Evaluating Coupling Models for Cloud Datacenters and Power Grids

Liuzixuan Lin
University of Chicago
Chicago, IL, USA
lzixuan@uchicago.edu

Victor M. Zavala
University of Wisconsin-Madison
Madison, WI, USA
victor.zavala@wisc.edu

Andrew A. Chien
University of Chicago & Argonne National Lab
Chicago, IL, USA
achien@cs.uchicago.edu

ABSTRACT

The rapid growth of datacenter (DC) loads can be leveraged to help meet renewable portfolio standard (RPS, renewable fraction) targets in power grids. The ability to manipulate DC loads over time (shifting) provides a mechanism to deal with temporal mismatch between non-dispatchable renewable generation (e.g. wind and solar) and overall grid loads, and this flexibility ultimately facilitates the absorption of renewables and grid decarbonization. To this end, we study DC-grid coupling models, exploring their impact on grid dispatch, renewable absorption, power prices, and carbon emissions. With a detailed model of grid dispatch, generation, topology, and loads, we consider three coupling approaches: fixed, datacenter-local optimization (online dynamic programming), and grid-wide optimization (optimal power flow).

Results show that understanding the effects of dynamic DC load management requires studies that model the dynamics of both load and power grid. Dynamic DC-grid coupling can produce large improvements: (1) reduce grid dispatch cost (-3%), (2) increase grid renewable fraction (+1.58%), and (3) reduce DC power cost (-16.9%). It also has negative effects: (1) increase cost for both DCs and non-DC customers, (2) differentially increase prices for non-DC customers, and (3) create large power-level changes that may harm DC productivity.

CCS CONCEPTS

• **Hardware** → **Power and energy**; • **Applied computing** → **Data centers**.

KEYWORDS

Power grid, Renewable energy, Carbon emissions, Data centers

ACM Reference Format:
Liuzixuan Lin, Victor M. Zavala, and Andrew A. Chien. 2021. Evaluating Coupling Models for Cloud Datacenters and Power Grids. In *The Twelfth ACM International Conference on Future Energy Systems (e-Energy '21), June 28–July 2, 2021, Virtual Event, Italy*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3447555.3464868

e-Energy '21, June 28–July 2, 2021, Virtual Event, Italy

ACM ISBN 978-1-4503-8333-2/21/06…$15.00
https://doi.org/10.1145/3447555.3464868

1 INTRODUCTION

Recent years have seen the rapid growth of power consumption (loads) by hyperscale cloud providers (e.g. Amazon, Microsoft, Google, Alibaba, Baidu) and also large-scale computing companies (e.g. Facebook, Apple, Tencent). While obscured by their rapid cannibalization of enterprise datacenters (DCs) in recent reports [38], a close look at the data documents their sustained growth in power consumption of 31% annually through 2018. Recent accelerants such as machine learning [27, 50], and Covid-19 driven digitalization [41, 49] may have increased this rate. With DCs exceeding 5% of power consumption in the Northern Virginia grid and 2% of US consumption today [2], their loads are already significant. Extrapolated to 2025, the power consumption of these hyperscale providers will exceed 5% of US power consumption.

Renewable generation is also increasing rapidly, driven by falling generation costs and growing climate concerns. The past decade (2010-2020) has seen wind and solar renewable power generation quadruple in the United States, reaching 11% nationwide, and higher fractions in California (33%) and Texas (21.5%), and across Europe. Ambitious renewable portfolio standard (RPS) goals have been set in many regions for the coming decade—including California (60%) [12], New York (70%) [43], Europe (33%) [6] all by 2030. This transformation of the power grid has created significant renewable integration challenges, including both decreasing renewable capacity credit and rapid fluctuation in generation mix (wind, solar, fossil-fuel), driven by ever larger variation in renewable generation as share of power grid generation increases [54] (see Figure 2). Specifically, a key challenge to overcome as we seek to enable deep grid decarbonization is the fact that our main renewable power sources (wind and solar) are intermittent, non-dispatchable, and are not synchronized with overall grid load dynamics. This issue is already creating inefficiencies in the system at moderate renewable adoption levels (e.g. negative prices and stranded power) [10].

Researchers in universities and cloud companies have proposed shifting computing load (power consumption) to save money [3, 28, 31, 36, 37, 45, 47, 53, 57] and also to reduce the carbon emissions of datacenters [13, 15, 16, 18, 19, 33, 34, 46]. This is a promising idea, but these studies do not consider the dynamic interaction with the power grid, viewing computing as too small a load to affect grid dynamics. With hyperscale cloud operators building gigawatt facilities, and now a significant and growing percentage of load in several power grids, we explore the importance of modelling grid dynamics on the viability of DC load shifting.

Some researchers have sought to exploit excess power embodied in curtailment and negative-priced power [10, 55, 56], shifting

computation to datacenters intermittently-powered by such zero-carbon power. Startups are commercializing these ideas [1, 48], and a number of power grid researchers have proposed the dynamic shifting of cloud power load to improve grid stability and to absorb more renewable generation (carbon-free power) [26, 32, 59, 60]. These studies consider simple power grids and datacenter loads, showing the theoretical promise of co-optimization.

In this paper, we consider multiple datacenters under intelligent control as a set of dynamic loads. We explore this coupling in the context of a detailed, realistic power grid dispatch model, exploring a range of cloud and grid renewable configurations. We study a wide range of hyperscale cloud configurations, ranging from 3.5% to 14% of the grid peak load, reflecting a realistic range for several US grids in 2022 [38] and nationally by 2027. For example, Northern Virginia was solidly in this range by 2018, passing 5% in 2019 [2, 14]. We also consider a wide range of renewable (wind) penetration, scaling from 15% to 60%, reflecting 2015 and aspirational 2050 in a US National Renewable Energy Laboratory (NREL) Wind Vision Report [42]. Across that space, we consider coupling approaches ranging from DCs as constant loads, local optimization of power cost using online dynamic programming (DP-ONL), and delegation of datacenter flexibility to the power grid (GC), allowing it to optimize social welfare for all. In each situation, we examine grid dispatch cost, renewable absorption, as well as both cloud datacenter and non-datacenter power prices. Our findings can help power grid operators understand benefits and challenges that arise from DC load shifting flexibility. Moreover, they can help both DC and grid operators understand how to best coordinate. Specific insights include:

- Greater datacenter (DC) load increases the quantity of renewable power absorbed, but due to temporal misalignment fixed DC loads reduce grid renewable fraction (RPS), retarding grid progress towards decarbonization. Directly, they produce a larger increase in fossil-fuel power consumption than prior RPS would suggest.
- Dynamic coupling of DC loads with the grid can improve alignment, which in the best cases reduces grid dispatch cost by 3% and increases RPS by as much as 1.58%, and also benefit DCs by reducing their power cost by 16.9%.
- The coupling model is important: independent, local optimization (DP-ONL) preserves DC autonomy, but aligns load less effectively, overshifting load and increasing both prices and emissions. Further, DC shifting can differentially increase prices for non-DC customers. Delegating load flexiblity to the grid (GC) outperforms DP-ONL in grid benefits and fairness, but creates rapid DC capacity variation—a challenge for DC productivity.

Overall, results indicate that the coupling model is significant for power grid and datacenter costs as well as carbon emissions.

The remainder of the paper includes background in Section 2. In Section 3 we consider grid impacts of growth in DC load and renewable generation, proposing our approach. Section 4 summarizes power grid model, metrics, and coupling models, and describes the coupling simulations. Sections 5 and 6 present simulations assessing additional DC load and renewable generation, and the impact of coupling models. In Sections 7 and 8, we discuss related work and present the conclusion and future directions.

2 BACKGROUND

2.1 Hyperscale Datacenter Power Growth

The history of cloud computing efforts in sustainability includes significant efficiency improvements (reducing power usage effectiveness (PUE) from over 2 to as low as 1.1 for hyperscale datacenters), and increasing the renewable power in grids by executing long-term power purchase agreements. Despite these efforts, the power consumption and associated carbon emissions of the largest cloud operators continue rapid growth. For example, the long-term trend for hyperscale cloud providers in North America is 31% growth per annum. Figure 1 shows the recent trajectory of hyperscale datacenter (cloud) power consumption in the past decade and extrapolation to 2025 [38]. This growth may be accelerating due to machine learning; at the current rate the hyperscale providers alone will exceed 10% of the world's power by 2030 [25].

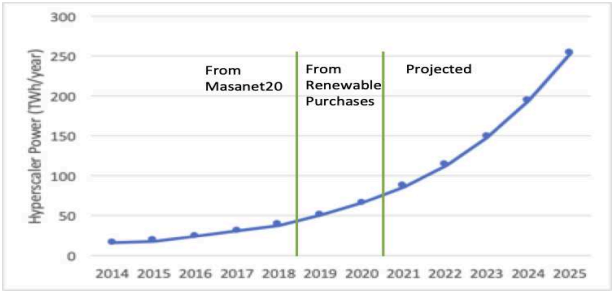


Figure 1: Hyperscale Datacenter Power growth (TWh) in North America 2010-18 [38], confirmed by renewable power purchases (2018-20), and extrapolated to 2025. 2% of US electric power in 2020 with 31% annual growth.

A key reason for this growth is Jevon's paradox (1865), which paraphrased tersely says "efficiency increases and price reductions only increase use, perhaps even faster" [24]. So computing's efficiency and price reductions drive the growth of computing use. Further, the cloud success in increased ease of software development, deployment and scaling, accelerates the use of computing for existing and new applications from search to social networks to worldwide videoconferencing.

In hyperscale cloud companies such as Amazon, Google, and Microsoft, datacenter sites have grown dramatically to increase cost-efficiency and to meet demand. Today (2020), many of these sites exceed 200 megawatts, and larger planned sites exceed 1 gigawatt [2, 4, 17, 39]. Our study explores modulating these large loads by 20, 40, even 60% to both increase power grid efficiency in absorbing renewable generation and reduce the carbon emissions of cloud computing, but also to enable datacenters (and other customers) to achieve lower prices.

2.2 Economic Dispatch, Fluctuating Generation, and Adaptive Load

Modern power grids use economic dispatch (lowest price power first) to select which generation is used, subject to transmission capacities and loads. Economic dispatch has enabled the integration

of thousands of renewable generators that bid their variable generation into power markets when it is available. These bids typically describe generation quantity for a specific period. Low-price bids enable renewables to displace other generators. Economic dispatch combined with transmission constraints produces the locational marginal prices (**LMP**s), the power prices at each grid location.

In the past decade, wind and solar renewable power generation has quadrupled in the United States, reaching 11% nationwide, and higher fractions in California (33%) and Texas (21.5%). Future RPS goals include California (60%), New York (70%), Europe (33%) all in 2030. Because of their variable generation, renewable absorption is challenging with failures producing curtailment and decreasing grid capacity factors [54]. For example, curtailment has been analyzed extensively in several US power grids [9, 11, 30] (1.5 TWh / year) and worldwide [5, 21, 22] (10's to as much as 100 TWh / year).

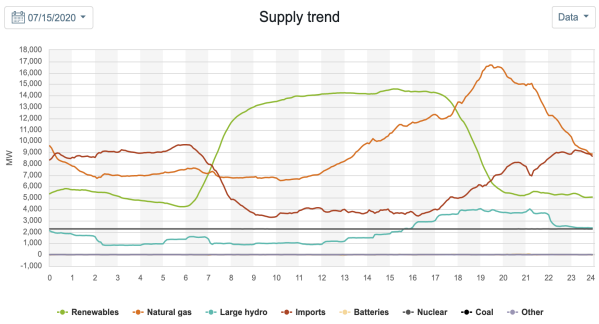


Figure 2: Generation Mix for CAISO Grid (California) (July 15, 2020). Renewable generation leads at noon, but it shifts to natural gas at 7pm and continues overnight.

This transformation of the power grid has produced fluctuation in generation mix (wind, solar, fossil-fuel) that grows with renewable generation. For example, California's power grid, California ISO (CAISO), reached an annual RPS of 30% in 2019, but on dozens of winter days, reached over 90% RPS for most daylight hours. These swings cause the carbon content of power in the grid to swing by 4x and more. The CAISO example shown in Figure 2 exhibits solar at 50% midday, with natural gas rising to 46% at 8pm. Similar phenomena occur in wind-heavy power grids (e.g ERCOT in Texas, Germany), but with different duration and periodicity.

3 PROBLEM AND APPROACH

Power markets and grids have complex dynamics, driven by transmission, generation, pricing, and ramp constraints. Small changes in load or supply can cause large effects (e.g. a $100/MWh price change). We consider multiple datacenters managed with intelligent control to support computation load shifting; this makes them dynamic loads. To explore combined dynamics of the datacenter control and power grid, we employ simulation, real load and generation profiles, optimal power flow (OPF) in the grid.

We explore questions such as: how does dynamic management of datacenter loads affect grid dynamics? Renewable absorption? Power prices? Using OPF, we simulate grid dispatch [23] for a detailed CAISO system model interconnected with the Western Electricity Coordinating Council (WECC), exploring varied configurations as below:

- Datacenters: 1.4–5.6 gigawatts (3.5 to 14%)
- Wind Penetration: 15% to 60% of grid consumption

The datacenter loads are 200 megawatts per site. While cloud power use varies widely by region, this range spans 2020 power levels to 5–10 year projections [38]. At these levels, datacenters are a critical component of power load. Wind penetration spans the 2015 level (15%) to the NREL Wind Vision [42] goal for 2050 (50%). Across this space, we explore intelligent techniques to shift load—under datacenter or grid control—using three datacenter-grid coupling models:

- Uncoupled: Datacenters are a constant power load
- Datacenter-local: Datacenters optimize cost independently, determining dynamic load (selfish-local)
- Grid-wide: Grid determines datacenter dynamic loads, optimizing overall dispatch cost (grid-global)

These models are illustrated in Figure 3. Uncoupled reflects fixed datacenter loads. Datacenter-local models independent datacenters, each with cloud resource managers time-shifting workload to optimize local power cost independently. Grid-wide optimization gives temporal-load flexibility to the power grid, allowing it to set datacenter power-levels to minimize grid-wide dispatch cost. Studies explore grid dispatch cost, renewable absorption, and power prices. We also examine datacenters' market power, characterizing impact on non-datacenter (non-DC) customers.

4 METHODS

We assess the impact of growing cloud load and renewable generation based on the direct-current optimal power flow (DC-OPF) model and notation in [26] and in Appendix A. The model minimizes the dispatch cost (generation and curtailment) subject to generation, load, transmission, power flow, and ramp constraints. One unusual feature (see Section 4.4) is that Datacenter-local uses a local algorithm to determine load dynamically.

4.1 Experiment Setup

We study a reduced model of the California power system (CAISO) which consists of 225 buses, 375 transmission lines, 130 thermal generation units with a total capacity of 31.2 GW, 40 loads, 11 non-wind renewable power plants, and 5 wind power plants. Imports flow into the system at 5 boundary buses. We use load and generation data from [40], and scale ramp rates for thermal power plants by 4-fold to reflect flexibility of the current CAISO fleet.

We run grid OPF optimization with a one-day time horizon and hourly intervals. The model minimizes total dispatch cost for the day, similar to day-ahead economic dispatch used in CAISO. We use load profiles (days) that correspond to typical weekday (WD) and weekend (WE) load for each season (Spring, Summer, Fall, Winter). The load ranges from 23,708 MW (WinterWE) to 31,089 MW (SummerWD), averaging 27,283 MW. Overall grid non-DC load and generation are shown in Appendix B.

For each season, there is one non-wind renewable generation profile and 1,000 wind scenarios from [26, 40]. These wind power scenarios average 15% of load (15% wind penetration). Weekdays

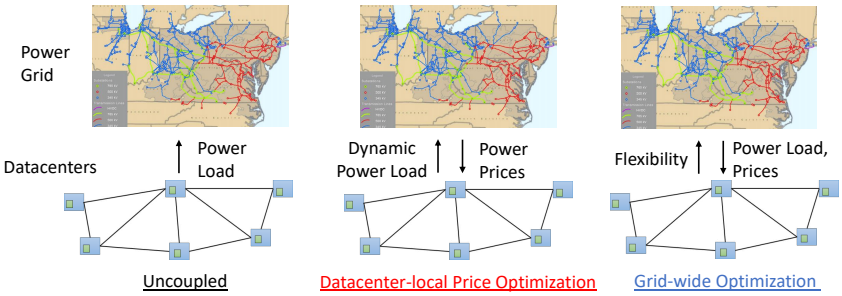
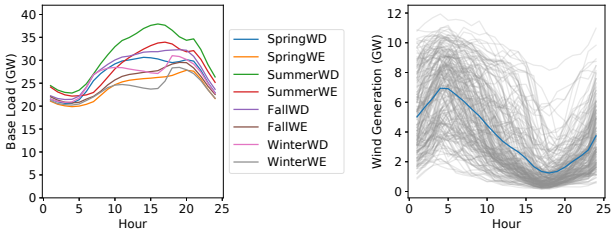


Figure 3: Three Models of Coupling: Uncoupled, Datacenter-local, and Grid-wide Optimization.



(a) (b)

Figure 4: Load Profile and Wind Supply in Coupling Model Studies. (a) Non-DC Load for Each Day Type. (b) Sampled Wind Scenarios for Summer, 15% Level (Blue Line: average).

(WD) and weekends (WE) in a season share the same set of wind scenarios. Figure 4a shows the varied load profiles (day type, season), and as an example, Figure 4b shows 200 of the summer wind production scenarios used. While the demand usually peaks around noon, wind production peaks in the late night or early morning, which is a mismatch and a reason for the absorption challenge. In all of our experiments, we add wind scenarios until the change in metrics is small, randomly selecting 200 scenarios from the 1,000 wind scenarios available for each season. For higher wind penetration, wind production levels for each site are scaled up equally, reflecting the assumption that existing wind locations are productive sites that could be scaled with larger turbines or site expansion [44].

As in [26], three levels of generation costs are used for conventional generation: 1 \$/MWh (nuclear), 2 \$/MWh (coal), and 4 \$/MWh (gas), corresponding to the fuel price. Curtailment penalties C_i^m (import), C_i^r (non-wind renewables), and C_i^w (wind) are 500, 1000, and 100 \$/MWh respectively. The value of lost load is set as 1000 \$/MWh for load shedding. As the locational marginal price (LMP) at a bus represents the marginal dispatch cost of adding 1 MW load there, we expect that the LMP may go negative when some type of curtailment arises. On the contrary, it can go very high when load shedding happens.

Datacenters are 200 MW at each site, and operate at 70% average power producing a $200 \times 0.7 = 140$ MW per hour average load in each 24 hours. This compute resource utilization is typical of cloud datacenters [2, 17, 39, 52], but the power level is well below the largest sites that already exceed 1.5 GW [2]. Our discussions with infrastructure planners at leading hyperscale cloud companies indicate future designs exceed 1 GW per site. In our simulations, the 200 MW datacenters are added to random buses (Appendix C) in the grid, and we ensure that the associated bus has sufficient capacity to support a 200 MW datacenter; this reflects cloud company site selection based on business considerations external to the power grid (e.g. tax breaks, jobs, internet hookups, etc.).

4.2 Implementation

The OPF model and coupling models in Section 4.4 are implemented in Julia 1.3.1 and solved with Gurobi Optimizer 8.1.1[20]. We parallelize the simulation instances with different parameter combinations using Python's Multiprocessing package.

Given a profile for a specific day type, a wind scenario, and a datacenter coupling model, we solve grid dispatch for that day. We find that sampling 200 wind production scenarios (see Figure 4b) is sufficient to create stable metric results in all cases, so all of the studies in Sections 5 and 6 reflect 200 wind scenarios.

4.3 Metrics

- **Grid Dispatch Cost (dollars)**—the objective function for OPF in power grid simulations. This is often called social welfare in electricity market clearing models [26].
- **Renewable Portfolio Standard (RPS)** is the ratio of absorbed renewable generation to total power consumed, inversely related to average carbon emissions per unit power.
- **Price (dollars)** is the average price (locational marginal price or **LMP**, power price at a location) across a set of locations and/or intervals, weighted by power demand. This is closely related to the prices that customers (both datacenters and non-datacenters) pay for electric power.
- **Capacity Variation (megawatts)** is defined as the average change in power level (load) of a datacenter between adjacent one-hour periods. More formally:

$$\frac{1}{23}\sum_{t=2}^{24}|l_{i,t} - l_{i,t-1}|$$

where $l_{i,t}$ denotes the power level of datacenter i at time t.

We report results that are the average of 8 day-types (weighted for the number of weekdays and weekend days) and varied wind scenarios. Standard deviations across the wind scenarios are shown as error bars, reflecting the range of results.

4.4 Datacenter-Power Grid Coupling Models

In the uncoupled case, each datacenter keeps the average power level (*avgLoad*) for all the 24 hours (i.e. constant load):

$$l_{i,t} = avgLoad, \ \forall i, t$$

For the other two coupling models, we assume that datacenters have some load flexibility, but must catch up within a 24-hour day. The two key parameters are:

- **Dynamic Range.** The magnitude of the interval over which datacenter load can be adjusted. Larger dynamic range reflects greater load flexibility.
- **Backlog.** The datacenter's deferred work relative to average progress (constant progress). Backlog is always positive, reflecting the fact that work can only be deferred as in [46].

Thus the formal flexibility constraints are:

$$load_{min} \leq l_{i,t} \leq load_{max}, \ \forall i, t \quad (1)$$

That is, all datacenter loads are always within the dynamic range. At time t, datacenter i's backlog update can be written as:

$$backlog_{i,t} = backlog_{i,t-1} + (avgLoad - l_{i,t}) \quad (2)$$

The backlog is always non-negative and must be zero at the end of the day to satisfy the average load constraint:

$$backlog_{i,t} \geq 0, \ \forall i, t \quad (3)$$

*4.4.1 Datacenter Local Optimization, using Online Dynamic Programming (**DP-ONL**).* Datacenters make local decisions to set load levels. As an exemplar of the many possible selfish algorithms, we use an online algorithm that makes hourly decisions that set datacenter load to minimize the expected power cost using the current price of power and a reference price.

Algorithm 1 Online Dynamic Programming (DP-ONL)

Input: $\text{LMP}_{i,t}$: current price, avgLMP_i: reference, $backlog_{i,t-1}$
Output: $l_{i,t}$: datacenter load, $backlog_{i,t}$

1: Select load level from $\{avgLoad + \frac{dr}{2}, avgLoad, avgLoad - \frac{dr}{2}\}$
2: to Minimize $l_{i,t} * \text{LMP}_{i,t} +$
$(backlog_{i,t-1} + avgLoad - l_{i,t}) * \text{avgLMP}_i$
3: $backlog_{i,t} = backlog_{i,t-1} + avgLoad - l_{i,t}$
4: **return** $l_{i,t}, \ backlog_{i,t}$

Each datacenter uses a local average price, $avgLMP_i$, as reference. The DP-ONL algorithm uses the current power price ($\text{LMP}_{i,t}$) and the reference ($avgLMP_i$, local average), weighted by the current load and future load to select a current load ($l_{i,t}$). The algorithm chooses from three levels—average load (70%), average load + half the dynamic range, average load - half the dynamic range. The catchup constraint at 24 hours can limit choice late in the day.

Algorithm 2 Grid Simulation with DP-ONL Datacenters

1: for all i,t: $l_{i,t} = avgLoad$, (Initialize Datacenter loads)
2: Solve grid OPF for $l_{i,t}$, defining initial prices ($\text{LMP}_{i,t}$) for all i,t (each DC for all 24 hours)
3: **for each** time interval t in [1,...,24] **do**
4: **for each** i **do**
5: $l_{i,t}, backlog_{i,t} = \text{DP-ONL}(\text{LMP}_{i,t}, \text{avgLMP}_i, backlog_{i,t-1})$
6: **end for**
7: Solve grid OPF using $l_{i,t}$ and setting new $\text{LMP}_{i,t}$ prices for future—$[t+1, ..., 24]$
8: **end for**

A coupled study must reflect the datacenters dynamic load changes in grid dispatch. To reflect the history and thereby enable dynamic programming decisions to affect the grid dispatch, we iterate forward over the hours of the day, computing the DP-ONL decisions for each datacenter, then power-grid dispatch for that hour.

*4.4.2 Grid-controlled Optimization (**GC**) of Datacenter Load.* In the GC model, the DCs fully delegate load flexibility to the power grid dispatch as in [59]. That is, grid dispatch sets the datacenter power level within the dynamic range for each hour, subject to the 24-hour average capacity and backlog constraints. Because the power levels are set by grid dispatch which maximizes social welfare (e.g. dispatch cost), it seeks to benefit the entire grid, not just the datacenters.

5 IMPACT OF GROWING DATACENTER LOAD AND RENEWABLE GENERATION

We consider datacenters as fixed loads in the base case. Each datacenter's load is 140 MW for all 24 hours.

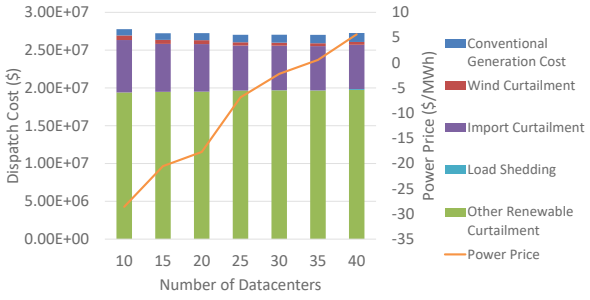
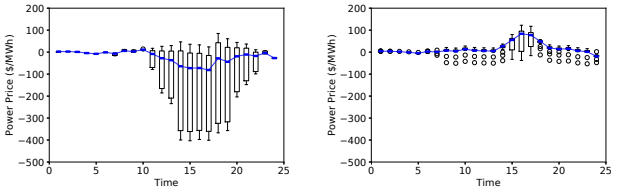


Figure 5: Dispatch Cost and Power Price by Datacenter Load Growth, 15% Wind.

5.1 Datacenter Load Growth

We first analyze the impact of adding 10–40 DCs (3.5-14% of average load) to the base system, under 15% wind penetration. Figure 5 shows how grid dispatch cost changes as we add datacenters. As the number of DCs increases, dispatch cost first decreases as more import and wind generation are consumed. From 10 to 25 DCs, the dispatch cost decreases by 2.7%, with 29.1% less wind curtailment and 14.2% less import curtailment. However, at the highest numbers of datacenters, transmission constraints become critical, and

the greater load is serviced by additional conventional generation, increasing dispatch cost by 0.9% from 35 to 40 DCs. In addition, significant renewable and import curtailment reflects the fact that absorption challenges still exist. Also shown in Figure 5, the average power price increases with the growth of datacenters, producing a $34.2/MWh increase, as we go from 10 to 40 DCs.



(a) 10 DCs (b) 40 DCs

Figure 6: Power Price across Datacenters, 15% Wind.

Figure 6 captures one day of power price variation across the different datacenter locations—spatial variation. For each hour of the day, the blue line shows the median price, and the box captures the interval between the first and third prices quartiles. In Figure 6a, the 10-DC scenario has excess renewables, especially when their production increases in the afternoon (coastal wind). This causes power prices decrease in some locations, producing a wider variation in power prices amongst the datacenters. In Figure 6b, the greater load associated with the 40 DCs reduces the excess, both reducing occurrence of negative prices and increasing the highest prices. The net effect is a much narrower spatial price variation.

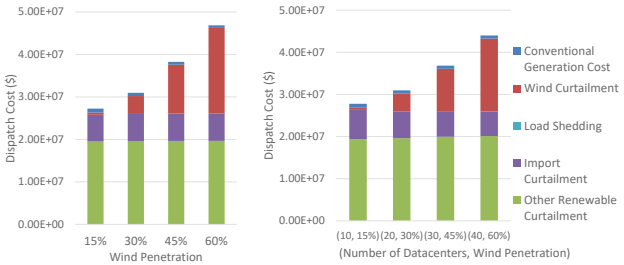
5.2 Renewable Generation Growth

Next we consider how the growth of renewable generation affects the grid. With a fixed number of DCs (20), we scale up wind generation from 15% to 60%, corresponding to the Wind Vision goal for 2050 [42]. Then we vary DCs and wind penetration together producing 4 scenarios: (10, 15%), (20, 30%), (30, 45%), (40, 60%).

5.2.1 Dispatch Cost. Figure 7a shows the change in dispatch cost as wind penetration grows. From 15% to 60% penetration, most elements of dispatch cost remain constant, and there is a small decrease in generation cost. However, wind curtailment (and associated curtailment penalty) increases sharply due to the grid's inability to absorb the increased wind generation. While the dispatch cost also increases sharply with combined wind and datacenter growth, the additional datacenters eliminate 11.3% and 14.3% wind curtailment in (30, 45%) and (40, 60%) cases respectively (Figure 7b).

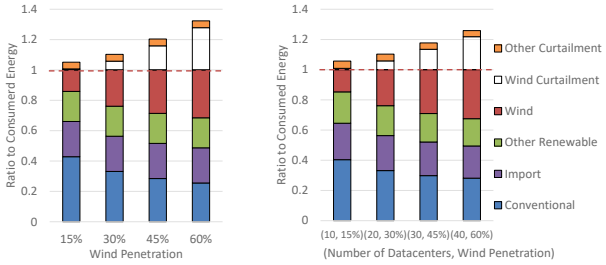
5.2.2 Generation Mix. In Figure 8a, as wind generation increases, it squeezes out the fossil-fuel based generation, ultimately displacing 40.4% of that generation. Over the same increase, the grid shows the diminishing absorption of wind generation: a 4-fold increase in generation nets only 2.2x increase in absorbed power. When both datacenters and wind generation are increased (Figure 8b), the added datacenter load increases wind absorption slightly but also conventional generation (fossil-fuel), producing a lower RPS.

Figure 9a details how each increment of datacenters reduces the achieved RPS. At 15% wind penetration, as fixed load datacenters



(a) 20 DCs, 15%-60% Wind (b) 10-40 DCs, 15%-60% Wind

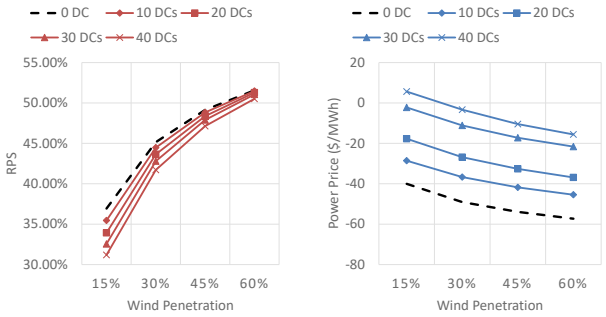
Figure 7: Dispatch Cost by Wind and Combined Growth.



(a) 20 DCs, 15%-60% Wind (b) 10-40 DCs, 15%-60% Wind

Figure 8: Generation Mix by Wind and Combined Growth.

are added (10, 20, 30, then 40), the grid RPS decreases by 1.47–5.74%. This decrease is because each increment of DC load is temporally misaligned with renewables. For higher wind penetrations, the grid RPS increases for all scenarios, but a penalty for added DC load remains. At 30% and 45% the penalty for 40 DCs is 3.44% and 2.01% respectively. At 60% wind, the range is 0.08–1.05% for 10–40 DCs. Note that a four-fold increase in wind has produced a 1.62-fold increase in RPS (at best!), due to growing wind curtailment.



(a) RPS (b) Power Price

Figure 9: RPS and Power Price by Wind Growth and Datacenter Load Growth.

5.2.3 Price. We present grid power prices for 10–40 DCs and 15%–60% wind penetration scenarios in Figure 9b. Each increment of DC

load increases the price of power across the range of wind levels. Again at higher levels of wind, excess generation produces wind curtailment and lower power prices.

5.3 Summary

Growing renewable generation presents absorption challenges for power grids, producing renewable curtailment that both increases costs and is a missed opportunity to reduce carbon emissions. Growing demand from datacenters as fixed loads can improve absorption, but does not eliminate curtailment. Worse, the addition of such loads decreases RPS due to temporal mismatch between new load and renewable supply. This frames an opportunity for improvement by coupling dynamic load of datacenters to the grid situation.

6 COUPLING DATACENTERS TO THE GRID

With the rapid growth of datacenter power consumption, the flexibility of computing load provides opportunities for co-optimization with the power grid. So, we consider three alternatives for coupling—no coupling, selfish DC cost-minimization, and global grid optimization by the power grid as defined in Section 4.4. We examine the impact on the power grid, the datacenters, and the other non-DC grid customers. The trends are summarized here, with detailed numerical results included in Appendix D.

6.1 Impact on Power Grid

6.1.1 Dispatch Cost. To highlight the impact of dynamic coupling, we consider DP-ONL (selfish DC) and GC (grid control) relative to the constant datacenter power case (no coupling). Figure 10 shows the dispatch cost impact of coupling models by wind penetration and dynamic range. First, the solid bars show a consistent 1% reduction in grid dispatch cost for both DP-ONL and GC (approximately $450,000 dollars/day) for a narrow dynamic range [0.6, 0.8]. GC performs slightly better. With a larger dynamic range [0.4, 1.0], the reduction in grid dispatch cost is over two times larger in nearly all cases, with reductions as large as 3% with GC. The one exception is where generation is tightest (15% Wind, 30 DCs), and DP-ONL (selfish) causes significant harm, increasing dispatch cost by 0.1%. We performed a deeper study and found that the harm is due to overshifting load in response to price signals, which also increases conventional generation (see Figure 12).

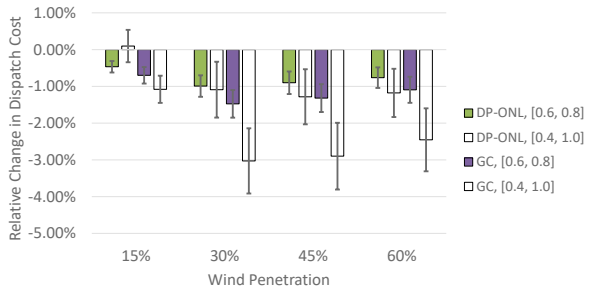
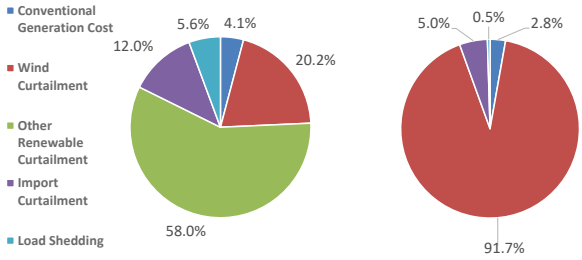


Figure 10: Change in Dispatch Cost by Wind Penetration and Dynamic Range, 30 DCs. (Error bars reflect standard deviation across wind scenarios.)

The primary benefit of coupling is to reduce renewable curtailment. This reduces dispatch cost by avoiding curtailment penalties, and also increases renewable absorption. As wind penetration increases, the mix of dispatch cost that coupling eliminates shifts. As shown in Figure 11, at 15% wind, the coupling eliminates curtailment penalties from a mix of wind (20%) and other renewables (58%), but there is a significant fraction from other sources. At 60% wind penetration, the curtailment penalties eliminated are nearly exclusively wind curtailment (92%). We examine these dynamics more closely in following sections.



(a) 15% Wind Penetration **(b) 60% Wind Penetration**

Figure 11: Contribution to reduction is dominated by reduced curtailment—GC examples (30 DCs, [0.4, 1.0]).

While distinct owners or operators might require independence, such independent control can create an overshifting problem (Figure 10). We examine this issue more closely, looking at the effect of DP-ONL shifting on conventional generation in Figure 12. The scatterplot illustrates how shifts by DP-ONL in the (30 DCs, 15%, [0.4, 1.0], SummerWD) scenario often fail to produce the anticipated benefit (shown as the red line). When DP-ONL increases the datacenter load (the right half of the x-axis), it expects that there is excess wind generation, but the grid effect is to increase conventional (fossil-fuel) generation. The datacenter load increase may not improve renewable absorption, relative to the unshifted baseline. When DP-ONL decreases load (left half of x-axis), it hopes to reduce conventional generation—by the corresponding amount. As the plot shows, sometimes conventional generation is reduced, but often by less than desired. In this example, uncoordinated, independent management at each of the 30 DCs by DP-ONL increases conventional generation by 1,821 MWh. The dynamic control incurs higher dispatch cost and even produces lower RPS than fixed-load datacenters. This result shows the importance of modeling grid dynamics in detail with multiple datacenters. Isolated study of datacenters without a grid model would have significantly overestimated DP-ONL's shifting benefits for price and carbon emissions.

6.1.2 Generation Mix. Let's consider generation mix, looking at how coupling affects overall grid renewable absorption as captured by RPS (or renewable fraction). In Section 5.2, we saw that the addition of datacenters as fixed loads can reduce RPS.

Both DP-ONL and GC improve RPS when compared to fixed datacenter loads as shown in Figure 13. RPS improvement from coupling increases with wind penetration, and GC (grid-wide) optimization gives clearly greater improvements at both low and high

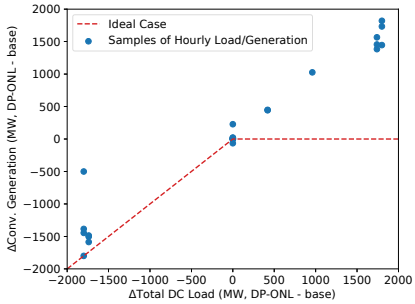


Figure 12: Overshifting can increase conventional generation: DP-ONL, 30 DCs, 15%, [0.4, 1.0], SummerWD, Wind Scenario No. 200.

dynamic range. DP-ONL's overshifting impacts RPS negatively at low wind levels, and particularly with large dynamic range. In contrast, GC gives robust RPS improvement that grows with both wind penetration and dynamic range. For 60% wind penetration, the RPS improvement with 40 DCs reaches 1.58% RPS, matching the RPS decrease we observed in adding the 40 DCs (Section 5.2). In short, coupling with GC would allow the addition of 40 DCs to be neutral (no negative effect) to grid RPS. This a positive highlight of the benefits from intelligently coupling datacenters with the grid.

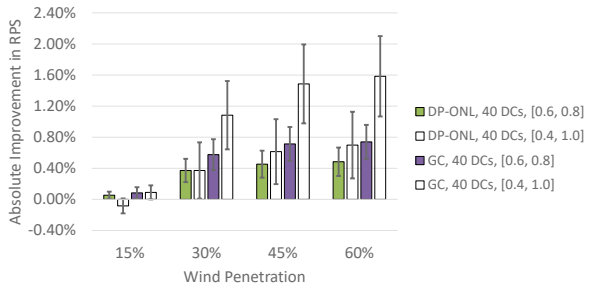


Figure 13: Improvement in Grid RPS by Wind Penetration and Dynamic Range, 40 DCs.

Next we focus narrowly on how coupling affects datacenter carbon emissions (RPS). We compute datacenter RPS as the weighted average of hourly grid RPS and datacenter hourly load. In Figure 14, the fixed-load datacenter RPS (red), increases with wind penetration. Dynamic coupling, using DP-ONL and GC, increases datacenter RPS by an additional 0.1%–1.9%, a benefit that increases with wind penetration. GC outperforms DP-ONL significantly, with a maximum benefit more than doubling that from fixed load. Here again we see DP-ONL's overshifting harm. DP-ONL's overshifting reduces datacenter RPS, underperforming both when the generation is tight and at high dynamic range.

6.1.3 Power Price. To assess the impact on grid customers, we consider the average power price. We first consider the entire grid in Figure 15a, and then break this down into DC and non-DC power prices. Starting from the left, increased wind penetration decreases

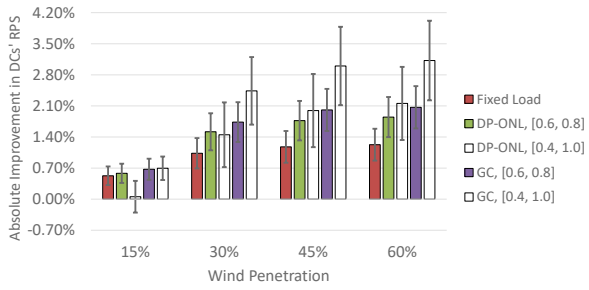


Figure 14: Datacenter RPS Improvement relative to Grid RPS by Wind Penetration and Dynamic Range, 40 DCs.

average grid power prices and increased datacenter load produces higher prices. At lower load and wind penetration, DP-ONL and GC appear to increase competition for power—working against the dynamics of the grid, producing higher prices. At higher wind fractions, both DP-ONL and GC have more room to work smoothly with the grid, shifting power to mitigate competition, and reducing average price. GC, with its global view does this more effectively.

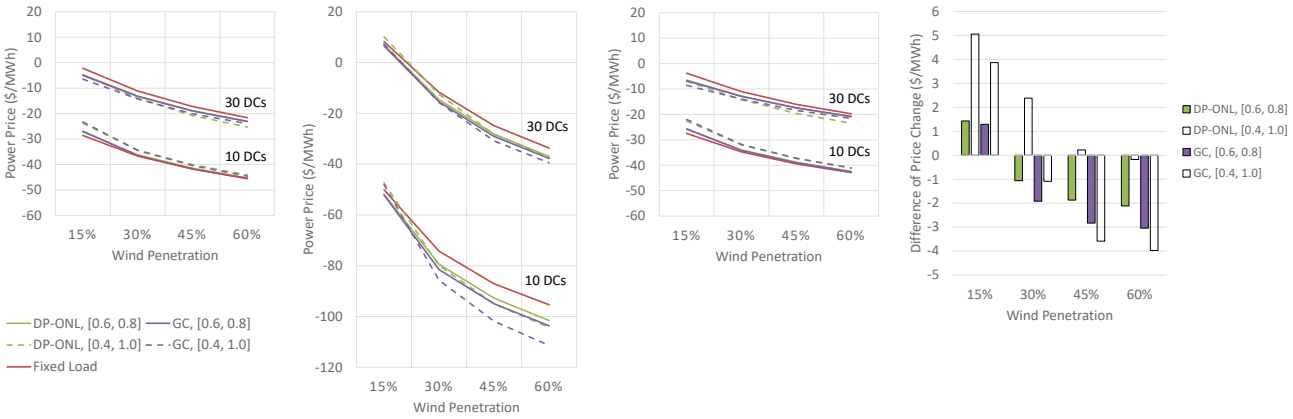
6.2 Differential Impact on Datacenter and Non-Datacenter Prices

Figure 15b shows the impact of dynamic management on datacenter power prices. In the 10-DC scenario, the DCs experience prices much lower than the overall grid, and the prices decrease faster than with fixed loads as wind penetration is increased, producing as much as a 16.9% decrease relative to fixed for (GC, 45% wind penetration, [0.4, 1.0]). For 30 DCs, prices start higher due to tight supply, but fall below overall grid prices with higher wind penetration.

As shown in Figure 15c, non-DC customers, the majority of grid customers, show trends similar to the overall grid average—they have much higher prices across the board for the 10 DC case, but when load is increased to 30 DCs, they get much lower prices, with the advantage diminishing as the wind penetration grows.

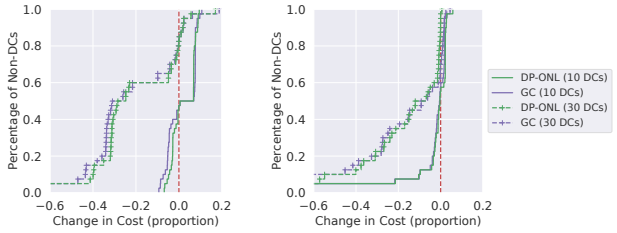
To highlight differences, we subtract the non-DC customers price change from datacenters price change in Figure 15d, presenting the differential price impact. With 30 datacenters and conservative dynamic range ([0.6, 0.8]), both coupling schemes benefit datacenters. However, when the datacenters have large dynamic range ([0.4, 1.0]), at low wind penetrations, their shifing harms themselves, producing a DC-nonDC price differential of as much as $5.1/MWh.

We then drill down into distributions of non-DC costs (Figure 16) to figure out the impact of coupling models and datacenter dynamic management, presenting the cumulative distribution function (CDF) of non-DC customers' cost changes (ranging from a 60% reduction to a 20% increase). At low datacenter load and wind penetration, for both DP-ONL and GC, nearly 50% of the non-DC customers experience harm (about 10% price increase). With 30 DCs and low wind, competition arises. In this case non-DC customers split, with some experiencing more harm and more gaining large benefit. GC outperforms DP-ONL slightly with similar harm but larger benefit. At higher (60%) wind penetration, the situation is much better due to less competition for power, with DP-ONL and GC producing



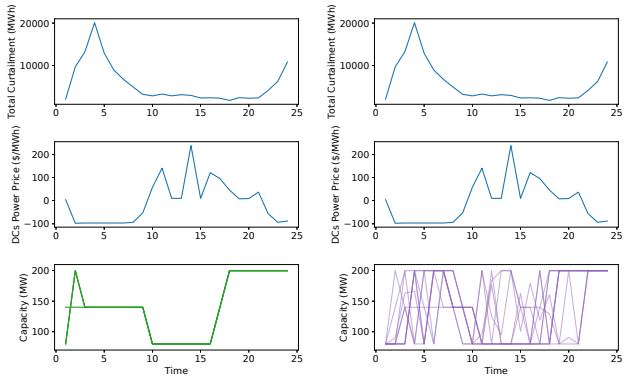
(a) Entire Grid (b) Datacenters (c) Non-DC Customers (d) Difference in Price Change, 30 DCs

Figure 15: Power Price by Wind Penetration, Number of Datacenters, and Dynamic Range. (a, b, c). Relative Change in Prices (DC - non-DC), 30 DCs. (positive: favors non-DC, negative: favors DC)



(a) 15% Wind Penetration (b) 60% Wind Penetration

Figure 16: Distribution (CDF) of Change in Cost for Non-DC Customers, [0.6, 0.8] Dynamic Range.



(a) DP-ONL (b) GC

Figure 17: Spatial and Temporal Capacity Variation (30 DCs, 60%, [0.4, 1.0] SummerWD, Wind Scenario No. 200). Top to bottom: total curtailment (fixed load), DC power price (fixed load), datacenter capacity (only 10 shown for readability).

similar effects. Essentially no customers are harmed, and 40% of them see slight harm (< 5%). Many non-DC customers even see more than 60% benefit with 30 DCs.

6.3 Datacenter Capacity Variation

Thus far, we have focused on the benefits of optimized power acquisition—lower power price or overall grid dispatch cost. However, dynamic coupling mechanisms increase or decrease the available power, changing the *computation capacity* of the datacenters, incurring a potential penalty. We illustrate this variation across time for 10 of 30 DCs in Figure 17 (SummerWD, 60% wind, 30 DCs, [0.4, 1.0])—the first two rows are curtailment and price. In the last row, the two coupling models produce varied power levels and distinct capacity variation patterns. With relatively uniform behaviors, DP-ONL defers workload to avoid high prices during daytime, but fails to create backlog and exploit low prices in the early morning. GC creates backlog and temporal load shifting, capturing more wind curtailment, exploiting low prices better. It exploits spatial diversity (see last row of Figure 17), all at the cost of frequent maximum capacity change, and large average capacity change.

To understand variability for individual datacenters, we summed the hourly capacity changes for each datacenter, showing the results in Figure 18 (30 DCs). DP-ONL creates significant hourly variation in capacity with hourly changes averaging 6–8MW (4–5%) with small dynamic range and 14–20MW (10–14%) with large dynamic range, and the capacity variation increases steadily with wind penetration. The situation for GC is more variable. For small dynamic range, GC produces larger hourly changes of 15 MW (11%), and with large dynamic range 40–42 MW (28–30%). Such large average changes roughly correspond to a random capacity distribution, and represent a challenge to cloud resource managers.

6.4 Summary

Our studies on intelligent management of power-level show that even modest numbers of datacenters and small dynamic range can change renewable absorption and power prices significantly.

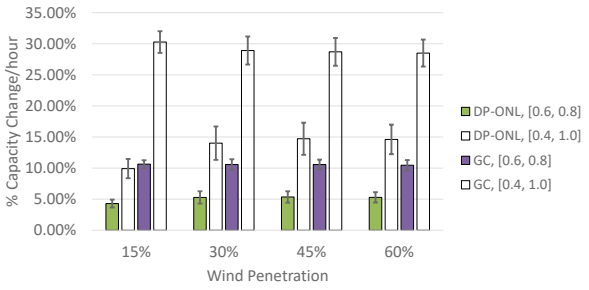


Figure 18: Average Hourly Capacity Variation (normalized by average capacity) by Wind Penetration and Dynamic Range, 30 DCs.

Both selfish-local (DP-ONL) or global grid (GC) coupling can improve renewable absorption and reduce power cost; GC is capable of restoring the damage to RPS that fixed datacenter loads create. However, the impact of dynamic management is sensitive to wind penetration, total load, and dynamic range. GC, delegating flexibility to the grid, with largest dynamic range gives largest benefits for RPS (absorption) and dispatch cost, but DP-ONL can also provide significant benefits. While both coupling approaches reduce datacenter power costs, because of their market power, they also produce a negative relative impact on power cost of other customers. The concerns about fairness that arise from this are also worthy of study. Local techniques (DP-ONL) can cause overshifting harm, reducing RPS and increasing power prices. All of these results show that understanding the effects of workload shifting must include detailed grid models to be accurate. While good for the grid, both DP-ONL and GC can cause datacenter capacity to fluctuate, challenging datacenter efficiency.

7 RELATED WORK

This study is the first to explore large-scale dynamic-management of datacenter power coupled with the power grid. We also consider independent optimization, as might occur with multiple datacenter operators. We review closely related work below.

Datacenters as Demand-Response. Researchers propose compute workload shedding or deferral to enable datcenters to reduce power consumption in response to a grid shortage. These approaches can reduce power costs, and priority and time shifting are used to minimize SLO (service-level objective) impact [29, 36]. Others have created markets that distribute the economic benefits of demand-response to flexible customers—both datacenters [35] or tenants colocated in a datacenter [7, 8, 51, 58]. In others, datacenter operators shift load across grids based on grid information about social welfare, optimizing social welfare across grids [61, 63].

Demand-response is one-way power management—directed by the power grid. Further, demand-response is designed for rare high load periods (typically < 10 periods/year), and target small power reductions (e.g. 10%). In contrast, our study considers continuous two-way coupling with large dynamic ranges up to 60%. Further, we consider action of multiple datacenters shifting load that is large enough to change economic dispatch in the power grid.

Datacenters as Dispatchable Loads. The closest related work is the Zero-carbon Cloud project (ZCCloud), which proposed DCs with 100% dynamic range, with power level determined by the grid (i.e. GC). ZCCloud DCs were only powered when excess zero-carbon power was available [10, 55, 56]. ZCCloud research showed that datacenters could increase grid RPS, reduce dispatch cost, and reduce renewable curtailment. GC in this paper can be thought of as moderate version of ZCCloud, allowing limited grid control. Aligned efforts exploring spatial and temporal load shifting have demonstrated grid benefits of reduced price variation, improved renewable absorption, and lower dispatch cost [32, 59].

Local Price or Carbon Optimization. Research efforts have explored power purchase to reduce energy costs, or utilize local and grid renewables. These approaches typically employ selfish optimization based on dynamic programming, optimal online control and optimization [15, 16, 31, 33, 47, 53, 57], prediction [37], and even machine learning [3] to determine the quantity of power purchase or charge/discharge energy storage. This work usually assumes datacenter load changes are small enough that they do not affect prices or grid dispatch. In contrast, we studied hyperscale cloud datacenters whose sizes can be 200 MW (and growing) and therefore can individually and collectively affect power markets and grid dispatch; consequently grid coupling is a primary focus.

Global Price or Carbon Optimization. With the same goals as local optimization, some efforts perform global optimization on a network of datacenters instead [19, 28, 34, 45, 60, 62, 64]. These efforts couple these decisions to internal datacenter management—employing load balancing or shifting across datacenters, a complementary focus to our work.

8 CONCLUSION AND FUTURE WORK

Growing renewable generation challenges the ability of power grids to absorb it. Further, adding fixed loads such as datacenters can damage RPS. We study load shifting, using a detailed grid model to accurately capture the dynamic impacts of cloud-scale datacenter shifts on power price and carbon-emissions. We study three coupling models—fixed loads, selfish local optimization, and grid orchestrated loads—and show that selfish-local methods give some benefits (improved grid dispatch cost), but their uncoordinated actions cause harm (overshifting that increases prices and carbon emissions). Larger grid benefits arise with grid control (GC) in grid dispatch cost, datacenter power prices, and improved renewable absorption. However, GC doubles datacenter capacity variation compared to DP-ONL, which may harm datacenter efficiency.

Looking forward, design of better coupling models presents challenges of autonomy, market fairness, and datacenter efficiency. Interesting future research directions include advanced local optimization techniques and coordination to avoid overshifting.

ACKNOWLEDGMENTS

Thanks to the reviewers, shepherd David Hutchison, the Zero-carbon Cloud team, and Arielle Rosenthal! This work is supported by NSF Grants CMMI-1832230, CMMI-1832208, and CNS-1901466; also thanks for support from Google, Intel, Samsung, and VMWare.

REFERENCES

[1] 2018. Lancium. https://www.lancium.com. A startup company, building zero-carbon cloud computing resources.

[2] 2019. Clicking Clean Virginia: The Dirty Energy Powering Data Center Alley. https://www.greenpeace.org/usa/reports/click-clean-virginia/.

[3] Sohaib Ahmad, Arielle Rosenthal, Mohammad H. Hajiesmaili, and Ramesh K. Sitaraman. 2019. Learning from Optimal: Energy Procurement Strategies for Data Centers. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems* (Phoenix, AZ, USA) *(e-Energy '19)*. Association for Computing Machinery, New York, NY, USA, 326–330. https://doi.org/10.1145/3307772.3328308

[4] Amazon Web Services (AWS). [n.d.]. Amazon Global Datacenters.

[5] Lori Bird, M Milligan, and Debra Lew. 2013. *Integrating Variable Renewable Energy: Challenges and Solutions.* Technical Report. NREL.

[6] Bloomberg. 2018. European Union Aims to Be First Carbon Neutral Major Economy by 2050. *Fortune* (November 2018).

[7] Niangjun Chen, Xiaoqi Ren, Shaolei Ren, and Adam Wierman. 2015. Greening multi-tenant data center demand response. *Performance Evaluation* 91 (2015), 229–254.

[8] Shutong Chen, Lei Jiao, Lin Wang, and Fangming Liu. 2019. An online market mechanism for edge emergency demand response via cloudlet control. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2566–2574.

[9] Andrew A Chien. 2020. Characterizing Opportunity Power in the California Independent System Operator (CAISO) in Years 2015-2017. *Energy and Earth Science* 3, 2 (December 2020). Also available as University of Chicago, Computer Science TR-2018-07, https://newtraell.cs.uchicago.edu/research/publications/techreports.

[10] Andrew A Chien, Richard Wolski, and Fan Yang. 2015. The Zero-Carbon Cloud: High-Value, Dispatchable Demand for Renewable Power Generators. *The Electricity Journal* (2015), 110–118.

[11] Andrew A. Chien, Fan Yang, and Chaojie Zhang. 2018. Characterizing Curtailed and Uneconomic Renewable Power in the Mid-continent Independent System Operator. *AIMS Energy* 6, 2 (December 2018), 376–401.

[12] CPUC. [n.d.]. California Public Utilities Commission (CPUC). http://www.cpuc.ca.gov/PUC/energy/Renewables.

[13] Wei Deng, Fangming Liu, Hai Jin, Bo Li, and Dan Li. 2014. Harnessing renewable energy in cloud datacenters: opportunities and challenges. *iEEE Network* 28, 1 (2014), 48–55.

[14] Dominion Energy. 2019. Rapid Growth of Datacenter Power requirements in Northern Virginia. Presentation at Power-Systems Engineering Research Center, Datacenters are 5% of power grid load, and growing fast.

[15] Hui Dou, Yong Qi, Wei Wei, and Houbing Song. 2017. Carbon-aware electricity cost minimization for sustainable data centers. *IEEE Transactions on Sustainable Computing* 2, 2 (2017), 211–223.

[16] Íñigo Goiri, William Katsak, Kien Le, Thu D Nguyen, and Ricardo Bianchini. 2013. Parasol and greenswitch: Managing datacenters powered by renewable energy. In *ACM SIGARCH Computer Architecture News*. ACM, 51–64.

[17] Google. [n.d.]. About Google Datacenters. https://www.google.com/about/datacenters/.

[18] Google. 2018. *Moving toward 24x7 Carbon-Free Energy at Google Data Centers: Progress and Insights.* Technical Report. Google.

[19] V. Gupta, P. Shenoy, and R. K. Sitaraman. 2018. Efficient solar provisioning for net-zero Internet-scale distributed networks. In *2018 10th International Conference on Communication Systems Networks (COMSNETS)*. 372–379.

[20] LLC Gurobi Optimization. 2020. Gurobi Optimizer Reference Manual. http://www.gurobi.com

[21] GWEC. 2016. *Global Wind Report: Annual Market Update.* Technical Report. Global Wind Energy Council. Documents curtailment around the world.

[22] Siqi Han. 2015. The Wind is Wasted in China. https://www.wilsoncenter.org/.

[23] M Huneault and FD Galiana. 1991. A survey of the optimal power flow literature. *IEEE transactions on Power Systems* 6, 2 (1991), 762–770.

[24] Willam Stanley Jevons. 1865. *The Coal Question; An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal Mines.* Macmillan and Company. London.

[25] Nicola Jones. 2018. How to Stop Data Centres from Gobbling up the World's Electricity. *Nature* (September 2018).

[26] Kibaek Kim, Fan Yang, Victor Zavala, and Andrew A. Chien. 2016. Data Centers as Dispatchable Loads to Harness Stranded Power. *IEEE Transactions on Sustainable Energy* (2016). DOI 10.1109/TSTE.2016.2593607.

[27] Will Knight. 2020. AI Can Do Great Things—if It Doesn't Burn the Planet. (January 2020). https://www.wired.com/story/ai-great-things-burn-planet/.

[28] Kien Le, Ricardo Bianchini, Thu D. Nguyen, Ozlem Bilgir, and Margaret Martonosi. 2010. Capping the Brown Energy Consumption of Internet Services at Low Cost. In *Proceedings of the International Conference on Green Computing (GREENCOMP '10)*. IEEE Computer Society, USA, 3–14. https://doi.org/10.1109/GREENCOMP.2010.5598305

[29] Tan N Le, Zhenhua Liu, Yuan Chen, and Cullen Bash. 2016. Joint capacity planning and operational management for sustainable data centers and demand response. In *Proceedings of the Seventh International Conference on Future Energy Systems*. 1–12.

[30] Liuzixuan Lin and Andrew A. Chien. 2020. *Characterizing Stranded Power in the ERCOT in Years 2012-2019: A Preliminary Report.* Technical Report TR-2020-06. University of Chicago.

[31] Minghong Lin, Adam Wierman, Lachlan LH Andrew, and Eno Thereska. 2012. Dynamic right-sizing for power-proportional data centers. *IEEE/ACM Transactions on Networking* 21, 5 (2012), 1378–1391.

[32] Julia Lindberg, Line Roald, and Bernard Lesieutre. 2020. The Environmental Potential of Hyper-Scale Data Centers: Using Locational Marginal CO2 Emissions to Guide Geographical Load Shifting. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. 3158.

[33] Zhenhua Liu, Yuan Chen, Cullen Bash, Adam Wierman, Daniel Gmach, Zhikui Wang, Manish Marwah, and Chris Hyser. 2012. Renewable and cooling aware workload management for sustainable data centers. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*. 175–186.

[34] Z. Liu, M. Lin, A. Wierman, S. Low, and L. L. H. Andrew. 2015. Greening Geographical Load Balancing. *IEEE/ACM Transactions on Networking* 23, 2 (2015), 657–671.

[35] Zhenhua Liu, Iris Liu, Steven Low, and Adam Wierman. 2014. Pricing Data Center Demand Response. *SIGMETRICS Perform. Eval. Rev.* 42, 1 (June 2014), 111–123. https://doi.org/10.1145/2637364.2592004

[36] Zhenhua Liu, Adam Wierman, Yuan Chen, Benjamin Razon, and Niangjun Chen. 2013. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. *Performance Evaluation* 70, 10 (2013), 770–791.

[37] Jianying Luo, Lei Rao, and Xue Liu. 2013. Temporal load balancing with service delay guarantees for data center energy cost optimization. *IEEE Transactions on Parallel and Distributed Systems* 25, 3 (2013), 775–784.

[38] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. Recalibrating global data center energy-use estimates. *Science* 367, 6481 (2020), 984–986. https://doi.org/10.1126/science.aba3758 arXiv:https://science.sciencemag.org/content/367/6481/984.full.pdf

[39] Microsoft Azure. [n.d.]. Azure Global Datacenters. https://azure.microsoft.com/en-us/global-infrastructure/.

[40] Anthony Papavasiliou and Shmuel S Oren. 2013. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research* 61, 3 (2013), 578–592.

[41] Brian Peccarelli. 2020. Three Ways COVID-19 is Accelerating Digital Transformation in Professional Services. (June 2020). https://bit.ly/34Qitb5, 37% growth.

[42] DOE Wind Program. 2015. *Wind Vision: A New Era for Wind Power in the United States.* Technical Report. DOE National Renewable Energy Laboratory, http://energy.gov/eere/wind/wind-vision.

[43] New York State Energy Planning Board. 2015. The Energy to Lead: 2015 New York State Energy Plan. http://energyplan.ny.gov/Plans/2015.aspx.

[44] SC Pryor, RJ Barthelmie, and TJ Shepherd. 2020. 20% of US electricity from wind will have limited impacts on system efficiency and regional climate. *Scientific reports* 10, 1 (2020), 1–14.

[45] Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs. 2009. Cutting the Electric Bill for Internet-Scale Systems. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication* (Barcelona, Spain) *(SIGCOMM '09)*. Association for Computing Machinery, New York, NY, USA, 123–134. https://doi.org/10.1145/1592568.1592584

[46] Ana Radovanovic. 2020. Our data centers now work harder when the sun shines and wind blows. https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows.

[47] Yuanyuan Shi, Bolun Xu, Baosen Zhang, and Di Wang. 2016. Leveraging energy storage to optimize data center electricity cost in emerging power markets. In *Proceedings of the Seventh International Conference on Future Energy Systems*. 1–13.

[48] Soluna Inc. 2018. 900 MW Wind-only Powered Data Center Project. https://www.soluna.io/.

[49] Staff. 2020. COVID-19 Accelerates Cloud Adoption, Market to Reach $1 trillion, IDC. *Equipment FA News* (October 2020). https://www.equipmentfa.com/news/31459/covid-19-accelerates-cloud-adoption-market-to-reach-1t-idc.

[50] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2020. Energy and Policy Considerations for Modern Deep Learning Research. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 9 (April 2020).

[51] Qihang Sun, Shaolei Ren, Chuan Wu, and Zongpeng Li. 2016. An online incentive mechanism for emergency demand response in geo-distributed colocation data centers. In *Proceedings of the seventh international conference on future energy systems*. 1–13.

[52] Muhammad Tirmazi, Adam Barker, Nan Deng, Md E. Haque, Zhijing Gene Qin, Steven Hand, Mor Harchol-Balter, and John Wilkes. 2020. Borg: The next Generation. In *Proceedings of the Fifteenth European Conference on Computer Systems* (Heraklion, Greece) *(EuroSys '20)*. Association for Computing Machinery, New York, NY, USA, Article 30, 14 pages. https://doi.org/10.1145/3342195.3387517

[53] Rahul Urgaonkar, Bhuvan Urgaonkar, Michael J Neely, and Anand Sivasubramaniam. 2011. Optimal power cost management using stored energy in data centers. In *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*. 221–232.

[54] Ryan H. Wiser, Andrew D. Mills, Joachim Seel, Todd Levin, and Audun Botterud. 2017. *Impacts of Variable Renewable Energy on Bulk Power System Assets, Pricing, and Costs.* Technical Report LBNL-2001082. A link to a webinar recorded on December 13, 2017 can be found at https://youtu.be/EMrFAklQnPI.

[55] Fan Yang and Andrew A Chien. 2016. ZCCloud: Exploring wasted green power for high-performance computing. In *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 1051–1060.

[56] Fan Yang and Andrew A. Chien. 2017. Large-scale and Extreme-Scale Computing with Stranded Green Power: Opportunities and Costs. *IEEE Transactions on Parallel and Distributed Systems* 29, 5 (December 2017).

[57] Lin Yang, Mohammad Hassan Hajiesmaili, Ramesh K. Sitaraman, Enrique Mallada, Wing Shing Wong, and Adam Wierman. 2019. Online Inventory Management with Application to Energy Procurement in Data Centers. *CoRR* abs/1901.04372 (2019). arXiv:1901.04372 http://arxiv.org/abs/1901.04372

[58] Linquan Zhang, Shaolei Ren, Chuan Wu, and Zongpeng Li. 2015. A truthful incentive mechanism for emergency demand response in colocation data centers. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2632–2640.

[59] Weiqi Zhang, Line A Roald, Andrew A Chien, John R Birge, and Victor M Zavala. 2020. Flexibility from networks of data centers: A market clearing formulation with virtual links. *Electric Power Systems Research* 189 (2020), 106723.

[60] Jiajia Zheng, Andrew A. Chien, and Sangwon Suh. 2020. Mitigating Curtailment and Carbon Emissions through Load Migration between Data Centers. *Joule* (October 2020). https://doi.org/10.1016/j.joule.2020.08.001

[61] Zhi Zhou, Fangming Liu, Shutong Chen, and Zongpeng Li. 2018. A truthful and efficient incentive mechanism for demand response in green datacenters. *IEEE Transactions on Parallel and Distributed Systems* 31, 1 (2018), 1–15.

[62] Zhi Zhou, Fangming Liu, Bo Li, Baochun Li, Hai Jin, Ruolan Zou, and Zhiyong Liu. 2014. Fuel cell generation in geo-distributed cloud services: A quantitative study. In *2014 IEEE 34th International Conference on Distributed Computing Systems*. IEEE, 52–61.

[63] Zhi Zhou, Fangming Liu, Zongpeng Li, and Hai Jin. 2015. When smart grid meets geo-distributed cloud: An auction approach to datacenter demand response. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2650–2658.

[64] Zhi Zhou, Fangming Liu, Ruolan Zou, Jiangchuan Liu, Hong Xu, and Hai Jin. 2016. Carbon-Aware Online Control of Geo-Distributed Cloud Services. *IEEE Transactions on Parallel and Distributed Systems* 27, 9 (2016), 2506–2519.

A OPTIMAL POWER FLOW FORMULATION

We use an optimal power flow formulation from [26]. For reviewer convenience, that formulation is reprinted below.

A.1 Notation

To begin with, the model notations are listed in the following table. The units for power/load, energy, and phase angle are megawatts, megawatt-hours, and degrees respectively.

Sets:

$\mathcal{D}; \mathcal{D}_n$ Demand loads; demand loads at bus n
$\mathcal{G}; \mathcal{G}_n$ Generators; generators at bus n
$\mathcal{I}; \mathcal{I}_n$ Import points; import points at bus n
$\mathcal{L}$ Transmission lines
$\mathcal{L}_n^+; \mathcal{L}_n^-$ Transmission lines to bus n; lines from bus n
$\mathcal{N}$ Buses
$\mathcal{R}; \mathcal{R}_n$ Renewable generators; Renewable generators at bus n
$\mathcal{T}$ Time periods
$\mathcal{W}; \mathcal{W}_n$ Wind-farm locations; wind farms at bus n
$DC; DC_n$ Datacenter locations; datacenters at bus n

Parameters:

B_l Susceptance of transmission line l
C_i Generation cost of generator i
C_j^d Load-shedding penalty at load j
C_i^w Curtailment penalty at wind farm i
C_i^m Curtailment penalty at import point i
C_i^r Curtailment penalty at renewable i
$D_{j,t}$ Demand load of consumer j at time t
F_l^{max} Maximum power flow of transmission line l
$M_{i,t}$ Power production of import i at time t
P_i^{max} Maximum power output of generator i
$R_{i,t}$ Power production of renewable i at time t
RU_i Ramp-up limit of generator i
RD_i Ramp-down limit of generator i
$W_{w,t}$ Power from wind farm w at time t
$\Theta_{n,t}^{min}$ Minimum phase angle at bus n at time t
$\Theta_{n,t}^{max}$ Maximum phase angle at bus n at time t

Decision variables:

$d_{j,t}$ Load shedding at load j at time t
$f_{l,t}$ Power flow of line l at time t
$m_{i,t}$ Curtailment at import i at time t
$p_{i,t}$ Power from generator i at time t
$r_{i,t}$ Curtailment at renewable i at time t
$w_{i,t}$ Curtailment at wind farm i at time t
$\theta_{n,t}$ Phase angle at bus n at time t

Others:

$l_{i,t}$ Power level (load) of datacenter i at time t

A.2 Optimal Power Flow Model

Given the parameters, the power grid solves the following direct-current optimal power flow model (called "economic dispatch" in [26]) to minimize the dispatch cost (objective 4a):

$$\min \quad \sum_{t\in\mathcal{T}}\left(\sum_{i\in\mathcal{G}} C_i p_{i,t} + \sum_{j\in\mathcal{D}} C_j^d d_{j,t} + \sum_{i\in\mathcal{I}} C_i^m m_{i,t} + \sum_{i\in\mathcal{W}} C_i^w w_{i,t} + \sum_{i\in\mathcal{R}} C_i^r r_{i,t}\right) \tag{4a}$$

$$\text{s.t.} \quad \sum_{l\in\mathcal{L}_n^+} f_{l,t} - \sum_{l\in\mathcal{L}_n^-} f_{l,t} + \sum_{i\in\mathcal{G}_n} p_{i,t} + \sum_{i\in\mathcal{I}_n}(M_{i,t} - m_{i,t}) + \sum_{i\in\mathcal{W}_n}(W_{i,t} - w_{i,t}) + \sum_{i\in\mathcal{R}_n}(R_{i,t} - r_{i,t}) = \sum_{j\in\mathcal{D}_n}(D_{j,t} - d_{j,t}) + \sum_{i\in DC_n} l_{i,t}, \quad \forall n\in\mathcal{N}, t\in\mathcal{T}, \tag{4b}$$

$$f_{l,t} = B_l(\theta_{n,t} - \theta_{m,t}), \quad \forall l = (m,n)\in\mathcal{L}, t\in\mathcal{T}, \tag{4c}$$

$$-F_l^{max} \le f_{l,t} \le F_l^{max}, \quad \forall l\in\mathcal{L}, t\in\mathcal{T}, \tag{4d}$$

$$\Theta_n^{min} \le \theta_{n,t} \le \Theta_n^{max} \quad \forall n\in\mathcal{N}, t\in\mathcal{T}, \tag{4e}$$

$$-RD_i \le p_{i,t} - p_{i,t-1} \le RU_i, \quad \forall i\in\mathcal{G}, t\in\mathcal{T}, \tag{4f}$$

$$0 \le p_{i,t} \le P_i^{max}, \quad \forall i\in\mathcal{G}, t\in\mathcal{T}, \tag{4g}$$

$$0 \le d_{j,t} \le D_{j,t}, \quad \forall j\in\mathcal{D}, t\in\mathcal{T}, \tag{4h}$$

$$0 \le m_{i,t} \le M_{j,t}, \quad \forall i\in\mathcal{I}, t\in\mathcal{T}, \tag{4i}$$

$$0 \le w_{i,t} \le W_{j,t}, \quad \forall i\in\mathcal{W}, t\in\mathcal{T}, \tag{4j}$$

$$0 \le r_{i,t} \le R_{j,t}, \quad \forall i\in\mathcal{R}, t\in\mathcal{T}. \tag{4k}$$

In this model, power is supplied from conventional thermal power plants (e.g. gas, nuclear, coal), non-wind renewables (e.g. hydro), imports, and wind power plants. Following [26], the imports and

renewables are non-dispatchable due to long-term commitments and goal of reducing carbon emissions, but they can be curtailed at the cost of C_i^m and C_i^r \$/MWh (C_i^w \$/MWh for wind) respectively. In addition, each unit of load shedding is at the cost of value of lost load (VOLL) C_j^d. Therefore, the dispatch cost (4a) consists of generation cost of conventional power plants and penalties of curtailment and load shedding.

The constraints are typical for power grid models. Constraints 4c–4e represent how the flow (4c) is determined given the line capacity (4d) and phase angle (4e) limits. Constraint 4f represents the ramp constraints limiting the rate of changing generation levels at conventional power plants. Constraints 4g–4k bound the generation, load shedding, and curtailments respectively.

B DEMAND AND WIND SUPPLY

The average generation and standard deviation of wind generation are shown in Table 1. The statistics of full set and subset are similar, which means our subset can represent the varied generation. Overall grid load and generation variability are shown in Table 2.

Table 1: Wind Generation Statistics: Full Set vs. Simulation Subset (15% Level, Unit: GWh, Avg. Stdev.: average standard deviation of 24-hour generations per sample)

Season	Type	Avg. Generation	Avg. Stdev.
Spring	Full	137.2	2.06
	Subset	135.6	2.05
Summer	Full	95.0	2.28
	Subset	91.8	2.17
Fall	Full	90.1	1.60
	Subset	86.3	1.55
Winter	Full	116.4	1.69
	Subset	115.1	1.68

Table 2: Average Load, Imports, Renewable Supply, and Net Load (Load - Import - Renewable) of the Base System (MW)

	Load	Imports	Renewable	Net Load
SpringWD	26,868	7,478	6,681	12,708
SpringWE	23,980	7,608	6,998	9,373
SummerWD	31,089	7,678	6,672	16,737
SummerWE	28,184	7,400	7,124	13,659
FallWD	28,055	7,675	6,657	13,722
FallWE	25,186	7,108	7,065	11,012
WinterWD	26,352	7,663	6,634	12,054
WinterWE	23,708	6,800	5,581	11,399

C DATACENTER LOCATIONS

The datacenters are introduced at randomly chosen locations in the CAISO topology as described in Section 4. The specific locations (buses) are :

DIABLO1	PINTO	VICTORVL	TERMINAL	CMAIN_GM
GRIZZLY6	LOSBANOS	IMPRLVLY2	GRIZZLY	STA_BLD
HAYNES3G	VICTORVL2	NAUGHT	GRIZZLYA	STA_E1
MISSION	SERRANO	RIVER	FOURCOR2	CHOLLA
CELILO	BURNS2	WESTWING	SAN_JUAN	NAVAJOGRP4
LUGO	MARTIN2	GRIZZLY3	VALLEY	COLGATE
VINCENT	PITSBURG2	DALLES	TRACY	STA_B
MEXICO	STA_F	RINALDI	HAYDEN	MOENKOP1

There are 40 locations in total, and the first n locations are used for simulations with n datacenters.

D DETAILED RESULTS OF COUPLING APPROACHES

The tables in this section correspond to the trends we summarize in Section 6 and contain the detailed numbers if only changes are provided in previous sections. Table 3 shows the dispatch cost in the uncoupled case and relative changes with DP-ONL and GC, corresponding to Figure 10.

Table 3: Dispatch Cost (\$) by Wind Penetration, Number of Datacenters, and Dynamic Range ("Wind"=Wind Penetration, "Range"=Dynamic Range). Relative changes are listed for DP-ONL and GC.

# of DCs	Wind	Range	Uncoupled	DP-ONL	GC
10	15%	None	2.78×10^7		
		[0,6, 0.8]		-0.96%	-1.09%
		[0.4, 1.0]		-2.34%	-2.75%
	30%	None	3.24×10^7		
		[0,6, 0.8]		-0.92%	-1.09%
		[0.4, 1.0]		-2.33%	-2.91%
	45%	None	4.01×10^7		
		[0,6, 0.8]		-0.75%	-0.89%
		[0.4, 1.0]		-1.93%	-2.41%
	60%	None	4.90×10^7		
		[0,6, 0.8]		-0.62%	-0.73%
		[0.4, 1.0]		-1.60%	-1.99%
30	15%	None	2.70×10^7		
		[0,6, 0.8]		-0.47%	-0.70%
		[0.4, 1.0]		+0.10%	-1.08%
	30%	None	3.00×10^7		
		[0,6, 0.8]		-0.99%	-1.48%
		[0.4, 1.0]		-1.09%	-3.03%
	45%	None	3.68×10^7		
		[0,6, 0.8]		-0.90%	-1.32%
		[0.4, 1.0]		-1.28%	-2.90%
	60%	None	4.51×10^7		
		[0,6, 0.8]		-0.76%	-1.09%
		[0.4, 1.0]		-1.18%	-2.45%

Table 4 presents the detailed renewable fractions, corresponding to Figure 13. The improvements in Figure 14 are relative to the "Uncoupled" at each wind penetration level.

Table 4: RPS by Wind Penetration, Number of Datacenters, and Dynamic Range. Absolute changes are listed for DP-ONL and GC.

# of DCs	Wind	Range	Uncoupled	DP-ONL	GC
10		None	35.46%		
	15%	[0,6, 0.8]		+0.04%	+0.06%
		[0.4, 1.0]		+0.10%	+0.15%
		None	44.47%		
	30%	[0,6, 0.8]		+0.08%	+0.14%
		[0.4, 1.0]		+0.21%	+0.38%
		None	48.83%		
	45%	[0,6, 0.8]		+0.07%	+0.13%
		[0.4, 1.0]		+0.20%	+0.38%
		None	51.48%		
	60%	[0,6, 0.8]		+0.07%	+0.12%
		[0.4, 1.0]		+0.18%	+0.35%
40		None	31.19%		
	15%	[0,6, 0.8]		+0.05%	+0.08%
		[0.4, 1.0]		-0.08%	+0.09%
		None	41.72%		
	30%	[0,6, 0.8]		+0.37%	+0.58%
		[0.4, 1.0]		+0.37%	+1.08%
		None	47.13%		
	45%	[0,6, 0.8]		+0.45%	+0.71%
		[0.4, 1.0]		+0.61%	+1.49%
		None	50.51%		
	60%	[0,6, 0.8]		+0.48%	+0.74%
		[0.4, 1.0]		+0.70%	+1.58%

Table 5 contains the detailed statistics of datacenter capacity, which suggests DP-ONL and GC are comparable in terms of standard deviation but DP-ONL outperforms GC in average capacity difference.

Table 5: Statistics of Capacity Variation (Unit: MW) by Wind Penetration and Dynamic Range, 30 Datacenters. (Avg. Local Stdev.: average standard deviation of 24-hour capacities per site)

Wind	Range	Coupling Approach	Avg. Hourly Difference	Avg. Local Stdev.
15%	[0.6, 0.8]	DP-ONL	6.01	16.48
		GC	14.89	17.86
	[0.4, 1.0]	DP-ONL	13.88	50.30
		GC	42.39	50.73
30%	[0.6, 0.8]	DP-ONL	7.39	16.74
		GC	14.81	18.17
	[0.4, 1.0]	DP-ONL	19.62	50.83
		GC	40.48	51.87
45%	[0.6, 0.8]	DP-ONL	7.49	16.41
		GC	14.81	18.22
	[0.4, 1.0]	DP-ONL	20.61	49.79
		GC	40.18	52.29
60%	[0.6, 0.8]	DP-ONL	7.41	16.24
		GC	14.64	18.23
	[0.4, 1.0]	DP-ONL	20.46	49.32
		GC	39.90	52.27