Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters

Bilge Acun
acun@meta.com
Meta
USA

Benjamin Lee
leebcc@seas.upenn.edu
University of Pennsylvania, Meta
USA

Fiodar Kazhamiaka
fiodar@stanford.edu
Stanford University
USA

Kiwan Maeng
kwmaeng@meta.com
Meta
USA

Udit Gupta
uditg@meta.com
Harvard University, Meta
USA

Manoj Chakkaravarthy
mchakkar@meta.com
Meta
USA

David Brooks
dbrooks@eecs.harvard.edu
Harvard University, Meta
USA

Carole-Jean Wu
carolejeanwu@meta.com
Meta
USA

ABSTRACT

Technology companies reduce their datacenters' carbon footprint by investing in renewable energy generation and receiving credits from power purchase agreements. Annually, datacenters offset their energy consumption with generation credits (Net Zero). But hourly, datacenters often consume carbon-intensive energy from the grid when carbon-free energy is scarce. Relying on intermittent renewable energy in every hour (24/7) requires a mix of renewable energy from complementary sources, energy storage, and workload scheduling. In this paper, we present the Carbon Explorer framework to analyze the solution space. We use Carbon Explorer to balance trade-offs between operational and embodied carbon, optimizing the mix of solutions for 24/7 carbon-free datacenter operation based on geographic location and workload. Carbon Explorer has been open-sourced at https://github.com/facebookresearch/CarbonExplorer.

CCS CONCEPTS

• **Hardware → Batteries**; **Renewable energy**; **Enterprise level and data centers power issues**.

KEYWORDS

Datacenter carbon footprint optimization, embodied and operational carbon footprint, renewable energy, load shifting, batteries

ACM Reference Format:
Bilge Acun, Benjamin Lee, Fiodar Kazhamiaka, Kiwan Maeng, Udit Gupta, Manoj Chakkaravarthy, David Brooks, and Carole-Jean Wu. 2023. Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '23), March 25–29, 2023, Vancouver, BC, Canada.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3575693.3575754

1 INTRODUCTION

Carbon-free energy is essential for environmental sustainability. The UN's *24/7 Carbon-Free Energy Compact* calls for an ambitious goal: "Every kilowatt-hour of electricity consumption is met with carbon-free electricity sources, every hour of every day, everywhere" [75]. In the US, the Department of Defense and the General Services Administration seek strategies for "supplying 24/7 carbon pollution-free electricity for the federal government" [76]. Computing must do its part to address this societal challenge. Electricity consumed by datacenters is significant and growing rapidly. Datacenters world-wide consumed 205 TWh in 2018 [58], exceeding the annual consumption of countries such as Ireland and Denmark [67]. Information and communication technology may account for 7% to 20% of global electricity demand by 2030 [3, 34].

Net Zero versus 24 / 7. At present, technology companies invest in renewable energy generation to offset datacenter energy consumption [18, 35, 39]. Amazon, Meta, and Google have collectively invested in 22 GW of renewable energy generation to meet Net Zero commitments. These investments align with broader efforts. In the United States, renewable energy generation is projected to increase from 20% in 2020 to 42% by 2050 as the nation pursues Net Zero [73]. Solar and wind comprise 47% and 34% of this renewable energy [77].

Yet 24/7 carbon-free computing remains challenging because renewable energy supply fluctuates. The broad deployment of solar and wind farms will lead to increasingly severe hourly and seasonal supply fluctuations. Figure 1 highlights fluctuations when renewable generation comprises 33% of the total. At times, the grid's supply of renewable energy may exceed demand, forcing inefficient curtailments that deactivate renewable energy generation in order to match supply with demand [6, 10, 11, 48]. At other times, solar and wind energy is scarce.
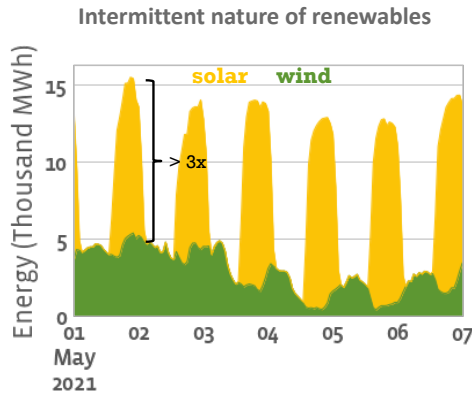
Intermittent nature of renewables

Figure 1: Hourly wind and solar energy generation in California grid during a week of time-frame.

Under these conditions, datacenters consume carbon-intensive energy despite achieving Net Zero. Investments in wind and solar farms issue a renewable energy credit for every MWh generated. Annually, datacenters claim Net Zero by matching credits against energy consumed by their computation. Hourly, however, datacenters continue to emit carbon by consuming carbon-intensive energy (from gas or coal) when carbon-free supply (from wind or solar) is insufficient. Thus, more must be done to eliminate computing's hourly carbon emissions.

Sustainability Design Space. 24/7 carbon-free computing requires re-visiting classic questions in datacenter architecture and management. Architects have studied power infrastructure [20] to maximize performance and utilization subject to power constraints [19, 29, 69]. They have developed schedulers in which jobs request resources and schedulers allocate them [33, 41], perhaps with fairness and sharing incentives [22, 84]. Architects must rethink these frameworks for carbon-free energy.

Architects must first define parameterized solutions for coping with intermittent renewable energy. And they require tools that reveal carbon-efficient combinations of these solutions. One solution is demand response, which schedules computation to align renewable energy supply and datacenter energy demand [5, 64, 79, 80, 86]. A second is energy storage, which reduces exposure to supply fluctuations [50, 54]. Finally, further investment in wind and solar offers more carbon-free energy.

These solutions have implications for infrastructure. Datacenters will need additional servers that perform extra computation when carbon-free energy is abundant [87], batteries much larger than those in today's power infrastructure, and investments in energy generation that reflect the datacenter's location and relative availability of wind and solar. All of this infrastructure incurs embodied carbon costs, an important but under-explored aspect of datacenter sustainability. Architects must pursue coordinated design and management—spanning hardware architecture, datacenter systems, distributed software—to identify the carbon-optimal mix of solutions.

Carbon Explorer. Carbon Explorer enables architects to optimize datacenters for 24/7 carbon-free computing (Section 2). It consumes data about datacenter energy demands and renewable

energy supplies across diverse geographic locations (Section 3). Moreover, it estimates the effect of investments in renewable energy using grid data, in energy storage using physically accurate battery models, and in server capacity needed for demand response scheduling using production datacenter traces (Section 4).

Carbon Explorer takes a holistic approach to datacenter design, balancing reductions in operational carbon against increases in embodied carbon incurred when manufacturing servers, batteries, and wind/solar farms. Although it seeks to maximize the number of hours using carbon-free energy (*i.e.*, coverage), Carbon Explorer occasionally discovers that embodied carbon costs outweigh operational carbon reductions and identifies a solution that does not achieve 100% coverage but is carbon-optimal. Specifically, we report several key findings (Section 5).

- **Site Selection.** Among 13 regions where Meta sites datacenters, Iowa and Nebraska (majorly wind) as well as hybrid regions (wind and solar) such as Texas are best for minimizing carbon costs because their supply valleys (days in the year with lowest levels of renewable energy supply) are shallowest.
- **Renewables Only.** Relying on renewable energy for coverage produces diminishing returns. Datacenters require 5× more renewables to increase coverage from 95% to 99.9% than from 0% to 95%. Accounting for embodied carbon in wind/solar farms, carbon-optimal designs achieve 37% to 97% coverage.
- **With Batteries.** Batteries permit datacenters to reach 100% coverage for four regions and 99% coverage for majority of the rest. Batteries must be large enough for a few hours of computation, but embodied carbon costs can be justified by reductions in operational carbon.
- **With Scheduling.** Demand response increases coverage by 1%–22% depending on the region. Scheduling requires 6% to 76% additional servers to support deferred computation, but their embodied carbon costs can be justified by reductions in operational carbon when 40% of workloads are flexible.
- **All Together.** Demand response reduces battery capacity required and makes 100% coverage optimal for five regions and above 99% for rest of the regions except OR.

To further research in carbon-aware datacenter design and management, the Carbon Explorer framework has been made available at https://github.com/facebookresearch/CarbonExplorer.

2 CARBON EXPLORER

Figure 2 illustrates *Carbon Explorer*, a design space exploration framework that takes a holistic approach to achieve 24/7 carbon-free computing. Carbon Explorer considers two important inputs: time-series data that details the power demand of large-scale datacenters and the intermittent nature of renewable energy generation at specific geographic locations (Figure 2-left). Next, Carbon Explorer characterizes a solution space that spans the following dimensions (Figure 2-center):

- Investments in varied types of renewable energy,
- Investments in varied amounts of energy storage,
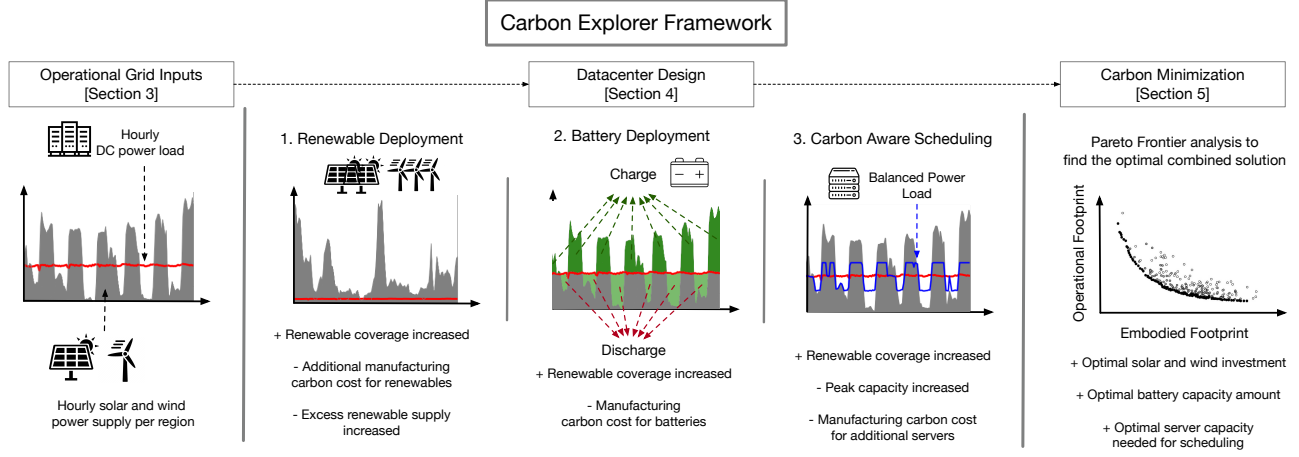- Scheduling that shifts varied amounts of computation.

Figure 2: Design Overview for Carbon Explorer. Carbon Explorer considers important characteristics, such as time-series power demand of large-scale datacenters and renewable energy availability on the power grids, as inputs. Carbon Explorer characterizes the design space across renewable energy investments, energy storage, and computation shifting. Carbon Explorer provides *quantitative measures for strategies to achieve carbon-minimum settings.*

Finally, Carbon Explorer models the datacenter design space and minimizes the carbon footprint, accounting for both operation **and** embodied carbon (Figure 2-right).

Carbon Explorer's inputs include two hourly time series. The first details power consumed by each of Meta's datacenters in various locations across the United States. The second details energy generation from the local grids in each datacenter location. Table 1 summarizes these locations and Meta's renewable energy investments. Section 3 characterizes datacenter energy demand and renewable energy supply, which has implications for a production datacenter's carbon footprint.

We evaluate three distinct solutions for 24/7 carbon-free datacenters. First, datacenters could offset their energy consumption with renewable energy generation. Operators invest in wind and solar farms on the power grids that supply their datacenters. Moreover, they implement power purchase agreements, which issue credits for renewable energy generated from those investments and offset datacenter energy consumed. This state-of-the-art solution has been central to hyperscale datacenters' pursuit of Net Zero goals.

Second, datacenters could install energy storage and batteries to handle the intermittent availability of renewable energy. Although today's datacenters do not yet deploy batteries to manage their operational carbon footprint, they do deploy batteries to ensure system resilience and shave power peaks [40, 56]. As lithium-ion batteries mature, they become cost-effective for deployment at scale. On-site energy storage enables a new strategy for 24/7 carbon-free datacenters.

Finally, datacenters could schedule computation in response to renewable energy supply. Such demand response likely requires investment in additional servers. A datacenter that defers tasks when renewable energy is scarce must compute for those tasks when renewable energy is abundant, generating demand for servers above and beyond typical loads. In effect, bursts of renewable energy generate bursts of computation and demand for servers.

The three solutions lead to trade-offs between operational and embodied carbon footprints. Renewable energy permits carbon-free operation but is constrained by energy availability and consistency, which varies with geography. Large batteries store carbon-free energy but incur carbon overheads from manufacturing. Additional servers permit demand response and scheduling but also incur carbon overheads from manufacturing [30].

Carbon Explorer defines a comprehensive design space for 24/7 carbon-free datacenters. Section 4 navigates trade-offs in the solution space with a quantitative approach. And Section 5 illustrates the solution space for various geographic locations, highlighting the impact of site selection for future datacenters.

3 OPERATIONAL GRID INPUTS: DEMAND AND SUPPLY CHARACTERISTICS

Carbon-aware datacenter design requires understanding datacenter energy demand and renewable energy supply, at fine granularity and in every region. Table 1 lists the locations of Meta's datacenters and renewable energy investment, each identified by the balancing authority for the electric grid [68]. Total investment in renewables is nearly six Gigawatts.

Our energy supply analysis draws data from the U.S. Energy Information Administration (EIA) Hourly Grid Monitor, which provides operating data for power grids in the lower 48 states [78]. Launched in 2019, the monitor provides hourly generation statistics by collecting data from balancing authorities (BAs). Each BA operates a grid and balances electricity flows, controlling electricity generation and transmission within its own region and between neighboring authorities.

3.1 Characterizing Datacenter Power Demand

Meta has built hyperscale datacenters across the globe with different capacities. These datacenters exhibit diurnal load patterns due to variations in user activity and exhibit peaks due to special
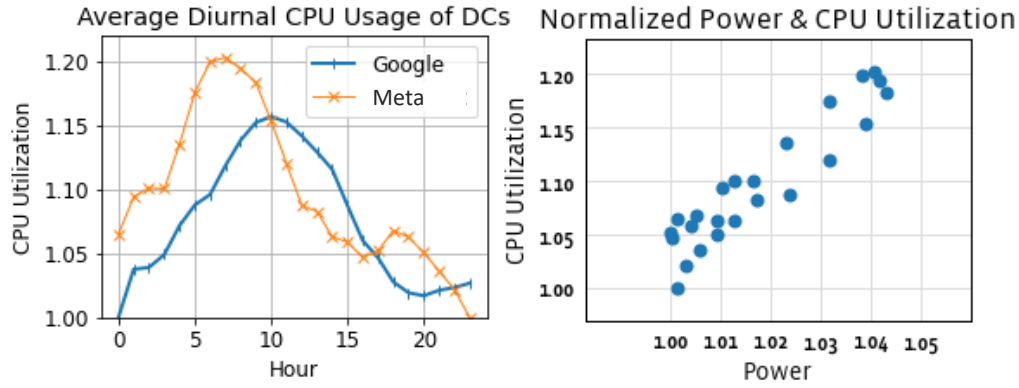
Figure 3: [left] Hourly DC CPU fluctuations of Meta and Google DCs. [right] Hourly CPU Utilization and Power correlation of Meta DCs.

Table 1: Meta's Datacenter Locations in the U.S. and Regional Renewable Investments [68]

Location	Balancing Authority	Renewable Investment [MW]		
		Solar	Wind	Total
1. Sarpy County, Nebraska (NE)	SWPP	0	515	515
2. Prineville, Oregon (OR)	BPAT	100	0	100
3. Eagle Mountain, Utah (UT)	PACE	694	239	933
4. Los Lunas, New Mexico (NM)	PNM	420	215	635
5. Fort Worth, Texas (TX)	ERCO	300	404	704
6. DeKalb, Illinois (IL)	PJM			
7. Henrico, Virginia (VA)	PJM	840	309	1149
8. New Albany, Ohio (OH)	PJM			
9. Forest City, North Carolina (NC)	DUK	410	0	410
10. Altoona, Iowa (IA)	MISO	0	141	141
11. Newton County, Georgia (GA)	SOCO	425	0	425
12. Gallatin, Tennessee (TN)	TVA	742	0	742
13. Huntsville, Alabama (AL)	TVA			
Total		1823	3931	5754

Table 2: Carbon Efficiency of Various Energy Sources

Type	gCO_2eq/kWh	Type	gCO_2eq/kWh
Wind	11	Natural Gas	490
Solar	41	Coal	820
Water	24	Nuclear	12
Oil	650	Other (Biofuels etc.)	230

with distinct service level objectives (SLOs). The highest priority, user-facing services require real-time response. Latency-tolerant workloads, such as batch and AI training jobs [71, 82], target specific SLO categories that include 4-, 8- and 24-hour completion times. Google has reported that flexible jobs with 24-hour completion SLOs make up about 40% of the Borg scheduler's jobs [74]. This flexibility permits carbon-aware workload scheduling.

3.2 Characterizing Datacenter Power Supply

In this section, we present two scenarios that describe how datacenters could consume energy from rapidly evolving power grids. In the first scenario, datacenter operators collaborate with utility providers to invest in renewable energy on the grids that power the datacenters by purchasing energy with sophisticated accounting frameworks that track renewable energy credits. This represent the state-of-the-art in reducing a datacenter's operational carbon footprint for Net Zero commitments [18, 35, 39]. In the second scenario, datacenter operates on renewable energy 24/7 by optimizing hourly supply and demand.

Net Zero. Datacenter operators invest in renewable generation, such as wind and solar, and implement power purchase agreements (PPAs) to reduce datacenter exposure to the grid's carbon intensity. PPAs link renewable energy credits (RECs) with a specific source of energy and issue, e.g., one certificate for every MWh generated [17, 25, 36]. With RECs, the energy consumed is much greener than the energy offered by the grid. Table 2 details the carbon intensity of different electricity sources in the grid. The grid's energy mix is determined by the utility provider's dispatch stack and portfolio of generating assets [9]. But the datacenter's energy mix is determined by its pre-negotiated PPAs, which deliver carbon-free energy.

events and holidays. Figure 3 shows diurnal usage for Meta and Google datacenters and illustrates how power usage correlates with processor utilization. Google analysis is done using the open-source Borg traces [26]. Meta's CPU utilization and power is averaged over three months. For Meta, CPU utilization swings by about 20% for an average datacenter and can swing by even more for an individual datacenter. For Google, the difference between the maximum and the minimum CPU utilization is 15%, on average [74].

However, diurnal patterns from interactive computation do not translate directly into power patterns. At datacenter scale, the difference between maximum and minimum energy demand is around 4%, on average, which is relatively insignificant compared to the swings in renewable energy supply. Thus, in today's datacenters, power variations will arise primarily from supply but not demand. Yet shifting computation to modulate datacenter power is possible because workloads exhibit different flexibility levels and come
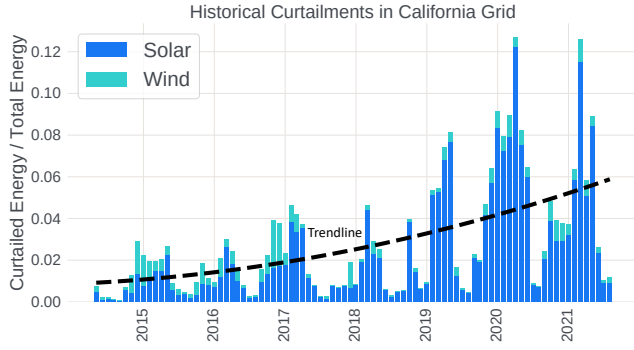
Figure 4: Wind and solar curtailments have been increasing with the renewables on the California grid [6].

With Net Zero matching, at the end of the month (or end of the year), the total amount of energy generated and credits issued is equal or greater than the total amount of energy consumed. Although on an hourly basis, the carbon intensity of the energy used can be as much as the grid's carbon intensity during the times when there is not enough renewable supply [27].

24/7 Carbon Free. In addition to installing renewable energy, 24/7 carbon-free datacenters must address variable, intermittent generation. Figure 5 highlights variability across geography with rows corresponding to three representative regions with distinct renewable energy profiles: (a) Oregon BPAT with wind; (b) North Carolina DUK with solar; (c) Utah PACE with a mix of wind and solar. More broadly, of the ten balancing authorities in Table 1, three offer primarily wind energy (BPAT, MISO, SWPP), three offer primarily solar energy (DUK, SOCO, TVA), and four offer a mix (ERCO, PACE, PJM, PNM).

Figure 5 also highlights variability across time with columns corresponding to summary statistics calculated over the year: (a) Yearly Average; (b) Histogram of Total Daily Generation [78]. On average, wind and solar installations provide significant supply, but averages obscure high variance across time. For BPAT, the best ten days of the year offer approximately 2.5 times more renewable energy than the average whereas the worst offer very little. Histograms quantify this variance and illustrate uncertainty in wind and solar supply.

As solar and wind farms proliferate, peaks and valleys in energy supply will become increasingly extreme. Utility providers will find it increasingly difficult to match its supply to consumer's demand. For example, California's renewable sources can generate much more electricity than needed in the middle of the day [6]. And curtailments are needed to manage excess supply and reduce renewable energy generation [10, 11, 48]. Figure 4 indicates that, since 2015 the curtailed gap between supply and demand has grown steadily as wind and solar capacity has increased. In 2021, curtailments reached 6% of the total generated renewable energy in the California grid, which has deployed a significantly more renewable electricity compared to the U.S. average (33% vs 20% in 2020 [7, 72]).

It is becoming increasingly complicated to fully consume peak renewable energy generation due to the high variance. When supply exceeds demand, only generators with the lowest prices can

supply energy to the grid. Prices can be zero or even negative because inputs to wind/solar farms are free and generators often receive government subsidies [15, 88]. As a result, grids may offer lower time-of-use energy prices and incentivize datacenters to defer computation to periods of abundant renewable energy.

Challenges in variable supply and curtailments require energy storage and demand response scheduling during periods of scarcity. Energy storage mitigates supply variations by providing carbon-free energy when solar and wind cannot [60, 61]. Demand response modulates datacenter demand for energy based on signals about renewable energy supply. Signals could come in the form of utility surcharges or credits when the datacenter consumes or reduces its energy demand during various times of day [55]. Signals can also come from utility providers' generation statistics that describe the mix of green and brown energy across time and geographic locations.

In summary, the Net Zero scenario describes how renewable energy investments significantly reduce the carbon intensity of datacenter operations. And the 24/7 scenario describes how additional investments in energy storage and demand response schedulers could further reduce carbon intensity on an hourly basis. Figure 6 compares the carbon intensity of these scenarios with that of the grid's energy mix.

4 DATACENTER DESIGN: STRATEGIES FOR CARBON FREE COMPUTING

A datacenter must implement a portfolio of complementary solutions to achieve its goal of using 24/7 carbon-free energy efficiently and robustly. Carbon Explorer considers renewable energy investments (Section 4.1), energy storage installations (Section 4.2), and carbon-aware scheduling (Section 4.3). In this section, we model and analyze these solutions and associated trade-offs in operational and embodied carbon footprints.

4.1 Renewable Energy

Carbon Explorer determines the solar and wind investments required for datacenters in different geographic regions to increase and achieve 100% hourly renewable coverage. We define *renewable coverage* as the percentage of hours in the year where datacenter power (P_{DC}) is covered by renewable power (P_{Ren}):

$$\left\{ 1 - \sum_{hour} \{P_{DC} - P_{Ren}\} / \sum_{hour} P_{DC} \right\} \times 100 \; \forall hour \in \text{DateRange}$$

Carbon Explorer projects hourly wind and solar energy supply by scaling EIA grid data in proportion to the desired renewable investment level. It takes the maximum generated solar and wind power throughout the year as the maximum capacity of the local grid. Then, the hourly generation data is linearly scaled to the desired renewable investment capacity. Finally, hourly renewable supply data is matched against hourly datacenter demand for every region to calculate renewable coverage. Figure 7 shows renewable coverage (z-axis) with different wind (x-axis) and solar (y-axis) investments from two regions served primarily by wind and solar, respectively.

Figure 7 reports Meta's existing renewable investments with black lines. While these investments help Meta achieve Net-Zero goals on a monthly or annual basis, coverage on an hourly basis is

a) Yearly Average of Hourly Generation
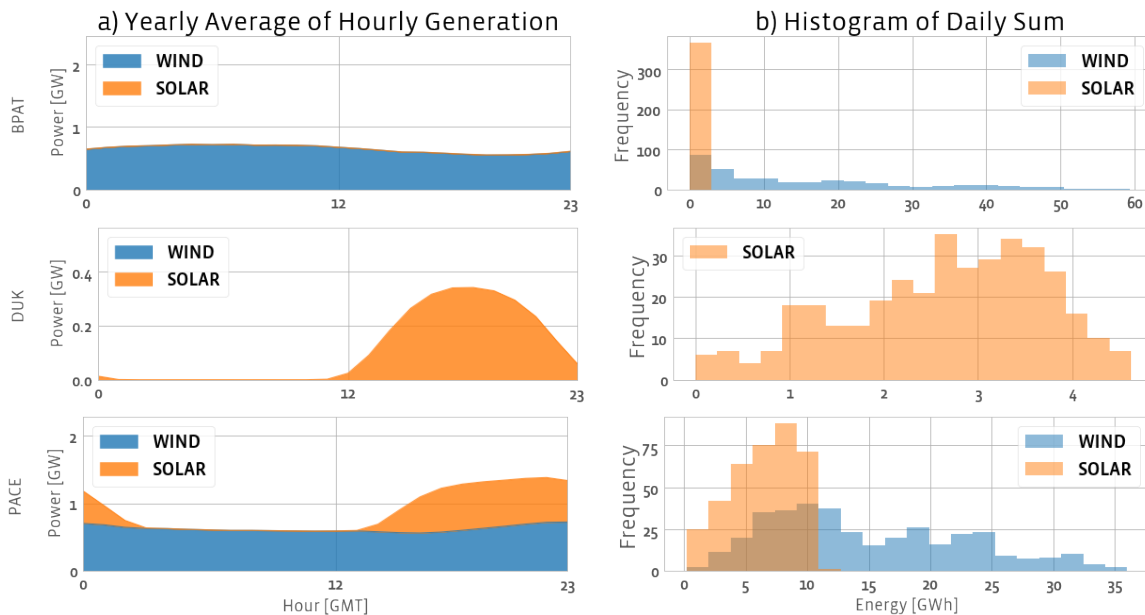
b) Histogram of Daily Sum

Figure 5: Figure shows hourly wind and solar generation of an average day in year 2020 (left) and histogram of daily sum to highlight day-to-day fluctuations of wind and solar generation in BPAT (in OR), DUK (in NC) and PACE (in UT) balancing authorities. The data is calculated over the entire year of 2020.
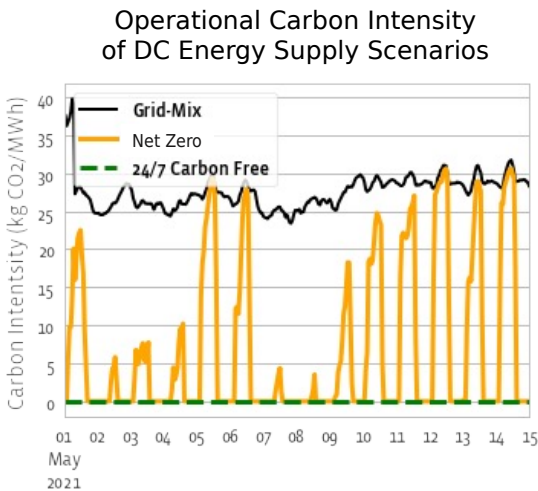

Operational Carbon Intensity of DC Energy Supply Scenarios

Figure 6: Comparison of hourly operational carbon intensity of different DC energy supply scenarios.

only 46% and 51% in the two regions. Each region tends to favor a particular type of renewable energy generation on its local grids and Meta's investments generally align with those profiles. One exception is Oregon, where Meta's investments emphasize solar despite the local grid's emphasis on wind.

For regions served primarily by wind energy, like Oregon, high day-to-day fluctuations increase the investment in wind generation needed to satisfy minimum energy needs. For regions that rely entirely on solar for renewable energy, it is impossible to increase

24/7 coverage much beyond 50% because solar energy is available only during the day. For regions that deploy a mix of solar and wind generated renewables, the tail is shorter and diminishing marginal returns in 24/7 coverage are less severe since wind and solar availability can complement each other.

There is a long tail to reach 100% renewable coverage. As coverage increases, curves flatten and indicate diminishing marginal returns from further investment in renewable generation. Figure 8 highlights the full length of the tail for Oregon's datacenter. It takes more than 5× more investments in renewable energy generation to go from 95% to 99.9% than to go from 0% to 95% coverage. Due to space limitations, we are unable to show profiles of every datacenter location. But this representative analysis shows that other solutions, such as energy storage and carbon-aware scheduling, are essential to complement renewable energy generation.

Note that accurate hourly energy supply data is crucial when making design decisions. Figure 8 shows that assuming wind and solar output to be same as average output every day leads to overly optimistic design conclusions. Under this assumption, achieving 100% coverage would require an order of magnitude less renewable investments. Thus, Carbon Explorer requires fine-grained time series supply and demand data when determining investments in renewable energy generation.

4.2 Battery Storage

The Carbon Explorer framework is designed to include a modular battery model that supports different storage technologies to be added through a simple API. In this work, we analyze lithium-ion batteries (LIB) as improvements in energy storage over the
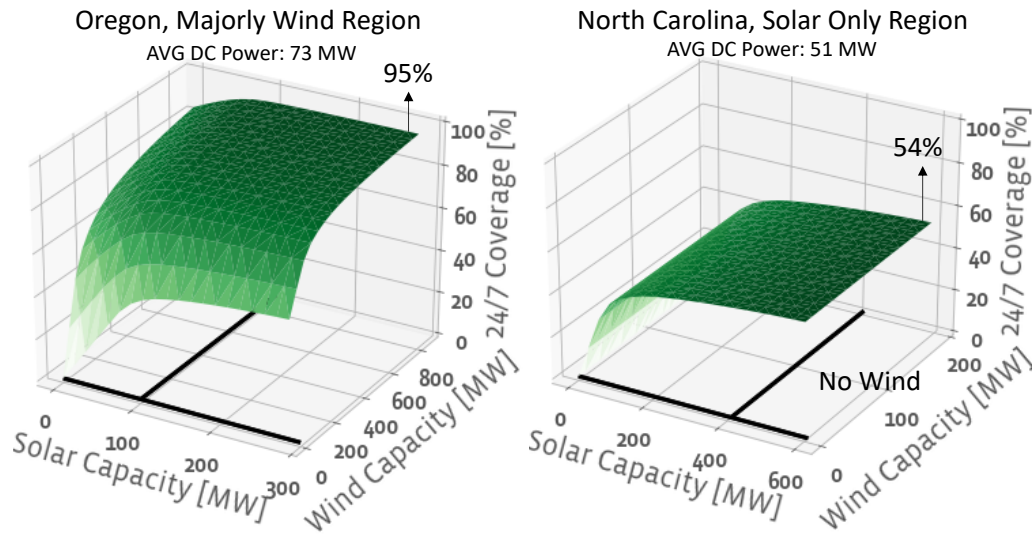
Figure 7: 24/7 coverage with varying amount of wind and solar investments. Black lines show Meta's renewable investment amount in the corresponding region.

last decade have led to LIB to offer high capacity and energy density [13]. As the technology has matured, LIB has become a common, cost-effective storage medium for renewable energy [83]. There is also emerging technologies such as sodium-ion (Na+) batteries, for which materials are easier to obtain and come with lower environmental impact [62]. For these reasons, batteries play an important role in 24/7 carbon-free computing. This section describes how Carbon Explorer evaluates the impact of batteries that are charged by renewable energy and discharged by datacenter servers.

Datacenters already deploy batteries to prevent the interruption of services during maintenance or power failures. Batteries

Figure 9: How much battery needs to be deployed for 24/7 renewable energy?

distributed throughout racks and clusters permit continuous operation when utility power fails and the datacenter must switch to diesel generator backups [57].

We envision batteries deployed on-site with the datacenter to reduce its carbon footprint. Batteries will be charged when there is excess renewable supply (i.e. when the amount of energy produced by the renewable deployment is larger than datacenter's demand). Batteries will be discharged to power the datacenter when there is a
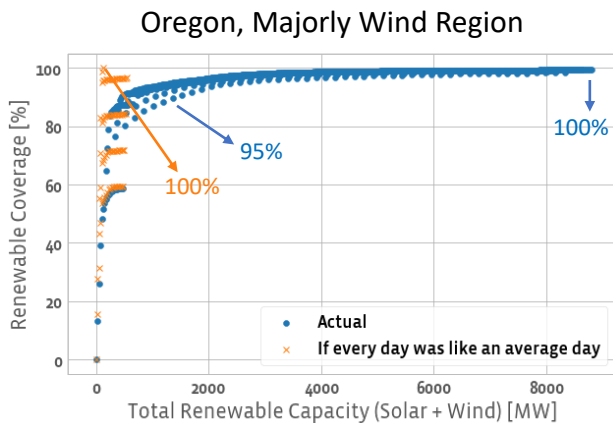

Figure 8: 24/7 coverage with different renewable investments highlighting the long tail. Each point represents a different solar + wind capacity combination.
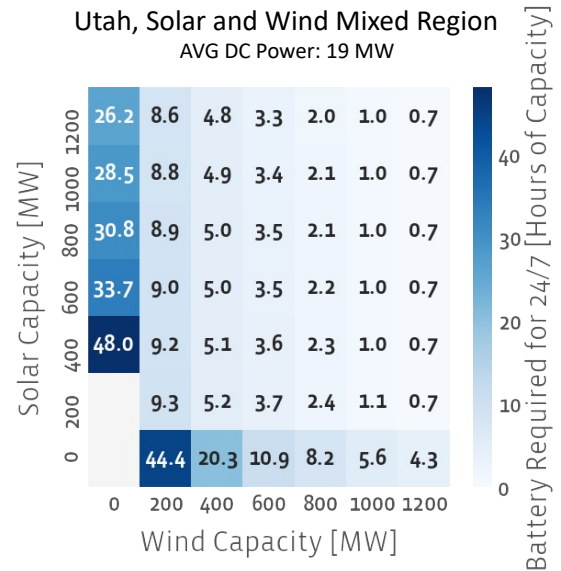
lack of renewable supply (i.e. when the amount of energy produced by the renewable deployment is smaller than datacenter's demand).

The battery model used in Carbon Explorer is the C/L/C model [38]; it explicitly models several characteristics of lithium-ion batteries, including energy content limits, efficiency loss, limits on the applied power with respect to the energy content, and linear charging/discharging rates with respect to battery capacity on minute basis. The model also includes a parameter for controlling the depth of discharge (DoD). Model parameters are tuned to represent a battery composed of Lithium Iron Phosphate cells (LFP) [1] – a cell type found often in large stationary storage applications.

Figure 9 shows the amount of energy storage capacity required to reach 24/7 renewable energy coverage at different solar and wind capacities for Utah datacenter. Capacity is reported in terms of computational hours (e.g., 2 hours for a 20MW datacenter corresponds to 40MWh of battery capacity). Regions with mixed solar and wind generation exhibit less variable, day-to-day fluctuations and can achieve 24/7 carbon-free compute with less battery capacity. By adding around five hours of battery capacity to its existing renewable investments in the region, Meta's Utah datacenter can reach 24/7 carbon-free operational energy. A battery of this size would be comparable to a utility-scale battery; the largest utility-scale energy storage project so far can offer 300 MW of power and 1,200-MWh of capacity [14].

In contrast, battery capacity requirements for 24/7 are greatest for regions that rely majorly on wind. Oregon suffers from extremely high day-to-day fluctuations and there are days with almost no wind power. Requirements are also high for regions that rely entirely on solar. For example, North Carolina datacenter requires 14 hours of battery-based compute.

4.3 Carbon Aware Scheduling

Carbon-aware scheduling (CAS) exploits delay tolerant workloads to achieve 24/7 carbon-free computing, shifting workloads from times when the carbon intensity of electricity sources is high to times when it is low. Hyperscale datacenter workloads are commonly organized into tiers based on their Service Level Agreements (SLAs). Higher tier jobs are latency sensitive and require high availability.

On the other hand, lower tiers can tolerate delays. Examples temporally flexible workloads include AI model training, data processing pipelines, and offline video processing. Google traces indicate a significant fraction of jobs submitted to the Borg scheduler are in the free and best-effort-batch tiers with weak SLAs [74]. Jobs at Meta exhibit similar characteristics. For example, the offline data processing workloads constitute about 7.5% of all the workloads in the fleet. Of these, about 87.4% of the workloads have SLOs that are greater than 4-hours with a majority having 24-hour SLOs. This provides great flexibility in workload time shifting to optimize carbon.

Carbon Explorer estimates the potential benefits of carbon aware workload scheduling using a greedy algorithm. The algorithm takes two customizable input constraints: datacenter capacity and flexible workload ratio for each hour of the day. Given these two constraints, flexible workloads are moved from times of highest carbon intensity
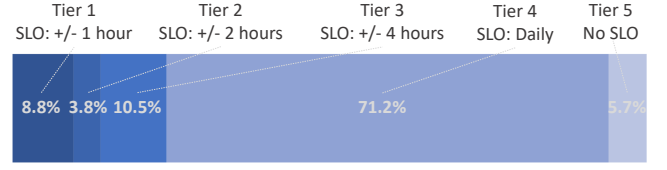

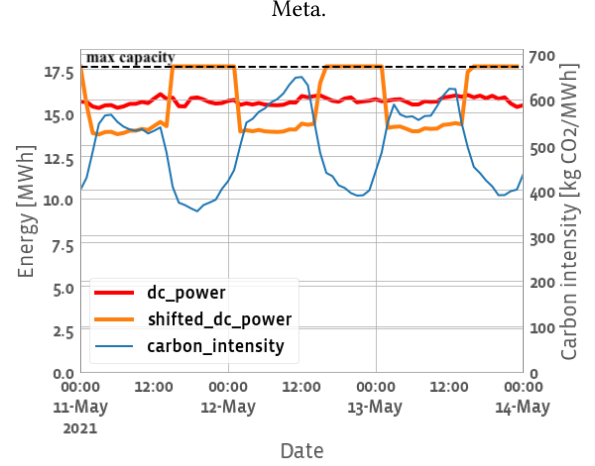Figure 10: Breakdown of data processing workloads by completion time SLO at Meta.

Figure 11: Carbon aware scheduling illustration for the Utah DC.

to times of lowest intensity until all flexible workloads have been moved or all datacenter servers have been used for the given hour.

Input Constraint 1: $P_{DC_{MAX}} = Maximum\ DC\ Capacity$
Input Constraint 2: $FWR = Flexible\ Workload\ Ratio\ (\%)$
Goal: For each day, minimize:

$$\sum_{h \in hour} \{P_{DC}(h) - P_{Ren}(h)\}$$

where

$P_{DC}(h) < P_{DC_{MAX}}$ and
$P_{DC}(h) \times FWR$ is allowed to shift

Figure 11 illustrates an example of a carbon aware scheduling over three days. The blue line shows how the grid's carbon intensity varies depending on the hour of the day. The red and orange lines shows datacenter power draw when carbon aware scheduling is and is not applied. In this example, the maximum allowed power capacity of the DC is assumed to be 17.6 MW and 10% of the workloads running every hour are flexible to finish within a day.

Additional Capacity. Shifting computation across time may require additional server capacity for sustained increases in computation when carbon-free/low-carbon energy is abundant. The need for surplus capacity reveals an interesting trade-off between operational and embodied carbon, which has not been considered by any existing work in the literature [5, 64, 79].

From an operations perspective, increasing the number of provisioned servers mitigates the data center's carbon footprint by permitting demand response and reducing the carbon intensity of its
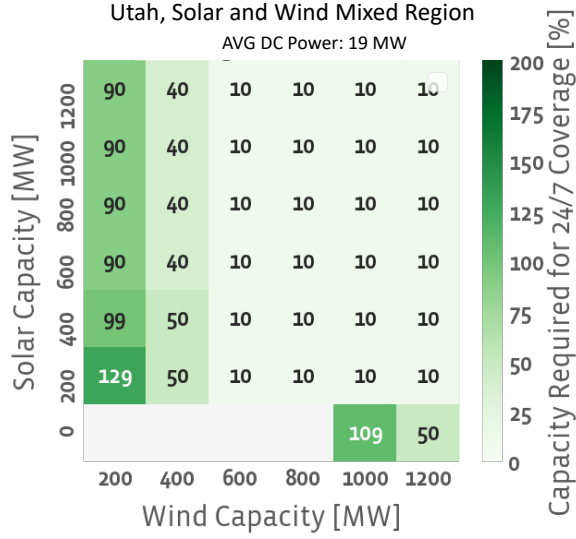
Figure 12: Scenario 3: Carbon aware scheduling for 24/7.

Figure 13: Carbon Explorer

energy. However, from the embodied perspective, over-provisioned servers increase embodied carbon emissions from hardware manufacturing [30]. Therefore, there is a fine balance between operational and capital expenditures.

Energy-proportional design is essential when over-provisioning datacenter servers [4]. Idle servers should draw little power, especially since hourly scheduling decisions provide ample time for servers to switch between power models. Indeed, server power can be accurately modeled as a linear function of utilization with the y-intercept denoting a server's idle power. Figure 3 illustrates energy-proportionality and correlation between Meta's datacenter power and its CPU utilization.

Figure 12 shows how much server capacity is required to achieve 24/7 carbon-free computation. Additional capacity is measured as a percentage of the datacenter's existing capacity. In this example, all workloads are assumed to be flexible to shift. Analysis shows that the additional capacity required to reach to 24/7 varies between 19% to over 100% (i.e. doubling the number of servers). Note that, as an alternative to deploying more servers, datacenters might Turbo Boost their current servers to increase compute throughput without increasing capital costs and embodied carbon.

5 CARBON MINIMIZATION: HOLISTIC DESIGN EXPLORATION

Reaching 24/7 carbon-free computing comes with non-negligible embodied carbon costs. Thus, Carbon Explorer must consider *both* operational and embodied carbon when minimizing the overall carbon footprint. Figure 13 presents the process of identifying an optimal datacenter design point from the carbon footprint's perspective.

First, Carbon Explorer requires inputs for its models of operational and embodied carbon. Operational inputs include hourly datacenter power demand and renewable power supply. Embodied inputs account fo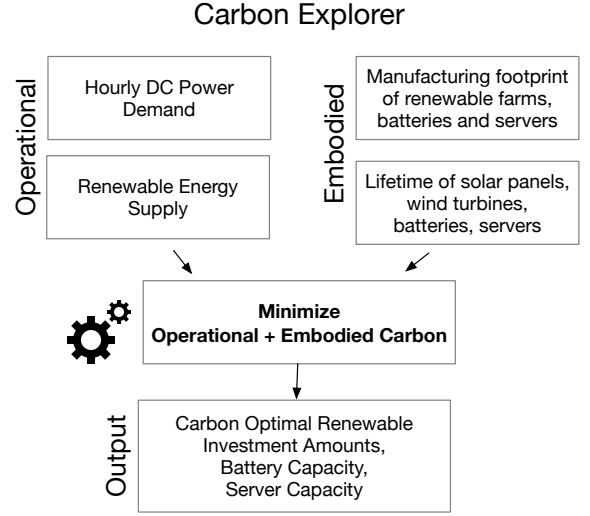r the carbon emissions from manufacturing and the expected lifetimes of solar and wind farms, lithium-ion batteries, and datacenter servers.

Carbon Explorer exhaustively searches the design space to minimize the sum of operational and embodied carbon. The design space includes the three solutions — renewable, battery, server investments — described in detailed in Section 4. Datacenter operators specify the bounds of the design space. Finally, Carbon Explorer outputs the carbon-optimal investments in renewable energy generation, battery capacity, and server capacity.

5.1 Embodied Carbon

Renewables. The manufacturing (or embodied) carbon footprint for wind turbines ranges from 10-15 grams of CO_2 per kWh whereas the footprint for solar farms ranges from 40-70 grams of CO_2 per kWh [31]. These numbers are derived from a life cycle analysis and accounts for manufacturing costs and the expected amount of energy generated over the asset's lifetime. The average lifetime for solar panels is 25-30 years and that for wind turbines is 20 years.

Batteries. The manufacturing footprint of lithium-ion batteries ranges from 74 to 134 kilograms of CO_2 per KWh of battery capacity [16, 66] and comprised of three major steps. (1) The majority of this footprint comes from production of the upstream battery materials, which is 59 kg per KWh and 44-80% of the total footprint. (2) Cell production and assembly is the part which uses the most electricity and hence the footprint ranges from 0-60 kg per KWh and is 0-44% of the total footprint depending on renewable energy usage during production. (3) End-of-life processing/recycling, which is a necessary and challenging task for batteries, generates 15 kg per KWh and is 11-20% of the total footprint [83]. The lifetime of the battery is calculated in terms of the number of discharge cycles. Utility-scale batteries, such as Tesla's Powerpack, last 3000-4000 cycles [70].

Lithium Iron Phosphate technology we model in this paper allows a large number of charge/discharge cycles. Charging rate (C-Rate) and Depth of Discharge are two of the important factors
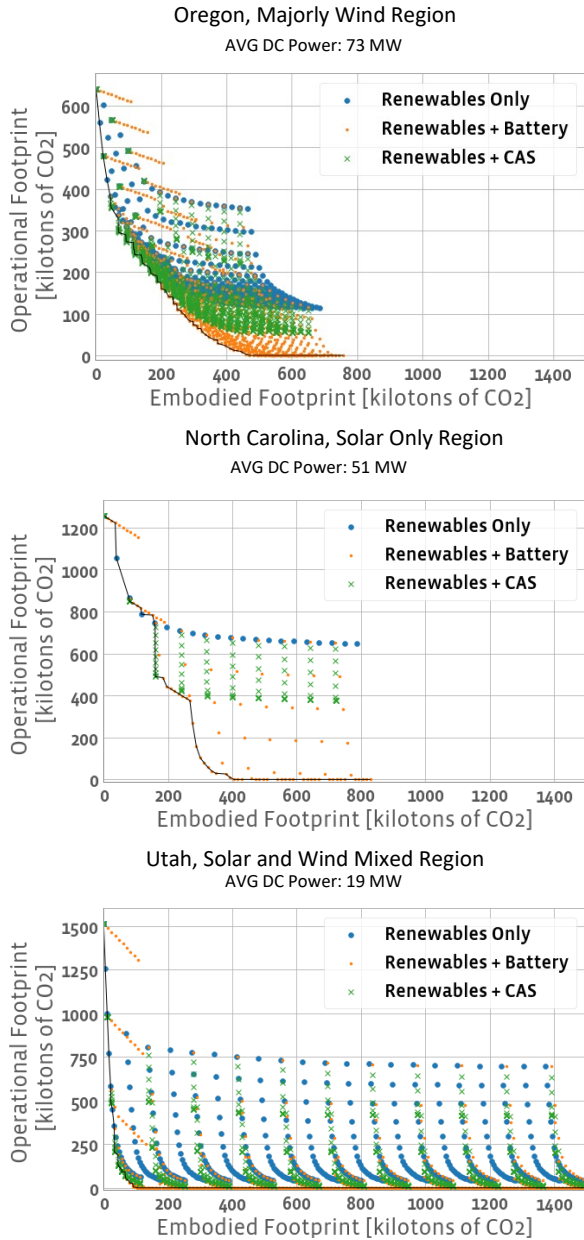
Figure 14: Operational and embodied footprint of the three solutions. Pareto frontier shows how the long tail to reach 100% renewable coverage can be shortened with complementary solutions.

effecting the lifetime of the battery. In this work, we assume C-Rate of 1C meaning full discharge or charge in one hour as our renewable energy data is on hourly basis. DoD provides a maximum usage limit over batter's total energy capacity, to extend it's lifespan. In standard environment with proper temperature, humidity, and for 1C-rate cycles, life cycle estimation for LFP batteries are 3000 cycles

at 100% DoD, and 4500 cycles at 80% DoD [63]. We studied these two scenarios.

Servers. The manufacturing footprint of servers is estimated to be 744.5 kg eq CO_2 [32], using an HPE ProLiant DL360 Gen10 server as a proxy. This server includes a single-socket CPU with 48 GB DRAM and has Thermal Design Power (TDP) of 85 Watts. Carbon measurements include its mainboard, SSD, daughterboard, enclosure, fans, transport and assembly. We estimate server lifetime of five years.

When demand response requires additional servers, to capture additional cost for floor-space and other infrastructure required, we added a parameterized surcharge cost. Even if new facility construction were required, initial estimates suggest carbon from construction is small compared to carbon from servers. For example, of Meta's 2019 Scope 3 carbon emissions, 42% and 7% is attributed to hardware infrastructure and construction, respectively [30]. In other words, construction's carbon footprint is only 16% of hardware's. This large difference might be explained by the relatively longevity of existing datacenters and rarity of new construction. A hyperscale datacenter's lifetime is 15 to 20 years whereas server hardware is typically three to five years. Therefore, Carbon Explorer models total embodied carbon by applying the multiplier to be 1.16x to server's embodied carbon.

However, sustainability reports carry several caveats. Hardware's carbon accounts for both refreshing existing datacenters and provisioning new ones. Construction's carbon is reported in the year incurred rather than amortized. Reports do not breakdown carbon incurred by a datacenter and its hardware. Further analysis will require better public data and carbon models for construction.

5.2 A Holistic Analysis

Figure 14 illustrates unique trade-offs between decreases in operational carbon (y-axis) and increases in embodied carbon (x-axis). In this evaluation, we assume 40% of datacenter workloads are delay-tolerant, a realistic flexible workload ratio [74], and can be deferred for carbon-aware scheduling. We examine four strategies: renewable energy generation alone, renewables with batteries, renewables with carbon-aware scheduling, and renewables with batteries and carbon-aware scheduling (Section 4).

The space includes solutions that can significantly reduce a datacenter's overall carbon footprint. Reductions in operational carbon incur increases in embodied carbon. Datacenter operators must be careful in its pursuit of 24/7 coverage because some solutions incur much higher embodied carbon costs than others. Renewable generation alone is insufficient and solutions that combine renewable energy generation with batteries and scheduling are needed.

The Pareto frontier indicates that any solution for 24/7 carbon-free operations (i.e., zero operational carbon) must include renewable energy and batteries. Moreover, as 24/7 coverage increases, solutions that include batteries will incur smaller embodied carbon costs than solutions that rely solely on renewable energy and/or deploying additional servers to support carbon-aware scheduling.

Unfortunately, the Pareto frontier exhibits a long tail, which indicates increasingly expensive solutions required to reach full 24/7 coverage. For example, in Oregon, investments in renewable energy and batteries can quickly reduce operational CO_2 from 600 to 400
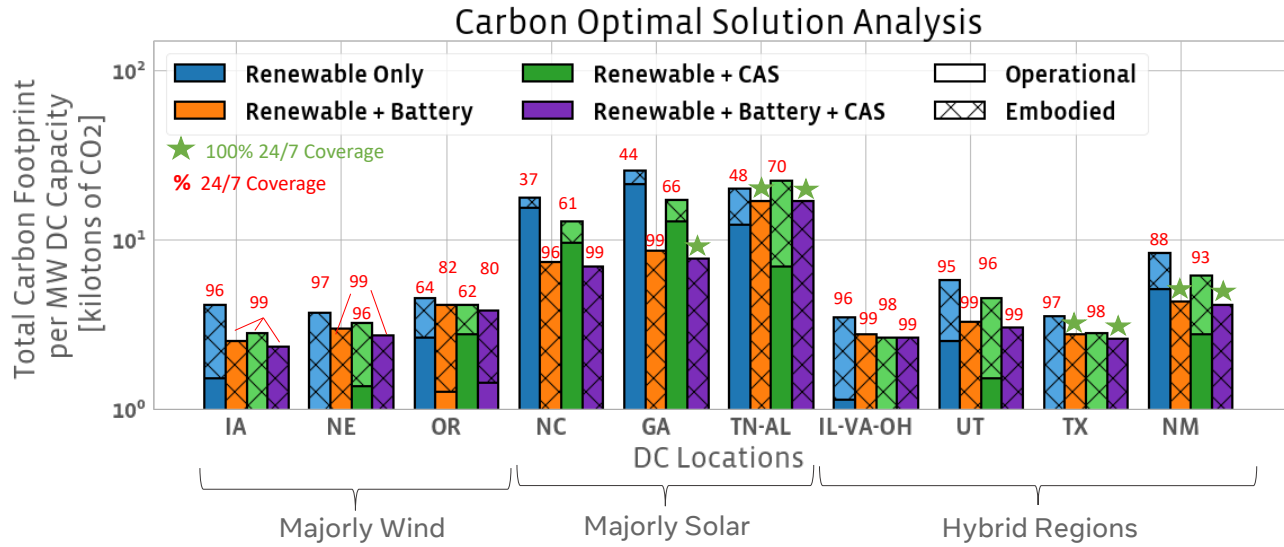
Figure 15: The total footprint of the carbon-optimal setting of each solution per MW DC capacity.

kilotons with minimal embodied cost. But eliminating the remaining 400 kilotons of operational carbon will require significantly larger embodied carbon cost.

Figure 15 details the most effective strategy for 24/7 carbon-free datacenter operation by geographic location and availability of renewable energy. We show the total carbon footprint, breaking down operational (solid) and embodied (cross-pattern) components. The carbon footprint is normalized relative to datacenter sizes, measured in MW of power capacity. We annotate each bar, identifying solutions that achieve full 100% 24/7 coverage (green stars) and those that make partial progress (red percentages).

Renewables Only. Relying solely on renewable energy generation incurs the highest embodied carbon costs in every geographic region. The 24/7 renewable coverage with renewables only ranges from 37% to 97% depending on geographical regions. Because renewable energy generation is intermittent, datacenter operators would need a large number of solar or wind farms to ensure sufficient supply during supply valleys. Even with significant investment, however, renewable energy supply fundamentally depends on weather and time of day, which leads to incomplete 24/7 coverage and higher operational carbon footprints that dominate the datacenter's total carbon footprint.

Figure 15 indicates the most effective renewable solutions include wind farms. A combination of wind and solar farms provides complementary generating assets and mitigates supply variance. Hybrid geographic regions, which use both wind and solar, achieve higher 24/7 coverage that ranges from 88% to 97%.

Regions that rely primarily on wind can also achieve high 24/7 coverage with careful datacenter site selection. Out of the thirteen

locations, datacenters in Nebraska (a majorly-wind region), Utah and Texas (wind and solar hybrid regions) stand out as the best locations to minimize total carbon footprint and achieve highest coverage. Valleys in energy supply are shallower in these windy regions compared to those in others. In contrast, regions that rely primarily on solar (i.e., NC, GA, TN, AL) struggle to achieve full 24/7 coverage and incur the highest carbon footprints since solar energy is only available during parts of the day.

Renewables + Battery. The addition of batteries reduces the total carbon footprint by an order of magnitude in all geographic regions. The reduction is most pronounced in regions that rely only on solar energy. For four of the thirteen datacenter regions (TN, AL, TX, NM), the most carbon-efficient solution deploys enough battery capacity to achieve 100% 24/7 coverage and completely eliminate the datacenter's operational carbon footprint. For the other regions optimal renewable energy coverage ranges from 82% to 99%. The battery capacity amount to achieve the optimal footprint ranges from 200MWh - 1800MWh for different datacenters and it would represent a utility scale battery capacity. Given hyperscale datacenters cost billions of dollars [59], this battery investment represents a small fraction of a data center's overall cost at current battery prices of $350/kWh [12].

Suppose we manage batteries such that the depth of discharge (DoD) is 100%. Under the carbon-optimal battery configuration, Figure 16 indicates batteries are often fully charged or fully discharged. Although this outcome arises naturally from our scheduling algorithm, which maximizes the battery usage to avoid carbon-intensive energy, high DoD can harm battery lifespan. Conversely, lower DoD
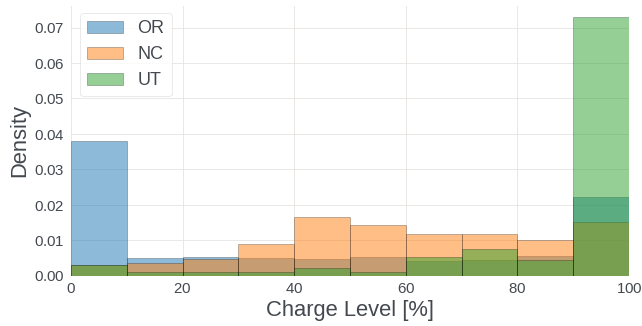
Figure 16: Battery charge level distributions

may benefit lifespan but reduce the batteries' usable capacities, requiring larger batteries or using carbon-intensive energy. Thus, we discover a trade-off between embodied and operational carbon.

For example, compare DoD of 100% and 80%. The lower DoD of 80% increases battery lifespan and the number of (dis)charge cycles by 50%. But because shallower discharges reduce the effective capacity of the battery, larger batteries are required and the associated embodied carbon increases by 43% in the carbon-optimal configuration. Evaluating the net impact, Carbon Explorer finds that 80% DoD lowers total carbon by 5% on average, across geographic regions.

Lowering DoD further increases battery lifetime but requires even larger batteries [63]. At some point, shallower DoD becomes counterproductive. A 60% DoD implies 10,000 daily (dis)charge cycles and 27-year battery lifespan. Other degradation factors would come in to play before reaching the 27-year lifespan. Overall, tuning DoD can lower total carbon by 3-9% across the DC regions.

Renewables + CAS. Carbon-aware scheduling provides an alternative to batteries, increasing 24/7 coverage by 1% to 21% across geographic regions. Carbon Explorer finds that deploying 6% to 76% additional server capacity allows the scheduler to move computation from periods when renewable energy is scarce to periods when it is abundant, thereby reducing the datacenter's overall carbon footprint. However, carbon-aware scheduling is constrained by the degree of workload flexibility and the number of provisioned servers available to process deferred jobs. Due to these constraints, scheduling alone is insufficient for full 24/7 coverage in regions characterized by many days with near zero renewable energy (e.g., wind in Oregon) or regions that rely exclusively on solar energy.

Renewables + Battery + CAS. A combination of battery deployments and carbon-aware scheduling offers additional improvements. We use a heuristic based solution where the priority is given to the workloads to minimize the runtime delays. Whenever there is lack of renewable supply, the energy stored in the battery is used first and workload shifting happens only if the energy stored in the batteries are not sufficient (at maximum DoD level). Whenever there is extra renewable supply, all available workloads are executed to use the available power first and batteries are charged with the remaining supply. This reduces the additional capacity required to reach 24/7 coverage compared with a battery only solution or

a CAS-only solution alone and makes 100% coverage the carbon-optimal solution in five of the regions. For rest of the regions except OR, carbon-optimal solution is at above 99%.

In summary, 24/7 coverage depends on renewable energy generation characteristics of the datacenter region, delay tolerance characteristics of the workloads and the possibility of using batteries. In several cases, achieving complete 24/7 coverage is neither feasible nor the most carbon-efficient solution. Solutions that complement investments in renewable energy are necessary. The addition of batteries and carbon-aware scheduling can reduce a datacenter's total carbon footprint by 15-65% depending on the region. However, there are other side-effects of using batteries and additional servers such as environmental factors and construction limitations. When designing carbon-aware solutions one must take the other factors into account as well.

Looking forward, to achieve a sustainable future for datacenters, understanding the design dimensions of carbon-free datacenter design is crucial. We expect the delay tolerance nature of computing to increase. For example, recent research on carbon-aware datacenter software for sustainability has started to emerge [2]. Delay-tolerant computing, such as AI training, is becoming more prevalent as well [81]. At the same time, the strong demand for a clean energy future has propelled significant efficiency improvement for both renewable infrastructures and energy storage technologies [8, 21]. Thus, in addition to computing's delay tolerance, a carbon-optimal datacenter design will also hinge on the trade-off between renewable energy utilization (as a result of operational efficiencies of energy generation and storage) and the associated embodied carbon cost. This is exactly the design space and optimization that Carbon Explorer enables.

6 DISCUSSION AND RELATED WORK

Renewable Energy. Prior academic research emphasizes renewable energy on-site at the datacenter [46]. Computation uses local, solar energy and minimizes energy consumed from the grid [23, 24]. The datacenter's power infrastructure is enhanced to switch between multiple types of local generators and microgrids [46, 47, 53]. And strategies are developed to deploy scale out servers and renewable generators in a modular fashion [43, 45]. These strategies seem sensible for edge and fog servers [44]. However, the hyperscale datacenters we study avoid many of these challenges. They do not need to manage local power generation because they have invested in renewable generation on the grid at scales that are unlikely on-site. Yet they improve sustainability through power purchase agreements. We study renewable energy across geographic locations and coordinate their installation with battery and server provisioning at scale.

Energy Storage. Batteries ensure datacenter availability but can also modulate the datacenter's demands for grid power [28, 29]. For datacenters that use renewable energy, batteries can mitigate intermittent supplies of solar and wind [51, 52]. Performance and efficiency vary with battery technology, motivating heterogeneous solutions [50, 54]. Battery aging can be mitigated by managing charge-discharge cycles and demand for stored energy [49]. We quantify energy storage required for 24/7 carbon-free computing and, without loss of generality, consider lithium-ion batteries for

their attractive downward cost trajectory and acceptable ten-plus year lifetimes under simulated usage.

Battery technologies will impact data center design and management. The price of lithium-ion batteries is falling significantly, declining by 80% from 2015 to 2020 [60, 61]. These batteries have been deployed at scale and, for example, can supply 28 MW for four hours. Such operational parameters align with hyperscale datacenters, which are provisioned for 20 to 40 MW. Four hours of battery operation could significantly reduce demand response requirements from job scheduling.

In terms of the implementation strategy of batteries, datacenter operators could collaborate with utility providers to invest in batteries on the grid just as they do for wind and solar farms. Alternatively, they could deploy batteries on-site at the datacenter. Datacenter may wish to implement custom battery charge-discharge policies, which have previously been explored at much smaller scales for uninterruptible power supplies [28, 29]. Whether these policies can be implemented in the form of contracts with grid operators is to be determined.

Finally, there are environmental and health risks associated with extraction of lithium from either the sea or the soil and the disposal of batteries. Spent LIBs contain toxic materials including heavy metals and flammable electrolytes, and therefore they need to be properly recycled and disposed in order not to cause contamination of the soil, water and air [65, 85]. This is another aspect that needs consideration when making large-scale battery deployments.

Carbon-Aware Scheduling. Time-series analysis accurately forecasts renewable supplies and datacenter demands for energy. Forecasts permit optimizing schedules of flexible jobs in response to energy supply [86]. Optimization objectives have accounted for electricity prices [55], carbon prices in cap-and-trade markets [42], the carbon-intensity of grid energy [64], and service quality [37]. Timely energy data is necessary for intelligent scheduling [5, 79]. We perform offline analyses to defer flexible computation and explore the design space for 24/7 carbon-free computing. A future implementation would benefit from prior schedulers.

Other Considerations. Note that Carbon Explorer quantifies the net impact on carbon emissions, the key contributor to climate change, and leaves broader considerations in sustainability for future work. Our view of lithium-ion batteries quantifies carbon from manufacturing and recycling but neglects the impact of lithium extraction or other side effects. Similarly, our view on servers and power do not consider the impact of electronic waste and water usage for cooling. Although important, these broader environmental impacts are beyond the scope of our study.

Moreover, Carbon Explorer emphasizes parameterized models because our understanding of carbon emissions in computing is still rapidly evolving. Operational emissions depend on the nature of renewable energy investments, the financial agreements for procuring this energy, and battery technologies. Embodied emissions depend on accurate accounting and transparent reporting throughout a massive supply chain. Carbon Explorer sets parameters based on the best publicly available data and these parameters can be tuned as better data becomes available.

7 CONCLUSION

This paper presents *Carbon Explorer* — a design space exploration tool to enable carbon-optimal investment strategies. Carbon Explorer determines carbon-optimal settings across the dimensions of *investments on various renewable energy types*, *the amount of energy storage*, and *carbon-aware computation shifting* by considering geographically-dependent renewable energy availability characteristics and computation demand patterns at the data center scale. Carbon Explorer demonstrates that, depending on graphical locations, carbon-optimal strategies vary and that when embodied carbon footprint is considered, 100% 24/7 operational carbon-free computing may not always be carbon-optimal. We hope Carbon Explorer can guide future sustainability investments to achieve operational and embodied carbon footprint optimality.

REFERENCES

[1] A123 Systems 2009. *High Power Lithium Ion APR18650M1A*. A123 Systems. LiFePO4 cell specifications.

[2] Thomas Anderson, Adam Belay, Mosharaf Chowdhury, Asaf Cidon, and Irene Zhang. 2022. Treehouse: A Case For Carbon-Aware Datacenter Software. *arXiv preprint arXiv:2201.02120* (2022).

[3] Anders SG Andrae and Tomas Edler. 2015. On global electricity usage of communication technology: trends to 2030. *Challenges* 6, 1 (2015), 117–157.

[4] Luiz André Barroso and Urs Hölzle. 2007. The case for energy-proportional computing. *Computer* 40, 12 (2007), 33–37.

[5] Noman Bashir, Tian Guo, Mohammad Hajiesmaili, David Irwin, Prashant Shenoy, Ramesh Sitaraman, Abel Souza, and Adam Wierman. 2021. Enabling Sustainable Clouds: The Case for Virtualizing the Energy System. *arXiv preprint arXiv:2106.08872* (2021).

[6] California ISO. 2021. Managing oversupply. http://www.caiso.com/informed/Pages/ManagingOversupply.aspx.

[7] California State Commission. 2020. 2020 Total System Electric Generation. https://www.energy.ca.gov/data-reports/energy-almanac/california-electricity-data/2020-total-system-electric-generation.

[8] Canadian Solar. 2021. Sustainability Report. http://investors.canadiansolar.com/static-files/e10bbede-2991-4365-b2a7-fd2da6111e22.

[9] Jeffrey S. Chase. 2014. Demand Response for Computing Centers. In *The Green Computing Book: Tackling Energy Efficiency at Large Scale*, "Wu chun Feng" (Ed.). "Chapman & Hall/CRC Computational Science", Chapter 8.

[10] A. Chien. 2020. Characterizing Opportunity Power in the California Independent System Operator (CAISO) in Years 2015-2017. *Energy and Earth Science* (2020).

[11] Andrew A Chien, Fan Yang, and Chaojie Zhang. 2018. Characterizing curtailed and uneconomic renewable power in the mid-continent independent system operator. *AIMS Energy* 6, 2 (2018), 376–401.

[12] Wesley Cole, A Will Frazier, and Chad Augustine. 2021. *Cost projections for utility-scale battery storage: 2021 update*. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).

[13] Boucar Diouf and Ramchandra Pode. 2015. Potential of lithium-ion batteries in renewable energy. *Renewable Energy* 76 (2015), 375–380.

[14] DOE Global Energy Storage Database. 2022. Vistra Moss Landing Energy Storage. https://sandia.gov/ess-ssl/gesdb/public/projects.html#1331.

[15] Erik Ela. 2009. *Using economics to determine the efficient curtailment of wind energy*. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).

[16] Erik Emilsson and Lisbeth Dahllöf. 2019. Lithium-Ion Vehicle Battery Production-Status 2019 on Energy Use, CO2 Emissions, Use of Metals, Products Environmental Footprint, and Recycling.

[17] Facebook. 2021. Advancing renewable energy through green tariffs. https://sustainability.fb.com/wp-content/uploads/2020/12/FB_Green-Tariffs.pdf.

[18] Facebook. 2021. How We're Helping Fight Climate Change. https://about.fb.com/news/2021/06/2020-sustainability-report-how-were-helping-fight-climate-change/.

[19] S. Fan, S.M. Zahedi, and B. Lee. 2016. The computational sprinting game. In *Proc. International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*.

[20] X. Fan, W.D. Weber, and L.A. Barroso. 2007. Power provisioning for a warehouse-sized computer. In *Proc. International Symposium on Computer Architecture (ISCA)*.

[21] First Solar. 2021. Sustainability Report. https://www.firstsolar.com/-/media/First-Solar/Sustainability-Documents/FirstSolar_Sustainability-Report_2021.ashx.

[22] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica. 2011. Dominant resource fairness: Fair allocation of multiple resource types. In *Proc. Networked Systems Design and Implementation (NSDI)*.

[23] Íñigo Goiri, William Katsak, Kien Le, Thu D Nguyen, and Ricardo Bianchini. 2013. Parasol and greenswitch: Managing datacenters powered by renewable energy. *ACM SIGPLAN Notices* 48, 4 (2013), 51–64.

[24] Íñigo Goiri, William Katsak, Kien Le, Thu D Nguyen, and Ricardo Bianchini. 2014. Designing and managing data centers powered by renewable energy. *IEEE Micro* 34, 3 (2014), 8–16.

[25] Google. 2013. Google's green PPAs: What, how and why. https://static.googleusercontent.com/media/www.google.com/en//green/pdfs/renewable-energy.pdf.

[26] Google. 2019. Borg cluster workload traces. https://github.com/google/cluster-data.

[27] Google. 2020. 24/7 by 2030: Realizing a carbon-free future. https://www.gstatic.com/gumdrop/sustainability/247-carbon-free-energy.pdf.

[28] Sriram Govindan, Anand Sivasubramaniam, and Bhuvan Urgaonkar. 2011. Benefits and limitations of tapping into stored energy for datacenters. In *2011 38th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 341–351.

[29] Sriram Govindan, Di Wang, Anand Sivasubramaniam, and Bhuvan Urgaonkar. 2012. Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters. In *Proceedings of the seventeenth international conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*. 75–86.

[30] U. Gupta, Y. Kim, S. Lee, J. Tse, H. S. Lee, G. Wei, D. Brooks, and C. Wu. 2021. Chasing Carbon: The Elusive Environmental Footprint of Computing. In *Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA)*.

[31] Garvin Heath. 2021. Life Cycle Assessment Harmonization. https://www.nrel.gov/analysis/life-cycle-assessment.html.

[32] Hewlett Packard Enterprise. 2021. HPE Product Carbon Footprint, HPE ProLiant DL360 Gen10 server. https://www.hpe.com/psnow/doc/a50002430enw?jumpid=in_lit-psnow-red.

[33] B. Hindman, A. Konwinski, M. Zaharia, A. Ghodsi, A. Joseph, R. Katz, S. Shenker, and I. Stoica. 2011. Mesos: A platform for fine-grained resource sharing in the data center. In *Proc. Networked Systems Design and Implementation (NSDI)*.

[34] Nicola Jones. 2018. How to stop data centres from gobbling up the world's electricity. *Nature* 561, 7722 (2018), 163–167.

[35] Lucas Joppa. 2021. Made to measure: Sustainability commitment progress and updates. https://blogs.microsoft.com/blog/2021/07/14/made-to-measure-sustainability-commitment-progress-and-updates/.

[36] Justine Calma. 2021. Microsoft is changing the way it buys renewable energy. https://www.theverge.com/2021/7/14/22574431/microsoft-renewable-energy-purchases.

[37] William Katsak, Íñigo Goiri, Ricardo Bianchini, and Thu D Nguyen. 2015. GreenCassandra: Using renewable energy in distributed structured storage systems. In *2015 Sixth International Green and Sustainable Computing Conference (IGSC)*. IEEE, 1–8.

[38] Fiodar Kazhamiaka, Catherine Rosenberg, and Srinivasan Keshav. 2019. Tractable lithium-ion storage models for optimizing energy systems. *Energy Informatics* 2, 1 (2019), 1–22.

[39] Ross Koningstein. 2021. We now do more computing where there's cleaner energy. https://blog.google/outreach-initiatives/sustainability/carbon-aware-computing-location/.

[40] Vasileios Kontorinis, Liuyi Eric Zhang, Baris Aksanli, Jack Sampson, Houman Homayoun, Eddie Pettis, Dean M. Tullsen, and Tajana Simunic Rosing. 2012. Managing distributed UPS energy for effective power capping in data centers. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*.

[41] Kubernetes. 2022. Production-grade container orchestration. https://kubernetes.io

[42] Kien Le, Ozlem Bilgir, Ricardo Bianchini, Margaret Martonosi, and Thu D Nguyen. 2010. Managing the cost, energy consumption, and carbon footprint of internet services. *ACM SIGMETRICS Performance Evaluation Review* 38, 1 (2010), 357–358.

[43] C. Li, Y. Hu, J. Gu, J. Yuan, and T. Li. 2016. Oasis: Scaling out datacenter sustainably and economically. *IEEE Transactions on Parallel and Distributed Systems* (2016).

[44] Chao Li, Yang Hu, Longjun Liu, Juncheng Gu, Mingcong Song, Xiaoyao Liang, Jingling Yuan, and Tao Li. 2015. Towards sustainable in-situ server systems in the big data era. In *2015 ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA)*. 14–26. https://doi.org/10.1145/2749469.2750381

[45] Chao Li, Yang Hu, Ruijin Zhou, Ming Liu, Longjun Liu, Jingling Yuan, and Tao Li. 2013. Enabling datacenter servers to scale out economically and sustainably. In *Proceedings of the 46th annual IEEE/ACM international symposium on microarchitecture*. 322–333.

[46] Chao Li, Amer Qouneh, and Tao Li. 2012. iSwitch: Coordinating and optimizing renewable energy powered server clusters. *ACM SIGARCH Computer Architecture News* 40, 3 (2012), 512–523.

[47] Chao Li, Ruijin Zhou, and Tao Li. 2013. Enabling distributed generation powered sustainable high-performance data center. In *2013 IEEE 19th International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 35–46.

[48] L. Lin and A. Chien. 2020. Characterizing Stranded Power in the ERCOT in Years 2012-2019: A Preliminary Report. *University of Chicago CS Tech Report* (2020).

[49] Longjun Liu, Chao Li, Hongbin Sun, Yang Hu, Juncheng Gu, and Tao Li. 2015. BAAT: Towards dynamically managing battery aging in green datacenters. In *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*. IEEE, 307–318.

[50] Longjun Liu, Chao Li, Hongbin Sun, Yang Hu, Juncheng Gu, Tao Li, Jingmin Xin, and Nanning Zheng. 2015. HEB: Deploying and managing hybrid energy buffers for improving datacenter efficiency and economy. In *2015 ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA)*. 463–475. https://doi.org/10.1145/2749469.2750384

[51] Longjun Liu, Hongbin Sun, Yang Hu, Jingmin Xin, Nanning Zheng, and Tao Li. 2014. Leveraging distributed UPS energy for managing solar energy powered data centers. In *International Green Computing Conference*. IEEE, 1–8.

[52] L. Liu, H. Sun, C. Li, Y. Hu, J. Xin, N. Zheng, and T. Li. 2016. RE-UPS: An adaptive distributed energy storage system for dynamically managing solar energy in green datacenters. *the Journal of Supercomputing* 72, 1 (2016), 295–316.

[53] Longjun Liu, Hongbin Sun, Chao Li, Yang Hu, Nanning Zheng, and Tao Li. 2016. Towards an adaptive multi-power-source datacenter. In *Proceedings of the 2016 International Conference on Supercomputing*. 1–11.

[54] Longjun Liu, Hongbin Sun, Chao Li, Tao Li, Jingmin Xin, and Nanning Zheng. 2018. Exploring highly dependable and efficient datacenter power system using hybrid and hierarchical energy buffers. *IEEE Transactions on Sustainable Computing* (2018).

[55] Zhenhua Liu, Iris Liu, Steven Low, and Adam Wierman. 2014. Pricing data center demand response. *ACM SIGMETRICS Performance Evaluation Review* 42, 1 (2014), 111–123.

[56] Sulav Malla, Qingyuan Deng, Zoh Ebrahimzadeh, Joe Gasperetti, Sajal Jain, Parimala Kondety, Thiara Ortiz, and Debra Vieira. 2020. Coordinated Priority-aware Charging of Distributed Batteries in Oversubscribed Data Centers. In *Proceedings of the International Symposium on Microarchitecture (ISCA)*.

[57] Sulav Malla, Qingyuan Deng, Zoh Ebrahimzadeh, Joe Gasperetti, Sajal Jain, Parimala Kondety, Thiara Ortiz, and Debra Vieira. 2020. Coordinated priority-aware charging of distributed batteries in oversubscribed data centers. In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 839–851.

[58] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. Recalibrating global data center energy-use estimates. *Science* 367, 6481 (2020), 984–986.

[59] Rich Miller. 2022. Google Has Invested $5 Billion in its Iowa Data Centers. https://datacenterfrontier.com/google-has-invested-5-billion-in-its-iowa-data-centers/.

[60] Kendall Mongird, Vilayanur V Viswanathan, Patrick J Balducci, Md Jan E Alam, Vanshika Fotedar, V S Koritarov, and Boualem Hadjerioua. 2019. *Energy storage technology and cost characterization report.* Technical Report. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

[61] National Renewable Energy Laboratory. 2020. Declining renewable costs drive focus on energy storage. https://www.nrel.gov/news/features/2020/declining-renewable-costs-drive-focus-on-energy-storage.html

[62] Prasant Kumar Nayak, Liangtao Yang, Wolfgang Brehm, and Philipp Adelhelm. 2018. From lithium-ion to sodium-ion batteries: advantages, challenges, and surprises. *Angewandte Chemie International Edition* 57, 1 (2018), 102–120.

[63] PowerTech Systems. 2022. Lithium Iron Phosphate (LFP or LiFePO4). https://www.powertechsystems.eu/home/tech-corner/lithium-iron-phosphate-lifepo4/.

[64] Ana Radovanovic, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, et al. 2021. Carbon-Aware Computing for Datacenters. *arXiv preprint arXiv:2106.11750* (2021).

[65] Mohammad Ali Rajaeifar, Pezhman Ghadimi, Marco Raugei, Yufeng Wu, and Oliver Heidrich. 2022. Challenges and recent developments in supply and value chains of electric vehicle batteries: A sustainability perspective. *Resources, Conservation and Recycling* 180 (2022), 106144. https://doi.org/10.1016/j.resconrec.2021.106144

[66] Robert Rapier. 2020. Estimating The Carbon Footprint Of Utility-Scale Battery Storage. https://www.forbes.com/sites/rrapier/2020/02/16/estimating-the-carbon-footprint-of-utility-scale-battery-storage/?sh=167c8bd97adb.

[67] Hannah Ritchie and Max Roser. 2020. Energy. *Our World in Data* (2020). https://ourworldindata.org/energy.

[68] S. VanLear, B. M. Depro, M. Q. Hogan, M. P. Gallaher, J. C. Woollacott, E. G. Brown and M. A. Gonzalez. 2021. Facebook's U.S. Renewable Energy: Impact Study. https://www.rti.org/publication/facebooks-us-renewable-energy.

[69] V. Sakalkar, v. Kontorinis, D. Landhuis, S. Li, D. DeRonde, T. Blooming, et al. 2020. Data center power oversubscription with a medium voltage power plan and priority-aware capping. In *Proc. International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS).*

[70] Zachary Shahan. 2020. Tesla Megapack, Powerpack, & Powerwall Battery Storage Prices Per KWh. https://cleantechnica.com/2020/10/05/tesla-megapack-powerpack-powerwall-battery-storage-prices/.

[71] Chunqiang Tang, Kenny Yu, Kaushik Veeraraghavan, Jonathan Kaldor, Scott Michelson, Thawan Kooburat, Aravind Anbudurai, Matthew Clark, Kabir Gogia, Long Cheng, Ben Christensen, Alex Gartrell, Maxim Khutornenko, Sachin Kulkarni, Marcin Pawlowski, Tuomas Pelkonen, Andre Rodrigues, Rounak Tibrewal, Vaishnavi Venkatesan, and Peter Zhang. 2020. Twine: A Unified Cluster Management System for Shared Infrastructure. In *14th USENIX Symposium on Operating Systems Design and Implementation (OSDI).* 787–803.

[72] The U.S. Energy Information Administration. 2021. Renewable energy explained. https://www.eia.gov/energyexplained/renewable-sources/.

[73] The White House. 2021. FACT SHEET: President Biden Renews U.S. Leadership on World Stage at U.N. Climate Conference (COP26). https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/01/fact-sheet-president-biden-renews-u-s-leadership-on-world-stage-at-u-n-climate-conference-cop26/.

[74] Muhammad Tirmazi, Adam Barker, Nan Deng, Md E Haque, Zhijing Gene Qin, Steven Hand, Mor Harchol-Balter, and John Wilkes. 2020. Borg: the next generation. In *Proceedings of the Fifteenth European Conference on Computer Systems.* 1–14.

[75] United Nations. [n. d.]. 24/7 Carbon-free Energy Compact. https://www.un.org/en/energy-compacts/page/compact-247-carbon-free-energy.

[76] U.S. Department of Defense (DoD) and the U.S. General Services Administration (GSA). [n. d.]. DoD, GSA Announce RFI to Gather Information for Supplying 24/7 Carbon Pollution-Free Electricity for Federal Government. https://www.defense.gov/News/Releases/Release/Article/2921646/dod-gsa-announce-rfi-to-gather-information-for-supplying-247-carbon-pollution-f/.

[77] U.S. Energy Information Administration. 2021. Annual Energy Outlook. https://www.eia.gov/outlooks/archive/aeo21/.

[78] U.S. Energy Information Administration. 2021. Hourly Electric Grid Monitor. https://www.eia.gov/electricity/gridmonitor/dashboard/electric_overview/US48/US48.

[79] Adam Wierman, Zhenhua Liu, Iris Liu, and Hamed Mohsenian-Rad. 2014. Opportunities and challenges for data center demand response. In *International Green Computing Conference.* IEEE, 1–10.

[80] Philipp Wiesner, Ilja Behnke, Dominik Scheinert, Kordian Gontarska, and Lauritz Thamsen. 2021. Let's Wait Awhile: How Temporal Workload Shifting Can Reduce Carbon Emissions in the Cloud. *arXiv preprint arXiv:2110.13234* (2021).

[81] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. 2022. Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems* 4 (2022), 795–813.

[82] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga Behram, James Huang, Charles Bai, Michael Gschwind, Anurag Gupta, Myle Ott, Anastasia Melnikov, Salvatore Candido, David Brooks, Geeta Chauhan, Benjamin Lee, Hsien-Hsin S. Lee, Bugra Akyildiz, Maximilian Balandat, Joe Spisak, Ravi Jain, Mike Rabbat, and Kim Hazelwood. 2021. Sustainable AI: Environmental Implications, Challenges and Opportunities. arXiv:2111.00364 [cs.LG]

[83] Yue Yang, Emineke G Okonkwo, Guoyong Huang, Shengming Xu, Wei Sun, and Yinghe He. 2020. On the sustainability of lithium ion battery industry–A review and perspective. *Energy Storage Materials* (2020).

[84] Seyed Majid Zahedi and Benjamin C Lee. 2014. REF: Resource elasticity fairness with sharing incentives for multiprocessors. *ACM SIGPLAN Notices* 49, 4 (2014), 145–160.

[85] Xianlai Zeng, Jinhui Li, and Lili Liu. 2015. Solving spent lithium-ion battery problems in China: Opportunities and challenges. *Renewable and Sustainable Energy Reviews* 52 (2015), 1759–1767.

[86] Chaojie Zhang and Andrew A Chien. 2021. Scheduling Challenges for Variable Capacity Resources. In *Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP).* Springer, 190–209.

[87] Chaojie Zhang, Alok Gautam Kumbhare, Ioannis Manousakis, Deli Zhang, Pulkit A Misra, Rod Assis, Kyle Woolcock, Nithish Mahalingam, Brijesh Warrier, David Gauthier, et al. 2021. Flex: High-availability datacenters with zero reserved power. In *2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA).* IEEE, 319–332.

[88] Jiajia Zheng, Andrew A Chien, and Sangwon Suh. 2020. Mitigating curtailment and carbon emissions through load migration between data centers. *Joule* 4, 10 (2020), 2208–2222.