# CURTAIL TO COMPUTE

## Siting Datacenters to Leverage California's Stranded Renewable Energy

NEXT 10

MARCH 2026

**NEXT 10 is an independent nonpartisan organization that educates, engages and empowers Californians to improve the state's future.**

Next 10 is focused on innovation and the intersection between the economy, the environment, and quality of life issues for all Californians. We provide critical data to help inform the state's efforts to grow the economy and reduce greenhouse gas emissions. Next 10 was founded in 2003 by businessman and philanthropist F. Noel Perry.


**ACKNOWLEDGMENTS**

This research was made possible by the generous support of Next10. We are grateful for their fnancial contributions, as well as their technical perspective and insights into the California context, which significantly shaped the direction and depth of this work.



**PRODUCED BY**

Next 10
F. Noel Perry
Marcia E. Perry
Stephanie Leonard
Grace Ochs

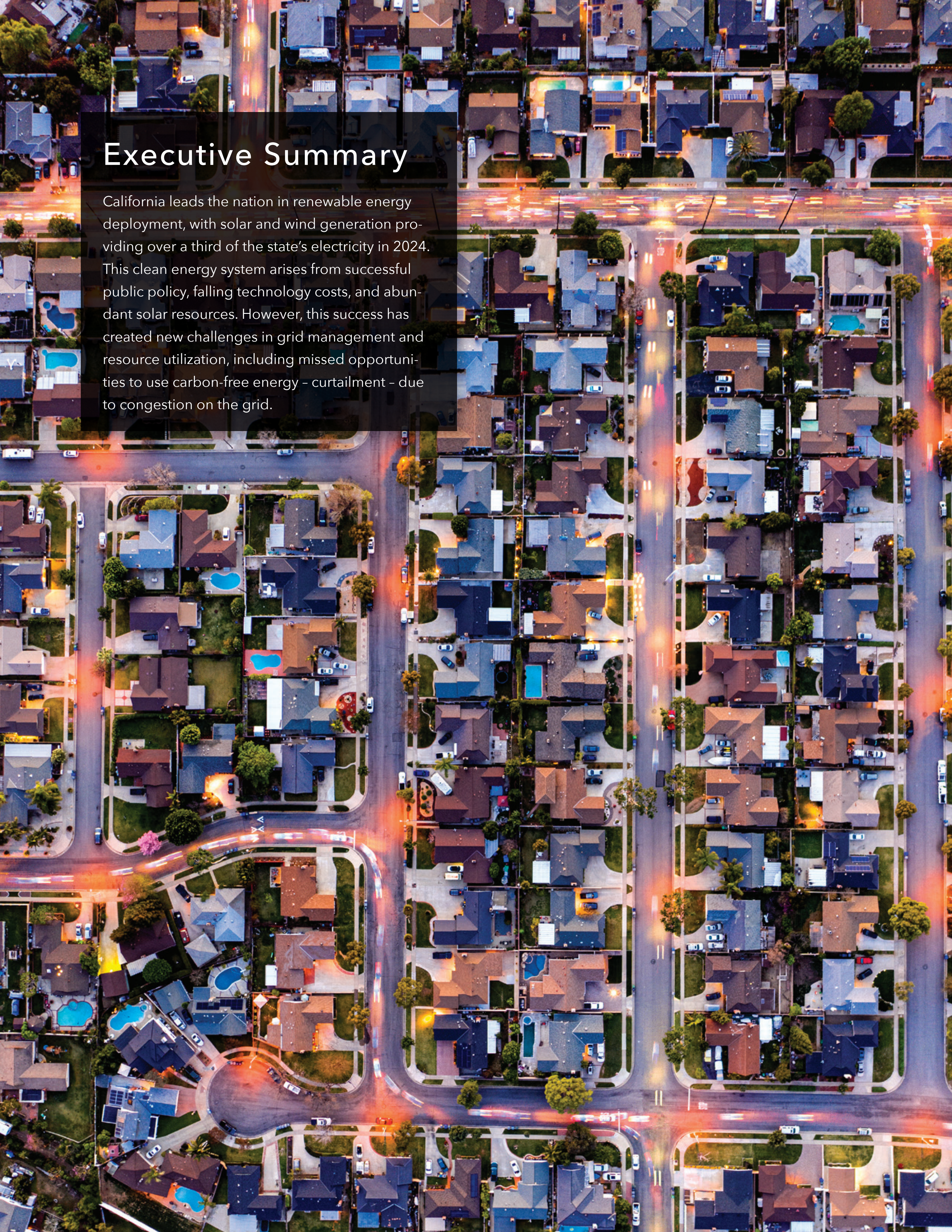
**PREPARED BY**

University of Pennsylvania, USA
Lindsey Yu
Jiali Xing
Abhinav Vadlamani
Benjamin C. Lee

# Table of Contents

# Executive Summary

California leads the nation in renewable energy deployment, with solar and wind generation providing over a third of the state's electricity in 2024. This clean energy system arises from successful public policy, falling technology costs, and abundant solar resources. However, this success has created new challenges in grid management and resource utilization, including missed opportunities to use carbon-free energy – curtailment – due to congestion on the grid.



In 2024, the California Independent System Operator (CAISO), the state's grid operator, curtailed approximately 3.4 terrawatt-hours (TWh) of renewable generation, a 29% increase from 2023,[1] and equivalent to the average annual usage of roughly 500,000 California households. This curtailment, occurring primarily during midday hours when solar generation peaks, represents wasted clean energy and lost economic value for renewable generators. The timing and geographic patterns of curtailment, concentrated in specific transmission-constrained zones, create opportunities for co-located energy users that can absorb this excess generation.

The conventional solution to curtailment is the construction of additional high-voltage transmission lines to ease congestion and provide additional capacity to move the energy from where it is generated to where it is needed. However, CAISO estimates that it would cost $700 million to $1.127 billion to build the needed high-voltage transmission lines serving the Bay Area. In this paper, we consider the alternative—rather than moving energy through wires, moving computation to the energy. A 20MW datacenter requires approximately $10 million in construction costs plus $15 million to $60 million in server costs, depending on whether the facility houses CPUs or GPUs.[2] Costs for datacenters are potentially lower than those for transmission upgrades, providing a non-wire solution for productively absorbing surplus generation while providing power for computation that would occur anyway in a way that eases stress on the grid. Strategically siting datacenters might provide California ratepayers significant relief through avoided transmission upgrade costs.

Generators must pay to offload this excess power, yet in other areas where demand is high energy prices rise. We find that these price differences can exceed $200 per megawatt-hour (MWh) between locations that are separated by less than 200 miles. In many instances, this cheap excess power is unable to go where it is needed due to congestion on the transmission lines so more expensive energy (such as electricity generated from natural gas) must be used to meet the higher demand in these other, usually more populated, areas.

Therefore, the root causes of California's curtailment challenge extend beyond these simple supply-demand imbalances to include complex constraints from transmission and grid topology. Our analysis of CAISO data reveals that over 70% of curtailments result from local transmission congestion rather than system-wide oversupply. The Path 15 corridor, the transmission line through Fresno and the Central Valley which connects Northern and Southern California's grids, is the most congested line in the state. It is projected to experience congestion for more than 7,300 hours annually by 2039, equivalent to 84% of all hours in a year. Congested transmission lines create persistent bottlenecks that prevent renewable energy generated from solar farms inland from reaching coastal load centers and the economic impact of grid congestion along Path 15 is forecasted to be $389.4 million by 2034.[3] These structural constraints, which will likely persist despite planned transmission upgrades, lead to pockets of stranded renewable energy that could be used for other purposes. The PG&E Fresno zone exhibits the highest curtailment rate and the rate is forecasted to continue to grow despite planned grid upgrades.

Simultaneously, the growth in artificial intelligence is driving unprecedented demand for computational infrastructure, with U.S. datacenters consuming 176 TWh in 2023 (4.4% of national demand) and potentially reaching 580 TWh by 2028. This report studies whether datacenters can economically transform stranded renewable energy into productive computational capacity.

Strategically siting datacenters to leverage California's curtailed renewable energy—electricity generated but prevented from reaching the grid due to oversupply or transmission constraints—would support both California's carbon neutrality goals and growing technology infrastructure needs.

1 Slade Johnson. 2024. California's solar electricity generation grew faster than any other state's last year. U.S. Energy Information Administration, Today in Energy. https://www.eia.gov/todayinenergy/detail.php?id=65364 Accessed: 2025-09-09.

2 Central Processing Units (CPUs) are processors designed to perform a wide-range of general computations whereas Graphics Processing Units (GPUs) are processors specially designed to perform the matrix calculations that dominate artificial intelligence and large language model computations. Compared to CPUs, GPUs perform calculations at far higher rates but use much more power. Datacenters have increased the number of GPUs they deploy in response to growth in AI computations.

3 California Independent System Operator (CAISO). 2025. ISO Board-Approved 2024–2025 Transmission Plan. Technical Report Board-Approved plan. California ISO. https://www.caiso.com/documents/iso-board-approved-2024-2025-transmission-plan.pdf

## Scenario Analysis

To evaluate the technological and economic feasibility of designing and siting datacenters to take advantage of curtailed energy in California, the report developed three scenarios. Each scenario makes distinct trade-offs between capital expenditure, operational costs, energy access, workload flexibility, and revenue diversification:

**Scenario A (Urban Baseline):** A traditional 20MW colocation[4] datacenter in Silicon Valley, representing the status quo. This facility prioritizes network latency and customer proximity, relies entirely on grid power at standard commercial rates, and cannot easily access curtailed energy due to transmission constraints. This scenario serves as our economic and operational baseline.

**Scenario B (Rural Curtailment-Only):** A new 20MW colocation datacenter sited in Fresno County, strategically positioned to consume locally curtailed solar energy. This design bypasses Path 15 congestion by consuming energy near its generation, accessing near-zero marginal cost power during curtailment windows, which account for 54% of hours in a year in our simulations. Without on-site energy storage, computation must be flexible and delay-tolerant.

**Scenario C (Rural + Battery Storage):** This is Scenario B's 20MW datacenter augmented with a 10MW/40MWh Battery Energy Storage System (BESS). Storage decouples computation from instantaneous generation. Batteries permit the datacenter to store curtailed energy during midday for computation in the evening, smooth intermittent supply for less-flexible computation, and perform energy arbitrage and grid services when demand for computation is low. This scenario explores whether the battery's capital costs can be justified by increased operational flexibility and revenue diversification.

**Table ES-1.** Summary of Total Modeled Costs by Scenario

| Scenario | Total CapEx ($M) | First-Year OpEx ($M) |
|---|---|---|
| Scenario A (Santa Clara) | $278.0 | $29.0 |
| Scenario B (Fresno) | $183.5 | $25.9 |
| Scenario C (Fresno + BESS) | $195.5 | $26.7 |

Table ES-1 summarizes our cost projections across the datacenter scenarios. The urban baseline, in Scenario A, has the highest capital and operating expenditures, reflecting premium real estate and electricity prices in Silicon Valley. The rural datacenter that uses only curtailed energy, in Scenario B, reduces operating costs by approximately 40% through access to near-zero marginal cost curtailed energy although computation is limited to times when curtailed energy is available.

Finally, augmenting the rural datacenter with batteries, in Scenario C, increases capital expenditures by $8 million to $12 million but extends computational availability. While the upfront costs are higher, batteries become viable if the datacenter were to generate additional revenue through energy trading and grid services or if batteries continue to become less expensive. Battery costs have declined from over $600 per kWh in 2019 to $250 in 2024 with projections reaching $150 by 2030.[5] Our simulations optimally size batteries, balancing gains in computational throughput against capital costs, and show how batteries can extend curtailment-only operations from 54% to 78% job completion while maintaining significant cost advantages over the urban datacenter.

Artificial intelligence's growing energy demands and California's growing renewable curtailment create both challenge and opportunity for sustainable computing. Our findings indicate that the approach of siting data-

4 Colocation datacenters are buildings, usually owned by real estate or other non-tech companies, in which multiple tenants rent space to keep their servers. These are compared to hyperscalers which are much larger and usually owned by a single tech company, such as Amazon or Google, to house their servers only.

5 National Renewable Energy Laboratory. 2024. 2024 Annual Technology Baseline: Utility-Scale Battery Storage. Data. National Renewable Energy Laboratory, Golden, CO. https://atb.nrel.gov/electricity/2024/utility-scale_battery_storage

centers in rural areas to harness surplus power could yield significant economic advantages, reducing break-even revenue from computation by nearly $14 million annually compared to conventional urban deployments. Building such datacenters will require significant capacity for on-site energy storage to support additional computation when curtailed energy is scarce or permit energy trading in the wholesale market.

## Key Findings

1. **Transmission Congestion, Not Oversupply, Drives Curtailment.** Over 70% of California's renewable curtailment stems from local transmission bottlenecks rather than system-wide excess generation. Path 15, the Central Valley corridor, will experience congestion in 7,343 hours of the year by 2039 (84% of hours), stranding solar generation even as consumers demand power. The PG&E Fresno zone exhibits the highest projected curtailment rate (11% by 2039), with solar capacity growing 83% faster than transmission expansion. This structural imbalance makes curtailment-based computing a durable opportunity rooted in grid topology rather than transient market conditions.

2. **Short-Term Bankability, Not Long-Term Profitability, Constrains Viability.** Rural datacenters that use otherwise curtailed energy could deliver high returns (28% versus 15% for urban facilities). These datacenters could reduce capital costs by 34% through lower land and construction costs while their tenants could benefit from 10% lower electricity costs through access to less expensive power. However, weak initial occupancy might require 23% higher rents to service debt, creating a short-term bankability challenge rather than a long-term profitability challenge.

3. **Workload Flexibility Determines Economic Viability.** Our simulations reveal trade-offs between cost and throughput. Computing continuously with grid power completes 100% of jobs with only modest cost savings (5–8%). Computing exclusively with curtailed energy significantly reduces costs (90%) but completes only 54% of jobs. Battery storage bridges this gap, enabling 78% completion while preserving cost advantages.

   These dynamics suggest differentiated computational tiers: baseline capacity for latency-sensitive workloads, opportunistic capacity for delay-tolerant batch jobs powered by curtailed energy, and battery-backed capacity for intermediate flexibility requirements.

To align datacenter deployment decisions with renewable energy generation and grid infrastructure, this paper recommends the following:

## Policy Recommendations

1. **Establish Curtailment Compute Zones:** Designate areas in high-curtailment regions (e.g., PG&E Fresno) for datacenters committed to computing with large quantities of curtailed energy. Support development with streamlined permitting and expedited grid interconnection. Account for Path 15 congestion points to maximize grid benefits while avoiding costly transmission upgrades.

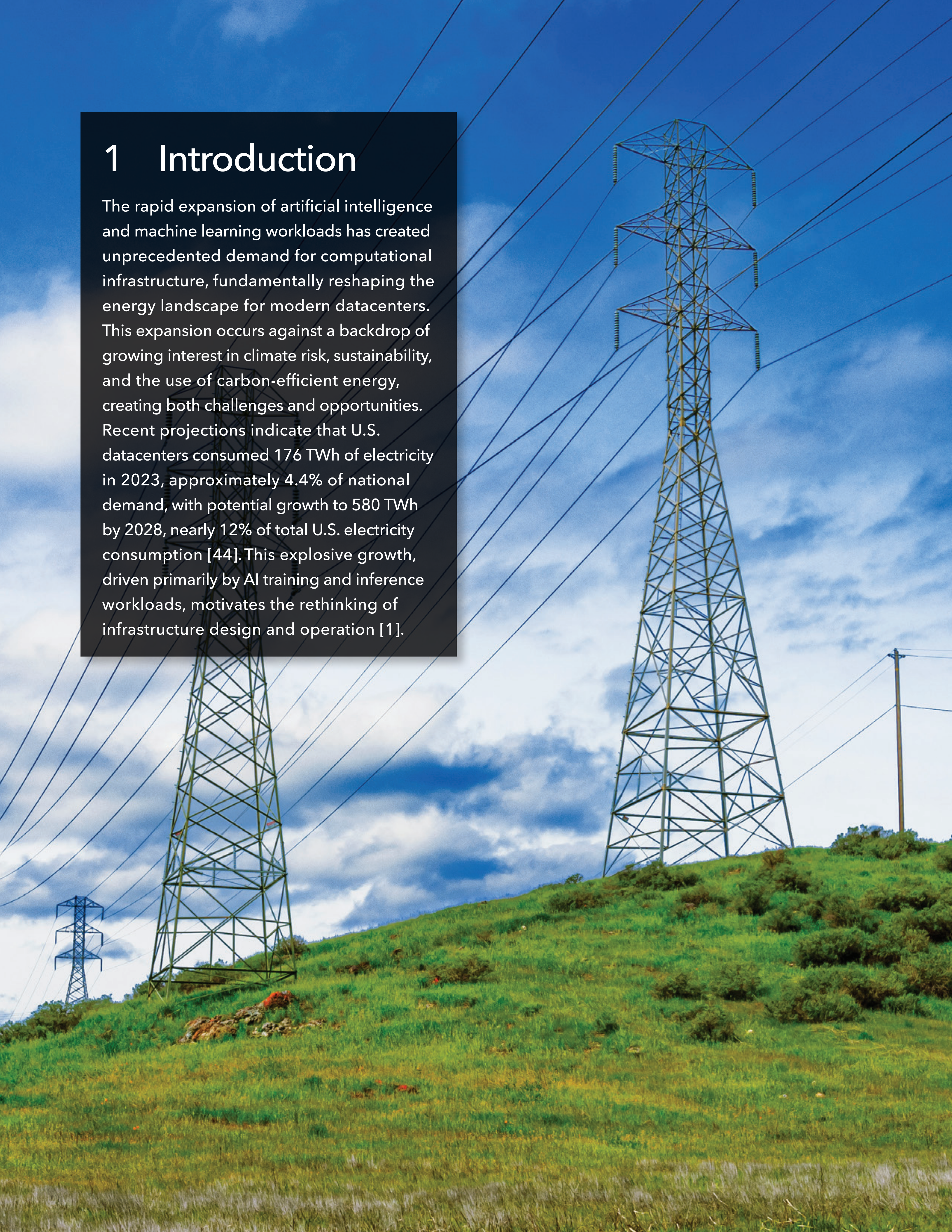
2. **Incentivize Multi-Tenant Flexibility:** Develop frameworks that allow colocation datacenters to deploy operator-owned compute infrastructure alongside tenant equipment, creating tiered service offerings aligned with energy availability. This addresses coordination challenges in multi-tenant facilities while monetizing the 33% unutilized capacity typical in colocation datacenters.

3. **Value Battery Grid Services:** Establish market mechanisms that incentivize datacenter operators to deploy and use batteries for frequency regulation, capacity reserves, and energy arbitrage. Our analysis shows batteries become economically viable when grid service revenues augment compute revenues, offsetting capital costs while enabling computational flexibility.

4. **Leverage Infrastructure Initiatives:** Coordinate the California Middle-Mile Broadband Initiative's 10,000-mile fiber deployment with curtailment zones, addressing the network connectivity gap that has historically disadvantaged rural datacenter sites and rural disadvantaged communities. Priority routing through Fresno and Central Valley curtailment zones could transform these regions into viable computational hubs.

# 1 Introduction

The rapid expansion of artificial intelligence and machine learning workloads has created unprecedented demand for computational infrastructure, fundamentally reshaping the energy landscape for modern datacenters. This expansion occurs against a backdrop of growing interest in climate risk, sustainability, and the use of carbon-efficient energy, creating both challenges and opportunities. Recent projections indicate that U.S. datacenters consumed 176 TWh of electricity in 2023, approximately 4.4% of national demand, with potential growth to 580 TWh by 2028, nearly 12% of total U.S. electricity consumption [44]. This explosive growth, driven primarily by AI training and inference workloads, motivates the rethinking of infrastructure design and operation [1].



Datacenter energy management is a critical sustainability challenge, requiring coordinated advances across hardware design, software scheduling, and infrastructure deployment [31]. Researchers have sought to reduce datacenter emissions by investing in renewable energy, energy storage, and workload scheduling [3]. Scheduling in response to grid signals can reduce emissions by exploiting computation's temporal and spatial flexibility [3, 48], while game-theoretic models attribute carbon costs to individual users and incentivize sustainable computing [3, 20]. However, these software optimizations operate within constraints imposed by datacenter location, grid topology, and renewable energy availability. Practical carbon-aware computing requires aligning datacenter siting decisions with grid dynamics and renewable energy generation patterns.

## 1.1 The California Context: Technology and Infrastructure

California presents a unique convergence of factors that raise interesting questions in sustainable datacenter computing. The state hosts significant datacenter infrastructure, with Silicon Valley and Northern California containing over 625MW of capacity across more than 160 facilities. Depending on the metric, which could account for total capacity, growth rate, and other factors, Silicon Valley ranks between 5th and 7th nationally [15]. This computational infrastructure supports not only California's $3.8 trillion economy but also serves as the foundation for global technology services. The geographic distribution of these datacenters, historically optimized for network connectivity and proximity to technology companies, now faces pressure to adapt to new constraints around energy availability, carbon emissions, and grid stability.

The state's datacenter landscape is evolving in response to AI computational demands. Traditional datacenter designs, optimized for CPU-based workloads with power densities of 3-12 kW per rack, are being replaced or retrofitted to support GPU-intensive AI training systems requiring 30-100 kW per rack. This transition requires not only increased power capacity but also enhanced cooling infrastructure and more sophisticated power management systems.

California's regulatory environment adds both complexity and opportunity to datacenter development. The state's ambitious carbon neutrality goals, codified in Senate Bill 100 requiring 100% clean electricity by 2045 [9], incentivize carbon-efficient computing. However, local air quality regulations restrict the deployment of diesel backup generators commonly used for datacenter reliability. These constraints have encouraged the exploration of battery storage, fuel cells, and grid-responsive operations that align with the approaches investigated in our study.

## 1.2 The California Context: Renewable Energy and Curtailments

California leads the nation in renewable energy deployment, with solar and wind generation providing over 25% of the state's electricity in 2023. The energy system arises from successful public policy, falling technology costs, and abundant solar resources. However, this success has created new challenges in grid management and resource utilization that impact the feasibility of sustainable datacenter operations.

Renewable energy curtailment is the intentional reduction of generated output from renewable energy sources due to grid constraints. In 2024, the California Independent System Operator (CAISO) curtailed approximately 3.4 TWh of renewable generation, representing a 29% increase from 2023 levels [25]. This curtailment, occurring primarily during midday hours when solar generation peaks, represents wasted clean energy and lost economic value for renewable generators. The temporal and spatial patterns of curtailment, concentrated in specific transmission-constrained zones, create opportunities for co-located computational loads that can absorb excess generation.

The root causes of California's curtailment challenge extend beyond simple supply-demand imbalances to include complex constraints from transmission and grid topology. Our analysis of CAISO data reveals that over 70% of curtailments result from local transmission congestion rather than system-wide oversupply. The Path 15 corridor through the Central Valley, connecting Northern and Southern California's grids, experiences congestion during more than 7,300 hours annually, creating persistent bottlenecks that prevent renewable energy from reaching coastal load centers. These structural constraints, which will likely persist despite planned transmission upgrades, lead to pockets of stranded renewable energy that could inform datacenter siting decisions.

Curtailments interact with California's locational marginal pricing (LMP) system to create large differences in energy prices that vary based on grid congestion. Curtailment-heavy nodes often experience negative prices as generators must pay to inject power into the grid while, simultaneously, load centers experience high prices. We find that these price differences can exceed $200/MWh between locations separated by less than 200 miles. Such inefficient resource utilization motivates placing more consumers of energy in curtailment zones.

## 1.3 Research Contributions

This paper advances sustainable datacenter deployment through four methodological and conceptual contributions.

First, we develop an analytical framework for characterizing curtailed renewable energy across temporal, geographic, and causal dimensions. We analyze CAISO operational data (2020–2025), transmission planning documents, and grid forecasts through 2039 to quantify curtailment magnitude, seasonal concentration, growth trajectories, and underlying causes. Our framework distinguishes between system-wide oversupply and local transmission congestion, enabling precise identification of zones where curtailment creates durable opportunities for co-located computation. This methodology extends beyond simple energy availability metrics to assess the structural persistence of curtailment despite planned infrastructure upgrades.

Second, we construct a comprehensive datacenter design and financial model that evaluates three deployment scenarios: urban baseline facilities, rural curtailment-optimized facilities, and rural facilities with battery storage. Our discounted cash flow framework explicitly models the dual constraints facing datacenter development—investor return requirements (15% target IRR) and lender debt service coverage ratios (1.0× minimum in Year 1, escalating to 1.3×). By incorporating differentiated lease-up dynamics, locational cost structures, and electricity price models, we reveal how binding constraints shift across deployment contexts and establish minimum viable rent as the key feasibility metric. Third, we develop an operational simulation framework that couples workload scheduling strategies with grid dynamics. Using scaled Azure traces and hourly CAISO data (marginal prices, carbon intensity, curtailment availability), we quantify the trade-offs between cost optimization and computational throughput under three scheduling policies: as-is baseline, curtailment-only, and carbon-aware. This framework enables the systematic evaluation of how battery storage mitigates the tension between intermittent energy availability and continuous computational demand.

Fourth, we establish a battery sizing methodology that balances capital costs against operational flexibility. Our optimization framework sweeps battery capacity from zero to datacenter load, calculating both total cost and per-job cost to identify U-shaped curves revealing optimal configurations. This approach provides concrete design guidance by quantifying the inflection point where marginal battery costs exceed marginal throughput benefits.

## 1.4 Key Findings and Implications

Our empirical analysis yields critical insights that challenge conventional assumptions about datacenter siting and renewable energy integration.

**Transmission Congestion, Not Oversupply, Drives Curtailment.** Over 70% of California's curtailment stems from local transmission bottlenecks rather than system-wide excess generation. Path 15 corridor congestion will reach 7,300 hours annually by 2039 (84% of all hours), stranding solar generation in the Central Valley even as coastal cities demand power. The PG&E Fresno zone exhibits the highest projected curtailment rate (11% by 2039) as solar capacity grows 83% faster than transmission expansion. These structural imbalances suggest curtailment-based computing could be a durable strategy for sustainable computing anchored in grid topology rather than transient market inefficiencies. Short-Term Bankability, Not Long-Term Profitability, Constrains Viability. Rural curtailment-optimized datacenters reduce capital costs by 34% and deliver higher returns of 28% (versus 15% for urban datacenters), but they require 23% higher rents. This paradox arises because weak initial occupancy forces datacenters to raise rents to service debt. Sensitivity analysis confirms initial occupancy dominates bankability, indicating the challenge is surviving lease-up rather than achieving profitability.

**Battery Storage Extends Computational Throughput but Requires Revenue Diversification.** Computing exclusively with otherwise curtailed energy reduces costs by 10% but the datacenter can only complete 54% of jobs due to the limited availability of that energy. With a battery that stores otherwise curtailed energy, the datacenter can complete 78% of jobs. However, the battery's capital cost is difficult to justify unless the battery permits ancillary grid services, energy arbitrage, or higher rents for datacenter tenants.

**Workload Flexibility Determines Economic Viability.** Scheduling strategies create fundamental trade-offs between cost and throughput. Computing continuously with grid power reduces cost minimally (5–8%) yet completes all tasks. Computing exclusively with inexpensive energy that would otherwise be curtailed costs significantly (90%) but can complete only 54% of jobs. Batteries bridge this gap, retaining the cost reductions while completing 78% of jobs. These trade-offs raise the possibility of differentiated computation tiers priced according to availability guarantees.

## 1.5 Paper Organization

Our study examines whether strategically sited datacenters could economically transform California's curtailed renew-able energy into productive computation. We develop detailed engineering and financial models across three scenarios, using current cost data and grid projections to assess feasibility.

Our analysis should be viewed as one plausible scenario built on reasonable point estimates rather than a definitive forecast. The electricity system is evolving rapidly and factors, including storage technology costs, transmission expansion, competing flexible loads, and policy changes could substantially alter the economics we describe. We aim to provide a rigorous framework for thinking about this opportunity while acknowledging uncertainty in long-term infrastructure planning.

The remainder of this paper is organized as follows. Section 2 characterizes California's curtailment patterns, trans-mission constraints, and pricing dynamics. Section 3 presents our datacenter design framework spanning urban baseline, rural curtailment-optimized, and battery-augmented scenarios. Section 4 develops our financial model examining capital costs, operational expenses, and minimum viable rent. Sections 5-6 present quantitative results including deployment case studies, battery optimization, and workload scheduling analysis. Section 8 situates our work within sustainable computing literature. Section 9 summarizes findings and discusses implications for policy and future research.

# 2 Characterizing Curtailed Energy

Curtailment prevents generated electricity, primarily from solar and wind farms, from reaching the grid. Curtailment is a response to oversupply during periods, such as the middle of a sunny spring day, when solar farms produce massive amounts of electricity that exceed demands from the grid's consumers. At these times, grids curtail excess supply to maintain stability and avoid overloading the system. Curtailment is also a response to grid congestion during periods when solar farms produce energy and consumers demand it, yet the grid lacks capacity in the transmission lines connecting the two. There is often insufficient physical capacity to move electricity from where it's generated (e.g., rural solar farms) to where it is needed (e.g., urban populations). As utilities reduce the carbon intensity of their energy assets with wind and solar farms, the amount of curtailed energy will grow and represent lost opportunities to consume carbon-free energy due to temporal and spatial misalignments between producers and consumers.

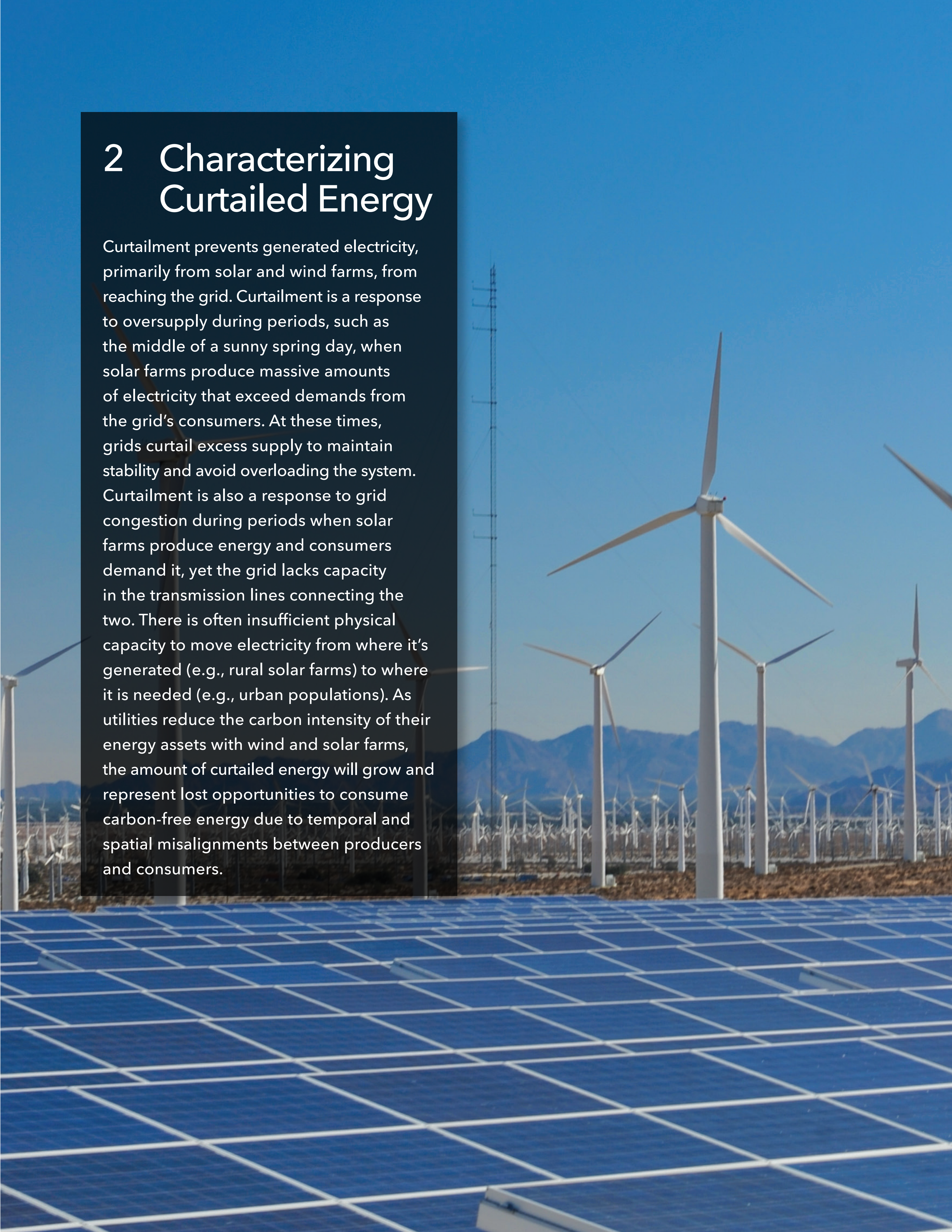
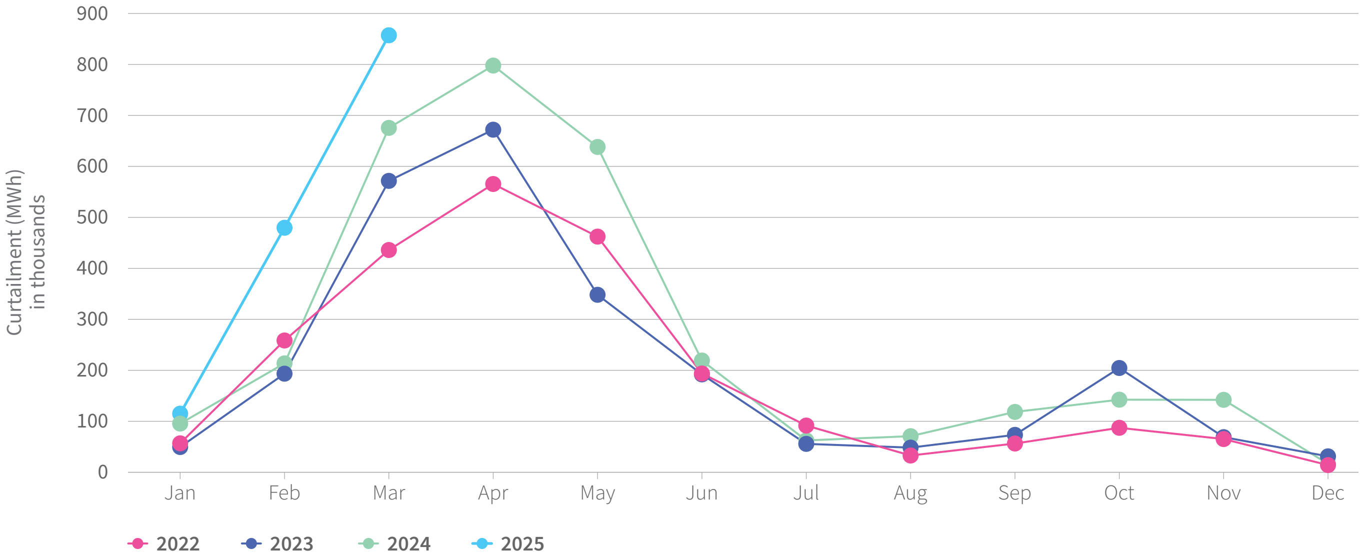
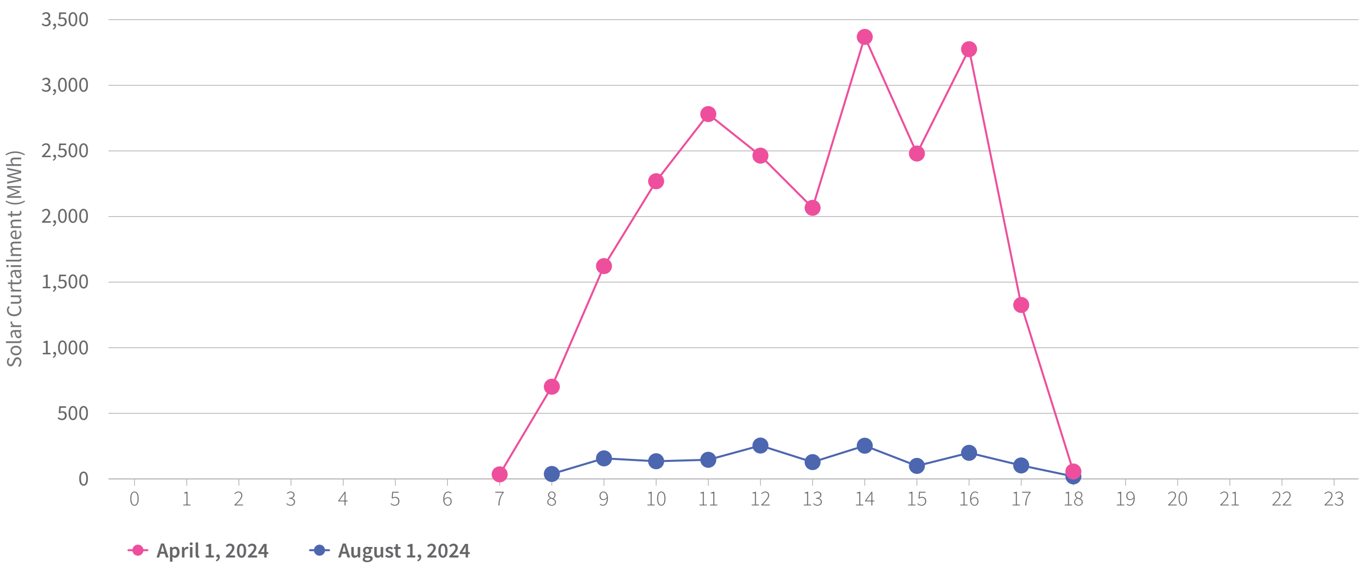


**Fig. 1.** Monthly Solar Curtailment by Year




Note: Curtailed energy is available in very large quantities in the spring months, but negligible quantities in the summer months. The seasonal differences are large and have grown significantly since 2020.

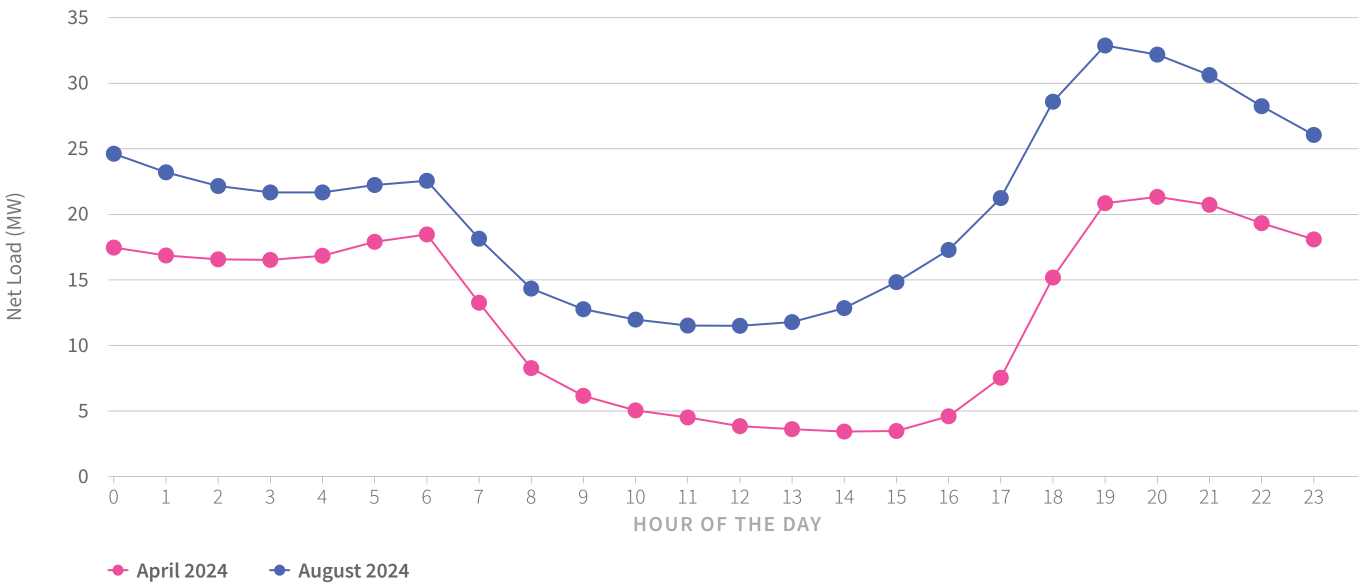
**Fig. 2.** Hourly Solar curtailment Comparison, April 1 vs August 1, 2024




Note: Curtailed energy is available in very large quantities in the spring months, but negligible quantities in the summer months. The seasonal differences are large and appear in fine-grained hourly, mid-day comparisons as well as coarse-grained monthly comparisons.

**Methodology.** Our analysis of renewable energy curtailment in California from 2020 to 2025 reveals significant and growing challenges that arise from structural mismatches between energy generation and consumption. We examine high-resolution data published by the California Independent System Operator (CAISO), which tracks curtailed power in megawatts (MW) at 5-minute intervals. For each 5-minute interval, we multiply each data point by 0.083 = 5 / 60 hours to obtain energy (MWh). We calculate daily totals of curtailed energy by summing these MWh values from the 288 5-minute intervals in a given day. These daily values are then aggregated to produce monthly and yearly totals for longitudinal analysis. Furthermore, we examine the grid's operational state by calculating the system's net load, which is defined as the total electricity demand minus the total electricity sup-

**Fig. 3.** Average Hourly Net Load, April vs August 2024




Note: The classic duck curve illustrates net load, the difference between demand and supply. The curve's dip motivates demand response frameworks in which consumers increase energy usage in the middle of the day."

plied by solar and wind assets. This metric, which is also derived from data tracked at 5-minute intervals, reveals periods of electricity oversupply that lead to curtailment.

**Seasonal Patterns.** Our analysis indicates curtailment exhibits strong, predictable seasonal patterns. Figures 1-2 show that quantities of curtailed energy are concentrated heavily in the spring months. From 2020 to 2025, average daily curtailment during the spring season, from February to June, is consistently and significantly higher than during the rest of the year. This seasonal peak occurs because sunny spring days lead to high solar generation, while milder temperatures result in relatively low midday demand for electricity. This mismatch between supply and demand forces operators to curtail solar production to maintain grid stability. While non-spring months exhibit a consistent baseline curtailment averaging 1,000 to 2,000 MWh per day, spring months exhibit a massive additional and predictable surplus of solar energy.

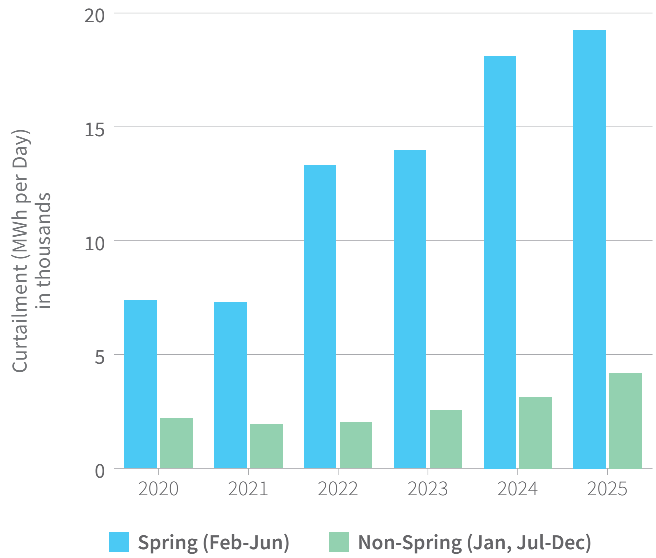
Spring's concentrated curtailment creates a utilization dilemma. Sizing a datacenter to absorb peak spring curtailment would leave expensive servers underutilized during other seasons, poorly amortizing capital costs. Similarly, utility-scale batteries sized for spring curtailment would sit underutilized for more than half the year. System architects must account for this seasonal pattern to ensure the economic feasibility of any infrastructure built to exploit curtailed energy.

**Diurnal Patterns.** Beyond seasonal variations, curtailment follows a distinct and predictable diurnal cycle that corresponds to the well-known Duck Curve, illustrated in Figure 3 by a shape defined by the grid's net load. We find that the Duck Curve is most prominent during spring months. In April, hourly data reveals a large increase in curtailed energy between 9:00AM and 4:00PM that often exceeds 190 MWh. A midday spike in energy due to abundant solar generation causes net load to fall to between 3,000 and 5,000 MW, creating risks of excess generation and requiring the grid to curtail solar assets. Although this is the effect that most researchers have sought to understand and exploit for computing, we note that the Duck Curve is not a consistent phenomenon. In August, hourly data indicates negligible quantities of curtailed energy that rarely exceed 20 MWh.

The predictable diurnal patterns in curtailed solar energy motivate sophisticated strategies for both workload management and energy arbitrage. Datacenters should be designed and managed to support a heterogeneous mix of workloads. High-priority, latency-sensitive tasks could run continuously on reliable grid power as a baseline while a queue of opportunistic, low-priority batches jobs (e.g., AI training, data processing) could be dynamically scheduled to absorb excess and less expensive mid-day energy. Furthermore, datacenters should deploy battery storage to partially decouple computation from generation and diversify revenue streams. If the queue of opportunistic jobs is empty, the datacenter could pivot to energy trading, buying and storing less expensive curtailed energy during the middle of the day and then selling it back to the grid during the evening.

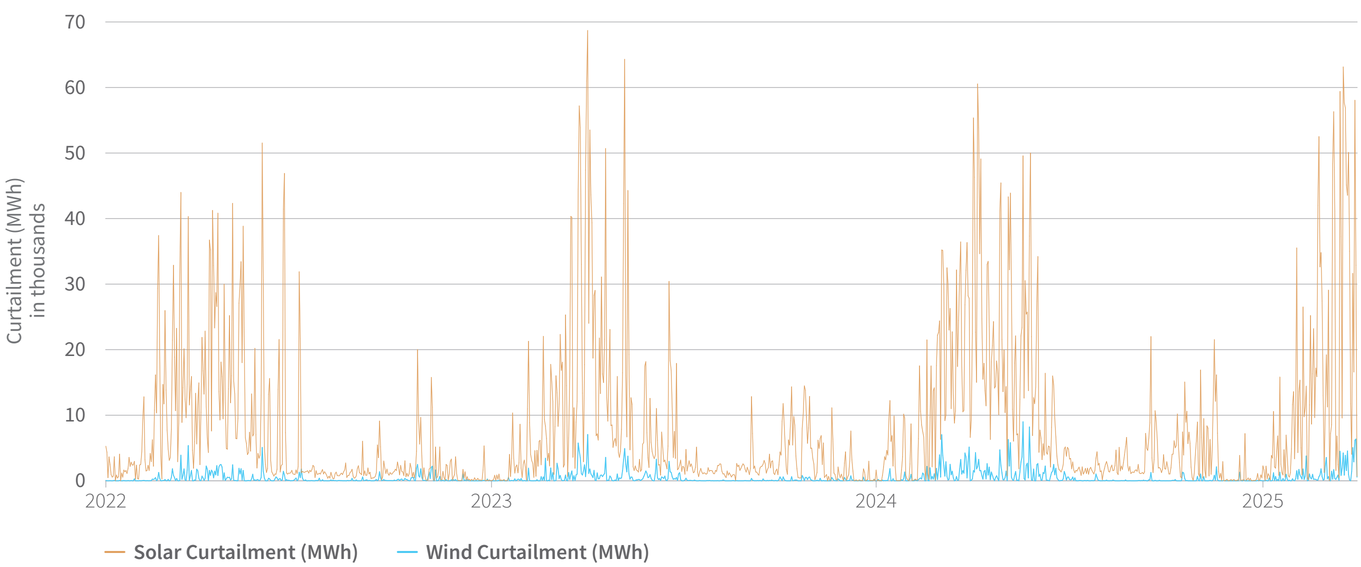
**Accelerating Growth.** Not only are quantities of curtailed renewable energy in California significant and growing, the growth is accelerating. In Figure 4, longitudinal analysis from 2020 to 2025 illustrates increasing spring curtailments and a growing structural mismatch between renewable energy generation and the grid's capacity to absorb it. Over this five-year period, the average daily curtailment during the spring season has more than tripled, reaching 21,000 MWh per day in 2025. The scale of this unused energy is

**Fig. 4.** Average Daily Curtailment by Season and Year



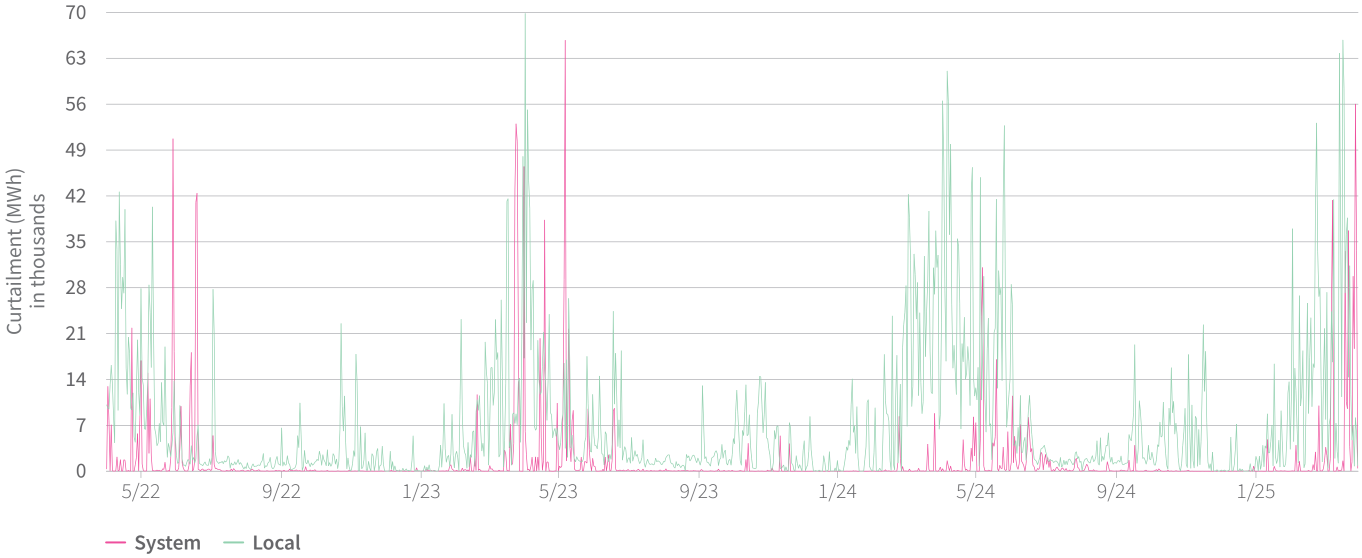

Note: Curtailed energy is available in increasing quantities as solar installations proliferate. Quantities of curtailed energy are increasing rapidly in the spring months but, unfortunately, remain unchanged in other months.

**Fig. 5.** Daily Wind vs. Solar Curtailment (MWh)




Note: Solar, not wind, energy is the predominant source of curtailed renewable energy.

**Fig. 6.** Daily Total Curtailment by Reason (MWh)

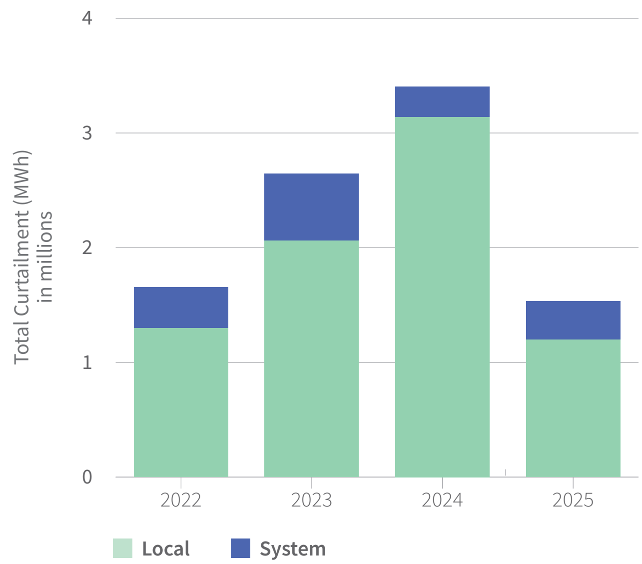



Note: A large majority of daily energy curtailments are due to local congestion rather than lack of system-wide demand.

substantial and equivalent to the energy required to power 1.5 to 2.0 million homes or operate hyperscale datacenters that require hundreds of megawatts of power. The growing magnitude of this curtailed energy aligns with growing demands for datacenter capacity.

**Solar Energy.** The vast majority of renewable energy curtailment in California is attributed to solar, not wind, energy. Figure 5 presents stacked daily cur`tailment data, illustrating how solar curtailment constitutes the overwhelming share of the total, particularly during seasonal peaks. In the spring months, curtailed solar energy is often 10 to 20 times larger than curtailed wind energy. While wind curtailment does occur, it is much smaller in magnitude and exhibits a less obvious seasonal pattern compared to the large, predictable spikes in solar curtailment. Interestingly, the fact that curtailed energy is produced predominantly by solar farms reduces risks in datacenter design and management. Solar, unlike wind energy, is more predictable and requires less complex workload schedulers.

**Fig. 7.** Yearly Curtailment by Reason (Stacked, MWh)




Note: Annual cumulative curtailments show local transmission congestion as the dominant cause, not system oversupply.

# 3 Designing Data Center Infrastructure

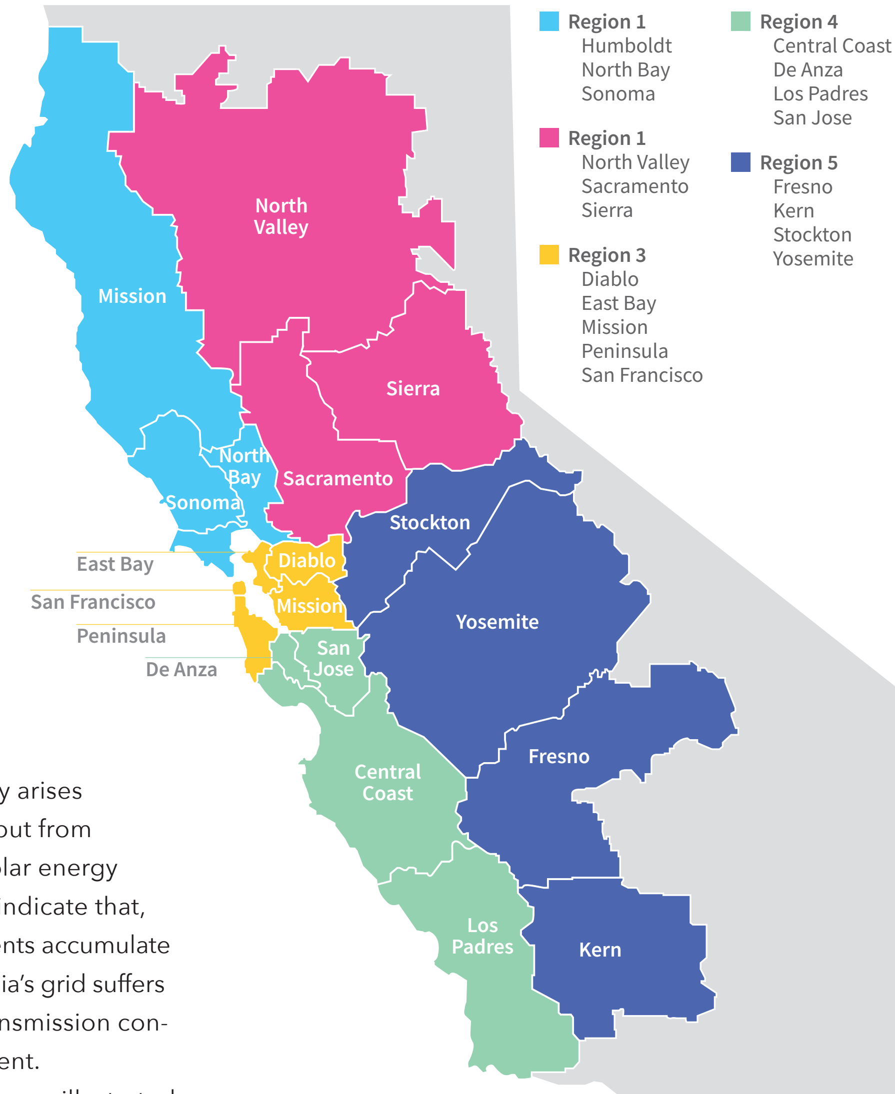
## 3.1 The Siting Problem

California curtails massive quantities of solar energy, but datacenter operators cannot simply determine how much energy is curtailed and could be made available for computation. Operators must also determine where this energy is generated and why it cannot reach existing computational infrastructure. Prior work on demand response and workload migration assumes datacenters can access curtailed energy through software-level optimization or geographic load balancing [39, 52]. However, these approaches overlook fundamental physical constraints from transmission congestion, which prevents curtailed energy from flowing to datacenters.

CAISO data indicates curtailed supply arises not from lack of system-wide demand but from transmission constraints that prevent solar energy from reaching consumers. Figures 6–7 indicate that, over the course of a year, local curtailments accumulate and exceed system curtailments. California's grid suffers from geographic misalignment and transmission constraints more than temporal misalignment.

Transmission Access Charge (TAC) zones, illustrated in Figure 8, correspond to major utility service areas and determine the rates customers pay for grid access. They provide a framework for identifying regions where transmission bottlenecks are most severe and where curtailed energy is stranded. Curtailed energy cannot easily be used to power computation without costly transmission upgrades to move the energy to servers or datacenter construction to consume that energy where it is generated.

**Fig. 8.** PG&E Service Territory



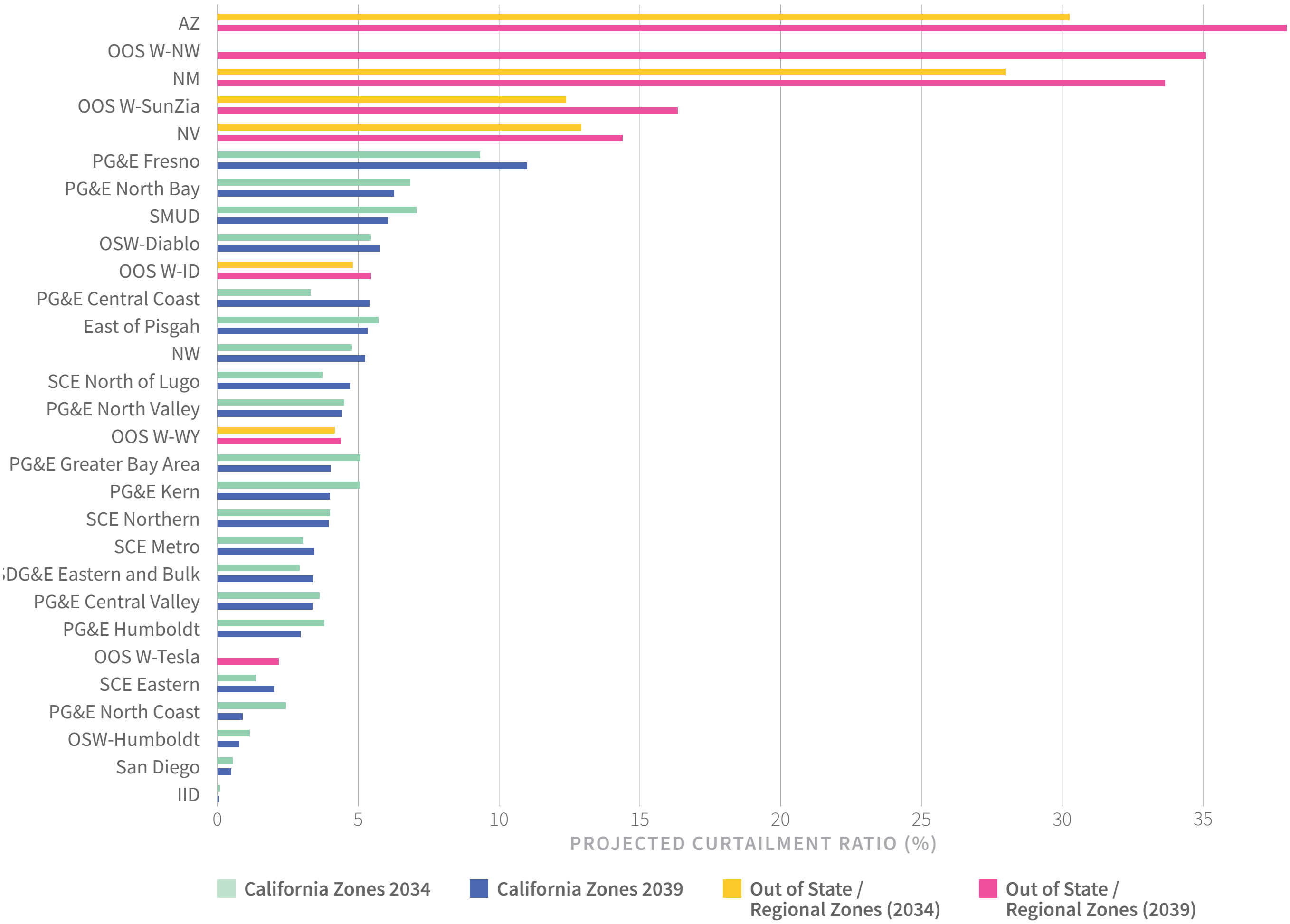

Note: PG&E service areas and transmission access charge (TAC) zones define the economic geography of California's grid.

## 3.2 Opportunities from Curtailed Energy

We identify where new datacenters could be sited to harness curtailed energy by analyzing CAISO's 2024-2025 Transmission Plan, which forecasts grid conditions through 2039 [6]. PG&E Fresno zone faces a projected curtailment rate of nearly 11% by 2039, illustrated in Figure 9, the highest in the state. Curtailment in Fresno and other top-ranked zones is projected to increase from 2034 to 2039, even after accounting for planned transmission upgrades and battery storage projects. Thus, grid congestion is not a transient problem but rather a structural imbalance that will intensify as solar installations outpace transmission capacity.

The primary cause of Fresno's curtailment challenge is localized congestion along Path 15, the transmission corridor connecting Northern and Southern California. As shown in Figure 10, this corridor runs through the Central

**Fig. 9.** Ranked Renewable Curtailment by Zone (2034 vs. 2039 Forecast)



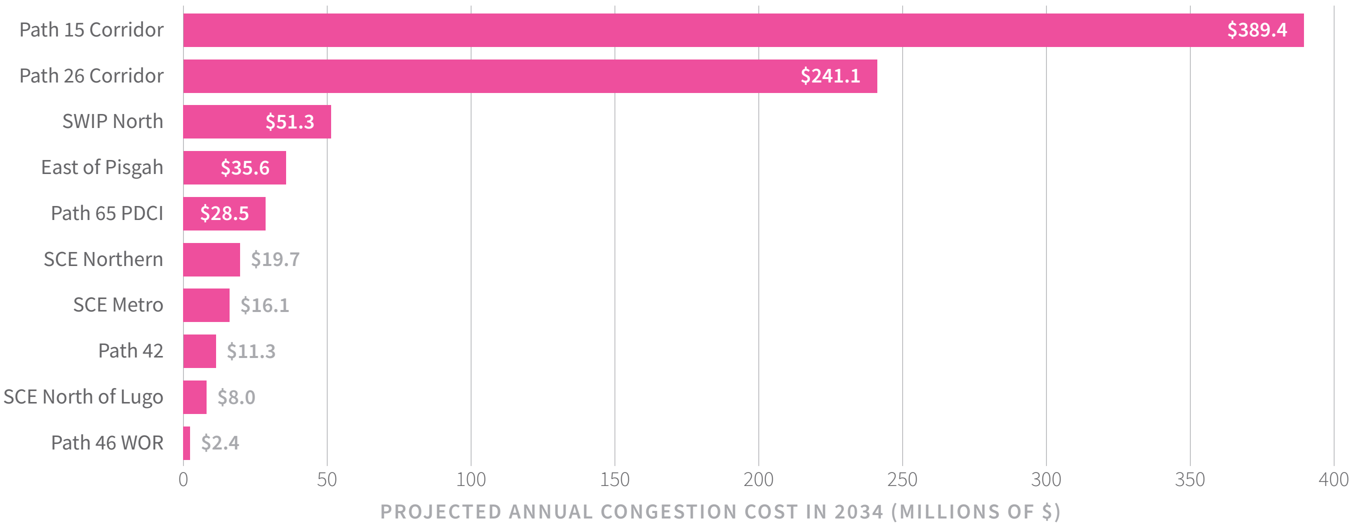

Note: Projected renewable energy curtailment across CAISO TAC zones in 2034 and 2039. PG&E Fresno zone exhibits the highest curtailment rate, which continues to grow despite planned grid upgrades.

Valley and is the state's most significant grid bottleneck. CAISO projects over 7,343 hours of annual congestion by 2039, which is equivalent to 84% of all hours in a year. Grid congestion results from the rapid expansion of solar energy in the region. Figure 11 indicates that Fresno's solar capacity will nearly double from 3,505 MW to 6,430 MW between 2034 and 2039, an 83% increase in just five years. Solar will dominate the region's generation portfolio, yet transmission capacity is insufficient to export this energy to coastal consumers.

**Negative Electricity Prices.** Grid congestion creates economic signals through Locational Marginal Pricing (LMP), which determines real-time electricity costs at specific geographic nodes. Figure 12 illustrates the economic consequence of Path 15 congestion. When Fresno's solar generation peaks during midday, local prices fall sharply and often becomes negative while Bay Area prices remain elevated. Negative prices mean generators must pay the grid to accept their electricity, a signal of stranded oversupply.

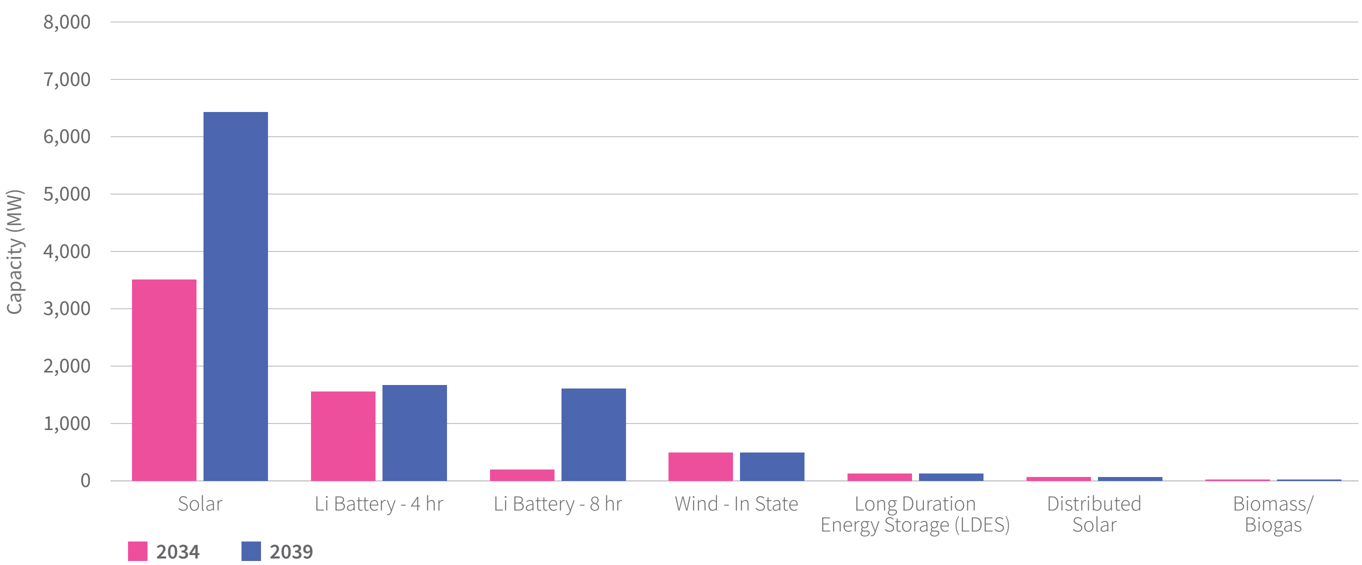
The conventional solution is the construction of high-voltage transmission lines, which CAISO estimates would cost $700M to $1.127B, for projects serving the Bay Area. In this paper, we consider the alternative. Rather than moving energy through wires, move computation to the energy. A 20MW datacenter requires approximately $10M in construction costs plus $15M–$60M in server costs, depending on whether the facility houses CPUs or GPUs. Costs for datacenters are potentially lower than those for transmission upgrades, providing a non-wire solution for absorbing surplus generation while providing value through computation.

**Fig. 10.** Ranked Economics Impact of Grid Congestion (2034 Forecast)




Note: Path 15 transmission corridor through the Central Valley will experience over 7,343 hours of annual congestion by 2039, stranding solar generation in the Fresno region.

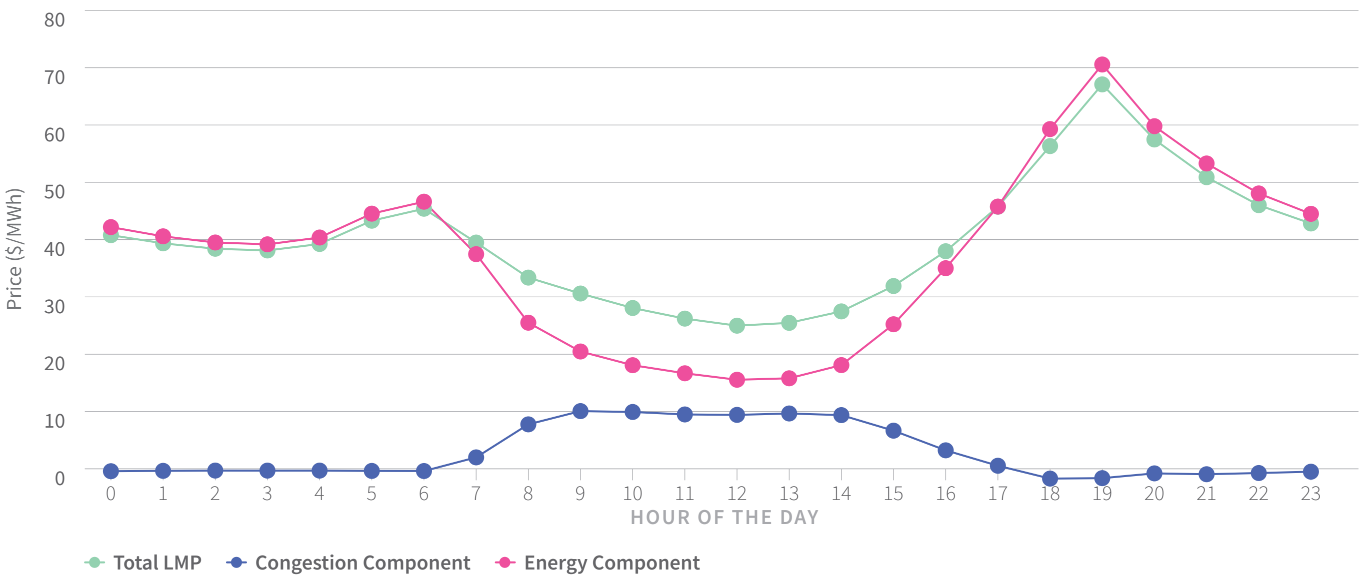
**Fig. 11.** Projected Generation Capacity Mix in PG&E Fresno Area (2034 vs. 2039)




Note: Solar generation capacity in PG&E Fresno region will nearly double by 2039, overwhelming local transmission capacity.

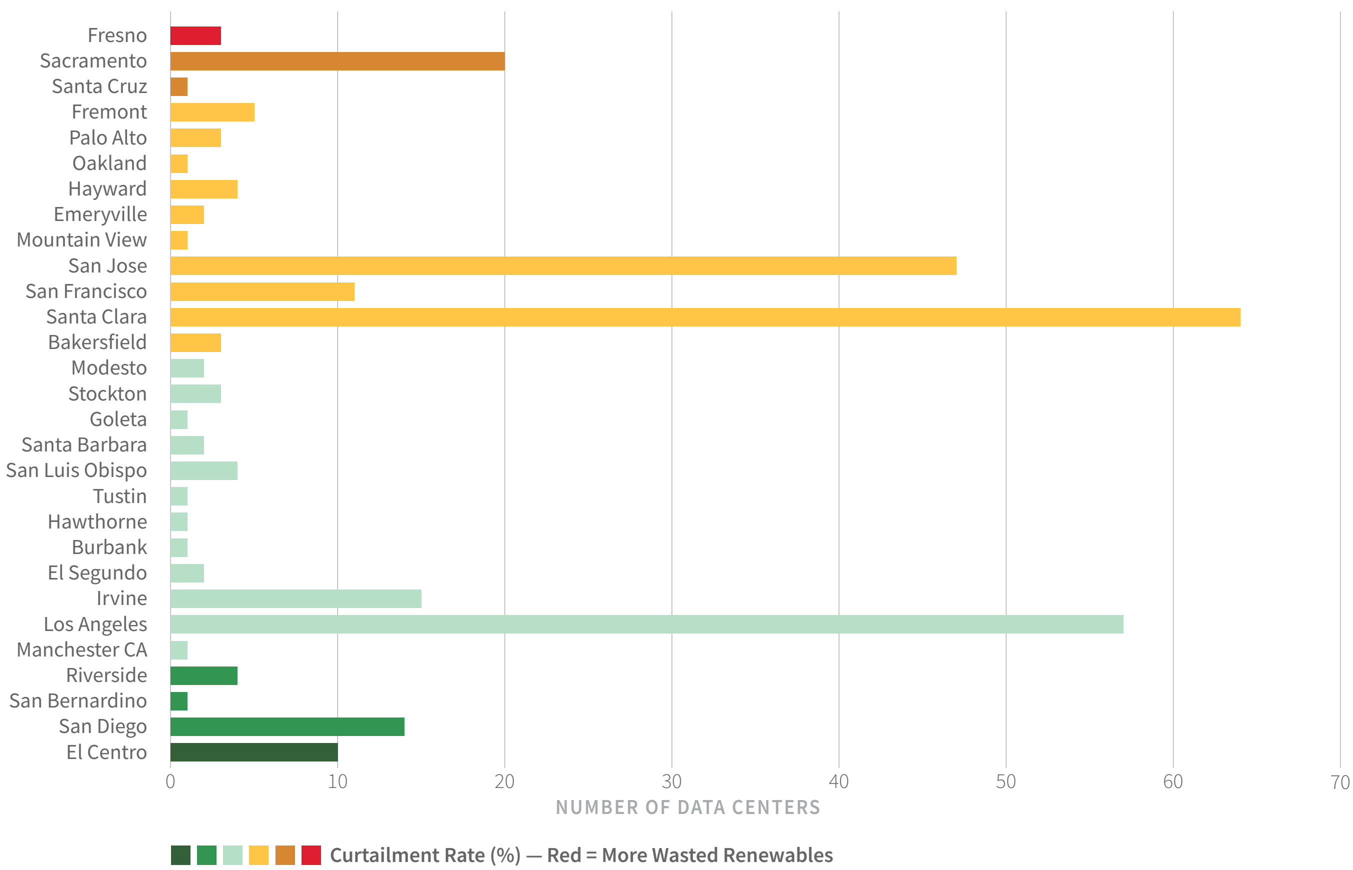
Our geographic analysis in Figure 13 visualizes the disconnect between California's computational and energy resources. The heat map indicates curtailment intensity (red indicates high waste while green indicates efficient utilization) overlaid with datacenter facility counts. Datacenters cluster in Santa Clara, San Jose, and San Francisco with more than 160 facilities accounting for more than 625MW where curtailment is moderate (orange-yellow). In contrast, Fresno and Sacramento experience high rates of curtailment (red) but host few datacenters. This misalignment may arise because datacenters prioritize proximity to Internet Exchange Points (IXPs) over proximity to energy resources; Fresno lacks the former.

**Fig. 12.** Average Hourly LMP and Components (The Duck Curve)




Note: Locational marginal prices diverge dramatically during curtailment periods. Fresno experiences negative prices (oversupply) while Bay Area prices rise (undersupply), demonstrating transmission bottleneck impacts.

**Fig. 13.** California Data Center Markets by Curtailment and Density




Note: California's datacenter infrastructure is geographically misaligned with curtailed energy resources. Major datacenter hubs (Santa Clara, San Jose) show moderate curtailment, while regions with highest curtailment (Fresno, Sacramento) have minimal datacenter presence.

## 3.3 Implementation Challenges and Constraints

Although a location may offer abundant curtailed energy, datacenter operators face several practical challenges that must be weighed against the economic benefits of accessing inexpensive power.

**Network Infrastructure Constraints.** Fresno's lack of a major Internet Exchange Point (IXP) creates both latency and bandwidth constraints. Without local peering facilities, traffic must traverse intermediate networks to reach coastal hubs and end users. This difference – zero IXPs in Fresno versus more than a dozen in the Bay Area [37] – impacts workload feasibility.

Interactive, latency-sensitive applications, such as web serving and real-time APIs, typically require end-to-end response times under 50–100ms [16, 42], making them poorly suited to facilities requiring additional network hops. However, this constraint is largely irrelevant for workloads that are most compatible with computation using curtailed energy. Machine learning training, batch analytics, rendering, and scientific computing operate on timescales of minutes to hours [49, 50], making the effects of additional network latency negligible. In the future, the California Middle-Mile Broadband Initiative may deploy 10,000 miles of state-owned fiber [7], improving Central Valley connectivity.

**Legacy Infrastructure Limitations.** Existing datacenters in the Fresno region were designed for traditional CPU-based workloads, supporting 3–12 kW per rack. Modern AI computation with GPU servers requires 30–40 kW per rack, increasing power density such that existing facilities cannot accommodate AI without significant retrofits to power distribution systems, cooling infrastructure, and physical floor loading capacity. This mismatch between legacy datacenters and emerging computational demands means that leveraging curtailed energy for AI workloads is more likely through the construction of new facilities rather than retrofits of existing ones.

**Multi-Tenant Coordination Challenges.** California's datacenter market is dominated by multi-tenant colocation facilities (313 facilities statewide) rather than hyperscale single-tenant datacenters. In this model, providers like Equinix, Digital Realty, and CoreSite supply physical infrastructure—rack space, redundant power, and cooling—while individual tenants control their own workloads through Kubernetes, Slurm, or VMware orchestration platforms. This decentralized control structure presents coordination challenges for demand response programs that require synchronized load reduction across multiple independent tenants.

However, multi-tenant facilities might successfully leverage curtailed energy. First, actual datacenter power consumption averages only 67% of provisioned nameplate capacity [8]. This utilization gap exists because facilities provision power infrastructure upfront to enable rapid tenant expansion, but physical capacity fills gradually over multiple years. Additionally, operators maintain N+1 redundancy reserves that create further headroom. Second, many tenants already deploy auto-scaling frameworks and run inherently delay-tolerant workloads such as batch analytics and model training—precisely the computational patterns suited to intermittent energy availability.

Rather than requiring complex tenant coordination, colocation operators could deploy their own compute infrastructure alongside tenant equipment and power that compute opportunistically when curtailed energy is abundant. This operator-owned compute could be offered to existing tenants as an additional service tier with dynamic pricing that reflects real-time energy availability. Such arrangements create natural incentives: operators monetize otherwise-idle infrastructure capacity, while tenants gain access to computation at rates substantially below standard colocation pricing.

## 3.4 Datacenter Design Scenarios

We use discounted cash flow analysis to systematically explore the datacenter design space. Rather than seek a single "optimal" configuration, we define three distinct scenarios that span conventional urban facilities (Scenario A) to curtailment-optimized rural facilities with energy storage (Scenario C). Each scenario makes different trade-offs between capital costs, operational costs, workload flexibility, and revenue diversification. Our analysis seeks to understand the conditions under which each design approach becomes viable.

**Scenario A (Urban Baseline).** A traditional 20MW colocation datacenter in Silicon Valley, representing the status quo. This facility prioritizes network latency and customer proximity, relies entirely on grid power at standard commercial rates, and cannot easily access curtailed energy due to transmission constraints. This scenario serves as our economic and operational baseline.

**Scenario B (Rural Curtailment-Only).** A new 20MW colocation datacenter sited in Fresno County, strategically positioned to consume locally curtailed solar energy. This design bypasses Path 15 congestion by consuming energy near its generation, accessing near-zero marginal cost power during curtailment windows, which account for 54% of hours in a year in our simulations. Without on-site energy storage, computation must be flexible and delay-tolerant.

**Scenario C (Rural + Battery Storage).** Scenario B augmented with a 10MW/40MWh Battery Energy Storage System (BESS). Storage decouples computation from instantaneous generation. Batteries permit the datacenter to store curtailed energy during midday for computation in the evening, smooth intermittent supply for less-flexible computation, and perform energy arbitrage and grid services when demand for computation is low. This scenario explores whether the battery's capital costs can be justified by increased operational flexibility and revenue diversification.

We study these scenarios to explore new opportunities for siting datacenters and provisioning their energy infrastructure. Siting in rural locations rather than urban centers may reduce the capital costs associated with land and datacenter construction. Siting near curtailed energy rather than Internet Exchange Points may reduce the operational costs associated with electricity. On-site batteries will increase capital costs, but they may also increase computational throughput and revenue from ancillary grid services. We develop parameterized models to more precisely assess these opportunities.

# 4 Modeling Economic Feasibility

We assess the feasibility of computing using curtailed energy with three complementary analysis frameworks. First, we develop an economic model that calculates the minimum viable rent required for each colocation datacenter to satisfy investor return targets and lender debt covenants. Second, we compare datacenter scenarios with sensitivity analyses, identifying infrastructure parameters that most significantly impact feasibility. Third, we examine operational dynamics, simulating workload scheduling and battery dispatch strategies.

The analysis indicates that the economic viability of rural datacenters that use curtailed energy hinges not only on long-term economic viability but also on the short-term bankability of the project during the initial lease-up period when infrastructure costs are incurred but utilization is low. While these datacenters must charge higher rents to tenants than urban facilities to satisfy lender covenants, they can deliver higher long-term returns.

## 4.1 Datacenter Scenarios

We define three datacenter infrastructure scenarios that span the design space from conventional urban deployment to curtailment-optimized rural facilities. Each scenario makes distinct trade-offs between capital expenditure, operational costs, energy access, workload flexibility, and revenue diversification.

**Scenario A (Urban Baseline).** A traditional 20 MW colocation datacenter in Silicon Valley (Santa Clara), representing the status quo. This facility prioritizes network latency and customer proximity, relies entirely on grid power at standard commercial rates, and cannot easily access curtailed energy due to Path 15 transmission constraints. The datacenter benefits from proximity to major Internet Exchange Points (IXPs), enabling high-margin interconnection services (modeled as 8% of base rent). This scenario serves as our economic and operational baseline against which rural alternatives are evaluated.

**Scenario B (Rural Curtailment-Only).** A new 20 MW colocation datacenter sited in Fresno County, strategically positioned to consume locally curtailed solar energy. This design bypasses Path 15 congestion by consuming energy near its generation point, accessing near-zero marginal cost power during curtailment windows. Our simulations indicate curtailment availability during 54% of hours annually, concentrated in spring months and midday hours. Without on-site energy storage, computation must be temporally flexible and delay-tolerant, limiting viable workloads to machine learning training, batch analytics, rendering, and scientific computing. The facility trades urban interconnection revenue for rural cost advantages: dramatically lower land costs ($150K vs. $2.5M per acre), reduced construction costs ($9M vs. $12M per MW), and access to inexpensive curtailed power.

**Scenario C (Rural + Battery Storage).** Scenario B augmented with a 10 MW / 40 MWh Battery Energy Storage System (BESS). Energy storage decouples computation from instantaneous curtailment availability, providing three operational benefits: (1) store midday curtailed energy for evening computation when curtailment is unavailable, (2) smooth intermittent supply to support less-flexible computational workloads, and (3) enable energy arbitrage and grid services (frequency regulation, capacity payments) when computational demand is low. However, the battery adds $12M in capital expenditure, increasing total project costs by 6.5%. This scenario explores whether battery capital costs can be justified by increased operational flexibility, higher computational throughput, and revenue diversification through grid services.

## 4.2 Parameters and Assumptions

Our analysis accounts for a 25-year horizon consistent with depreciation schedules and loan amortization periods in infrastructure finance [17]. Terminal value, representing the asset's worth at sale, is calculated by applying a terminal multiple to final-year net operating income. We use different multiples for our datacenter scenarios, specifying 10.0× for Scenario A (Metro) and 8.5× for Scenarios B and C (Rural). These multiples reflect higher perceived risks for facilities in less-established markets. The higher Metro multiple also captures the long-term value of interconnection service revenues, which are concentrated in network-dense IXP hubs.

**Financing.** We assume 55% of datacenter projects are financed with debt; 60–70% leverage is common [4]. The interest rate is set to 7.5% and is aligned with industry reports on colocation datacenters where tenants assume all major property operating costs (taxes, insurance, maintenance) [4]. This lease structure creates predictable net cash flows that justify lower borrowing rates. The loan amortizes over 25 years with an Interest-Only (I-O) period during lease-up that we assume is 2 years for Scenario A and 3 years for Scenarios B and C. This I-O period is standard and ensures the project need not make full interest and principal payments before revenue stabilizes [24].

**Table 1.** Key DCF Model Parameters and Assumptions.

| Parameter Category | Parameter | Description & Assumption |
|---|---|---|
| *Global Project Scope* | scale_mw | Total IT load capacity of the data center (20.0 MW). |
| | years | Full investment horizon and analysis period (25 years). |
| | discount_rate | Annual rate for discounting future cash flows to present value (8.0%). |
| | rent_escalator | Assumed annual contractual increase in base rent (2.5%). |
| | opex_escalator | Assumed annual inflation for non-power operating costs (3.0%). |
| *Financing & Valuation* | debt_ratio | Loan-to-Cost (LTC); percentage of CapEx financed with debt (55.0%). |
| | debt_rate | Annual fixed interest rate on the development loan (7.5%). |
| | debt_term_years | Total amortization period of the loan (25 years). |
| | io_years | Interest-Only period; aligns with lease-up (2 for A; 3 for B/C). |
| | terminal_multiple | Exit multiple applied to final year's NOI to calculate sale price (10.0x for A; 8.5x for B/C). |
| *Scenario-Specific Costs* | total_capex | Total all-in capital expenditure to build (Varies: A, B, C; see Table 7). |
| | opex_y1 | Year 1 non-power operating expenditure (Varies: A, B, C; see Table 7). |
| | capex_incentive_pct | Upfront reduction of CapEx from incentives (0% for A/C; 5% for B). |
| *Revenue & Lease-Up* | initial_utilization | Occupancy rate in Year 1; reflects pre-leasing success (70% for A; 30% for B/C). |
| | target_utilization | Stabilized, long-term occupancy rate (85.0% for all). |
| | lease_up_years | Time (in years) to ramp from initial to target utilization (2 for A; 3 for B/C). |
| | service_revenue_pct | Ancillary (e.g., interconnection) revenue as a % of base rent (8% for A; 0% for B/C). |
| | renewables_uplift_pct | Rent premium for renewable-aligned facility (0% for A; 2% for B/C). |
| | bess_revenue_y1 | Annual Year 1 revenue from BESS grid services ($140,000 for C only). |
| *Bankability Targets* | TARGET_IRR | Minimum required 25-year Levered IRR for equity investors (15.0%). |
| | MIN_YIELD | Minimum required steady-state yield (Final Year NOI / CapEx) (9.0%). |
| | DSCR_RAMP | Minimum Debt Service Coverage Ratio (NOI / Debt Payment) for Y1 (1.00x), Y2 (1.15x), Y3+ (1.30x). |

**Revenue Model and Lease-Up Dynamics.** We assume rents increase at 2.5% per year. This is more conservative than market reports indicating 3–5% increases and much lower than overall growth in market rents, which exceeded 12% per year in 2024 [13, 24]. Beyond leasing space and power, colocation operators could generate revenue from ancillary services such as interconnection, the physical and virtual links between cloud providers, carriers, and other tenants. Interconnection revenue is most concentrated in established IXP-dense hubs. We model this revenue as 8% of base rent for Scenario A, a conservative assumption given that market leaders like Equinix report interconnection alone contributes approximately 17% of total revenue [27]. For Scenarios B and C, we conservatively assume 0% service revenue, reflecting Fresno's lack of a major IXP. Finally, we model a modest 2% rent premium for Scenarios B and C to reflect growing tenant willingness to pay for sustainability credentials [18].

A new datacenter does not open at full capacity. We model differentiated lease-up curves based on market maturity. For Scenario A (Metro), we assume 70% initial utilization in Year 1 and ramp to 85% over 2 years, which reflects pre-leasing dynamics in established, high-demand markets like Santa Clara where low vacancy rates cause tenants to commit to leases 18 to 24 months before datacenter construction completes [13, 24]. In contrast, for Scenarios B & C (Rural), we assume 30% initial utilization and ramp to 85% over 3 years. This conservative assumption reflects risks of an unproven market. Fresno lacks a major IXP and presents potential latency disadvantages, requiring the project to establish new datacenter demand rather than simply capture excess demand.

**Table 2.** Total Capital Expenditure (CapEx) Breakdown by Scenario

| Component | Scenario A | Scenario B | Scenario C |
|---|---|---|---|
| *Total Land Cost* | $37,500,000 | $2,250,000 | $2,250,000 |
| *Total Construction Cost* | $240,000,000 | $180,000,000 | $180,000,000 |
| *Total Site Prep Cost* | $300,000 | $600,000 | $600,000 |
| *Total Fiber Cost* | $150,000 | $600,000 | $600,000 |
| *Total Grid Interconnection* | $15,000,000 | $10,000,000 | $10,000,000 |
| *Total BESS CapEx* | $0 | $0 | $12,000,000 |
| **Final CapEx** | **$277,950,000** | **$183,450,000** | **$195,450,000** |

**Operating and Power Cost Structure.** Colocation datacenters often pass power costs through to tenants [12, 34]. Tenants receive separate power bills, and the datacenter operator does not pay for their electricity use. The operator incurs operational costs only for staffing, maintenance, taxes, and transmission access charges (TACs). Under this cost structure, we expect a rural datacenter that uses inexpensive curtailed energy offers two distinct benefits. First, the datacenter operator will incur lower capital and non-power operating costs due to the rural location. Second, the datacenter tenant will incur lower electricity bills due to the inexpensive curtailed energy.

We assume that operating costs increase by 3.0% per year to account for inflation. This rate is intentionally above the 2.5% per year increase in rents. This assumption is conservative because it implies gradual reductions in profit margins.

## 4.3 Capital Expenditure Breakdown

**Land.** Location creates an order-of-magnitude difference in land costs. Urban datacenter sites in Santa Clara cost $2.0–$2.5M per acre [29], reflecting competition for scarce industrial land in Silicon Valley. Rural sites in Fresno cost $150K per acre [30], a 10–17× reduction. A 20 MW datacenter may require 15 acres and incur costs of $37.5M (Scenario A) versus $2.25M (Scenarios B and C). Urban sites benefit from existing infrastructure that requires minimal preparation ($20K per acre) while rural sites require grading, fencing, utility extensions, and environmental studies ($40K per acre). Yet total site development is cheaper in rural locations.

**Construction.** We model all-in construction costs that include the facility shell, electrical infrastructure, HVAC systems, security, and fire suppression. We estimate this cost at $12M and $9M per MW for urban and rural facilities, respectively. The urban premium reflects higher labor costs in the Bay Area and logistics of construction in dense urban areas. These costs align with industry benchmarks reporting a U.S. national average of $11.7M per MW with higher costs in metro areas [33].

**Grid Interconnection.** Urban facilities are geographically close to substations but suffer from insufficient grid capacity. Utilities, such as PG&E, require developers to fund required transmission infrastructure upfront under Rule 30 to avoid multi-year interconnection delays [38]. We estimate $15M for urban substation upgrades and redundant feeds required for Tier III reliability. We estimate $10M for rural interconnection costs due to less substation congestion but longer connection distances.

**Network Connectivity.** Urban datacenters benefit from proximity to dense fiber infrastructure, incurring minimal costs such as $150K for 1 mile of lateral connection. Rural datacenters have historically faced prohibitive trenching costs to reach backbone fiber. However, the California Middle-Mile Broadband Initiative, which deploys 10,000 miles of state-owned fiber with over 3,000 miles under construction, is transforming rural connectivity [7]. We model 4 miles of fiber connection at $150K per mile, totaling $600K for rural scenarios, which represents the cost of reaching Middle-Mile Initiative access points.

**Battery Energy Storage (Scenario C Only).** The Battery Energy Storage System adds $12M in capital costs. For a 20MW datacenter, we model costs for 40 MWh of battery capacity in the form of a 10MW battery with a 4-hour duration. Additionally $8 per kW is incurred for safety systems and $2M is incurred for power conversion systems. Battery costs have declined from over $600 per kWh (2019) to $250 (2024) with projections reaching $150 by 2030 [35]. Table 2 summarizes total capital costs by component and scenario. The urban facility (Scenario A) faces a $94M premium over the rural, curtailment-only facility (Scenario B), driven primarily by land costs ($35M difference) and construction premiums ($60M difference). Adding battery storage (Scenario C) increases rural capital costs by $12M, a modest increment relative to the urban-rural difference but significant in terms of financing burden during the initial lease-up period.

**Table 3.** Total Annual Operational Expenditure (OpEx) Breakdown by Scenario

| Component | Scenario A | Scenario B | Scenario C |
|---|---|---|---|
| *Annual Electricity Cost (tenant)* | $19,970,000 | $18,040,000 | $17,900,000 |
| *Annual TAC Cost* | $2,444,040 | $2,444,040 | $2,444,040 |
| *Annual Staffing + Remote* | $3,500,000 | $2,750,000 | $3,250,000 |
| *Annual Property Tax* | $3,052,500 | $2,278,125 | $2,278,125 |
| *Annual BESS O&M Cost* | $0 | $0 | $300,000 |
| *Annual BESS SaaS + Insurance* | $0 | $0 | $325,000 |
| *Annualized Risk of Downtime Cost* | $2,000 | $400,000 | $200,000 |
| **Total Annual OpEx** | **$28,968,540** | **$25,912,165** | **$26,697,165** |

## 4.4 Operational Expenditure Breakdown

**Electricity Costs.** Power represents the largest operational expense for datacenters, but its treatment in our model requires careful explanation. In the colocation business model, operators do not pay for tenant power consumption and tenants receive separate utility bills [12, 34]. Thus, our model excludes tenant electricity costs and includes only power-related expenses incurred directly by the operator such as common area power and standby generation fuel costs.

Electricity costs in Table 3 quantify the tenant's power costs, which matter for our analysis because they represent the value proposition to tenants Scenario B and C tenants pay 9.7% less for electricity ($85–$86 per MWh) compared to Scenario A tenants ($95 per MWh), providing a financial advantage that is independent of colocation rent. This cost advantage stems from access to curtailed energy during the 54% of hours in a year when locational marginal prices in Fresno approach zero or turn negative.

**Transmission Access Charges.** TAC is a significant operational cost borne by the datacenter operator. California utilities assess an annual fee per kilowatt of contracted demand to recover transmission infrastructure costs. PG&E's current TAC rate is $10.84 per kW-year [11], resulting in $2.44M annual costs for a 20 MW facility (20,000 kW × $10.84). This methodological choice conservatively estimates higher TAC costs for infrastructure finance. Specifically, it assumes datacenter operators budget for the full reservation of grid capacity at nameplate demand to ensure reliability regardless of actual energy use. Large industrial customers, like datacenters, often face billing structures, such as PG&E's Rule 30, that include fixed demand charges or upfront capital contributions to recover the cost of transmission infrastructure deployed and reserved for their use. By applying TAC at full contracted capacity, our analysis establishes a conservative financial baseline.

This assumption is conservative for the rural datacenter scenarios we evaluate. Under actual volumetric billing, Scenario B's 54% operational rate would incur TAC costs that are nearly half those of a datacenter operating continuously, further improving its already favorable rate of return. Thus, our fixed-cost approach shows that curtail-to-compute strategies are feasible even under the most conservative transmission cost assumptions.

**Staffing and Maintenance.** Colocation facilities require 24/7 staffing for security, facility management, and technical support. We model annual staffing costs at $3.5M for Scenario A, $2.5M for Scenario B, and $3.0M for Scenario C. Urban facilities face higher labor costs due to Bay Area wages but benefit from larger workforces. Rural facilities offer lower base wages but incur an additional $250K annual premium for on-call specialist travel and wages required to attract technicians and contractors to more distant locations [14, 53]. Scenario C incurs a further $500K in staffing costs to support battery operations, monitoring, and grid service dispatch. Maintenance costs beyond labor are modeled as 1.0% of total CapEx annually, a standard industry assumption.

**Table 4.** Bankability Analysis by Scenario with Differentiated Lease-Up Risk

| Scenario | Min Viable Rent ($/MW-yr) | Levered IRR | Min DSCR (Year 1) | Binding Constraint |
|---|---|---|---|---|
| *A (Metro)* | $2,000,732 | 15.32% | 1.85× | Investor IRR |
| *B (Rural Flex)* | $2,459,961 | 27.76% | 1.00× | Lender DSCR |
| *C (Rural BESS)* | $2,708,984 | 27.75% | 1.00× | Lender DSCR |

**Property Taxes.** We model effective property tax rates of 1.10% (urban) and 1.25% (rural), with the rural premium reflecting higher local assessment millage rates in counties dependent on commercial property tax revenue. Property taxes are assessed on total CapEx value, resulting in annual costs of $3.05M (Scenario A), $2.28M (Scenario B), and $2.44M (Scenario C).

**Insurance and Risk Costs.** Datacenter operators carry multiple insurance policies: property insurance (covering facility replacement), business interruption insurance (covering revenue loss during outages), and liability insurance. Base insurance costs are modeled at $200K–$300K annually. Scenario C adds a $200K premium for battery-specific coverage, addressing fire risk, thermal runaway, and potential grid interconnection liabilities associated with energy storage systems.

We also model expected downtime costs, calculated as the product of annual outage probability and revenue loss per hour. Urban facilities benefit from redundant grid connections and robust utility infrastructure, resulting in negligible expected downtime costs ($2K annually, representing 0.01% outage risk). Rural facilities face higher grid reliability risks. We model a 20% annual probability of experiencing a 2-hour outage in Scenario B, and 10% in Scenario C as the battery provides backup power, with downtime valued at $250K per hour due to lost tenant revenue and SLA penalties. This results in annualized risk costs of $400K (Scenario B) and $200K (Scenario C).

**Battery Operating Costs (Scenario C Only).** Battery systems incur ongoing operating expenses beyond initial capital costs. Operations and maintenance (O&M) is modeled at 2.5% of battery CapEx annually ($300K) to include cell monitoring, thermal management, and periodic capacity testing. Energy Management System (EMS) software is needed for grid service dispatch and market participation ($125K annually). Additional insurance for battery fire risk is required ($200K, noted above). These recurring costs sum to $625K annually and must be justified by battery-enabled revenues. Table 3 summarizes total annual operating costs by component and scenario. Note that Year 1 costs serve as the baseline and we assume costs increase at 3% annually. Rural scenarios (B and C) achieve 10.6–7.8% OpEx reduction versus urban (A), driven primarily by lower staffing costs and reduced property taxes. However, these savings are partially offset by TAC (which applies equally to all scenarios) and higher premiums for maintenance and downtime risk.

## 4.5 Minimum Viable Rent (MVR) Solver

A bankable project delivers industry-standard returns to equity investments while satisfying lender debt service requirements. The investor requirement ensures an internal rate of return (IRR) that ensures long-term profitability on the equity invested. We set a target IRR of 15%, a conservative baseline consistent with industry benchmarks for stabilized datacenter assets [21, 40]. Development projects with higher risk profiles must offer higher returns between 25–40% IRR [2].

The financing constraint specifies a debt service coverage ratio (DSCR) that ensures short-term solvency by requiring cash flow to cover annual debt payments. We set targets that require net operating incomes to exceed debt service payments by an increasing margin. In year one, DSCR ≥ 1.00× and cash flows must be at least equal to debt payments. In years two and three, DSCR ≥ 1.15×

and DSCR ≥ 1.30×, creating 15% and 30% cash flow buffers for increasingly stable operation. This tiered structure is standard practice [40], ensuring the project can survive low-revenue lease-up years (Year 1) while requiring a healthy cash flow cushion once stabilized. The Year 1 covenant is particularly critical for speculative developments in unproven markets, where weak initial occupancy can strain debt service capacity.

Our model determines the minimum rent (in $/MW-year) that satisfies both investor IRR and lender DSCR thresholds. The solver operates as follows. First, initialize base rent at a low value (e.g., $1.5M/MW-year). Determine 25-year cash flows given model parameters and calculate IRR and DSCR values. If both IRR ≥ 15% and annual DSCRs clear thresholds, record as candidate MVR. Otherwise, increment rent and repeat.

This procedure reveals which constraint binds. If the solver stops at low rent with IRR near 15%, the investor constraint bound while the lender was easily satisfied. If the solver requires high rent to meet DSCR thresholds, the lender constraint bound and the investor receives windfall IRR above 15%, which could be viewed as compensation for accepting the project's high lease-up risk. A lender-constrained datacenter faces short-term survival risk during lease-up but may offer high long-term returns. An investor-constrained datacenter is financially stable but may offer lower returns. Thus, understanding which constraint is binding reveals feasibility in two dimensions.

# 5 Comparing Datacenter Scenarios

Table 4 presents the minimum viable rent (MVR), revealing fundamentally different financial constraints for each datacenter design scenario.

**Scenario A.** The urban baseline achieves an MVR of $2.00M per MW-year, which is set by the investor requirement. Strong initial occupancy (70% in Year 1) generates robust early cash flow, allowing the facility to achieve Year 1 DSCR of 1.85×, far exceeding the 1.00× lender requirement. Because the lender is easily satisfied, the binding constraint becomes the investor's requirement for 15% returns. The solver stops at $2.00M per MW-year because charging higher rent would generate excessive returns without improving project viability.

**Scenario B.** The rural curtailment-only facility requires an MVR of $2.46M per MW-year, determined entirely by lender constraints rather than investor return targets. Weak initial occupancy (30% in Year 1) constrains early cash flow, forcing premium rents to achieve the 1.00× minimum DSCR and avoid loan default. This lender-driven requirement generates an IRR of 27.76%, nearly double the investor's 15% target. Thus, the challenge is short-term bankability, not long-term profitability. While Scenario B delivers 81% higher returns than urban facilities, low initial occupancy requires higher rents to mitigate lender risk.

**Scenario C.** The battery-augmented scenario exhibits the same lender-constrained dynamics as Scenario B but requires even higher rent of $2.71M per MW-year. The $12M battery investment increases annual debt service by $1.2M (assuming 55% debt financing at 7.5%), further straining Year 1 cash flow and forcing a $249K per MW-year premium over Scenario B. Under our conservative assumption of $140K annual battery revenue from grid services, this creates a $109K shortfall that must be recovered through higher rent.

Three additional developments could make Scenario C viable. First, battery revenue from energy arbitrage, frequency regulation, and capacity payments, could exceed our conservative $140K baseline. Second, more computation enabled by energy storage could support workloads that pay higher rents for more consistent power. Third, continued battery cost declines could halve required capital costs and reduce rents.

Figure 14 visualizes cash flow dynamics for each datacenter scenario at its respective MVR. The left panels show cumulative IRR over time, while the right panels track cumulative equity value. For Scenario A, the IRR line (blue) asymptotically approaches the 15% target (red dashed line), illustrating how the investor constraint is binding. For Scenarios B and C, the IRR lines cross the 15% target early, around Year 5, and finish well above it, illustrating how the lender constraint forced colocation datacenter rents higher than necessary for investor returns.

## 5.1 Calibration Against Market Rates

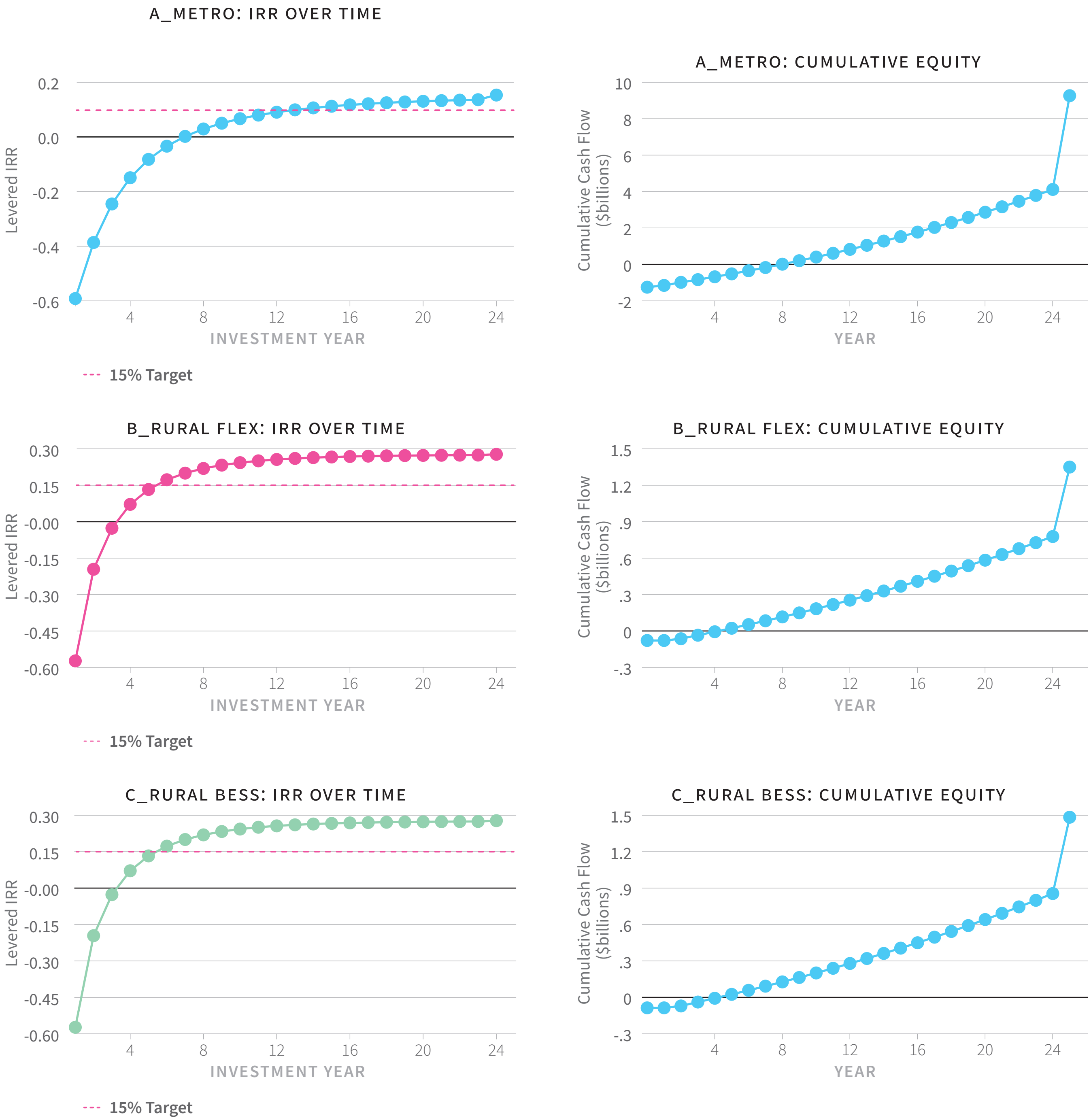
Current rates in metropolitan markets are $160–$170 per kW-month [5], equivalent to $1.92–$2.04M per MW-year. Scenario A's MVR of $2.00M per MW-year falls squarely within this range, which validates our model. In contrast, Scenarios B and C require rents 24% and 37% above metro rates, respectively. This premium presents a challenge as tenants must be willing to pay $2.46–$2.71M per MW-year for rural capacity despite lacking the interconnection services and latency benefits of urban facilities. Yet in return for this premium, tenants could realize several benefits.

First, tenants might reduce their electricity bills. As detailed in §4.4, tenants in rural locations pay $85–$86 per MWh for electricity versus $95 per MWh in metro locations, a 9.7% reduction. A 1 MW tenant consuming 8.76 GWh annually would save $78K–$87K per year, partially offsetting higher rents.

Second, tenants might pay a premium for carbon-free computing and we model a modest 2% sustainability premium for Scenarios B and C. Industry surveys suggest enterprises are willing to pay 5–10% premiums for renewable energy [18]. If tenants value sustainability at the upper end of this range, the net rent difference narrows substantially. For AI training workloads, where a single model can consume megawatt-hours and generate significant Scope 2 emissions, access to curtailed solar energy could justify premium pricing as companies face increasing pressure to decarbonize computation.

Third, our 30% initial occupancy in rural colocation datacenters may be pessimistic. A committed anchor

**Fig. 14.** Internal Rate of Return and Cumulative Equity by Scenario




Note: Cash flow analysis at the minimum viable rent (MVR), illustrating internal rate of return and cumulative equity for each datacenter scenario. A's MVR is set by the 15% investor constraint while Scenarios B and C's MVRs are set by lender's 1.00x DSCR constraint.

tenant that commits to 50–70% of capacity prior to construction would eliminate the Year 1 DSCR constraint and lower the required rent toward Scenario A's levels. Recent market dynamics, where 73–83% of capacity is pre-leased 18–24 months prior to delivery [13, 24], suggest this is possible if datacenter operators target AI workloads that are delay-tolerant and align with the intermittent availability of curtailed energy.

## 5.2 Sensitivity Analysis

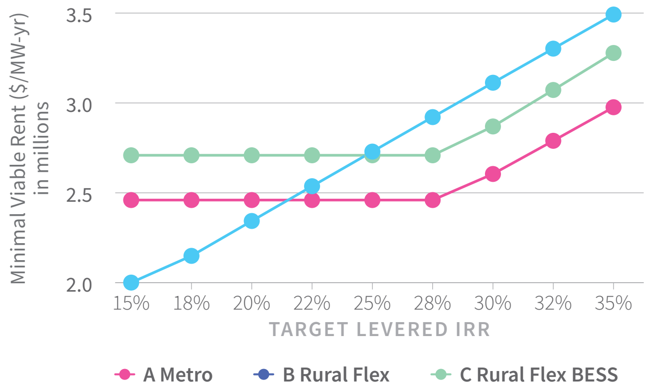
**Sensitivity to Investment Return Targets.** Figure 15 demonstrates how minimum viable rent varies with target IRR across scenarios. We sweep the target IRR from 15% to 35%, representing the range from stabilized asset returns to high-risk development project returns [2, 21].

For Scenario A (Metro), MVR increases linearly with target IRR from the 15% starting point. This confirms the investor constraint is binding: every incremental required return translates directly to higher required rent. The line begins at $2.00M per MW-year at the 15% target and rises to $3.08M per MW-year at the 35% target.

For Scenarios B and C (Rural), MVR remains flat at approximately $2.46M and $2.71M per MW-year, respectively, until the target IRR exceeds the project's floor IRR of 27.7%. This flat section confirms these scenarios are lender-constrained. Specifically, rent cannot decrease further without violating the Year 1 debt covenant regardless of the investor's willingness to accept lower returns. Only when target IRR exceeds 28% does the investor constraint begin to bind and required rent begins rising.

This analysis reveals how binding constraints create distinct regimes across return expectations. Below 25% target IRR, Scenario A dominates through its lowest absolute rent ($2.00M per MW-year at 15% target) due to strong pre-leasing. A critical inflection occurs between 25–30% target IRR, making rural deployment cost-competitive. Scenario A must charge $2.50–$2.79M per MW-year to deliver these returns while Scenario B's $2.46M per MW-year rent generates 27.76% IRR as a consequence of lender constraints. Beyond 30% target IRR, all scenarios require higher rent to satisfy investor requirements though rural facilities maintain structural cost advantages from lower costs throughout this range.

**Fig. 15.** Minimum Viable Rent vs. Target IRR




Note: The plot visualizes the different binding constraints. The MVR for Scenarios B and C remains "flat" at its DSCR-floor level (at $2.46M and $2.71M, respectively) until the Target IRR exceeds the project's "floor" IRR of 27.7%. In contrast, Scenario A's MVR is determined by the Target IRR from the 15% start, increasing linearly from $2.00M/MW-yr.

**Sensitivity to Model Parameters.** To identify which parameters most strongly influence project viability, we sweep each parameter ±20% around its baseline value. Table 5 ranks parameters by their total impact on two key metrics: leveraged IRR (long-term profitability) and Year 1 DSCR (short-term bankability). Total impact depends on the metric's range of values when sweeping parameters.

**Table 5.** Sensitivity Rankings (Scenario B Baseline, ±20% Sweep).

| Parameter | IRR Impact | DSCR Impact |
|---|---|---|
| *base_rent_per_mw_y1* | *13.66%* | *0.84x* |
| *target_utilization* | *12.08%* | *0.00x* |
| *total_capex* | *10.85%* | *0.42x* |
| *debt_ratio* | *6.82%* | *0.42x* |
| *lease_up_years* | *5.44%* | *0.00x* |
| *opex_y1* | *3.12%* | *0.44x* |
| *debt_rate* | *2.24%* | *0.42x* |
| *service_revenue_pct* | *1.69%* | *0.10x* |
| *rent_escalator* | *1.65%* | *0.00x* |
| *initial_utilization* | *1.62%* | *0.84x* |
| *capex_incentive_pct* | *0.55%* | *0.02x* |
| *io_years* | *0.49%* | *0.00x* |
| *opex_escalator* | *0.39%* | *0.00x* |
| *renewables_uplift_pct* | *0.27%* | *0.02x* |
| *terminal_multiple* | *0.11%* | *0.00x* |

Source: Model Output. Impact is the total range of the metric based on the parameter sweep.

Only base rent and total capital cost parameters impact both long-term returns and short-term bankability. Base rent emerges as the single most important parameter because it determines cash flows that impact returns over twenty-five years and lender constraints in the first year. Capital costs rank second because higher construction costs increase debt service and reduce equity returns. These findings validate our focus on rural colocation datacenters and their potential to reduce capital costs.

Several parameters impact return but not bankability. Higher stabilized occupancy after the lease-up period will increase returns while longer lease-up periods will erode returns. These parameters do not impact bankability, which depends primarily on initial occupancy in early years. Similarly, parameters that shape future value, such as rent escalators, exit multiples, and interest-only periods, affect long-term returns but not bankability.

Another set of parameters impacts bankability but not returns. Initial occupancy impacts debt service in early years but not long-term returns. This finding confirms one of our key insights: the 30% initial occupancy assumption for rural colocation datacenters is a short-term finance issue rather than a long-term profitability issue. Rural developers could ease concerns about initial occupancy through committed anchor tenants (e.g., hyperscale companies such as Microsoft and Google) or discounted launch pricing that boosts early occupancy.

We illustrate the top six drivers for each metric. The IRR plot (Figure 16) illustrates proportional impacts (e.g., a 20% increase in rent increases IRR by 20%). The DSCR plot (Figure 17) exhibits steeper slopes for cost parameters, reflecting constraints in early years where small cost increases can cause the datacenter operator to breach the 1.00× lender covenant.

The parameter sensitivity analysis suggest different strategies for building colocation datacenters. Metro datacenters, constrained by target returns not debt service, should optimize for long-term profitability through stabilized occupancy, accelerated lease-up, and controlled costs. Rural datacenters face different challenges. Because lender constraints are restrictive, these operators must prioritize initial occupancy perhaps at lower returns, then pivot to higher tenant rents after the lease-up period ends.

## 5.3 Model Limitations and Assumptions

**Lease-Up and Bankability Risk.** Our model's central finding depends on the differentiated lease-up curve: 70% initial occupancy for Scenario A versus 30% for Scenarios B and C. The 30% assumption is a proxy for the risk of building a datacenter in a location without a major IXP. Our sensitivity analysis (Table 5) confirms this parameter is the primary factor for debt service in year one. The model's findings are highly sensitive to this assumption and a faster-than-expected lease-up in rural areas would significantly lower the MVR for Scenarios B and C, making them more competitive.

**Ancillary Network Services.** We model a differentiated service based on location. For Scenario A, we assume an 8% revenue uplift from services, which is conservative because high-margin revenue for interconnection is typical in IXP-dense hubs like Santa Clara. For Scenarios B and C, we assume 0% service revenue for Scenarios B and C. This reflects the paper's key finding that the rural Fresno site lacks a major IXP and must therefore compete on cost (rent and power), not on low-latency interconnection services. This assumption, however, could be subject to change as rural development improves network connectivity in these areas.

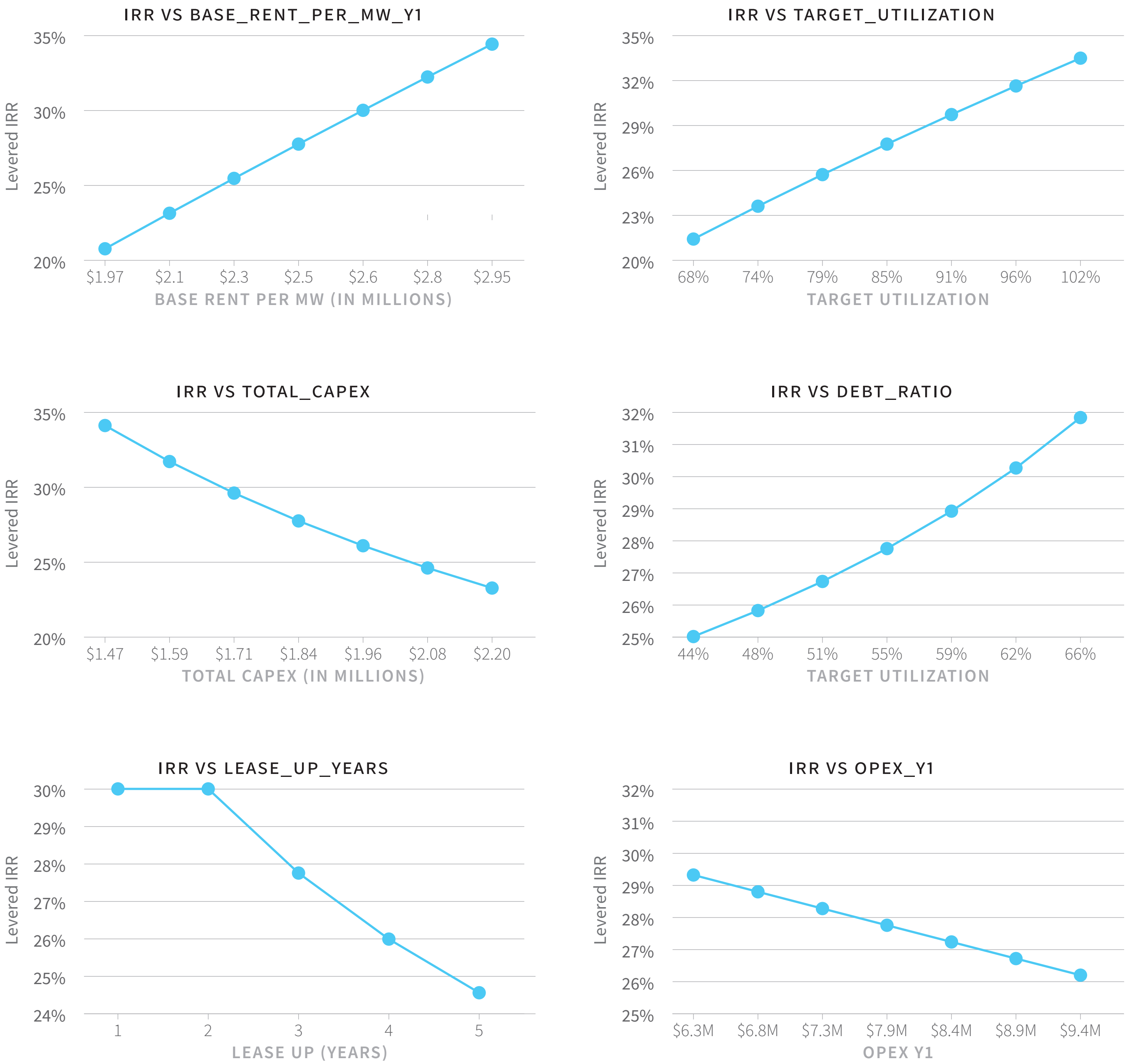
**Ancillary Energy Revenue.** The revenue from ancillary battery services in Scenario C is a significant variable with high uncertainty. We conservatively estimate $140,000/year and neglect larger, more complex potential for energy arbitrage. This minimal revenue is insufficient to offset the battery's capital cost, making Scenario C's MVR approximately $249k/MW-yr more expensive than B's. A full evaluation of the BESS's economic impact would require a more sophisticated model that captures volatile energy price spreads.

**Electricity Cost.** We use CAISO Locational Marginal Pricing (LMP) data and standard utility tariffs. But datacenter operators often negotiate Power Purchase Agreements (PPAs) or special utility programs for lower, predictable rates [34, 45]. The actual cost of power could differ from our estimates.

## 5.4 Discussion and Implications

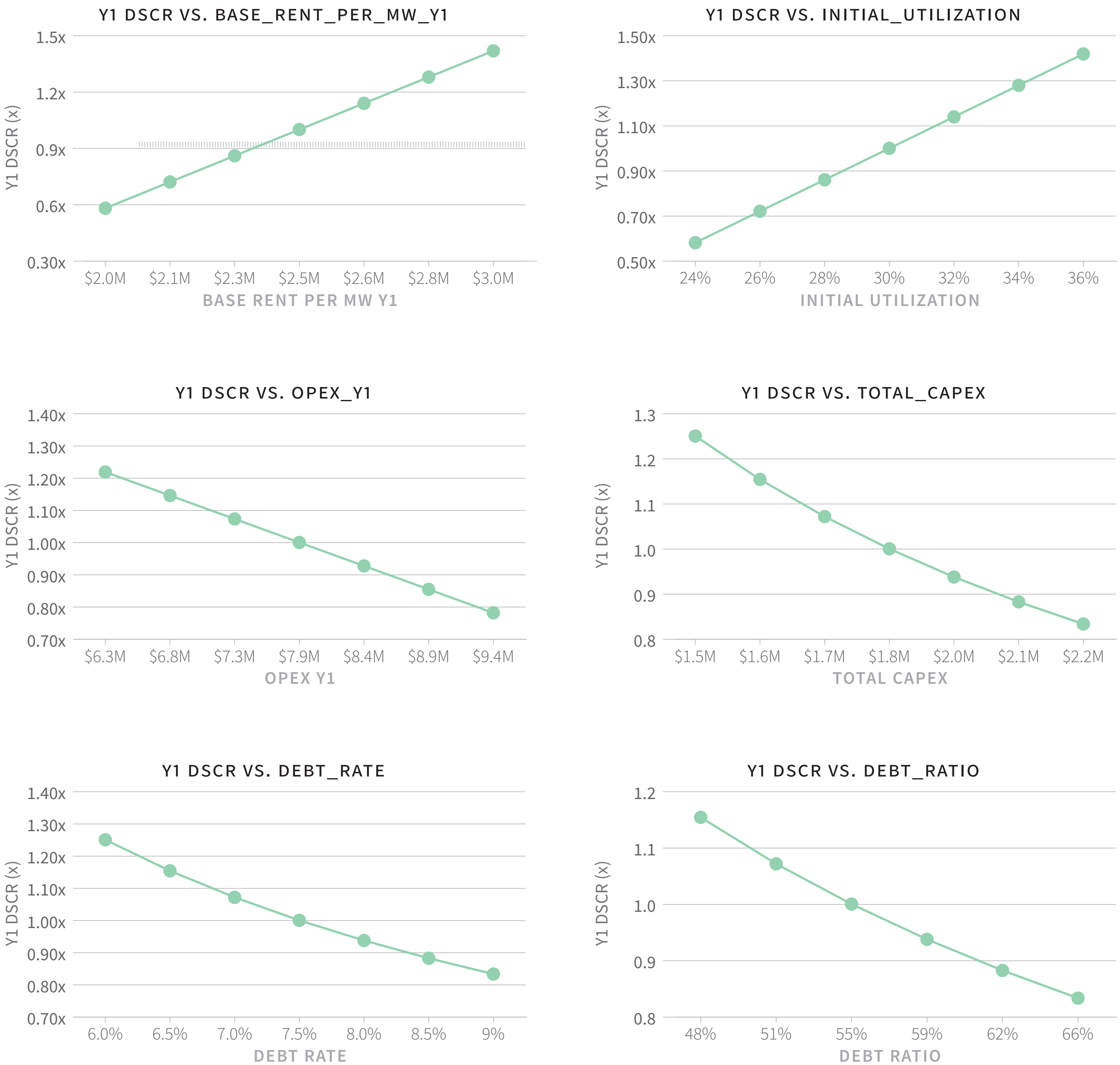
Our analysis reveals that rural curtailment-optimized datacenters are viable when they can overcome a short-term bankability barrier. These datacenters can deliver higher returns (28% IRR versus 15% for urban facilities) but requires higher rents to satisfy Year 1 debt covenants. The bankability barrier diminishes when operators secure strong pre-leasing commitments that increase initial occupancy, if costs for curtailed energy are even lower than expected, or if sustainability mandates support higher rents for green computing. However, urban datacenters will

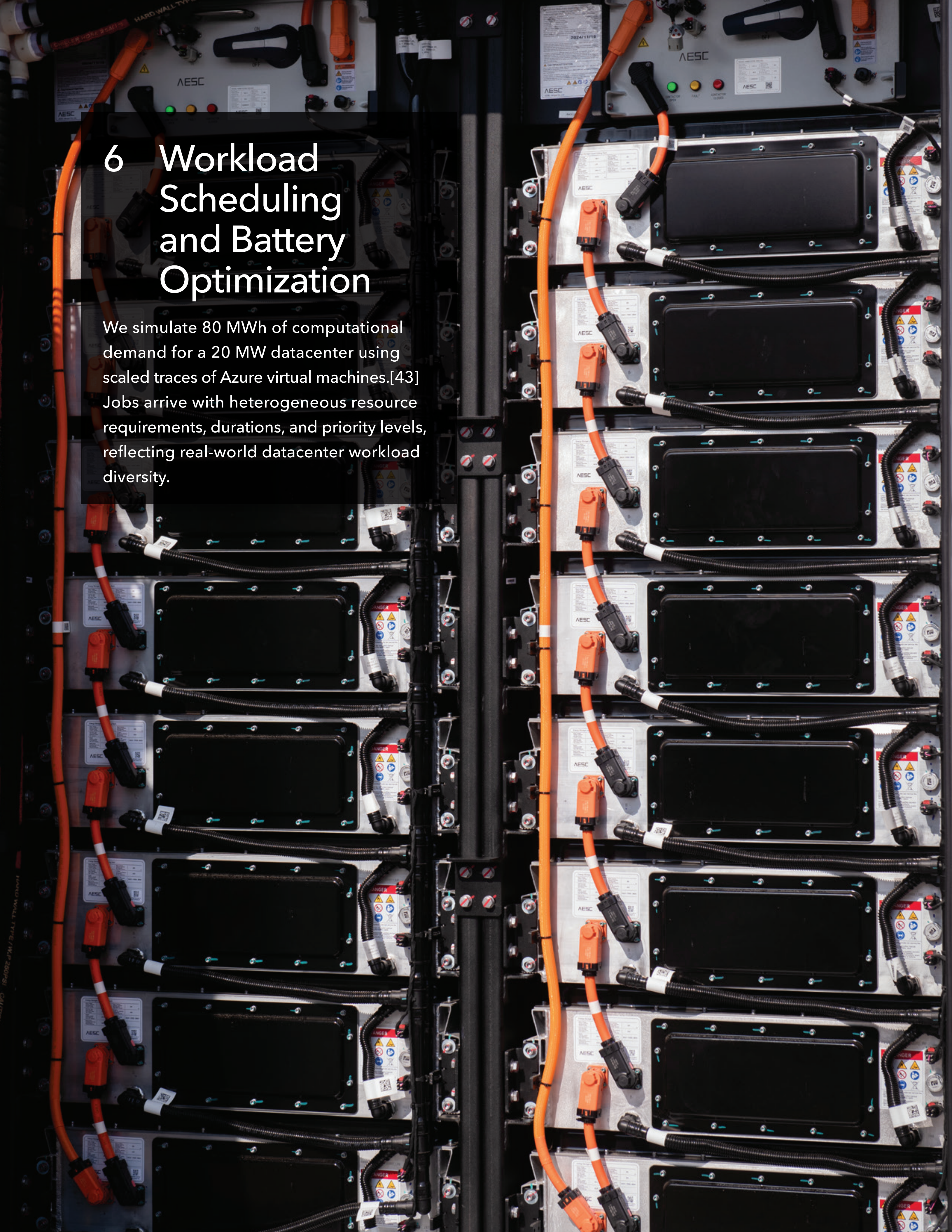
**Fig. 16.** Top 6 Most Sensitive Parameters for Profitability (Levered IRR)




likely remain preferable for workloads that are latency-sensitive or require ancillary data services that rural locations cannot provide. Building datacenters with batteries increases capital costs that our baseline analysis cannot justify. But batteries become viable if the datacenter were to generate additional revenue through energy trading and grid services or if batteries continue to become less expensive. Our simulations optimally size batteries, balancing gains in computational throughput against capital costs, and show how batteries can extend curtailment-only operations from 54% to 78% job completion while maintaining cost advantages.

**Fig. 17.** Top 6 Most Sensitive Parameters for Bankability (Y1 DSCR)






# 6 Workload Scheduling and Battery Optimization

We simulate 80 MWh of computational demand for a 20 MW datacenter using scaled traces of Azure virtual machines.[43] Jobs arrive with heterogeneous resource requirements, durations, and priority levels, reflecting real-world datacenter workload diversity.

**Fig. 18.** Weekly Data Center Operation by Scheduling Policy


Datacenter demand under three scheduling strategies over a week, with and without a 40 MW battery, showing impacts of grid carbon intensity, curtailment, and battery.

To ground the analysis in realistic grid conditions, we use three hourly data sources for the CAISO NP-15 node throughout 2024. First, Locational Marginal Prices from CAISO [10] capture real-time pricing. Second, marginal carbon intensity from the WattTime API captures the emissions rate of the next MWh consumed. Finally, renewable energy curtailment reported by CAISO at 5-minute resolution, of which we assume 40% is accessible given the datacenter's geographic proximity to generation sources.

We implement three workload scheduling strategies. The as-is baseline executes jobs upon arrival, representing conventional datacenter operation with no demand response. The curtailment-only strategy executes jobs exclusively during curtailment windows, dropping all jobs that arrive when curtailment is unavailable, representing the most aggressive curtailment-optimization approach.

The carbon-aware strategy executes all jobs but shifts them to lower-carbon hours when capacity permits, balancing job completion with emissions reduction.

For scenarios with energy storage, we model a 20 MW / 40 MWh Battery Energy Storage System with 90% round-trip efficiency, consistent with today's lithium-ion systems. The battery charges when curtailed energy is available or during low-price hours. It discharges to power computation or provide grid services during high-price hours.

## 6.1 Impact of Scheduling Strategy

Figure 18 illustrates datacenter operation over a representative week, comparing the three scheduling policies. The as-is and carbon-aware policies operate similarly, both completing 100% of jobs. They differ only in timing: carbon-aware shifts jobs to lower-carbon hours when possible, reducing carbon footprint by 12% with negligible impact on completion.

The curtailment-only policy exhibits fundamentally different behavior, operating only during curtailment windows (gray shaded regions). This results in significantly lower computational throughput, completing only 54% of jobs in the simulated week but achieving 90% lower electricity costs per job by accessing near-zero marginal cost curtailed energy. Batteries transform operational dynamics, charging during midday curtailment hours and discharging during evening hours, boosting job completion from 54% to 78% while maintaining access to inexpensive energy. For as-is and carbon-aware policies, the battery enables energy arbitrage, charging when electricity is less expensive or carbon-intensive, discharging otherwise, but does not fundamentally alter job scheduling. These policies achieve modest cost reductions (5–8%) through arbitrage without changing computational throughput.

The analysis reveals inherent tensions in aligning datacenter operations with grid conditions. Curtailment-only scheduling maximizes energy cost savings (90% reduction) but suffers poor utilization (54% job completion) due to intermittent energy availability. Conversely, as-is and carbon-aware strategies maximize utilization (100% completion) but pay standard grid rates, limiting cost savings to modest arbitrage opportunities. Battery storage provides intermediate flexibility, enabling curtailment-only operations to achieve 78% completion by bridging gaps between curtailment windows, allowing operators to balance cost optimization against throughput requirements. These dynamics validate our DCF model's cost assumptions—Scenario B achieves 10% lower electricity costs than Scenario A—while revealing the operational complexity required to capture these savings.

## 6.2 Optimal Battery Sizing

Battery capacity represents a fundamental capital-versus-throughput trade-of where larger batteries extend operational hours but incur proportionally higher capital costs. Figure 19 quantifies this trade-of by sweeping battery capacity from zero to 20 MW and evaluating both total cost and per-job cost across our three scheduling policies, revealing starkly different economic profiles.

For as-is and carbon-aware strategies that complete all jobs regardless of curtailment availability, batteries provide no economic benefit. Although energy arbitrage reduces operating costs modestly (5–8%), these savings are overwhelmed by battery capital amortization, causing per-job cost to rise monotonically with battery size. These policies achieve full job completion without storage, making batteries economically unjustifiable.

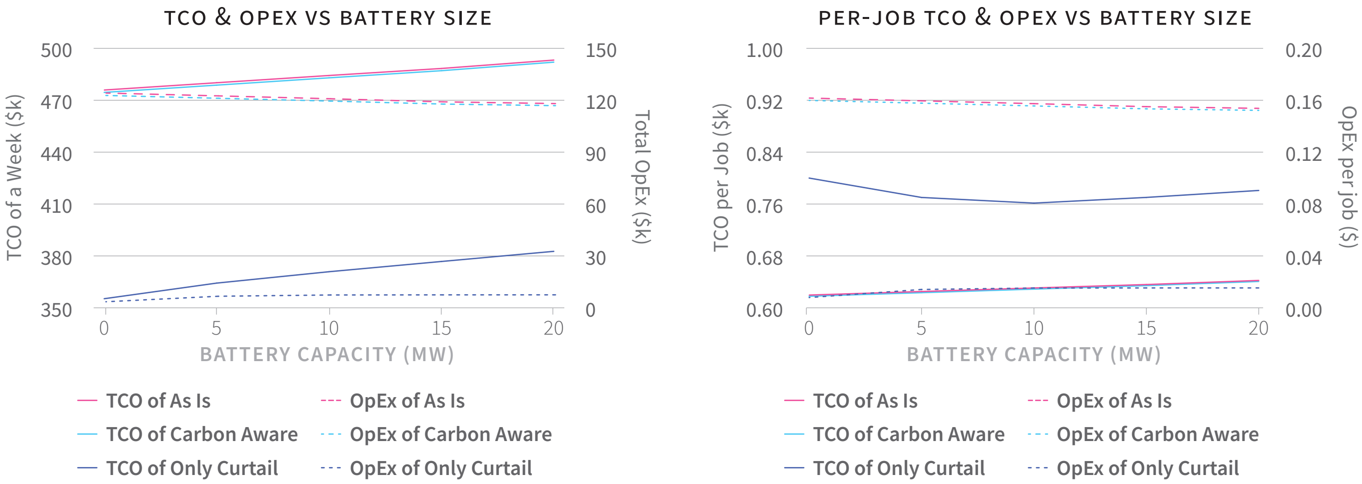
Conversely, the curtailment-only strategy exhibits a U-shaped per-job cost curve that reveals an optimal battery size. Without storage, the facility completes only 54% of jobs, creating high per-job costs through poor amortization of fixed datacenter capital. As battery capacity increases, more jobs are completed, spreading fixed costs across more useful work and initially reducing per-job cost. The curve reaches a minimum at 10 MW battery capacity—a 0.5× ratio to the 20 MW datacenter load, providing 40 MWh storage—where marginal battery costs begin exceeding marginal throughput benefits. At this optimum, curtailment-only operations achieve 78% job completion while maintaining 90% lower operating costs per job compared to grid-powered alternatives, though total per-job costs remain 21% higher due to incomplete server utilization.

This 10 MW optimal sizing provides a concrete design target for Scenario C by balancing three competing requirements: sufficient storage duration (4 hours) to bridge typical curtailment gaps between midday solar peaks, manageable capital burden ($12M for 40 MWh versus $24M for 80 MWh), and sufficient computational throughput to justify fixed datacenter costs. The finding validates our Scenario C parametrization while exposing the fundamental tension between cost minimization through curtailment-only operation and throughput maximization through continuous availability.

## 6.3 Operational Implications for Financial Model

These operational dynamics validate and extend our financial analysis in three critical ways. First, the curtailment-only strategy's 90% operating cost reduction per job aligns closely with our DCF model's 10% electricity cost differential between Scenarios B and A ($18.0M versus $20.0M annually), confirming that our financial projections capture realistic operational economics. Second,

**Fig. 19.** Optimal Battery Sizing by Scenario



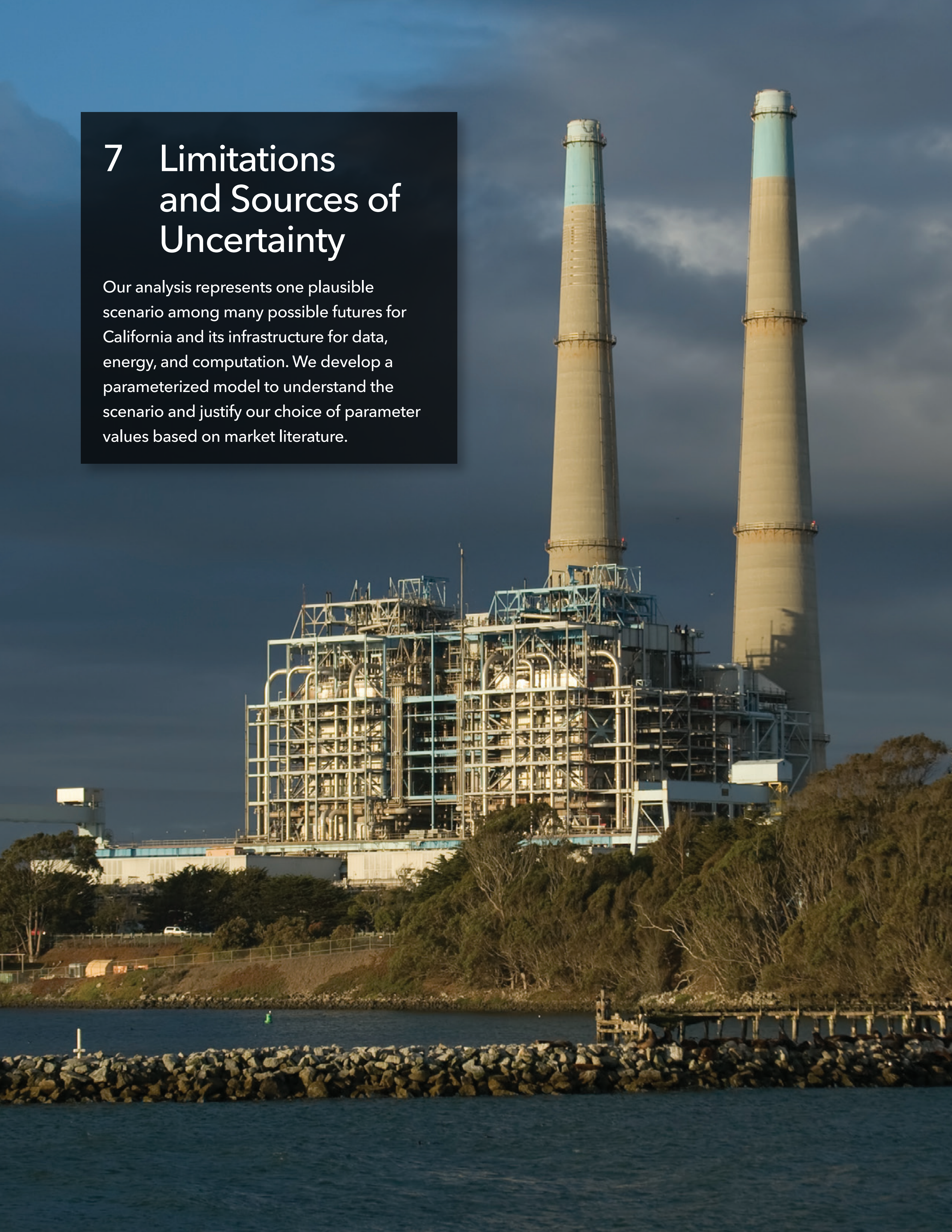

Note: While as-is and carbon-aware strategies show increasing cost and decreasing operational costs with battery capacity, curtailment-only minimizes cost at 10 MW.

our simulation quantifies only computational throughput gains from batteries, neglecting any revenue from buying at negative curtailment prices and selling at peak evening rates and from grid services such as frequency regulation and capacity payments. If these additional streams prove substantial (analysis of grid revenue service models is beyond the scope of this paper), Scenario C's $249K/MW-year rent premium over Scenario B is justified and the 10 MW battery sizing provides a reasonable starting point.

Furthermore, simulation reveals a hybrid operating model that resolves the multi-tenant coordination challenge. Rather than forcing all tenants to accept intermittent power availability, colocation operators can deploy three operational tiers simultaneously: baseline capacity running latency-sensitive workloads on reliable grid power (as-is scheduling), opportunistic capacity running delay-tolerant batch jobs exclusively on curtailed energy (curtailment-only scheduling), and battery-backed capacity supporting intermediate workloads requiring flexibility without continuous availability (carbon-aware scheduling with battery buffers). This tiered architecture could permit differentiated pricing, aligning the operator's desire to monetize slack capacity with tenant needs for tiered access to compute. This would transform the rural datacenter from a binary curtailment-or-grid choice into a portfolio of operational modes that adapt dynamically to both energy availability and computational demand.

# 7 Limitations and Sources of Uncertainty

Our analysis represents one plausible scenario among many possible futures for California and its infrastructure for data, energy, and computation. We develop a parameterized model to understand the scenario and justify our choice of parameter values based on market literature.



**Market Dynamics.** This paper examines the idea of datacenters with large batteries to exploit curtailed energy, taking a static snapshot of market conditions today. But markets are dynamic and cost assumptions may be impacted by other entities in the market. The economic benefits identified in our analysis assume that utility-scale, long-duration battery costs continue to decline significantly or access to low-cost curtailed energy will persist.

But if these benefits prove real and durable, they would likely attract other flexible loads to transmission-constrained regions. Hydrogen electrolyzers, behind-the-meter storage, electric vehicle charging hubs, and other demand-side entities could compete with datacenters for the same curtailed generation. This competition would bid up local energy prices and reduce the cost advantages available to any single participant, including datacenters.

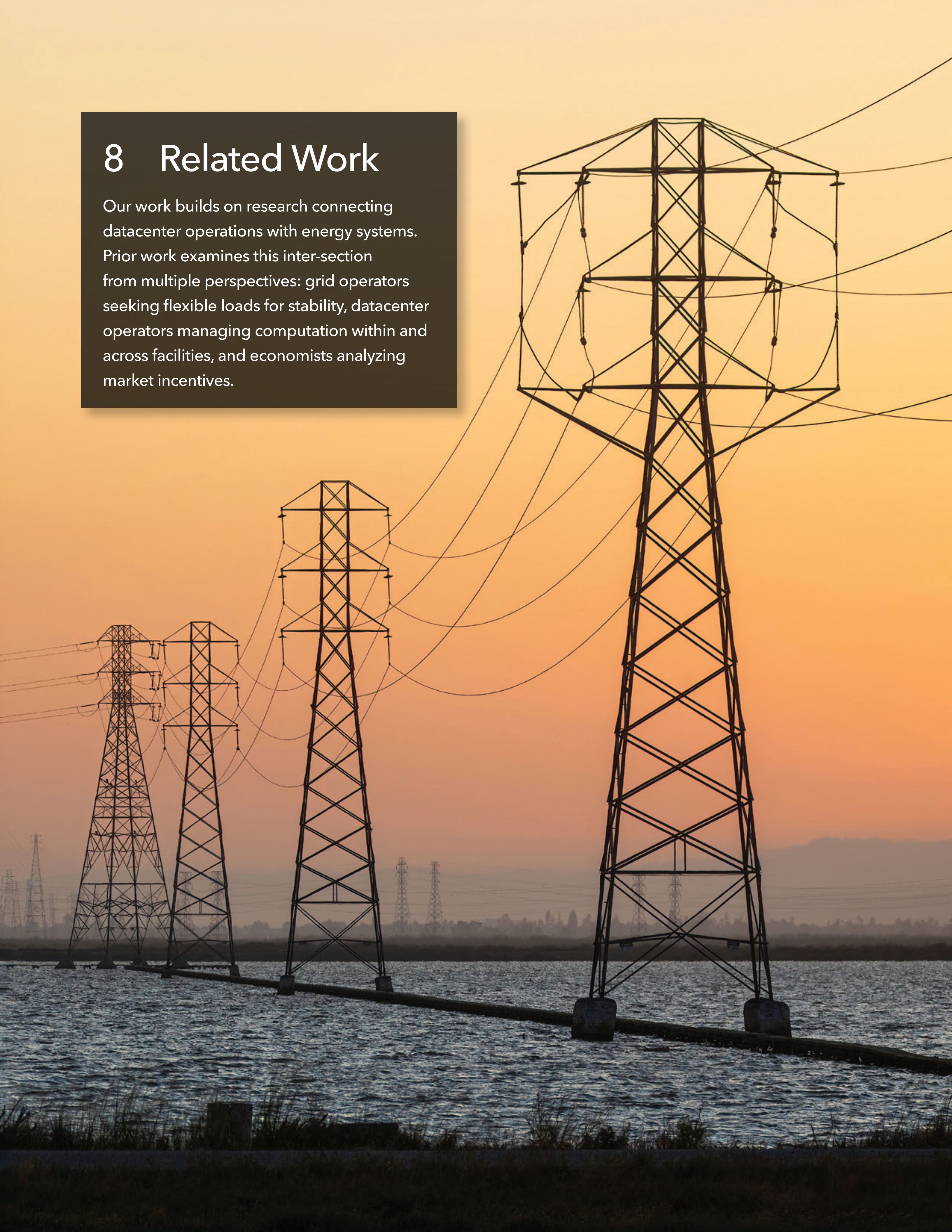
**Transmission Planning.** Our analysis draws on CAISO's transmission plan. While these projections include the grid operator's best assessment based on today's conditions, these plans may change based on evolving renewable energy deployments, load growth, policy proposals, and regulatory environment. Transmission planning is inherently uncertain.

**Investment Decisions.** Infrastructure investment decisions of this magnitude typically involve risk and probability assessments for key variables rather than single point estimates. Investors would evaluate a range of optimistic and pessimistic energy prices, curtailment volumes, technology costs, and policy environments. We will release our parameterized model so that researchers can define and explore other scenarios.

We acknowledge there is risk and uncertainty in our analysis and conclusions. Yet our analytical framework remains valuable for understanding the potential scope of using curtailed energy to compute under well researched assumptions.

# 8 Related Work

Our work builds on research connecting datacenter operations with energy systems. Prior work examines this inter-section from multiple perspectives: grid operators seeking flexible loads for stability, datacenter operators managing computation within and across facilities, and economists analyzing market incentives.



From the grid operator's perspective, researchers have studied flexible datacenter loads and opportunities to improve stability. From the datacenter operator's perspective, researchers have studied strategies for managing computation within a single shared datacenter as well as across multiple geographically distributed datacenters. Many of these studies acknowledge the role of electricity markets and incentives, which inform our own financial models and feasibility assessments when powering datacenter computation with curtailed energy.

For a broader perspective, researchers have surveyed challenges and opportunities associated with the electricity used by artificial intelligence and datacenters. The National Academies provides a broad perspective on challenges and opportunities, surveying AI electricity usage and its implications for datacenter and grid infrastructure [1]. Lee et al. provide a broad, multi-disciplinary research agenda for sustainable computing, identifying key challenges across the entire hardware and software stack, from embodied carbon in manufacturing to the operational energy of AI models [32]. And Wu et al. provide a holistic characterization of the carbon footprint of AI, spanning the entire model development cycle and system hardware life cycle [46].

**Grid Stability and Demand Response.** Several studies establish the significant opportunities for grid stability when large datacenter loads are flexible. These datacenters could absorb curtailed or stranded renewable energy. Norris et al. identify opportunities to integrate large, flexible datacenter loads into the U.S. grid by exploiting existing capacity [36]. Their analysis indicates that regions like CAISO could integrate approximately 5.9GW of new datacenter capacity without new generation if workloads could curtail power usage by up to 1% annually and tolerate an average pause of 2.5 hours per curtailment event. Similarly, Kim et al. use stochastic optimization to analyze the optimal placement of dispatchable datacenters to harness stranded power [26]. If flexible datacenters are strategically sited, they can reduce curtailment and improve grid efficiency. Yet Xing and Lee review the practical challenges that have hindered the widespread adoption of datacenter demand response, arguing that many existing frameworks are too complex, break critical abstraction layers, or provide insufficient incentives [48].

Our work builds on similar ideas but differ in scope and granularity. Whereas Norris et al. analyze capacity at the balancing authority level, we go a step further and overlay localized data on transmission congestion and existing datacenter sites so that we can more accurately assess datacenter siting and retrofitting opportunities. Whereas Kim et al. examine the Western Electricity Coordinating Council and stranded wind energy, our work examines California and curtailed energy, which is predominantly from solar generation. We use simpler methods for scheduling, combined with a financial model, to assess feasibility rather than solve a large optimization problem.

**Managing the Datacenter's Computation.** Researchers optimize computation within a single datacenter shared by multiple, independent tenants, each with their own computational needs. Xing et al. propose Carbon Responder, a framework for coordinating demand response within a private datacenter by allocating power reductions across heterogeneous workloads in response to grid carbon signals [47]. Guo and Pan address the challenge of energy management within multi-tenant colocation datacenters [19]. Their key contribution is a decentralized optimization algorithm that coordinates energy consumption between the datacenter operator and individual tenants to reduce overall costs, essentially creating an internal energy market. Thus, Guo and Pan optimize energy use within an existing, operational colocation facility in response to grid pricing. Similarly, Islam et al. address the challenge of uncoordinated power management among tenants in a colocation datacenter [22]. Their RECO system uses financial rewards to incentivize tenants to collectively reduce power consumption, thereby lowering the colocation operator's overall energy costs, particularly by managing peak demand charges. Our work examines how and where to build or retrofit datacenters to harness curtailed renewable energy, complementing these frameworks that decide which computations within the datacenter are allocated that energy for computation.

**Managing the Datacenter's Energy Assets.** A significant body of work focuses on optimizing energy resources attached to a single datacenter. Acun et al. propose the Carbon Explorer framework to analyze the trade-offs between operational and embodied carbon to optimize for 24/7 carbon-free datacenter operation. They co-optimize a mix of solutions such as renewable energy sources, energy storage, and workload scheduling, identifying unique mixes based on geographic location and workload characteristics. Kwon develops a model for mathematically optimizing operations for a single datacenter with co-located solar generation and battery storage [28]. They propose a two-stage stochastic program that optimizes server provisioning and power procurement to minimize energy costs while satisfying specific targets for renewable energy use and quality of service.

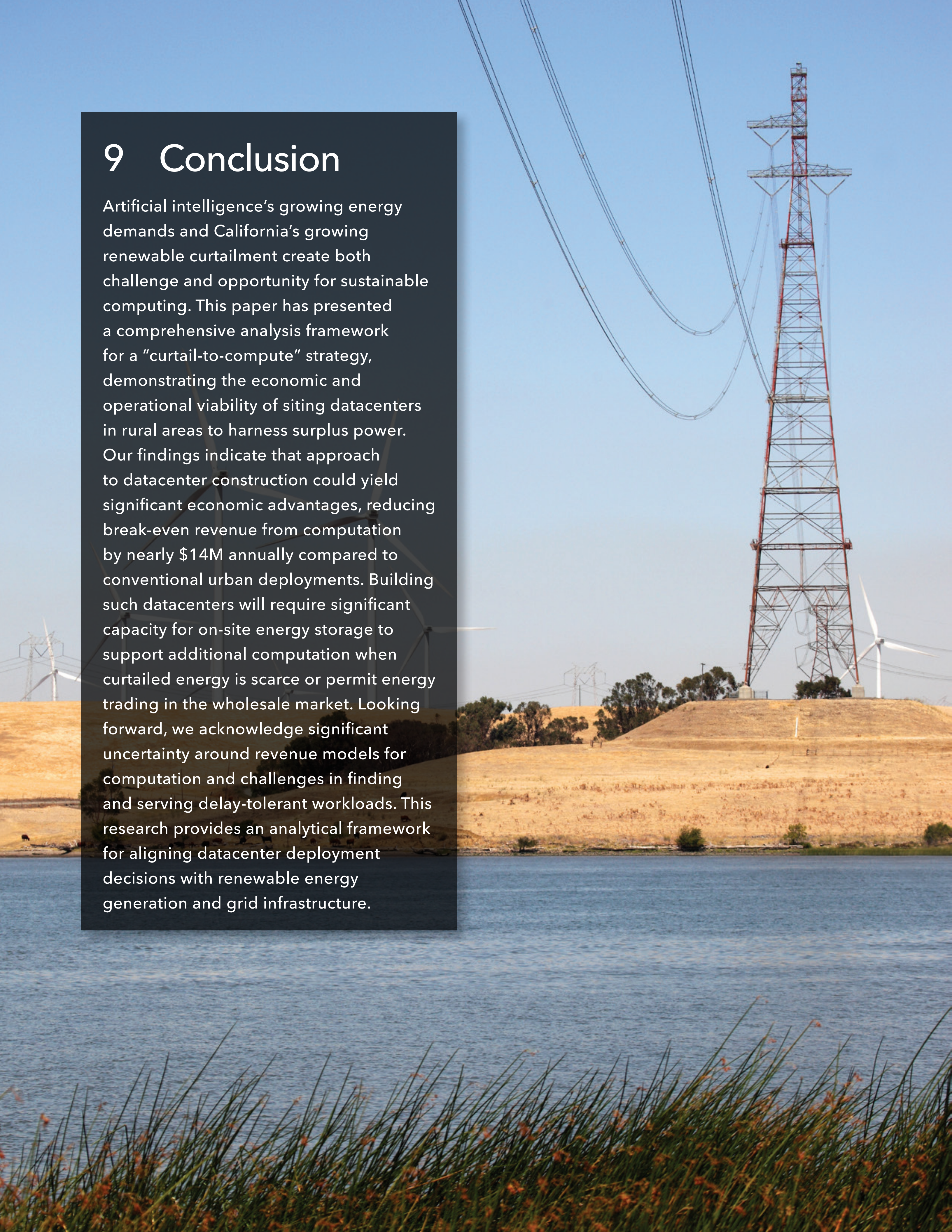
Zhang et al. observe that datacenters often deploy batteries for resilience [51]. But datacenters install all of a datacenter's battery capacity at time of construction while incrementally installing its server capacity over time, leading to a gap between the energy storage that is available and that is required for resilience. Surplus battery capacity could be dispatched to provide ancillary services that help stabilize the grid. Niaz et al. explore the use of curtailed energy to power a combined datacenter and hydrogen refueling station in California. They formulate a mixed integer linear programming problem to optimize the dispatch and allocation of curtailed energy. Although we share their goal of exploiting carbon-free, curtailed energy, our work differs significantly in scope and methodology. Our analysis focuses specifically on powering datacenter servers for computation and identifying physical and geographical constraints. We use simpler methods for scheduling, combined with a financial model, to assess feasibility rather than formulate and solve a large optimization problem.

While we think colocated generation is an interesting avenue for future work, our work currently focuses on harnessing curtailed energy from the broader grid. We do not necessarily seek to meet a target for renewable energy use or carbon footprint. Rather we seek to maximally use an otherwise stranded or wasted energy resource. Our work examines batteries but differs fundamentally in its objective and approach. Whereas Zhang et al. maximize revenues from underutilized backup batteries by dispatching them for grid stability, we seek to minimize costs from fully utilized batteries by dispatching them for useful computation.

**Migrating Computation between Datacenters.** Another line of research explores workload migration across geographically distributed datacenters, seeking to exploit the availability of renewable energy. Zheng et al. study strategies for migrating workloads between datacenters to exploit renewable energy that would otherwise be curtailed [52]. An illustrative case study quantifies benefits when shifting workloads from a region that uses fossil fuels heavily (e.g., PJM) to one that uses renewables heavily (e.g., CAISO) during periods of curtailment. Similarly, Jin et al. propose a model for a company with multiple, geographically distributed datacenters to participate in regional peer-to-peer (P2P) energy markets [23]. They develop an algorithm and optimization framework that permits a company with multiple datacenters to shift workloads between them and take advantage of favorable P2P market conditions.

Our work shares the foundational goal of exploiting curtailed energy for datacenter computation, but we find that datacenters should be built near sources of renewable energy generation due to the challenges of moving energy and data. While we agree with Zheng et al.'s observation that it might be easier to move data rather than energy, we would prefer to move neither. As a result, we suggest building or retrofitting datacenters in geographic locations where curtailed energy is abundant rather than assume a coordinated inter-regional load balancing scheme exists. We focus on managing computations within a single datacenter, seeking to maximally use readily accessible curtained energy in a specific geographical location.

**Economics and Incentives.** Finally, our work is informed by economic and regulatory frameworks that could incentivize datacenter load flexibility. Satchwell et al. provide a comprehensive overview of evolving electricity rate designs for large loads from a utility and regulatory perspective [41]. They analyze various tariff structures, including demand charges, real-time pricing, and curtailable service rates, that can incentivize load flexibility. Our work anticipates these differentiated electricity rates and assesses opportunities for computing with potentially less expensive curtailed energy. While Satchwell et al. focus on the design of economic signals and incentives, we propose datacenter infrastructure that is positioned and capable of responding to these signals.



# 9 Conclusion

Artificial intelligence's growing energy demands and California's growing renewable curtailment create both challenge and opportunity for sustainable computing. This paper has presented a comprehensive analysis framework for a "curtail-to-compute" strategy, demonstrating the economic and operational viability of siting datacenters in rural areas to harness surplus power. Our findings indicate that approach to datacenter construction could yield significant economic advantages, reducing break-even revenue from computation by nearly $14M annually compared to conventional urban deployments. Building such datacenters will require significant capacity for on-site energy storage to support additional computation when curtailed energy is scarce or permit energy trading in the wholesale market. Looking forward, we acknowledge significant uncertainty around revenue models for computation and challenges in finding and serving delay-tolerant workloads. This research provides an analytical framework for aligning datacenter deployment decisions with renewable energy generation and grid infrastructure.

## References


1. 2025. *Implications of Artifcial Intelligence Related Data Center Electricity Use and Emissions: Proceedings of a Workshop*. The National Academies Press, Washington, DC. https://doi.org/10.17226/29101
2. Accordant Investments. 2024. The Numbers Supporting 30%+ IRR Returns in Hyperscale Data Center Development. Blog. https://www.accordantinvestments.com/blog/the-numbers-supporting-30-irr-returns-in-hyperscale-data-center-development
3. B. Acun, B. Lee, F. Kazhamiaka, K. Maeng, U. Gupta, M. Chakkaravarthy, D. Brooks, and C. J. Wu. 2023. Carbon Explorer: A Holistic Framework for Designing Carbon-Aware Datacenters. In *Proceedings of the 28th International Conference on Architectural Support for Programming Languages and Operating Systems* (ASPLOS).
4. Alantra. 2024. *Data Centre M&A Review 2024*. Technical Report. Alantra. https://www.alantra.com/wp-content/uploads/2025/01/Data-center-activity-2024-vF.pdf
5. Brightlio Research. 2024. Data Center Statistics and Market Insights 2024. https://brightlio.com/data-center-stats/#pp-toc-huja89rng0qk-anchor-0 Accessed October 2025.
6. California Independent System Operator (CAISO). 2025. *ISO Board-Approved 2024–2025 Transmission Plan.* Technical Report Board-Approved plan. California ISO. https://www.caiso.com/documents/iso-board-approved-2024-2025-transmission-plan.pdf
7. California Department of Technology. 2025. California Middle-Mile Broadband Initiative. https://middle-mile-broadband-initiative.cdt.ca.gov/. Accessed on September 8, 2025.
8. California Energy Commission. 2024. *Data Center Load Forecasts, 2024 - 2040*. Technical Report. California Energy Commission. Accessed on September 8, 2025.
9. California Energy Commission, California Public Utilities Commission, and California Air Resources Board. 2021. *2021 SB 100 Joint Agency Report: Achieving 100 Percent Clean Electricity in California: An Initial Assessment.* Technical Report. California Energy Commission. https://www.energy.ca.gov/sb100
10. California Independent System Operator (CAISO). 2025. Day-Ahead Locational Marginal Price (LMP) Data for Node NP-15. CAISO OASIS (Open Access Same-Time Information System). Data retrieved on September 8, 2025.
11. California Independent System Operator (CAISO). 2025. High Voltage Access Charge Rates Efective January 1, 2025. https://www.caiso.com/documents/high-voltage-access-charge-rates-effective-jun-01-2025.pdf. Accessed on September 8, 2025.
12. CBRE Investment Management. 2024. Decoding Data Centers: A Deep Dive into a Dynamic Sector. Insights Article. https://www.cbreim.com/insights/articles/decoding-data-centers
13. CBRE Research. 2024. North American Data Center Trends 2024. Cited for 20% year-over-year rent increases in tight markets.
14. Chris Crosby. 2019. *The Inconvenient Truth About Edge Datacenter Operating Costs*. Technical Report. Data Center Frontier. Accessed on September 8, 2025.
15. Cushman & Wakefeld. 2023. *2023 Global Data Center Market Comparison.* Market Report. Cushman & Wakefeld. https://cushwake.cld.bz/2023-Global-Data-Center-Market-Comparison Fourth edition analyzing 63 global markets.
16. Jefrey Dean and Luiz André Barroso. 2013. The Tail at Scale. *Commun*. ACM 56, 2 (February 2013), 74–80.
17. Digital Realty Trust, Inc. 2023. Q4 2023 Financial Results and Supplemental Information. https://investor.digitalrealty.com. Accessed October 2025.
18. DLA Piper and Inframation. 2022. *The meteoric rise of the data centre: Global Data Centre Investment Outlook.* Technical Report. DLA Piper. Accessed on September 8, 2025.
19. Yuanxiong Guo and Miao Pan. 2015. Coordinated energy management for colocation data centers in smart grids. In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 840–845. https://doi.org/10.1109/SmartGridComm.2015.7436409
20. L. Han, J. Kakadi, B. C. Lee, and U. Gupta. 2025. Fair-CO2: Fair Attribution for Cloud Carbon Emissions. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*.
21. Houlihan Lokey. 2024. Real Estate Market Update: Data Centers – Summer 2024. https://cdn.hl.com/pdf/2024/real-estate-market-update-data-centers-summer-2024.pdf Accessed October 2025.
22. Mohammad A. Islam, Hasan Mahmud, Shaolei Ren, and Xiaorui Wang. *2015. Paying to save: Reducing cost of colocation data center via rewards. In 2015 IEEE 23rd International Symposium on Quality of Service (IWQoS).* 159–168. https://doi.org/10.1109/IWQoS.2015.7404746
23. Tianyu Jin, Linquan Bai, Mingyu Yan, and Xinyu Chen. 2025. Unlocking Spatio-Temporal Flexibility of Data Centers in Multiple Regional Peer-to-Peer Energy Transaction Markets. *IEEE Transactions on Power Systems* 40, 5 (2025), 3914–3926. https://ieeexplore.ieee.org/document/10852611
24. JLL Research. 2024. Global Data Center Outlook 2024. Cited for annual rent escalations and market absorption trends.
25. Slade Johnson. 2024. California's solar electricity generation grew faster than any other state's last year. U.S. Energy Information Administration, Today in Energy. https://www.eia.gov/todayinenergy/detail.php?id=65364 Accessed: 2025-09-09.
26. Kibaek Kim, Fan Yang, Victor M. Zavala, and Andrew A. Chien. 2017. Data Centers as Dispatchable Loads to Harness Stranded Power. *IEEE Transactions on Sustainable Energy* 8, 1 (2017), 208–218. https://doi.org/10.1109/TSTE.2016.2593607
27. Data Center Knowledge. 2023. How Equinix Decides When to Build a Data Center. (2023). https://www.datacenterknowledge.com/hyperscalers/how-equinix-decides-when-to-build Accessed October 2025.
28. Soongeol Kwon. 2020. Ensuring renewable energy utilization with quality of service guarantee for energy-efcient data center operations. *Applied Energy 276* (2020), 115424. https://doi.org/10.1016/j.apenergy.2020.115424
29. LandSearch. 2025. Santa Clara County, CA Industrial Land for Sale. https://www.landsearch.com/industrial/santa-clara-county-ca. Accessed on September 8, 2025.
30. LandWatch. 2025. Fresno County, CA Land for Sale. https://www.landwatch.com/california-land-for-sale/fresno-county/acres-11-50. Accessed on September 8, 2025.
31. Benjamin C. Lee, David Brooks, Arthur van Benthem, Mariam Elgamal, Udit Gupta, Gage Hills, Vincent Liu, Linh Thi Xuan Phan, Benjamin Pierce, Christopher Stewart, Emma Strubell, Gu-Yeon Wei, Adam Wierman, Yuan Yao, and Minlan Yu. 2025. A View of the Sustainable Computing Landscape. *Patterns* (2025). https://www.sciencedirect.com/science/article/pii/S2666389925001448
32. Benjamin C Lee, David Brooks, Arthur van Benthem, Mariam Elgamal, Udit Gupta, Gage Hills, Vincent Liu, Linh Thi Xuan Phan, Benjamin Pierce, Christopher Stewart, Emma Strubell, Gu-Yeon Wei, Adam Wierman, Yuan Yao, and Minlan Yu. 2025. A view of the sustainable computing landscape. *Patterns 6*, 3 (2025), 101296. https://doi.org/10.1016/j.patter.2025.101296

33. John McWilliams, Ethan Tribble, Adrian Conforti, and Jason D'Orlando. 2025. *Data Center Development Cost Guide 2025*. Technical Report. Cushman & Wakefeld. Accessed on September 8, 2025.
34. Latitude Media. 2024. The Rules Around Data Center Cost Allocation Are Getting Clearer. (2024). https://www.latitudemedia.com/news/the-rules-around-data-center-cost-allocation-are-getting-clearer/ Accessed October 2025.
35. National Renewable Energy Laboratory. 2024. *2024 Annual Technology Baseline: Utility-Scale Battery Storage.* Data. National Renewable Energy Laboratory, Golden, CO. https://atb.nrel.gov/electricity/2024/utility-scale_battery_storage
36. Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell. 2025. *Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems. Technical Report NI R 25-01. Nicholas Institute for Energy, E*nvironment & Sustainability, Duke University, Durham, NC. https://nicholasinstitute.duke.edu/publications/rethinking-load-growth
37. PeeringDB. 2024. PeeringDB: The Interconnection Database. https://www.peeringdb.com. Accessed: January 2025.
38. PG&E Corporation. 2025. PG&E Accelerating Connection of New Data Centers throughout Northern and Central Califor-nia. https://investor.pgecorp.com/news-events/press-releases/press-release-details/2025/PGE-Accelerating-Connection-of-New-Data-Centers-throughout-Northern-and-Central-California/default.aspx Accessed on September 8, 2025.
39. Ana Radovanovic, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, MariEllen Cottman, and Walfredo Cirne. 2021. Carbon-Aware Computing for Datacenters. arXiv:2106.11750 [cs.DC] https://arxiv.org/abs/2106.11750.11750
40. RCLCO Real Estate Consulting. 2024. Data Center Market Analysis: Target Returns and Investment Trends. Industry report summarizing stabilized leveraged IRRs of 12–19% depending on risk profle.
41. Andrew Satchwell, Natalie Mims Frick, Peter Cappers, Sanem Sergici, Ryan Hledik, Goksin Kavlak, and Glenda Oskar. 2025. *Electricity Rate Designs for Large Loads: Evolving Practices and Opportunities*. Technical Report. Lawrence Berkeley National Laboratory. https://escholarship.org/uc/item/19m1k8vr
42. Eric Schurman and Jake Brutlag. 2009. Performance Related Changes and Their User Impact. Presentation at Velocity Web Performance and Operations Conference.
43. Mohammad Shahrad, Rodrigo Fonseca, Íñigo Goiri, Gohar Chaudhry, Paul Batum, Jason Cooke, Eduardo Laureano, Colby Tresness, Mark Russinovich, and Ricardo Bianchini. 2020. Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. USENIX Association, 205–218. https://www.usenix.org/conference/atc20/presentation/shahrad
44. Arman Shehabi, Sarah J. Smith, Alex Hubbard, Alex Newkirk, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor. 2024. *2024 United States Data Center Energy Usage Report*. Technical Report. Lawrence Berkeley National Laboratory.
45. REIT.com Editorial Staf. 2024. Data Centers: High Demand, Slow Deliveries, Big Development Profts. (2024). https://www.reit.com/news/articles/data-centers-high-demand-slow-deliveries-big-development-profits Accessed October 2025.
46. 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. 2022. Sustainable AI: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems 4* (2022), 675–703.
47. Jiali Xing, Bilge Acun, Aditya Sundararajan, David Brooks, Manoj Chakkaravarthy, Nikky Avila, Carole-Jean Wu, and Benjamin C Lee. 2023. Carbon Responder: Coordinating Demand Response for the Datacenter Fleet. *arXiv preprint arXiv:2311.08589* (2023).
48. Jiali Xing and Benjamin C. Lee. 2024. Datacenter Demand Response for Carbon Mitigation: From Concept to Practicality. In *2024 IEEE International Green and Sustainable Computing Conference (IGSC)*. 1–8. https://ieeexplore.ieee.org/document/10765828
49. Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh. 2020. Large Batch Optimization for Deep Learning: Training BERT in 76 Minutes. In *8th International Conference on Learning Representations (ICLR 2020)*. Addis Ababa, Ethiopia (Virtual). https://openreview.net/forum?id=Syx4wnEtvH
50. Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, Ali Ghodsi, Joseph Gonzalez, Scott Shenker, and Ion Stoica. 2016. Apache Spark: A Unifed Engine for Big Data Processing. *Commun. ACM 59*, 11 (November 2016), 56–65. https://dl.acm.org/doi/10.1145/2934664
51. Yingbo Zhang, Hong Tang, Hangxin Li, and Shengwei Wang. 2025. Unlocking the fexibilities of data centers for smart grid services: Optimal dispatch and design of energy storage systems under progressive loading. *Energy* 316 (2025), 134511. https://doi.org/10.1016/j.energy.2025.134511
52. 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. https://doi.org/10.1016/j.joule.2020.08.001
53. ZipRecruiter. 2025. Data Center Operations Salary in California. https://www.ziprecruiter.com/Salaries/Data-Center-Operations-Salary--in-California. Accessed on September 8, 2025.