.
$/KW	125	250	100
$/KWhr	200	5000	10000
Float Life (yrs)	4	12	12

Table 3: Cost and Float life of different ESDs

FW/UC ESDs based on the maximum power needs to support an entire 5 minutes (maximum time needed to transition to Diesel Generators) of operation for PB. Table 3 summarizes the important relevant parameters for these two technologies, and the reader is referred to [12] for more details on these ESD technologies. The batteries in these configurations are sized for DR as explained earlier in the hard-partitioned options for DR.

Figure 3 compares TCOs of different configurations with different DR power caps. Heterogeneous/Hybrid configurations ("Hybrid-Battery+FW" and "Hybrid-Battery+UC") reduce TCOs significantly (by more than half for 30% and 35% power caps), compared to the most cost effective homogeneous options. On the other hand, "Hybrid-Battery+FW" and "Hybrid-Battery+UC" do not compromise on the ability to handle a power outage at all, since the two purposes are clearly separated and handled by different technologies.

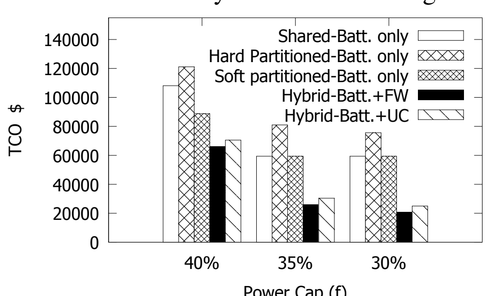

Figure 3: TCO Comparison between Homogeneous and Heterogeneous ESDs. "Hybrid-Battery+FW": Flywheel for power outage, battery for demand response. "Hybrid-Battery+UC": ultracapacitor for power outage, battery for demand response.

4 Summary

In this paper, we have examined provisioning energy storage for two important purposes - Power Backup (PB) and Demand Response (DR) - in the datacenter. These two purposes have very different characteristics in terms of their needs (power vs. energy), mandates (hard guarantees vs. best effort), and frequency of usage (a few times a year vs. multiple times in a day). Using actual power traces of a datacenter, we have evaluated different configurations that consider dual purposing batteries for both PB and DR, having separate sets of batteries, and soft-partitioning them. We have also considered under-provisioning the batteries, and the effect of battery wear on overall-TCO. The study gives the following key insights that are useful for a datacenter designer:

- Despite the differences, dual-purposing is a better option than maintaining separate sets of batteries, when batteries are the only choice.
- This is mainly because (peak) power demands to handle an outage for PB really dominates when provisioning. The consequent energy, that we literally get for free, is more than adequate for power capping in DR.
- Such dual-purpose provisioning does not compromise on the ability to handle a power outage before the load can be transferred to Diesel Generators.
- Given the non-orthogonality between the power and energy dimensions of batteries, and the different needs of PB and DR, hybrid ESD technologies which use a power-efficient option (such as Flywheels and Ultracapacitors) for PB, and use batteries just for DR, are much more cost-effective than any battery-only solution. They can reduce TCO by half compared to the dual-purposed battery-only options that have been studied until now, without compromising on the ability to handle power outages

5 Acknowledgements

This work was supported, in part, by NSF grant 1302225 and a research gift from Google. We would also like to thank Sriram Govindan from Cloud Server Infrastructure Team in Microsoft for his help in collecting datacenter power traces.

References

[1] APC White paper: Comparing UPS System Design Configurations, 2008.

[2] APC Whitepaper: Comparing Data Center Batteries, Flywheels, and Ultracapacitors, 2011.

[3] S. D. Downing and D. F. Socie. Simple rainflow counting algorithms. *International Journal of Fatigue*, 4, 1982.

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

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

[6] V. Kontorinis, L. E. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. P. tis, D. M. Tullsen, and T. S. Rosing. Managing Distributed UPS Energy for Effective Power Capping in Data Centers. In *Proceedings of ISCA*, 2012.

[7] J. F. Manwell and J. G. McGowan. Lead acid battery storage model for hybrid energy systems. *Solar Energy*, 50(5):399, 1993.

[8] National Survey of Datacenter Outages, 2010. http://www.inquirere.com/wp-content/uploads/2010/11/National-Survey-on-Data-Center-Outages.pdf.

[9] The Cost of Power Disturb. to U.S. Businesses, 2001. http://www.onpower.com/pdf/EPRICostOfPowerProblems.pdf.

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

[11] D. Wang, C. Ren, S. Govindan, A. Sivasubramaniam, B. Urgaonkar, A. Kansal, and K. Vaid. ACE: Abstracting, Characterizing and Exploiting Datacenter Power Demands. In *Proceedings of IISWC*, 2013.

[12] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy. Energy storage in datacenters: what, where, and how much? In *Proceedings of SIGMETRICS*, 2012.