

Towards Energy Simulations for Proportionally Designed and Controlled Data Centers

Authors List (Hidden for Review)

Institution (Hidden for Review)

Department (Hidden for Review)

Abstract

Data centers are the backbone of today's internet based economy. From an energy perspective, in the U.S data centers consume around 2% of the total energy produced, with 10 times or more power density than conventional office buildings. Through the COVID-19 pandemic, data centers across the globe have seen a disruptive shift in traffic and workload patterns. Given this timely example, it is warranted that data center operators be able to simulate their facilities to verify that their infrastructure can handle these types of shifts in traffic. This research proposes that dynamic computational workloads of data centers can be sufficiently and accurately characterized with current building energy modeling tools to allow such simulations.

The methodology presented in this article couples a network model indicative of internet traffic with the building power load profiles. The effectiveness of such a model is demonstrated by extending the model to five data centers dispersed globally and comparing their results. The framework can be used by architects and designers to assess building energy models of data centers and to provide an indicator of opportunistic thermal headroom during non-design day conditions.

Introduction

In 2020 U.S data centers are projected to consume 73 billion kWh [Shehabi, 2016]. This value is significantly curtailed from the trends documented in a 2007 Report to Congress [Kooimey, 2007]. The report provided evidence of unsustainable energy demands by data centers as the industry expanded, based-on the industry's growth observed from the early 2000s. Kooimey's report was a medium that spread awareness of the problem. Since then various opportunities for energy reduction have been identified and implemented by IT equipment and facility systems architects.

Two specific engineering choices have had profound benefits towards data center energy efficiency. First, proportional power and thermal control of data center components is now enabled across all dimensional

scales. An example of a millimeter scale proportional control is of central processing units (CPU) with dynamic frequency voltage scaling (DFVS) [O'Sullivan et al., 2015], [Barroso et al., 2018], [Joshi and Kumar, 2012].

To appreciate the DFVS correlation across the distinct technical domains of computer systems and buildings it should be noted that legacy hardware operations did not proportionally utilize power or cooling with their information technology (IT) workloads. Most intuitively, this meant that an IT device consumed nearly the same amount of power whether it was idle or being 100% utilized. The latest generation of IT equipment have much better proportional control enabled by variable fan speeds, variable power draw, and variable heat dissipation resulting from workload dependent power draw enabled with DFVS.

Second, power and thermal constraints have been relaxed compared to legacy practices. As an example, led by hyper-scale early adopters like Facebook, dual conversion power typologies are now being replaced with bypass systems [Park, 2015]. The removal of dual conversion power systems not only saves capital costs associated with complex uninterruptible power distributions, but it also saves operational expense by requiring power rectification at the IT point of connection only.

Omission of dual conversion power distribution systems is one example that has been enabled by deep integration across IT and building systems power architectures. Another example is the collaboration between thermal architects of IT and buildings that led to the expansion of the operational environment's thermal psychometric window for IT equipment. An expanded psychometric window increases the opportunity to exploit free cooling modes across various climate zones [ASHRAE-TC9.9, 2016]. In more aggressive implementations, some data center operators have gone beyond the opportunistic use phase power usage effectiveness optimality and have discarded compressor-based equipment from their cooling plants completely. Compressor-less cooling plants offer significant capital costs savings as-well as the operational energy and maintenance costs [Mulay,

2018].

The power usage effectiveness metric (PUE) as defined by Equation 1 has been the primary focus for data center facilities architects.

$$PUE = \frac{\text{Total Power Used at Facility}}{\text{Power Consumed by IT Equipment}} \quad (1)$$

In operations, the facilities and IT parameters change simultaneously together in a continuous sequence, lending to accurate real time evaluations of the PUE. However, the continuous states of IT workloads and ambient environmental conditions makes design time PUE calculations complicated. The complications drive data center building designers to quantify PUE at discrete step-wise workload values. Like the PUE calculations, data center building energy models are also reasoned about through these coarse step loads. The step-wise inputs are sufficient for two things. First, for sizing equipment which must support the nominal data center capacity in any possible operational condition. Second for producing energy models which are suitable for prescriptive compliance reviews; allowing municipalities to normalize across different data centers in their jurisdiction.

However, without accounting for continuous workload profiles these coarse models mis-represent the part-load operations of the data center and inaccurately quantify total energy demands over a time period. The mis-representation has significant impacts throughout the life of the hardware deployed in these data center facilities. For example without realistic coincident alignment of workloads and ambient conditions, the thermoelectrical equipment whose energy use is dependent on workloads and ambient temperatures may be over-sized for significant portion of their operational times. More explicitly, in conditions where the ambient enthalpy is lower then the design day, the additional capacity of the cooling equipment are stranded.

The proposed simulation in this research will allow for a more realistic evaluation of data center energy use; in-turn exposing the facility's power and cooling equipment headroom for non-design day conditions. These evaluations can facilitate data center operators to opportunistically oversubscribe the IT loads as an extension to the electrical over-subscription schemes presented by Li [Li et al., 2018].

The contribution of this work is an IT load aware building energy model that shows how a dynamic time-series of IT power can be incorporated with EnergyPlus. This method can further be expanded by exploiting the real-time headroom at coincident IT load and environmental conditions vs. the design-day capacities of the facilities equipment for opportunistically increasing the supportable IT workloads that are constrained by power; leading to free capacity.

As an example of free capacity, consider a design day dry-bulb temperature of 110°F. When a peak

in IT workload occurs on a more favorable day, say 95°F, the computer room air conditioner heat rejection equipment needs less than 50% of the designed power value [Liebert, 2020]. With a slight modification in the electrical infrastructure, this difference can opportunistically be allocated to IT workloads. However slight the the electrical modification, the decision is most cost effectively made during the design phase of the the facility.

This article first provides a background for the need of IT coupled data center building energy models and discusses similar works. Then a novel simulation methodology is proposed for extending the existing data center modeling capabilities of EnergyPlus to allow for dynamic IT workload profiles in it. Using the proposed method, two simulations are performed and their comparisons are discussed. The first of the two models serves as a baseline, where the peak IT load is constant throughout the year. In the second simulation, a dynamic IT workload is expressed. A comparison of the results are then presented in the Results from Proposed Method section. The article concludes with a summary of the work and outlines future research that leans on this model.

Background

Over the last 15 years, the deployment paradigm of data centers has transformed from monolithic deployments of high-end devices to heterogeneous deployments of cost-efficient commodity equipment that dwarf the their predecessor in all dimensional scales. With an appropriate title, in Warehouse Scale Computer Barroso et al. describe their data center design and operational experiences in this new paradigm [Barroso et al., 2018]. The work provides details from the software platforms to the industrial sized power and cooling plants found at mega-Watt scale data center campuses. This work provides compelling financial incentives for designs and operations that prioritize proportional energy use at these massive scales. Furthermore, Barroso provides explicit workload profiles from real data centers that lend themselves to opportunistically oversubscribe the capacity of the mechanical and electrical infrastructure in proportion to the IT workloads.

It is evident by the preceding discussions that energy efficiency of data centers is dependent on the proportional power management across all components found in them. However, the PUE metric presented above does not capture the end to end electro-mechanical efficiency of data centers as it misses the IT inefficiencies. For IT devices, several methods for enhancing the proportional power management have been summarized by O'Sullivan [O'Sullivan et al., 2015]. O'Sullivan presents design solutions for those methods and demonstrates that through workload proportional computing, it is technically feasible to optimize the end to end electro-mechanical efficiency

of the hardware. The control feedback loops that O'Sullivan describes also lend themselves to be leveraged by building systems. As an example, return air temperature readings at air handlers can be supplemented with input power demand values to more effectively control air handling fan speeds and their cooling stages.

Awareness of the total power demands jointly by the building system's and the IT telemetry leads to another area of integration, power capping. Power capping throttles the CPU power cycles, similar to load shedding schemes implemented by power utility companies for consumer energy use. A novel power capping strategy is proposed by Li [Li et al., 2018]. Li's power capping methods provide a means for building systems to actively exploit the device level power management capabilities of Intel's CPU Node Manager feature ([Intel, 2015]. Intel's CPU Node Manager enables the data center infrastructure telemetry to observe building telemetry (ie from power meters on rack power strips) and throttle the CPU when the workloads demands encroach the facilities power, risking an upstream circuit-breaker trip.

Given the various operational interdependence between building and IT systems, traditional building energy modeling techniques have failed to suffice for data centers [Beatty, 2015]. Beatty's critique resonates with the three influential data center energy evaluation methods. Those methods either are directly enforced or adopted through reference by authorities having jurisdiction across the US.

The first and most influential method is ASHRAE 90.4, the Energy Standard for Data Centers (ASHRAE, 2019). Compliance with the standard can be achieved through prescriptive or performance-based methods. EnergyPlus simulations are one of the approved methods for performance-based compliance along with spreadsheet-based methods. Regardless of the method, the standard allows various subsystems to be isolated for compliance based on whether they are mechanical or electrical load components (MLC/ELC) which fall back on prescriptively defined threshold values.

The second influential method is a part of the certification process for the US Green Building Council's LEED program. LEED requires prescriptively defined performance calculations to validate compliance [USGBC, 2016]. Compliance with the LEED method is done through a spreadsheet tool, where designers input their data center IT characteristics along with the properties of the IT power distribution system. The strict prescriptive boundary set by LEED accounts for the electro-mechanical efficiency at discrete design points of specific devices but does not account for the computational workloads. The computational workloads are known to significantly influence IT equipment power draw [Marculescu, 2019].

The third method, adopted by many power utility

companies for compliance with their rebate program comes from the Pacific Gas and Electric Company (PGE). PGE method mandates an documented IT load profile to indicate the part load conditions, where the profile is established with 1-month of trend data for new construction. In industrial practice however, it takes several years to fully populate a data center with equipment and data center workloads, like the weather, have seasonal periods of cyclic behavior [Zhuang et al., 2015].

Similar Works

As described above, PUE is not a full proof metric for data center energy efficiency. By acknowledging that PUE does not account for IT equipment efficiency, Perolini provides a coordinated control strategy across IT workload and building infrastructure based on network queuing theory [Parolini et al., 2012]. Perolini logically composes a set of IT systems and supporting infrastructure as a singular cyber-physical system network graph. Hierarchically structured quality of service (QoS) values is common among Perolini and Li [Li et al., 2018]. Both dynamically tune IT with the facilities infrastructure systems. The fundamental difference between these two works is that Li considers the IT workload as the controlled variable bound by the facility's systems and Perolini considers the cooling equipment as the controlled variable bound by the IT workload. In their recognition of some data center workloads as being preemption tolerant and others as time sensitive, these methods allow over-subscription of both power and thermal capacities without any service level performance degradation.

Data center cooling systems have been algorithmically coupled with data center IT workloads by Wei in [Wei et al., 2017]. They use a reinforcement learning agent that is aware of the data center's thermal state in real-time. The agent acts as a controller of the HVAC equipment and respective IT traffic balancer, where the balancer migrates work between two or more data centers. In the work they demonstrated that when given a state-action pair and an objective function; the agent takes a control action in a given state to reduce power while allowing the IT system to meet it's performance objective in terms of processing latency. The state-action pairs are optimized with an objective function that values the power demand along with a IT performance metric at a given time-step. Wei's work was limited to data centers located in mixed use building, where the cooling system had some fungibility with other parts of the building, which allowed waste energy from the data center to be reclaimed to heat the other parts. Nonetheless, their work is similar to this research in that it also integrated EnergyPlus together with Python.

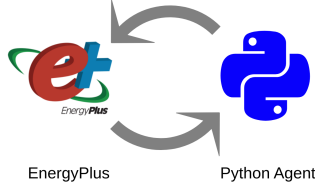


Figure 1: Dynamic Process Flow between EnergyPlus and the Python Agent.

Methodology

In this section a dynamic workload-profile based energy simulator is introduced and a simulation is described in detail. The simulator allows Python algorithms to interact with EnergyPlus through the Building Controls Virtual Test Bed (BCVTB) [Wetter and Noudui, 2016]. The BCVTB interface from Gym-EPlus [Zhang et al., 2019] is used to implement the interaction. Functionally, the Python agent handles programmatic logic about the IT workloads. Python is chosen for the interface function due to its flexibility and the granular level of program control and EnergyPlus is chosen due to its established efficacy with building modeling.

EnergyPlus is the US Department of Energy's state-of-the-art building modeling software. It features an extensible physics-based building thermodynamic and psychrometric modeling engine to simulate sub-hourly energy demands and occupant comfort values. Data-centers have been explicitly supported in EnergyPlus since release of Version 8.3.0. As shown in Figure 1, the proposed simulator uses Python as an external interface to reset the Data center load for each time-step. In the subsequent modeling time-step, the EnergyPlus models runs with the externally revised variables. The basic process loop of the simulator is illustrated in Figure 1.

The data center building is configured in EnergyPlus with an IT load density of $2 \frac{kW}{m^2}$. The remainder of the model is based on Moriyama's data center model from [Moriyama et al., 2018]. Moriyama uses a full suite of mechanical cooling equipment and configures them with the SequentialLoad setting. By using SequentialLoad in the EnergyPlus IDF, each piece of equipment ramps up in the order listed.

Without the Python feedback loop, EnergyPlus is a hermetic suite in which simulation run-parameters are defined in a configuration file; namely the IDF. In the default case, EnergyPlus modeling workflow serially runs the simulations and then users get to review the results of each simulation at completion. Iterating on the parametric values requires editing of the IDF file and re-running of the entire simulation. This proposed simulation specifically defines a loop in Python which iterates over a list of load coefficients derived from the network traffic corresponding to the cooling load for each simulation time step.

The traffic profile of an internet service is an intrinsic

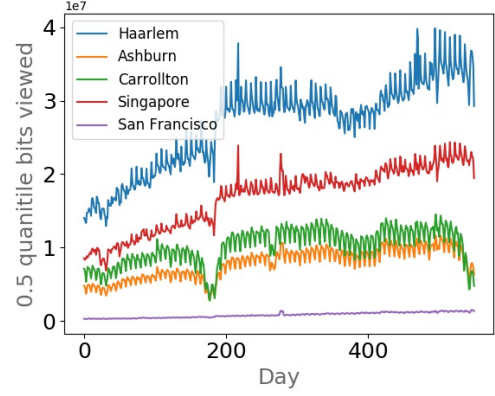


Figure 2: Sum of English, French, German, Mandarin, Spanish, Russian, and Japanese traffic to each of the 5 data center locations.

property of the service. Characterizing the traffic profiles requires granular data that spans across a wide temporal range. However, due to the competitive nature of the internet industry such data that allows for building level traffic profiles is not publicly available. Therefore, in this research a network traffic simulator is created. The objective for the simulator was to represent a globally distributed online platform indicative of real world user demand. To represent a global service, the load factors in this article reflect visits to 145,063 Wikipedia pages over an 18-month period as provided by Kaggle [Kaggle, 2017].

One issue with the Kaggle data set is that it just provides a time series of page visits that is agnostic of source location (ie user location) or serving data center. For evaluating data center level workloads, the pages were mapped to a data center as part of a pre-processing step of this research. The pre-processing required parsing the raw URL strings and parsing out its language hash markers. Then, countries where these languages are the official languages were mapped to the nearest Wikipedia data center to indicate the traffic to each data center. This resulted in the seven languages being routed by a minimum distance function from any country that they are the official language to the nearest data center site. Treating each of the seven languages as a service allows this research to have an abstraction of the (Moriyama, et al., 2018) service which is intuitive to reason about as indicated in Figure 2.

To validate the impacts of resetting the proportional load factors, two specific simulations were run for each of the data centers from Figure 2. The first is the novel dynamic load simulation and second configured as in [Moriyama et al., 2018]. In both simulations nominal power density of the data center is set to $2kW/m^2$ and the auto-sizing feature is enabled for all HVAC equipment.

In the proportional load simulation, the traffic load factor is pulled in by the Python program and at each

time step it is passed to EnergyPlus. The load factor that is passed represents the CPU loading of servers, where CPU loading is a dominate energy consumer of IT equipment but its not the only power consumer. In EnergyPlus IT equipment always consumes a near constant rate of energy for hard disks, random access memory, fans and other parts regardless of the load. This constant rate is approximately 30% of the power density, in line with industry practices [Barroso et al., 2018].

Algorithm 1 indicates how the the service demand profile is generated. It is based on Kaggle’s web-traffic data-set (K) which is a time-series of the count of visits (v) to 145,063 Wikipedia pages distributed across seven languages (l) at daily intervals [Kaggle, 2017]. In this work pages (p) are aggregated together by language and classified as a common service. P_l is a list of all pages with the language l . The subscript d represents individual data centers. The demand for the service at a particular data center is obtained by the average quantile visits to the pages of the respective language as indicated by $P_l[d]$. Figure 2 illustrates the service profile for the data range and notes the seven languages. The y-axis is the number of page views on a $1e^8$ scale, the x-axis spans the date range, and the curves represent each of the seven languages.

Algorithm 1 Service Profile

Require: K

- 1: **for** p in K **do**
 - 2: $P_l = \text{empty}[]$
 - 3: $P_l[l] \leftarrow p_l == l$
 - 4: **for** d in $\text{range}(K_d)$ **do**
 - 5: $P_l[d][\text{quantile}(v)] \leftarrow \sum P_l[l][d]$
 - 6: **end for**
 - 7: **end for**
-

Next we discuss the results from the simulations.

Results

Two simulations for each of the five data center locations from Figure 2 were run. The total annualized energy use of the two models is plotted for comparison in Figure 3. The purple bars indicate the simulations in which the CPU loads are reset at each time-step (RL), and the green bars indicate simulations without interactive resetting of the loads (NoRL). For Singapore (sin), San Francisco (sfo), and Ashburn (ash) the RL simulations result in higher annualized energy compared to the NoRL model. While for Amsterdam (ams) and Carrolton, TX (car), the RL model results in lower energy demand compared to the statically configured load from the NoRL simulation.

The variations in the relative energy demands between the two models can be attributed to the CPU load profiles as illustrated in Figure 4. The figure illustrates time series plot of each of the five data centers analyzed through the RL simulations. The x-

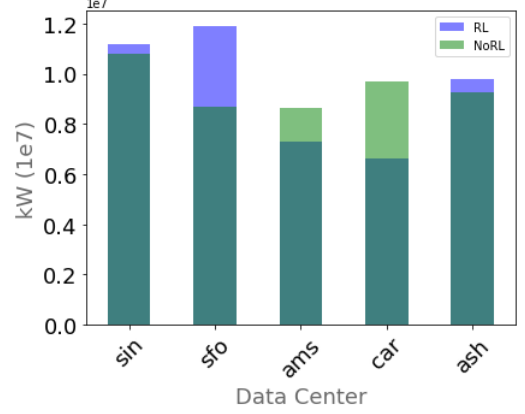


Figure 3: Total Site Energy Difference between statically configured IT load and dynamically reset IT loads.

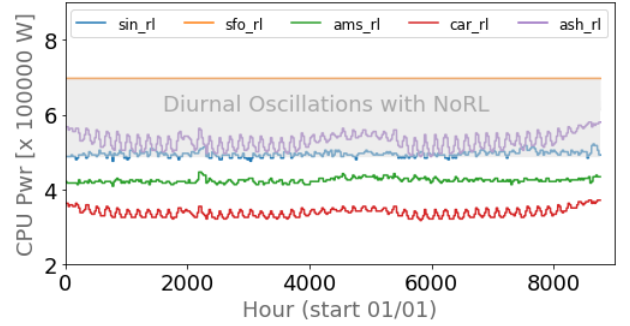


Figure 4: Time series of the total CPU power at each data center (see legend). Gray area indicates the daily demand cycle that the IT equipment creates with NoRL.

axis indicates the hour of the year and the y-axis indicates the CPU load for each respective data center. Examining the profiles for each data center they all appear quite distinct from one another. Carrolton and Ashburn appear to have strong seasonal patterns. Amsterdam and Singapore data centers show irregular oscillations in demand throughout the year. The San Francisco data center has a constant line. This constant demand is due to the load not varying much from the peak capacity in the traffic profile that was used for the site.

In contrast to the distinct CPU load profiles of the RL loads as discussed above, with static schedule configuration of IT load profiles used in the NoRL models, the CPU loads at all the data centers are identical to each other. This configuration of the NoRL models is based on default data center load schedule as used by [Moriyama et al., 2018]. The static property of the NoRL model is also indicated in 4 by the gray band. The gray band is actually composed of line curves that oscillate diurnally.

Given this understanding of the cyclic loads, the relative variance between the RL and NoRL simulations total energy results can be reasoned about by con-

sidering the profile of San Francisco data center as an extreme example. San Francisco's RL model does not vary the CPU load, whereas the NoRL model does. With the constant peak load, the San Francisco data center's RL simulation indicates higher utilization than the NoRL model therefore demanding more power.

Discussion

To reason about the EnergyPlus RL results from Figure 3 for total energy, the San Francisco site again serves as an example. In the figure, the total site energy used for the annual model is 1.19×10^7 kWh. Through the course of 8,760 hours the power demand for the data center is 1,359 kWh on average; translating to an annualized PUE value of approximately 1.36 due to the base IT load density of $2\text{kW}/\text{m}^2$ (with a network load coefficient set to 1 for all hours). The PUE values of the other sites are more complex to reason about without further processing of the data, as they all operate at part load conditions.

In this research the network traffic serves as a proxy for CPU utilization. In practice the network traffic and CPU have complex correlations that can't be generalized with the presented approach. In production data center environments representative correlations can be obtained between network traffic, the IT workloads, the traffic instantiate, and the workload's CPU demand. However, proving such correlations is out of scope of research.

This research assumes that all the facilities are of the same size in terms of power. From looking at Figure 2, San Francisco site has a much lower traffic profile than Haarlem. To claim that the same power capacity handles this range of traffic another assumption is made. This assumption is that the CPU-Network ratio varies from site to site. For example a bit (digital bit) of network traffic generates 40 times as much CPU workload in San Francisco as it does in Haarlem. This is a realistic condition in data center fleets were have different generations of IT hardware in operations at various data centers. Or this condition can be encountered when software optimizations to specific workloads are done for targeted consumer markets that are geographical segregated. Nonetheless, since the network coefficient is bound between 0 and 1, and network-workload-CPU model can be supported in the same framework.

Conclusion

There are two notable benefits of resetting the IT load values outside of the IDF file. First, defining load values for each hour in the IDF file is toilsome and error prone. The toil of entering values for each simulation time-step would require 8,760 entries in the file. This increases the chance of introducing errors into the simulation. Second, the external definition allows more sophisticated logic to control the time-step

variables. In this article, only the IT load factors were reset, none-the-less the same interface can reset many other simulation run-parameters with similar logic. An example of a feedback logic is reinforcement learning, which can be framed to globally optimize the systems and is an topic for future research. The traffic profile used in this article are representative of a globally distributed services user facing workloads and suffices for purposes of demonstrating the dynamics of data center operational loads. Each real-world service will have a unique profile resulting from a combination of critical and opportunistic back-end workloads. The modular construction of this model is conducive for incorporating the other workloads and being inclusive of back-end workloads.

References

- Barroso, L. A., U. Hozle, and P. Ranganathan (2018). *The Datacenter as a Computer: Designing Warehouse-Scale Machines 3rd ed.* Morgan and Claypool.
- Beatty, D. (2015). *Gaps in Modeling Data Center Energy.* ASHRAE Journal.
- Intel (2015). *Intelligent Power Node Manager 3.30 External Interface Using IPMI Document Number 332200-001US.* <https://www.intel.com/content/dam/www/public/us/en/documents/technical-specifications/intel-power-node-manager-v3-spec.pdf>.
- Joshi, Y. and P. Kumar (2012). *Energy Efficient Thermal Management of Data Centers.* Springer-Verlag New York.
- Kaggle (2017). *Web Traffic Time Series Forecasting.* Retrieved: January 08, 2020.
- Koomey, J. (2007). *Estimating Total Power Consumption by Servers in the US and the World.* Berkeley, CA: US Department of Energy, Lawrence Berkeley National Labs.
- Li, Y., C. Lefurgy, M. Rajamani K., Allenware, G. Silva, S. Ghose, and D. Heimsoth (2018). *CapMastro: Exploiting Power Redundancy, Data Center-Wide Priorities, and Strained Power for Boosting Data Center Performance.* Austin, TX: IBM Research Division.
- Liebert (2020). *Liebert DS Thermal Management System,* Retrieved: May 14, 2020. https://www.vertiv.com/globalassets/products/thermal-management/roomcooling/liebertds28105kw830-tonssystemdesignmanual_01.pdf.
- Marculescu, D. (2019). *EMC2 Workshop at NeurlPS.* Retrieved: January 05, 2020.

- Moriyama, T., D. Magistris, T. G., P. M., T. H., A. Munawar, and R. Tachibana (2018). Reinforcement learning testbed for power-consumption optimization.
- Mulay, V. (2018). *StatePoint Liquid Cooling system: A new, more efficient way to cool a data center. (Facebook Engineering)* Retrieved: January 02, 2020. <https://engineering.fb.com/data-center-engineering/statepoint-liquid-cooling/>.
- O’Sullivan, N., U. Ammu, and M. Yazdani (2015). Energy proportional power management for data center applications. *DesignCon*.
- Park, J. (2015). *Data Center Handbook, First Edition*. Menlo Park, CA: John Wiley & Sons.
- Parolini, L., B. Sinopoli, K. Bruce, and Z. Wang (2012). A cyber-physical systems approach to data center modeling and control for energy efficiency. *Proceedings of the IEEE* 100(1), 254–268.
- Shehabi, A. (2016). United States Data Center Energy Usage, United States Department of Energy. Technical report, Berkeley, CA.
- USGBC, U. G. C. (2016). *Minimum Energy Performance Data Center Calculator for LEED Certification*.
- Wei, T., S. Ren, and Q. Zhu (2017). Deep reinforcement learning for joint datacenter and hvac load control in distributed mixed-use buildings. *IEEE Transactions on Sustainable Computing*.
- Wetter, M. and T. Noudui (2016). *Building Controls Virtual Test Bed User Manual Version 1*. Berkeley, CA.
- Zhang, Z., A. Chong, Y. Pan, Z. Chenlu, and L. P. Khee (2019). Whole Building Energy Model for HVAC optimal control: A practical framework based on deep reinforcement learning. *Energy and Buildings* 199, 473–490.
- Zhuang, Z., H. Ramachandra, C. Tran, S. Subramaniam, C. Botev, C. Xiong, and B. Sridharan (2015). *Capacity Planning and Headroom Analysis for Taming Database Replication Latency*. ICPE’15 Jan 31-Feb 4.
- ASHRAE-TC9.9 (2016). *Thermal Guidelines for Data Processing Environments, Fourth Edition 2016*, Retrieved: January 02, 2020. <https://www.ashrae.org/File>