ZCCloud: Exploring Wasted Green Power for High-Performance Computing

Fan Yang and Andrew A. Chien*
Department of Computer Science, University of Chicago
{fanyang, achien}@cs.uchicago.edu
**also Math & Computer Science, Argonne National Laboratory*

***Abstract*—In supercomputer centers, available power, cooling, or carbon footprint often limits supercomputer performance. We propose a new approach to continue scaling that avoids many of these limits, augmenting a traditional system with another that employs only "wasted" renewable power, stranded power. This excess power cannot be economically distributed through grid, and is only intermittently available. We call this approach Zero-carbon Cloud (ZCCloud).**

We explore the potential benefits of unreliable resources with production DOE HPC workloads using a simple periodic model, and identify job types that benefit most (capability jobs and on-time jobs). The benefits scale with duty factor and resource quantity. Next, to create realistic models of "stranded power" we study 28 months of Mid-continent Independent System Operator (MISO) power market history (1,259 generators, 77 million 5-minute intervals). We find that opportunity varies, but the best single wind site can provide 80% duty factor, and 20MW average stranded power. Combining sites further improves duty factor. With resource volatility models from the MISO study, we simulate production DOE HPC workloads and find that stranded power HPC, ZCCloud, can provide significant benefit, decreasing average job-wait time by 50%.

***Keywords*-High-Performance Computing; Cloud; Power Grid; Batch Scheduling; Power limits;**

I. Introduction

With the end of Dennard scaling [1], [2], power has become an increasingly important concern for supercomputing centers. Today's 10-petaflop systems draw more than 10 MW [3], [4], and near-exascale machines announced by the US DOE continue this trend [5], [6]. A study examining hardware challenges for Exascale computing [2] proposed a limit of 20 MW as maximum acceptable for an exascale system. While this limit is likely to be exceeded, there is broad agreement that power is a critical challenge for these systems because of the energy cost, cooling cost, facilities cost, and the associated environmental impact [7].

Supercomputing centers now routinely are limited by electrical power and cooling. For example, power limits are routinely the cause of expensive expansions, relocations, and new buildings [8], [9]. Numerous centers operate under power "caps" due to local utility contracts or concerns about carbon footprint [10]. In other recent cases, centers have had forced power reductions due to regional and national emergencies [11]. In short, power limits are an important reality for supercomputing systems, and as the energy appetites of these systems continue to grow, are likely to be an even more prominent concern.

On the other hand, the large power consumption of computing facilities also creates a significant carbon footprint and leads to negative environmental impacts. Overall, Information and Communications Technologies (ICTs) contribute to over 2% of global carbon emissions, but are amongst the most rapidly growing parts [12]. Expanding computing in conventional fashion to meet future demands would further increase the damaging carbon emissions of ICT.

We are exploring new options for scaling supercomputing facilities. To escape the aforementioned limiting factors, a number of characteristics are desirable. First, scaling supercomputers should not compete with other power grid loads for generation resources – particularly at peak demands. Second, the most attractive approaches to scaling supercomputing will minimize increases in carbon footprint. And finally, it is desirable to have the ability to scale to extreme scale and performance.

We propose Zero-carbon Cloud (ZCCloud), an approach to expanding traditional supercomputers with volatile computing resources that are powered exclusively by wasted power generated from intermittent renewable sources. ZCCloud exploits power that in "grid markets" is not worth transmission and use. Such "stranded power" (SP) occurs in contemporary power grids regularly because although the use of renewable power generation is desirable, it is both unreliable and generated in widely distributed fashion. Such power is easily identified in market-based dispatch systems – it has a negative price – the grid would literally pay the producer to not generate the power.

ZCCloud computing resources are co-located with renewable generators and deployed incrementally in units of containers [13], [14]. This approach eliminates building costs, enables free-cooling to lower cooling costs, and easy scaling. Co-location eliminates transmission. ZCCloud resources are activated only when stranded power is available, and will be shutdown whenever there is insufficient power. As a result, their power consumption does not compete with normal grid loads, and is truly zero-carbon.

Of course, the challenge for ZCCloud is that stranded power is not always available, yet HPC users expect nearly 100% availability. The intermittent availability of stranded power is due to a complex combination of variable genera-

1530-2075/16 $31.00 © 2016 IEEE
DOI 10.1109/IPDPS.2016.96

IEEE computer society

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

tion (wind and sun availability), transmission resources, and complex markets with long-term purchase contracts, day-ahead trading, and real-time trading. The net effect is a vast, untapped stranded power resource. For example, in 2014 the MISO power grid spilled (or down-dispatched) 2.2 terawatt-hours of power from wind resources alone [15]. Stranded power is even larger than that! Our challenge in ZCCloud is to understand, and learn how to translate this unreliable power into a useful computing resource.

To explore the utility of volatile resources, we use production workload traces from the Argonne Leadership Facility (ALCF) Mira system [16] to study of periodically available resources. This study investigates when and how intermittent resource are a benefit. Next, to explore the phenomenon of stranded power, we analyze 28 months of MISO [17] power grid dispatch records, characterizing power price and availability. Finally, we combine these, using the stranded power to determine ZCCloud's pattern of availability. The result is an assessment of ZCCloud's benefit, using only stranded power. Contributions of the paper include:

- A novel approach to extend HPC scalability that employs intermittent renewable energy (stranded power), called Zero-carbon Cloud (ZCCloud).
- Study of the utility of periodically-available resources that shows job wait times reductions of over 80%. Larger benefits accrue to capability (>8192 nodes) and on-time jobs (complete before uptime end)
- Detailed characterization of stranded power, including quantity as well as temporal and spatial distribution. Analysis of 28 months of Midcontinent Independent System Operator (MISO) market logs shows that a single wind farm can have >20 MW stranded power at 80% duty factor. Combining multiple sites increases both quantity and duty factor.
- Evaluation of ZCCloud system benefit. Study using MISO statistics with varied deployments and workloads shows that SP-based resources can reduce job wait times by >50%, and even outperform periodic resources at high duty factors.

The rest of the paper is organized as follows. Section II summarizes power grid and HPC background. In Section III, we describe the ZCCloud approach. Section IV explores periodically-available resources, and analysis of grid "stranded power" is covered in Section V. We combine these in Section VI to explore the benefits of ZCCloud. We discuss our results and related work in Section VII. Section VIII summarizes and points out promising research directions.

II. Background

We briefly summarize a few key elements of background, including the rise of renewable power, how power grids dispatch generations, and basic elements of the batch computing model typically used for high-performance computing.

A. Growing Renewable Power and RPS Goals

Growing concerns about carbon emissions and its long-term impacts on climate change have created a world-wide consensus on the transformation of power generation in general, and in particular a major promotion of renewable power generation. The primary two sources considered viable for significant increase are wind power and solar power. These sources have been the subject of numerous government programs to encourage their deployment and use, including "feed-in" tariffs in Germany, Spain, and other nations in Europe, as well as "production tax credits" in the United States. In the US, together solar and wind comprised 5.2% of overall power generation in 2014 [18]. California has been a leader for setting Renewable Portfolio Standards (RPS), requirements for power generation mix, reaching a 20% renewable mix in 2010, and on track to reach its 33% target for 2020 [19] for wind and solar power. In September 2015, California adopted an RPS goal of 50% renewable by 2030 [20]. Other states across the midwest (included in the MISO power grid) have adopted a range of standards ranging from 25% (2015) in Illinois, 25-31% (2025) in Minnesota, and 55% (2017) in Vermont. Other large states include 50% by 2030 in New York, and 10GW by 2025 in Texas.

Obama's "Clean Power Plan", issued August 2015, calls for a national reduction in carbon emissions by 2030, with renewable power as a critical element of this. And, the US Department of Energy released a landmark report, "Wind Vision 2015", that describes how the US could achieve a 35% RPS for wind alone by 2050. In practical terms, this means regions with relatively more wind resources, such as Texas, are achieving RPS of over 50% by 2050.

Renewables such as wind and solar have time-varying productivity. Solar follows a diurnal cycle, but has significant variation within that. Wind power has higher variability over long periods (weeks or months), but much less over short periods (hours). Both wind and solar, particularly in the US, are deployed in distributed fashion which when combined with variation creates major power grid scheduling and transmission challenges, giving rise to the phenomenon of stranded power [21], [22], which is widely viewed as growing in magnitude with RPS [23].

B. Fundamentals of Energy Markets and Grids

While energy markets vary, modern ISO's in the US dispatch generation and price power purchases based on a fast-moving dynamic market system. California's ISO using a market with 12-minute intervals, and the MISO market we studied is even faster with 5-minute intervals. The MISO real-time market sets the transaction prices for megawatt-hour (MWh) of power, determining the prices paid by utilities to power generators.

The markets generate locational marginal pricing (LMP), that is a price for electricity at each point in the transmission network, including generators, intermediate nodes,

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

and egress to utility distribution. These algorithms attempt to achieve "merit order", purchase first from generators of lowest cost, but cannot do so perfectly. This is because as generation and load vary at each point in the grid, there can be transmission congestion, the LMP can vary widely by location at a given point in time (from +$1,000 to -$1,000 per MWh), as well as at a given location at even adjacent 5-minute intervals [24].

Power grid management is much more difficult than data network scheduling because the grid must match producers and loads instantaneously – power is generally not queued or stored in significant quantities. Furthermore, the power grid must accommodate sudden increases due to weather, sudden changes, or even equipment failures, or outages occur [25]. While much research and technology is being explored to advance the use of power storage and power switching, such technologies are not a significant factor in the management of today's power grids. Thus, if power cannot be productively transmitted to a load, then it is wasted. We term this phenomena "stranded power".

C. Batch Computing and High-Performance Computing

Batch submission is the primary means of scheduling parallel applications at leading supercomputing sites. Batch enables the schedulers to schedule processors, memory, and sometimes network and input/output to both minimize inter-application interference and maximize resource utilization [26], [27]. In batch systems, a user submits a job, specifying the resources needed, including number of compute nodes, CPUs, memory, and network. The batch jobs are queued, and are executed when the schedule identifies resources for the job appropriate to overall resource optimization. Because batch job submission naturally creates an asynchronous interaction interface, it is delay-tolerant and fits our volatile resource model conveniently.

III. Stranded Power and ZCCloud Approach

Traditional supercomputing deployments use dedicated machine-room facilities, creating infrastructures with physical limits on power consumption, heat, size, etc. (see Figure 1). Power limits can arise from transformer limits, regional transmission, and local transmission. Often supercomputer centers create large power demands, leading to new infrastructure requirements. Heat removal for machine rooms are limited by chilled water and air systems as well as heat exchangers. Physical spaces are of definite size. Other concerns include the price of grid power and the externalities of consuming it (environmental impact).

Power, heat, and space serve as a practical and increasingly important limit on the scale of supercomputers. Our approach escapes these limits by exploiting zero-cost stranded green power, using containerized off-site deployment, and free cooling.

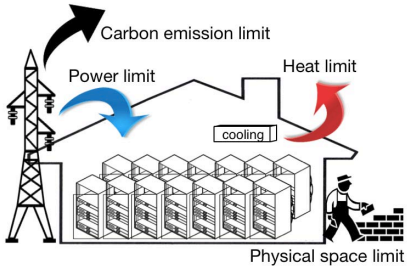

Figure 1. Supercomputer deployment limits (power, heat, size, carbon)

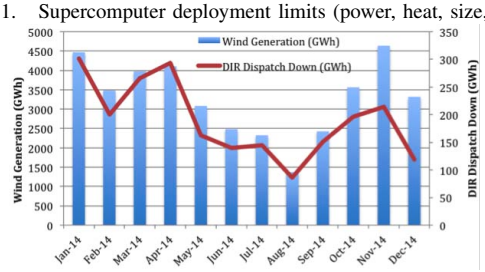

Figure 2. Wind generation and down dispatch of MISO in 2014.

A. Stranded Power

The high variability and intermittency of renewable resources create challenges for power grids that must match generation and load continuously. The situation is further exacerbated by distributed renewable generation (small plants, generally), and limited transmission. Directly, the variation in renewable generation capacity and location make it difficult to reliably meet electric demand [28].

While economic mechanisms such as the MISO real-time market dispatch generation and price power in 5-minute increments [17], there can be oversupply and transmission congestion that prevents purchase in "merit order" (lowest price sources). This is particularly important for renewables that have zero fuel cost, and therefore often have zero incremental generation cost – and thus want to sell power whenever they can generate it.

In situations of congestion or excess generation, these power markets will price some generator's outputs with a low or negative price. In short, the generated power is not economically useful to the power grid. Such negative pricing signals are called "economic curtailment" or dispatch down [29]. The corresponding power that is generated, but cannot be used is called "stranded power". Stranded power is easily identified in market-based dispatch systems, as having a negative price – the grid literally would pay the producer not to generate the electricity. The only power that ZCCloud uses is stranded power.

Figure 2 shows the monthly spilled power wind power (not allowed into the grid) of Mid-continent Independent System Operator (MISO) in 2014. This amounts to 2.2 terawatt-hours overall, and to an average of 183 MW for the entire year. Adding uneconomic generation power (negative price) increases this number to over 7 terawatt-hours Comparable waste exists in other ISO's such as Eastern Region

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

Figure 3. ZCCloud containers below wind turbines. Turbines can be several MW. and wind farms >250 MW.

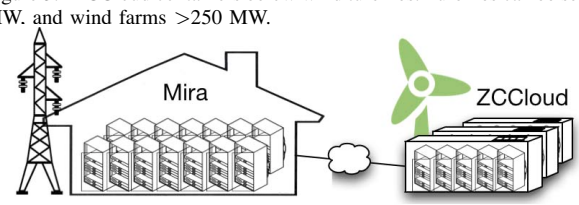

Figure 4. System overview of Mira-ZCCloud.

Coordinating District of Texas (wind) and are projected for California ISO (solar and wind) [30], and many European countries such as Denmark, Germany, Ireland, and Italy [21].

B. ZCCloud Approach

We propose Zero-carbon Cloud (ZCCloud), which uses stranded power to enable scaling to higher levels of capability and performance [31], [32]. ZCCloud computing resources are deployed on generation sites and placed in containers that provide physical protection and cooling, and also achieve efficient computing density [13], [14]. Figure 3 shows a ZCCloud deployment on a wind farm, where computing resources are directly connected to the power backbone of wind turbines. This avoids any need for transmission, and makes possible the exploitation of stranded power. Modern industrial wind turbines being installed are generally over 2 MW, and wind farms >250MW, so large-scale computing infrastructures can be deployed on a single site. ZCCloud computing resources only work when stranded power is available, and otherwise are fully powered off, producing intermittent resources.

ZCCloud can be scaled up by deploying more containers under a set of generators. Generators are often ganged in parallel, so flexible access to power from a large number of turbines is achievable. In contrast to traditional supercomputer deployments, the containers are modular constraints of power consumption, heat, and size. With more containers, all of these attributes scale easily.

Because ZCCloud resources are intermittent, they are not a replacement, but rather a complementary extension of traditional supercomputers, such ALCF's Mira system. Figure 4 depicts our combined Mira-ZCCloud configuration that connects these resources with high-speed network. We can add more ZCCloud resources for further scaling.

IV. Periodic Intermittent Resources

To understand the utility of intermittent resources, we consider the addition of periodic intermittent resources to a traditional supercomputer. These new resources are available once each day at the same time, and for the same duration. We explore the benefits of the periodic resource, varying workload and system properties, and report job wait time and system throughput.

A. Experimental Methodology

System Model: We use ALCF's Mira system [16] as the model for traditional supercomputing resources. Mira is a 10 PFlop IBM Blue Gene/Q system consisting of 49,152 nodes, 786,432 cores, and 768 terabytes of memory. In Mira, programs are allocated nodes (a bundle of computing and memory). We use the Mira only system as a baseline. We compare this to systems that combine a base Mira system and intermittent resources (ZCCloud). We denote this combination "Mira-ZCCloud" (abbreviated M-Z). To match ZCCloud's use of stranded power, our intermittent resources are all "on" or all "off" on a periodic basis.

Job Scheduling: We simulate job service using version 0.99 of Cobalt integrated with Qsim [33]. Cobalt is used for production scheduling and resource management at ALCF [27]. We use a job trace collected from Mira, as described in Table I. In Mira-ZCCloud, a single job scheduler distributes jobs equally across Mira and ZCCloud resources when ZCCloud is powered on. Our simulations give the scheduler start and end times of ZCCloud availability intervals, scheduling only jobs that can complete before ZCCloud shutdown. Long-running jobs whose duration exceeds ZCCloud uptime are only assigned to Mira resources. We also assume that all jobs are delay-tolerant.

Workload: To understand how Mira-ZCCloud performance varies with different workload and intermittent resource properties, we vary workloads and configurations as shown in Table II. We define *duty factor* as the fraction of time the resources are available. For example, compute resources with a duty factor of 50% are available one-half the time in a single interval from 20:00 to 8:00 the next day.

Metrics: We use job wait time and system throughput as the metrics to quantify the Quality of Service (QoS) and performance of our system. *Wait time* is defined a jobs

Table I
ALCF Workload Trace

Parameters	Values
# Jobs	78,795
Time Period	12/31/2013—12/30/2014
Runtime (hrs)	0.004—82, Avg 1.7, StDev 3.0
# Nodes	1—49,152, Avg 1,975, StDev 4,100
Rsc Utilization	84% of Mira (100% availability)
App Domains	Physics, Chemistry, Material Science, Computer Science, Biology, Engineering

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

Table II
EXPERIMENT PARAMETERS

Workload Parameters	
Node hours	[N]xWorkload, N=1+Duty_Factor*Resources
Shape	Uniform, Burst
Resource Configurations	
System Resources	Mira, Mira-ZCC(1xMira), Mira-ZCC(2xMira), Mira-ZCC(4xMira)
Duty Factor	25%, 50%, 100%

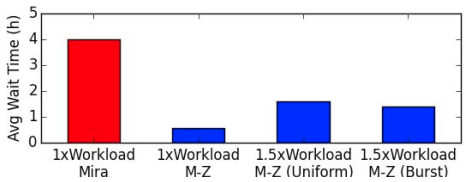

Figure 5. Average wait time vs. Workload Size (Mira-ZCCloud = Mira + 1xMira Intermittent, 50% duty).

starting time minus submission time. *Throughput* is defined as the number of batch jobs completed in a given time.

B. Experiments

To examine the benefit of adding periodic intermittent resources, we begin with a configuration that pairs Mira and ZCCloud with 1x Mira resources and 50% duty factor. We vary the workload size (and therefore system utilization) by adding more jobs with the same attribute distribution to the workload. We denote original ALCF trace *1xWorkload*, and scale its total node hour (i.e. the summation of runtime*nodes) to 1.5x to match Mira-ZCCloud utilization with Mira utilization, denoting it *1.5xWorkload*.

Basic Comparison: Figure 5 compares the performance of Mira and Mira-ZCCloud. With same utilization (1xWorkload on Mira and 1.5xWorkload on M-Z), Mira-ZCCloud reduces average wait time by more than half. There are two explanations for such improvement: (1) according to the M/M/1 queue model [34], wait time is reduced exponentially with increasing computing resource size, and (2) the resource intermittency provides "clearing" effect, which frees an entire half of the Mira-ZCCloud system every 24 hours, giving the scheduler the opportunity to optimize better. To test sensitivity to resource utilization, we also compare the performance of Mira and Mira-ZCCloud for the same quantity of workload (1xWorkload), where Mira-ZCCloud reduces average wait time by more than 80%.

Workload Shape: We also consider different workload shapes, including uniform distribution of node hours (*Uniform*), and 1x node hours during downtime + 2x node hours during uptime (*Burst*). The idea is that uniform shape might favor the traditional supercomputers, and the burst shape matches the periodic Mira-ZCCloud system. Our results (see

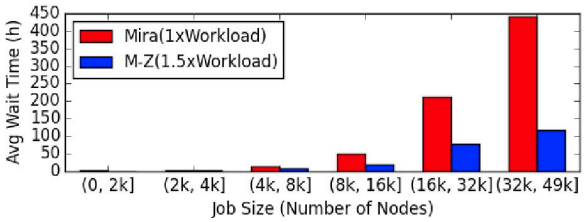

Figure 6. Average Wait Time vs. Job Size (Mira-ZCCloud = Mira + 1xMira Intermittent, 50% duty, same utilization with Mira).

Figure 7. Average wait time vs. on-time metric (Mira-ZCCloud = Mira + 1xMira Intermittent, 50% duty, same utilization with Mira).

Figure 5) show that burst shaped workloads achieve only slightly better wait times than the uniform workload for two reasons: (1) the burst load does not match uptime perfectly, some jobs are "late" - cannot finish in current uptime, (2) in the ALCF workload average wait time is dominated by large jobs, which are hard to schedule. Overall, it appears the shape of the workload need not match resource availability to deliver good performance; resource availability awareness is not required.

Job Size: Figure 6 compares the average wait time for Mira and Mira-ZCCloud across job sizes with same utilization. On average, jobs that require less than 8k nodes usually start within 10 hours after their submission to Mira; jobs that require more than 8k nodes (we call them *capability jobs*) have an average wait time longer than 50 hours. Capability jobs improve significantly on Mira-ZCCloud, with average wait time reduced by approximately 75%.

Classification by Submission: We observed that the benefit of ZCCloud correlates with a simple job classification. *On-time jobs*, those that can complete within the current uptime of intermittent resources, and *Late jobs*, those that arrive too late to do so, and therefore must wait for the next uptime. In Figure 7, we divide jobs into the these two categories. The average wait time for on-time jobs and late jobs decreases by 80% and 55% respectively. The on-time jobs benefit more from adding intermittent resources to Mira than late jobs. The on-time/late classification discriminates, placing longer jobs disproportionately in the late category (average 2.3 times longer). We believe this is one reason they receive smaller benefit, but other factors may be involved.

Varied System Properties: We explore the performance impact of varying system properties, studying ZCCloud with 1x, 2x, and 4x Mira resources, denoted *1xMira, 2xMira, and 4xMira*. We vary duty factor of ZCCloud from 25% to 100%. Systems with various sizes and duty factors also have different available node hours. To make a comparison fair, we

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

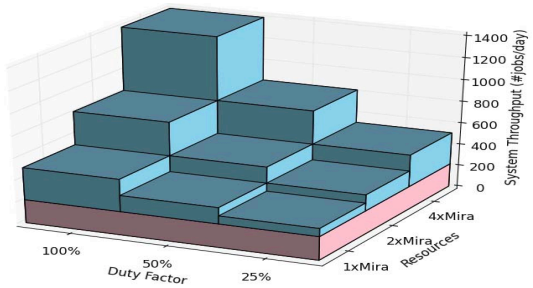

Figure 8. Blue: system throughput vs. duty factor vs. system size for Mira-ZCCloud (all with same utilization). Pink: system throughput of Mira.

adjust workload for system configuration to achieve a similar resource utilization. For example, we run 2xWorkload on a system with 4xMira resources and 25% duty factor, which has the same utilization with 1.5xWorkload on a system with 1xMira resources and 50% duty factor.

Results for a range of duty factors and system sizes (see Figure 8) show that system throughput increases with both increases resources and duty factor in the expected proportion. That is doubling duty factor gives approximately the same benefit as doubling ZCCloud resources. Specifically, {1xMira, 50% duty factor} and {2xMira, 25% duty factor} achieve similar throughput. Both increasing duty factor and adding intermittent resources to Mira-ZCCloud improve system throughput.

V. Stranded Power in MISO

We analyze power grid behavior to understand properties of stranded power, including quantity, as well as temporal and spatial distribution. Quantity enables ZCCloud resource scale, and durations determine the achievable duty factor.

A. Power Grid Data

We use a large collection of records from the Midcontinent Independent System Operator (MISO) [17] as described in Table III. The MISO market manages power for a large geographic area, covers most of ten states (Illinois, Indiana, Iowa, etc.), as well as parts of Texas, Montana, Missouri, and Manitoba. Specifically, we use the real-time market cleared offers (RTOs) that provide precise information on the local marginal price (LMP), offered power, cleared power, offered price, and a wealth of other data in 5-minute increments from 1/1/2013 to 4/14/2015 (see Table IV). The largest source of renewable power in MISO is wind generation (approximately 10% of the overall power market), and is our focus here.

Table III
MISO Market Data Set

Parameter	Value
Period	1/1/2013—4/14/2015
Generation Sites	1,259 Total, 200 Wind
5-minute Intervals	76,937,135 Total, 36,617,860 Wind
Total GWh	1,188,528 Total, 88,571 Wind
Total $'s Power	$39.7 B Total, $1.7 B Wind

Table IV
Real-Time Cleared Offers

Dimensions	Description
LMP	Local Marginal Price (5-minute intervals)
Delivered MW	Delivered power (5-minute intervals)
Economic Max	Offered power (by generator)
Time	Time of offer (hour)

B. Defining Stranded Power (SP)

Informally, stranded power (SP) is power that has little or no economic value to the grid. With this backdrop, there are many possible definitions of stranded power. Here we give two families of definitions, and use them to analyze the MISO market records to derive stranded power quantity as well as its temporal and spatial distribution.

1) *Instantaneous Stranded Power:* $LMP < C(C = \$0, \$1, , \$5)$ for a 5-minute interval. $5 is 5x lower than the average MISO power price. This is the power that has little economic value at some point.
2) *Net Price Stranded Power* $NetPrice < C(C = \$0, \$1, , \$5)$, which means the average price of power in a period is very low.

$$NetPrice = \frac{\sum_{period} LMP \cdot Power}{\sum_{period} Power}, Power \text{ in MWh} \tag{1}$$

With varied models, each of these models give rise to a class of SP definitions (see Table V). The period during which SP is available is a *stranded power (SP) interval*.

We focus on two key properties of Stranded Power: (1) temporal and spatial distribution (duty factor, interval sizes), i.e. the availability across time and site; and (2) quantity (megawatts). These properties determine the capability of a site to support ZCCloud compute resources. The cheapest, most efficient ZCCloud deployment is at a single site, so we focus on single sites first, and then explore multiple sites for higher availability.

Metrics include:

- *Duty Factor (site)*: Fraction of time SP is available.
- *Cumulative Duty Factor (N)*: Fraction of time SP is available across N sites.
- *Average SP (site)*: Average Stranded Power in MW.

Table V
Stranded Power (SP) Models

Model	SP Defn	Description
LMP	LMP[x]	SP available when LMP<$x
NetPrice	NetPrice[x]	SP available when NetPrice<$x

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

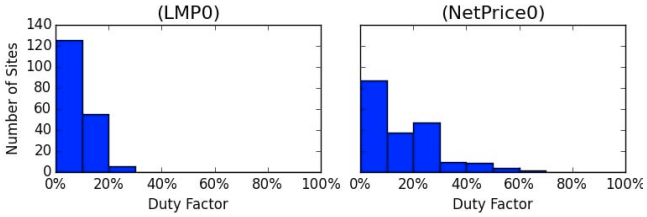

Figure 9. Distribution of generation sites vs. duty factors for LMP0 and NetPrice0.

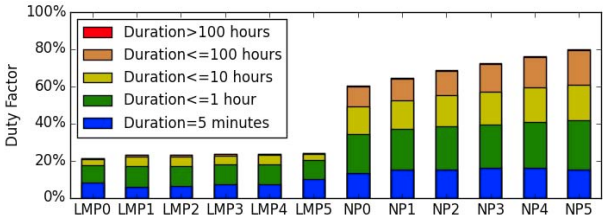

Figure 10. Best single-site duty factors vs. SP models. Each sub-bar represents the fraction of intervals of different duration. (NP=NetPrice)

C. Analysis

From the cleared power records for each generation site, we compute the duty factor of single sites (Figure 9). The difference between the two SP models is significant. For LMP model (instantaneous stranded power), most sites have a duty factor less than 20%, and none have duty factor larger than 21%. For NetPrice, dozens of sites have duty factor $> 30\%$. And several have duty factors greater than 60%. In short, long periods of negative or extremely low-priced power are interrupted by LMP's greater than $5. But high duty factors can be achieved by using average price.

To explore the difference further, we look at SP interval duration distributions (see Figure 10). For the LMP model, most SP intervals are less than 1 hour. The NetPrice model has much longer SP intervals. Nearly half of the NetPrice SP intervals have durations longer than 1 hour, and the duty factor improves significantly to 60%. This is because NetPrice masks fluctuations in the price, and large negative prices enable NetPrice to extend through some positives, increasing both interval length and duty factor.

Next, we consider multiple sites (see Figure 11). For LMP0, cumulative duty factor is 20% for a single site. Increasing threshold of LMP from $0 to $5 does little to increase this.[1] Adding more generation sites increases cumulative duty factor slowly, reaching 30% for 2 sites, and 50% for 7 sites. NetPrice yields much higher duty factors from 60-80% for a single site, and more than 80% for three sites. Increasing threshold of NetPrice improves duty factor significantly, especially for a small number of sites. However, even large numbers of sites with a NetPrice5 model do not achieve 100% duty factor. There are significant periods with no stranded power in the grid.

[1]We treated 4 generation sites with the same pricing behavior – attached to the same power grid node as a single site for these graphs.

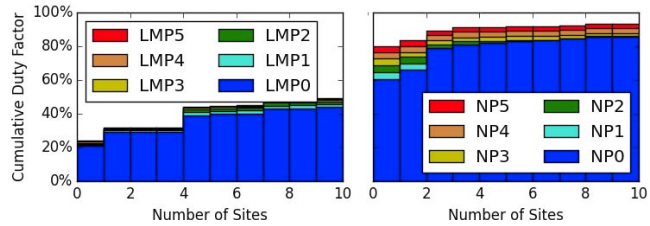

Figure 11. Cumulative Duty Factor vs. Number of Generation sites with highest duty factor. (NP=NetPrice)

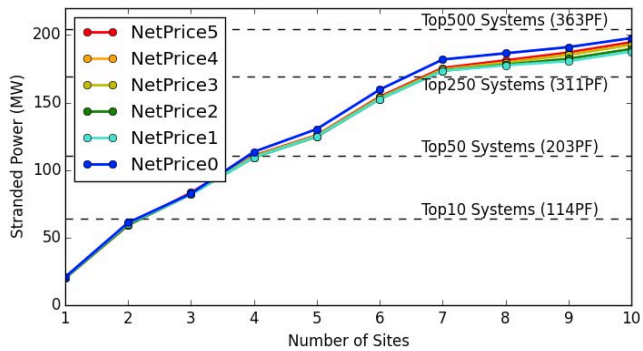

Figure 12. Average Stranded Power vs. Number of Generation sites with highest duty factor, compared Top500 supercomputing systems.

Figure 12 compares the average stranded power of wind generation sites to the total power consumption of Top500 supercomputing systems [35]. For NetPrice model, a single site has enough stranded power to support the Top system with 20MW. As generation sites are added, two sites are sufficient for the Top 10 systems, 4 sites sufficient for the Top 50 systems, and 7 sites are sufficient for the Top 250 supercomputing systems. Only two generation sites can power all Top10 systems and provide compute power of more than 100 PFlops. Note that these generation sites are all within one subregion of the United States, while the Top 500 are drawn from around the world.

VI. Computing on Stranded-Power

In Section IV, we explored the potential performance benefits of periodic intermittent resources, showing that capability jobs and on-time jobs gain significantly. In Section V, our analysis of the MISO grid operation revealed stranded power single-site duty factors as high as 80% with average power of 20MW. This is promising, however, stranded power interval has widely varying durations, compared to periodic intermittent resources. Here we explore that impact, directly evaluating the utility of ZCCloud resources on stranded power, simulating real workloads on computing resources whose availability is determined by stranded power.

A. Experimental Methodology

We use the simulation environment and resource models described in Section IV. The only modification is that

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

Table VI
<WIND SITE, MODEL> WITH HIGHEST DUTY FACTOR

SP Model	Region	Unit	Duty Factor	Avg MW
LMP0	West	7663	21.1%	8.1
LMP5	West	7663	23.9%	8.9
NetPrice0	West	3007	60.4%	21.3
NetPrice5	West	3007	80.1%	20.7

Table VII
EXPERIMENTAL PARAMETERS

Parameters	Options
SP Model	LMP0, LMP5, NetPrice0, NetPrice5
Workloads	1x, 1.25x, 1.5x, 1.75x
Resources	1xMira, 2xMira, 3xMira, 4xMira

resource availability is determined by MISO stranded power based on four SP models (LMP0, LMP5, NetPrice0, NetPrice5) (see Section V). For each SP model, we pick a generation site with the highest duty factor (Table VI) for the ZCCloud deployment. We also adjust the workload to explore increased system utilization and system size. Our parameters are documented in Table VII.

B. Experiments

We first compare SP-based resources with periodic intermittent resources with the two stranded power models, LMP and NetPrice (Figure 13). For LMP models, the availability of additional ZCCloud reduces the average wait time by over 50%. Compared to periodic intermittent resources at the same duty factor (21% and 24%), the LMP stranded power resources perform comparably, but slightly worse than periodic resources. This is already encouraging because it means that despite their volatility stranded-power based resources can be useful. LMP generates many short intervals that are less useful than the single availability interval provided by the periodic resource in each 24 hour period.

Next we explore NetPrice models that achieve much higher duty factors. Compared to periodic intermittent resources at the same duty factor (60% and 80%), the NetPrice resources perform comparably, and surprisingly even better for the 80% duty factor case. After some analysis, we discovered that the NetPrice stranded power system's

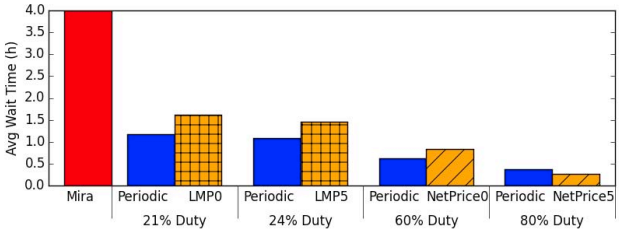

Figure 13. Periodic resource vs. SP-based resources with same duty factor (1xMira, 1xWorkload).

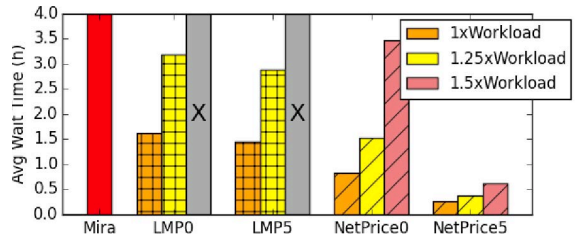

Figure 14. Average wait time vs. workload vs. SP models (1xMira).

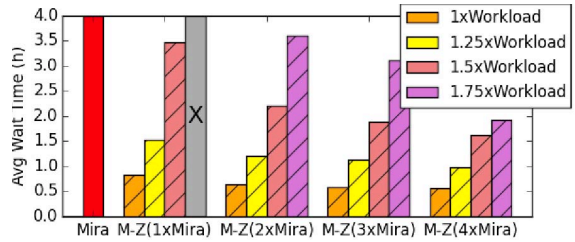

Figure 15. Average wait time vs. Workload vs. System Size (NetPrice0, Duty factor = 60%)

variable intervals sometimes have durations much longer than 24 hours (the limit for the periodic resources). These longer intervals increase system efficiency, despite their rarity, whereas the periodic intermittent resources experience an interruption at least every 24 hours.

Focusing on the stranded power systems, we study performance under heavier workloads (see Figure 14). Workloads that fail to complete due to insufficient node hour capacity are denoted as "X" (e.g. 1.5xWorkload on Mira-ZCCloud with 1xMira ZCCloud resources, LMP models). For both LMP models, the Mira-ZCCloud system outperforms the base Mira at the same utilization. For the NetPrice models, the performance is better across the board, but the high duty factors provided by NetPrice stranded power mean that even for highest loads (1.5xWorkload), the system utilization is only 70%, below the Mira system's 84%. The average wait time improvements vary from 20% to as large as 90%.

Next we explore the addition of resources. Exploring system performance at growing scale, we study the average wait time of Mira-ZCCloud (Figure 15). As with periodic resources, increasing resources improves performance, particularly for high utilization workloads. Each successive set of bars lowers all of the average wait times. At duty factor of 60%, the M-Z(2xMira) has enough resources to absorb the 1.75xWorkload, and beyond that we see modest performance improvements. As before, in all cases, even at the same utilization, ZCCloud outperforms the base system.

VII. DISCUSSION AND RELATED WORK

Our results show that ZCCloud can create useful resources based on stranded power. The scale of such power is substantial, and enables compute as large as the entire Top 250 supercomputing systems. The temporal and geographic structure of stranded power in the Midwest enables long intervals and consequently good application performance -

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

up to high resource utilization and scale. In short, ZCCloud is a promising path to enable continued scaling of high-end supercomputing systems.

With power emerging as a critical obstacle for continue performance scaling [2], [36], power-efficiency has become a major objective of numerous research programs, including DARPA's UHPC, OHPC, and PERFECT [37] programs which established goals of 75 Gigaflops/watt (system-level), the International Exascale Software Roadmap [38], and a core element of the recently launched US Extreme Computing Initiative (ECI). These efforts to increase energy efficiency have produced a wealth of research on how to eliminate power waste. Mainstream techniques include improving power distribution, improving cooling systems, and using free cooling. At the computing level, a host of techniques to dynamically match power to the computing need – not a fixed target. For example, under-provisioning power for average demand rather than peak [39], [40] or scheduling techniques that select power resources based on the load [39], [41], and job scheduling strategies, such as migrating jobs and shifting peak load [42]. Generally these approaches improve power efficiency by lowering CPU frequency and deferring jobs while attempting to minimize loss of system performance and quality of service. ZCCloud differs from these techniques in its focus on tapping new, carbon-neutral sources of power, and using them to enable new levels of performance.

A number of efforts exploit renewable resources to power supercomputers and data centers in order to reduce their carbon impact. For example, Google claims carbon-neutrality since 2007. Microsoft, Facebook, Apple all buy renewable power through long-term purchase contracts. A summary of these efforts are collected in an annual study [43] by Greenpeace. Numerous research efforts have studied the addition of renewables to data centers [42], [44], or more ambitiously "follow the sun" [45], [46]. All these designs depend on reliable grid power, and use renewables only to reduce grid power consumption. Further, they all contemplate buying renewable generation (either building plant or long-term purchase) to provide green power. In contrast, ZCCloud exploits un-economic grid power, stranded power. And its model creates dispatchable loads that consume excess generation can benefit grid stability, cost, and management [31], [32].

ZCCloud provides intermittent compute resources that, while difficult to use, are not a new idea. Peer-to-peer [47]–[49], desktop grid [50], [51], and Condor [52] are all well-known volatile resource management systems. Cloud providers have begun to provide unreliable/revocable resources including Amazon's spot instances [53] and Google's recent preemptible VM's [54]. However all of these models have unreliable individual instances/VM's. With ZCCloud, we can provide *many* resources with correlated availability, perhaps 100,000's of nodes.

Different from previous efforts on scalable, energy-efficient, and green computing, ZCCloud not only scales easily without new physical or power grid infrastructure, but it also improves system throughput and performance. Finally, ZCCloud enables the use of stranded grid power to create carbon-free computing.

VIII. Summary and Future Work

Scaling compute power beyond limits in power, cooling, and other physical infrastructure is a challenging and costly task. Reducing carbon footprint is yet another challenge. ZCCloud is a promising approach to continued scaling of the largest supercomputers, using stranded renewable energy. This approach has a number of benefits, including incremental scalability of infrastructure, low power cost, and negligible carbon footprint.

Our assessment shows not only that sufficient stranded power is available, but that it can be usefully applied to production HPC workloads, providing significant wait time and throughput benefits. While many questions remain to be answered, for the central question of "can these intermittent resources be useful?", the answer is clearly yes!

Promising directions for future work include: studying storage/network performance in decentralized datacenters, exploring multi-site ZCCloud systems, and additional ISO's with different renewable mixes, market structures, and demand patterns. Here we have focused exclusively on utility; an important direction is to assess the costs and economics of stranded-power based computing. Another interesting direction is the use of prediction, and also the other computing services beyond batch parallel computing. We are also interested in exploring the reliability implication of power in the hardware on and off.

Acknowledgements

This work was supported by in part the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Award DE-SC0008603 and Contract DE-AC02-06CH11357, and the National Science Foundation under Award CNS-1405959. We thank Paul Rich and William Allcock from Argonne Leadership Computing Facility who provided insight and expertise that greatly assisted the research.

References

[1] S. Borkar and A. A. Chien, "The future of microprocessors," *Commun. ACM*, vol. 54, May 2011.

[2] J. Shalf *et al.*, "Exascale computing technology challenges," in *VECPAR 2010*, 2011, pp. 1–25.

[3] "TIANHE-2, the #1 Top 500 Supercomputer," http://www.top500.org/system/177999, June 2015, 17 MW power.

[4] "Blue Waters: Awesome Power, Awesome Efficiency," http://www.datacenterknowledge.com/archives/2010/06/24/blue-waters-awesome-power-awesome-efficiency/, June 2010, 15 MW power.

[5] "Aurora-OLCF," https://www.nersc.gov/assets/Uploads/OLCF-HEP-ALCF-workshop.pdf, 2015, 13 MW power.

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.

[6] "Summit-OLCF," https://www.nersc.gov/assets/Uploads/OLCF-HEP-ALCF-workshop.pdf, 2015, 10 MW power.

[7] A. Gore *et al.*, *An Inconvenient Truth*. Paramount, 2007.

[8] "Blue waters," https://en.wikipedia.org/wiki/Blue_Waters, June 2010.

[9] "Computational and Theory Research Facility," https://www.nersc.gov/assets/Uploads/CRT-for-NUG.pdf, 2 2015.

[10] e. a. N. Bates, "The electrical grid and supercomputing center: An investigative analysis of emerging opportunities and challenges," in *3rd D-A-CH Conference on Energy Informatics*. Zurich, Switzerland: Springer, 2014.

[11] O. Kimura *et al.*, "Saving Electricity in a Hurry: A Japanese Experience after the Great East Japan Earthquake in 2011," *ACEEE Summer Study on Energy Efficiency in Industry*, 2013.

[12] H. Dunn, "The carbon footprint of ICTs," *Global Information Society Watch*, 2010.

[13] "HPE Moduler Datacenter," https://www.hpe.com/us/en/integrated-systems/pods.html.

[14] "AST Moduler Datacenter," http://www.astmodular.com/solutions/family/modular-data-centers_1.

[15] Bird, et. al., "Wind and Solar Energy Curtailment: Experience and Practices in the United States," National Renewable Energy Laboratory, Tech. Rep., 2013.

[16] "Mira-ALCF," https://www.alcf.anl.gov/mira.

[17] "MISO," https://www.misoenergy.org/.

[18] "Renewable Energy13.4% of US Electricity Generation in 2014," http://cleantechnica.com/2015/03/10/renewable-energy-13-4-of-us-electricity-generation-in-2014-exclusive/.

[19] "California Public Utilities Commission (CPUC)," http://www.cpuc.ca.gov/PUC/energy/Renewables.

[20] C. Megerian and J. Panzar, "Gov. Brown signs climate change bill to spur renewable energy, efficiency standards," Los Angeles Times Newspaper, September 2015.

[21] Lew, et. al., "Wind and Solar Curtailment," in *the International Workshop on Large-Scale Integration of Wind Power Into Power Systems*, October 2013.

[22] Bird, et. al., "Integrating Variable Renewable Energy: Challenges and Solutions," National Renewable Energy Laboratory (NREL), Tech. Rep., 2013.

[23] "Wind Vision," http://energy.gov/eere/wind/wind-vision.

[24] "California ISO (CAISO), Locational Marginal Pricing (LMP): Basics of Nodal Price Calculation," CAISO Market Operations, Tech. Rep., 2005.

[25] "Power outages affect thousands in San Diego," http://www.10news.com/news/power-outages-affecting-thousands-in-san-diego-092015.

[26] A. B. Yoo *et al.*, "Slurm: Simple linux utility for resource management," in *Job Scheduling Strategies for Parallel Processing*. Springer, 2003, pp. 44–60.

[27] "Cobalt," https://www.alcf.anl.gov/cobalt-scheduler.

[28] R. Fares, "Renewable Energy Intermittency Explained: Challenges, Solutions, and Opportunities," Tech. Rep., 2015.

[29] P. Brown, "US renewable electricity: how does wind generation impact competitive power markets," in *Congressional Research Service. Report for Congress*, 2012, pp. 7–5700.

[30] E3, "Investigating a higher renewables portfolio standard in california: Executive summary," Report from Energy and Economics, Inc., Tech. Rep., January 2014, http://ethree.com/.

[31] A. A. Chien *et al.*, "Zero-Carbon Cloud: High-value, Dispatchable Demand for Renewable Power Generators," *Electricity Journal*, October 2015.

[32] F. Yang and A. A. Chien, "Scaling Supercomputing with Stranded Power: Costs and Capabilities," University of Chicago, Tech. Rep., 2016, http://newtraell.cs.uchicago.edu/files/tr_authentic/TR-2016-04.pdf.

[33] "Qsim," https://trac.mcs.anl.gov/projects/cobalt/wiki/qsim.

[34] J. Abate and W. Whitt, "Transient behavior of the m/m/l queue: Starting at the origin," *Queueing Systems*, 1987.

[35] "Top500 supercomputers," 2015, http://www.top500.org/.

[36] C. National Research Council, *The Future of Computing Performance: Game Over or Next Level?* National Academies Press, 2011, available from www.nap.edu.

[37] "DARPA wants a more PERFECT processing engine with a 75x improvement in GFLOPS/Watt." 2012, https://eda360insider.wordpress.com/2012/02/17/darpa-wants-a-more-perfect-processing-engine-with-a-75x-improvement-in-gflopswatt-want-to-play/.

[38] J. Dongarra *et al.*, "The international exascale software project roadmap," *IJHPCA*, 2011.

[39] L. Liu *et al.*, "HEB: deploying and managing hybrid energy buffers for improving datacenter efficiency and economy," in *ISCA'15*. ACM, 2015, pp. 463–475.

[40] X. Zhou *et al.*, "Underprovisioning the Grid Power Infrastructure for Green Datacenters," in *ICS 2015*. ACM, 2015.

[41] C. Li *et al.*, "Chameleon: Adapting throughput server to time-varying green power budget using online learning," in *ISLPED'13*. IEEE Press, 2013, pp. 100–105.

[42] M. E. Haque, Í. Goiri, R. Bianchini, and T. D. Nguyen, "GreenPar: Scheduling Parallel High Performance Applications in Green Datacenters," 2015.

[43] Greenpeace, "Clicking Clean: How Companies are Creating a Green Internet," April 2014.

[44] R. Bianchini *et al.*, "Parasol: A Solar-Powered Micro Datacenter," in *EuroSys 2012*, 2012.

[45] "Follow-the-sun data centers," Tech. Rep., 2015, http://whatis.techtarget.com/definition/follow-the-sun-data-center.

[46] J. L. Berral *et al.*, "Building green cloud services at low cost," in *ICDCS'14*. IEEE, 2014, pp. 449–460.

[47] A. Oram, Ed., *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly, 2001.

[48] L. Napster, "Napster," *http://www.napster.com*, 2001.

[49] I. Stoica *et al.*, "Chord: A scalable peer-to-peer lookup service for internet applications," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4, pp. 149–160, 2001.

[50] A. Chien *et al.*, "Entropia: architecture and performance of an enterprise desktop grid system," *JPDC*, vol. 63, no. 5, pp. 597–610, 2003.

[51] D. P. Anderson, "Boinc: A system for public-resource computing and storage," in *Grid Computing 2004*. IEEE, 2004.

[52] "HTCondor," http://research.cs.wisc.edu/htcondor/.

[53] "Spot Instance," http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html.

[54] "Google Compute Engine: Preemptible VM," https://cloud.google.com/compute/docs/instances/preemptible.

Authorized licensed use limited to: Riva Laughlin. Downloaded on February 21,2024 at 20:20:28 UTC from IEEE Xplore. Restrictions apply.