Navigating Dennard, Carbon and Moore: Scenarios for the Future of NSF Advanced Computational Infrastructure

Andrew A. Chien
University of Chicago and Argonne National Laboratory
aachien@uchicago.edu

Mark J. Dietrich
Bloodstone Solutions Inc.
mjdietrich@bloodstonesolutions.org

ABSTRACT

After a long period of steady improvement, scientific computing equipment (SCE, or HPC) is being disrupted by the end of Dennard scaling, the slowing of Moore's Law, and new challenges to reduce carbon, to fight climate change. What does this mean for the future? We develop a system and portfolio model based on historical NSF XSEDE site systems and apply it to examine potential technology scenarios and what they mean for future compute capacity, power consumption, carbon emissions, datacenter siting, and more.

CCS CONCEPTS

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

KEYWORDS

High-performance computing, power consumption, energy efficiency, carbon emissions, total-cost of ownership, Moore's Law, Dennard Scaling, data centers

ACM Reference Format:

Andrew A. Chien and Mark J. Dietrich. 2022. Navigating Dennard, Carbon and Moore: Scenarios for the Future of NSF Advanced Computational Infrastructure. In *Practice and Experience in Advanced Research Computing (PEARC '22), July 10–14, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491418.3535156

1 INTRODUCTION

The importance of computing for science, education, and knowledge discovery has never been greater and continues to grow. This is reflected by deep penetration of science and engineering disciplines, and the expansion from modeling and simulation to data-driven discovery. Reflecting this importance, investment by governments and universities in computing infrastructure has continued to increase [1], and marquee supercomputer systems with price tags exceeding $500 million [2] to reach an Exaflop. Some estimate that the high-end computing available for research in US universities has grown by 30-fold in the past twenty years. But in the face of a rapidly changing technology landscape, can this rapid growth in computing capability be sustained and how should we shape this investment?

Recent years have seen major technology disruptions. End of Dennard heralded a major shift for software to parallelism, first to multicore and recently to GPUs. Slowing of Moore's Law, with a looming end (or shift to "More than Moore, heterogeneous integration") threatens fast exponential improvement in computing cost and density. Both trends threaten to slow improvement in compute per watt (power density). And, computing has grown to such a vast application that its power consumption, and associated carbon impact is measurable and significant in the global fight to slow climate change. In this disrupted landscape, how are we to navigate these changes in Dennard, carbon, and Moore to deliver the continued rapid growth in scientific computing that is essential to fuel education and discovery? More specifically: can we sustain our continued rapid growth in computing capacity? How should upgrade cycles change? What are the critical costs? Power? Space? Equipment?, and can we avoid dramatic increases in carbon emissions?

In this paper we construct a model based on ten large-scale scientific computing systems (SCE) deployed at major NSF XSEDE sites, using them to create a compute, space, power, and cost model for scientific computing systems. From these systems and sites, we derive both trends and a model for capital, TCO, power, and more. Using this model, we consider a baseline scenario, projecting its implications through two decades of technology and systems evolution. To explore the sensitivity to parameters, we consider a set of twelve scenarios around the baseline, exploring slower and faster power density improvement, compute improvement and more. The key results of our study include:

- Continued performance improvement drives replacement cycles similar to historical patterns (4-7 years). This was counter to our expectations and reflects the continuing attraction of future systems based on improving technology. The future system economics and capabilities are too compelling to slow down.
- Growing power consumption of SCE systems is a challenge if we are to sustain growth in computing capability. This power consumption presents challenges in growing operational cost, carbon impact, and facilities requirements. These challenges may require changes in academic computing and NSF program strategies.
- Fueled by continuing compute density improvements and advanced cooling systems (e.g. water, immersion), the datacenter space requirements for SCE do not increase dramatically. This was also counter to our initial expectations.
- The growing power consumption of SCE systems drives an increasing Scope 2 (direct CO2 emissions) impact. This motivates new approaches that increase the ability to use low-carbon (and low-cost) power, as electricity cost becomes

PEARC '22, July 10–14, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9161-0/22/07…$15.00
https://doi.org/10.1145/3491418.3535156

an increasing fraction of TCO. Fortunately, total carbon emissions from power, and quantity of power consumed, can be decoupled by the use of renewable power, and adopting green siting for systems. Such a strategy is an attractive approach for the future.

The full results of this study are presented in [3].

2 METHODOLOGY

We gathered data on ten SCE systems installed at major NSF XSEDE sites (Indiana University, University of California - San Diego, University of Texas - Austin, Carnegie-Mellon University) between 2015 and 2020. Eight were funded by the NSF; these were the largest NSF systems and account for nearly one-third of its SCE funding. Computational capacity of these systems ranges from 92 teraflops to 41 petaflops with power ranging from 29 kilowatts to 3.4 megawatts, and system cost from $2 to $50 million. System sizes vary by several orders of magnitude, so our analysis and conclusions should be scale independent. Data was also collected on the four institutions' hosting facilities, including data center power utilization efficiency (PUE), average power costs ($/kWh), nominal space costs ($/rack-year) and carbon-emissions (metric tonnes/kWyear). See Appendix [4] for detailed statistics for these systems, Tables A.1 and A.2.

Building on prior research that explores hosting SCE at new "green" data centers with low-cost renewable energy [5], here we consider the opportunity to reduce power cost and carbon emissions. In the Appendix [4] Table A.3 summarizes three commercial offers described.

2.1 Models and Scenarios

We model SCE systems as a combination of performance, capital costs, and power and space requirements, allowing us to forecast their joint evolution over time. When all three parameters are normalized against compute performance, the resulting ratios exhibit log linear relationships. We use minimum least squares to fit an exponential curve[1] to the calculated values of each ratio (capital cost/Rpeak, power/Rpeak, space/Rpeak). The calculated trends represent the baseline scenario for our analysis and are used to project future capital and operating costs. This data can also be used to determine the optimum economic replacement cycle for SCE, for a given scenario.

We consider several alternative future scenarios by adjusting the rate of change of the three ratios (for each of capital cost/Rpeak, power/Rpeak, space/Rpeak) to reflect different technology and market developments. These alternative scenarios are intended only to explore the economic implications of those alternative scenarios. Moore's Law addressed the number of transistors that could be produced per unit area of integrated circuit, and by extension the relationship between both system cost and system size and total compute capacity. We model both relationships together as "compute density improvement" ("CD") with two scenarios, CD0 (baseline) and CD- (slower improvement), ranging from 28.5%/yr to 15.4%/yr. Similarly, Dennard scaling addressed the power required by each transistor, and by extension the power required to provide a given amount of compute capacity. We model this relationship as "power density improvement" ("PD") with four scenarios, PD0 (baseline), PD- (slower improvement) and PD+ and PD++ (faster improvement), ranging between 8.8%/yr (PD-) and 30.7%/yr (PD+).

2.2 Determining the Economically Optimum Replacement Cycle

We determine the optimal replacement cycle by modeling a series of SCE systems purchased sequentially. The replacements must have a fixed cadence (i.e. annually, with range from 3 to 9 years), and the objective to be optimized is total cost per unit of compute capacity (dollars per Gigaflops/s per year) for the series of computing systems. The average costs and capacity are "totalled" by calculating the net present value (NPV)[2] of annual capital plus operating expenditures in current dollars, and of annual capacity measured in GFlops/s-Years. The lowest average cost per unit capacity identifies the economically optimum SCE replacement cycle.

This calculation specifically depends on conditions incorporated in the analysis, including expected changes in the parameters of the SCE model, economic discount rate (10% is used in this presentation), and cost factors for the selected hosting facility, as well as the period of analysis itself. The calculation is varied to reflect different technical assumptions as well as constraints on budgets for capital and other ownership costs. Since capital expenditures are independent of the hosting facility, we separately calculate the impact of hosting facility choice on replacement cycle. Other component TCO costs are handled in a similar fashion.

3 RESULTS

Analysis of operating parameters and cost factors for the ten sample SCE systems yields baseline improvement trends that drive the model. Capital cost ($/Rpeak) decreased by 28.5% per year (correlation coefficient, R2=0.39). Measured price/performance improvement is consistent with a doubling of transistor density every two years. The CD0 baseline is set at 28.5%, and CD- projects slower improvement, doubling every four years (15.4%/yr).

Power density (kW/Rpeak) decreased by 16.8% per year (R2=0.47). The PD0 baseline is set at 16.8%/yr, while PD- doubles the period of improvement (8.8%/yr). PD+ halves the improvement period (30.7%/yr), and PD++ applies a one-time improvement of 47%, followed by annual improvement similar to PD+ (28.3%).

Space required (racks/Rpeak, aka 1/performance-density) decreased by 26.8% per year (R2=0.56), consistent with the rate of improvement in capital cost/Rpeak and a trend like Moore's Law. This is captured by CD0 and CD-.

Cost is a critical constraint in SCE, and we model two investment scenarios. "I0" is the baseline, fixing annual investment at $1,000,000, regardless of delivered capacity. "I1" increases capital investment as needed to yield annual compute capacity increases of 40%/yr (consistent with historical experience).

[1]Curve of the form y=B*mx, where x is the system in-service date, the calendar year plus a fraction representing the month of the in-service date, at which ratio y is measured.

[2]Net present value is a financial calculation that allows a series of future costs or payments to be aggregated in a way that reflects the "time value of money", where a payment today costs more than a payment in the future. Net present value (NPV) is the sum of multiple payments, each of which has been discounted to the present using a specified discount rate. The NPV of one series of payments can be compared to the NPV of a different series of payments.

Table 1: Projection of Baseline Scenario (PD0, CD0, I0)

Year	2022	2027	2032	2037	2042	20-year growth
Capital Investment	$1,000,000	$1,000,000	$1,000,000	$1,000,000	$1,000,000	1.00x
Petaflops/s (PF)	0.9	4.5	24.2	129.1	688.3	809x
System Power (kW)	65	139	297	634	1,352	20.7x
Cost of power ($/kWyear, 1st year of operation, average campus site)	$764	$921	$1,110	$1,337	$1,612	2.11x
Power costs (annual, average campus site)	$49,907	$128,272	$329,689	$847,378	$2,177,955	43.6x
Scope 2 CO2 Emissions (MT, annual, average campus site)	250	533	1,137	2,426	5,175	20.7x
System Size (Racks)	1.8	2.0	2.3	2.5	2.8	1.56x
Space costs (annual, average campus site)	$10,026	$11,208	$12,529	$14,005	$15,656	1.56x
TCO (annual)	$1,059,933	$1,139,480	$1,342,218	$1,861,383	$3,193,611	3.01x
TCO/PF	$1,245,840	$251,129	$55,465	$14,422	$4,640	0.00372x

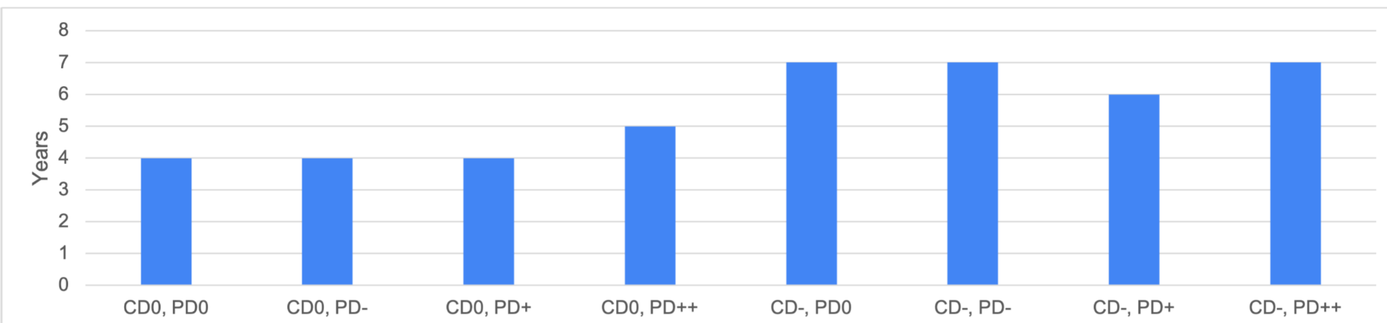

Figure 1: Optimum Replacement Cycles as a function of "Compute Density" and "Power Density" Scenarios. We have explored many additional scenarios, but the key element is the continued drive by Moore's law – independent of power density improvement rates (PD scenarios). As long as the number of transistors per dollar continues to increase rapidly (Moore's law), then it will make most economic sense to continue to upgrade our systems at a rapid cadence.

Table 1 illustrates the operation of the model by projecting technical and annual financial parameters for SCE systems purchased over a 20 year period, using fixed annual capital investment. For the baseline, capacity growth at historic rates does not require increased capital budgets.

3.1 Continued Cost/Performance Improvement Drives Ongoing Replacement

We, like many others, thought that, with the end of Dennard scaling and the concomitant slowing of single-core performance, the capability and TCO-driven cycle of system replacement would slow down. Our analysis, which optimizes TCO $ per unit compute delivered (see Figure 1), shows that the continued reduction of cost per unit compute drives continued upgrade cycles of 4-7 years.

3.2 Growing Power Requirements and Costs are Growing Concerns

A critical consequence of the end of Dennard scaling is that many believe the increasing energy efficiency of computing hardware is slowing (and likely to do so to a greater degree in the future). Our PD0 baseline, and even more pessimistic PD- scenario, reflect this possibility. These scenarios will see large system power increases, and power cost as an increasing fraction of TCO in the near term (see gray and red bars in Figure 2), and eventually become prohibitive.

Optimistic scenarios such as PD+ and PD++, which posit that recent large energy-efficiency gains due to increased use of GPUs can be sustained as a long-term technology trend, would produce more benign power costs and enable continuation of the status quo.

3.3 Space Requirements are Not a Growing Concern

With the rapid growth of cloud datacenters (size, number) and the apparent increase in size of many campus datacenters, another concern is that SCE computing systems will require larger physical spaces in the future. And that larger space will become a significant TCO cost. Our model results show this is not the case (see Figure 3); for most scenarios space requirements do not increase rapidly. The

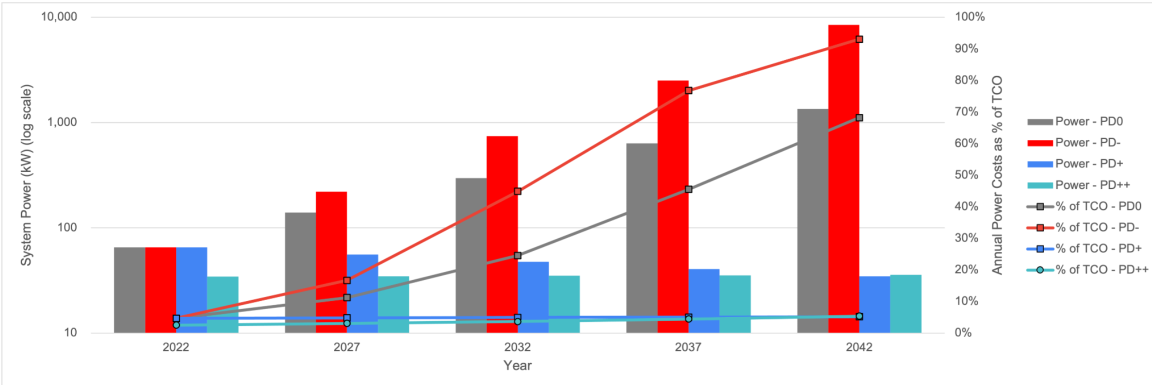

Figure 2: System Power and Power Cost as % of TCO, varied "Power Density" improvement Scenarios

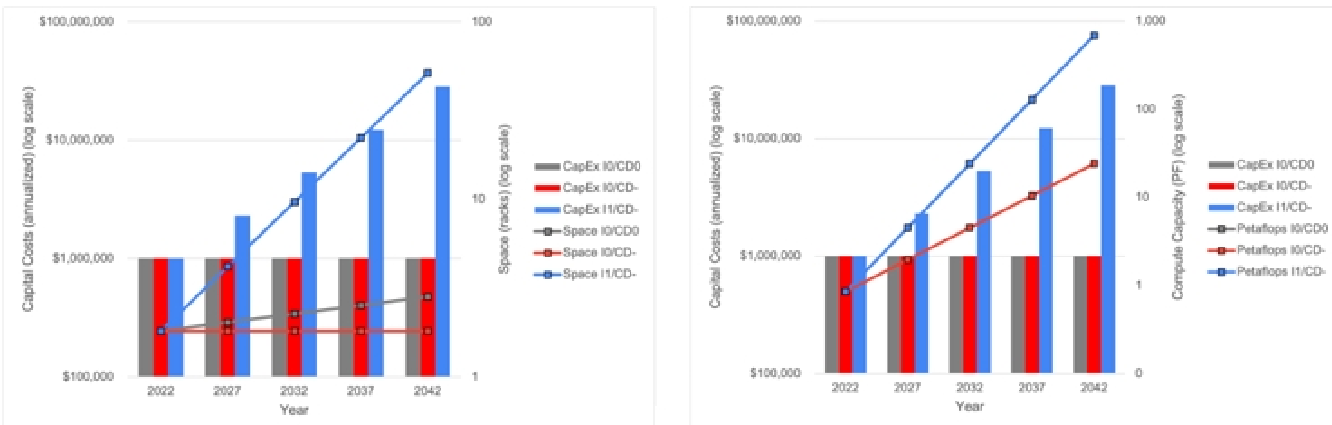

Figure 3: Evolution in Capital Costs & Space [l], Capital Costs & Compute Capacity [r], Different Scenarios

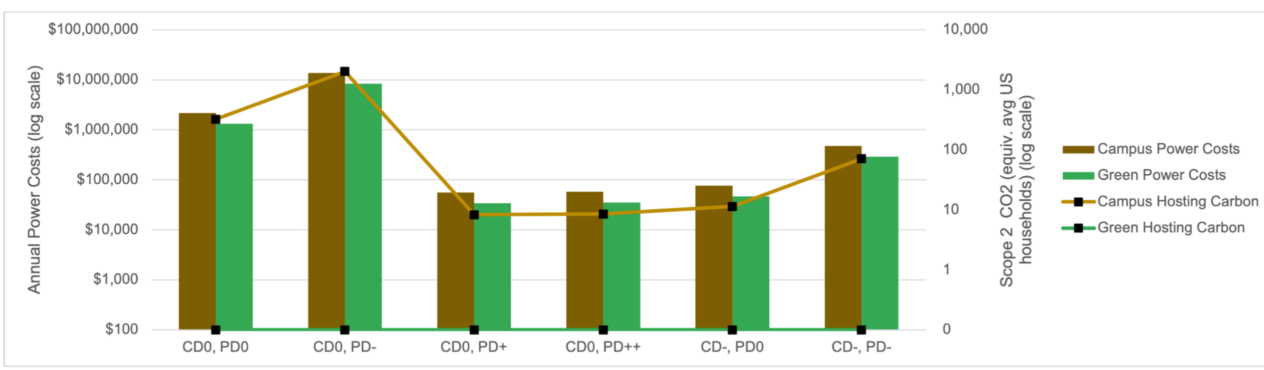

Figure 4: Power Cost and Scope 2 Carbon in 2042, Campus vs. Green Hosting, Different Scenarios

few scenarios where it does (all CD-) are tied to slowing Moore's law and cost/performance improvement, space costs growth and compute cost grow rapidly together. In short, these scenarios are cost-prohibitive (and therefore unlikely).

3.4 Carbon is Closely Related to Power, but Can be Decoupled

Our baseline scenario highlighted growing power consumption. Even at fixed power cost/MWh, the PD0 and PD- scenarios produce annualized power costs exceeding $1 million/year for NSF-scale systems (see left, Figure 4).

In the optimistic scenarios (PD+, PD++ where the improvements from GPUs are sustained for decades), power costs remain manageable. In all cases, recent novel approaches show how green hosting can decouple power consumption from carbon-emissions [6]. The green hosting options demonstrate that flexibility in siting can both reduce the cost of power, and effectively eliminate its carbon emissions.

4 SUMMARY AND FUTURE WORK

This study explores a range of trajectories for the future scientific computing systems (SCE), based on the changing technology landscape, highlighting directions of concern and change. Future work might extend these results with a larger portfolio of model inputs, and explore different specific system design points (leadership, capacity, diverse capability), highlighting the trends and challenges.

A DATA COLLECTION AND DATA USED TO CREATE MODEL

We gathered data on ten SCE systems installed between 2015 and 2020, purchased by four US universities:

- Indiana University (Site 1 in the table below),
- University of California - San Diego (SDSC – Site 2),
- University of Texas - Austin (TACC – Site 3) and
- Carnegie-Mellon University (PSC – Site 4).

Of these 10 systems, eight were funded by the NSF. Our sample is small, but during this period these systems included the largest systems funded by the NSF and account for almost 1/3rd of its SCE funding.

Data from public sources was assembled into summaries for each of the ten systems and presented to representatives of the four institutions for confirmation. Technical specifications were reviewed against publicly reported Rpeak values, to validate the configurations used in the analysis. Two institutions provided detailed confirmations of both technical details and capital costs and confirmed the accuracy of, or corrected, the pre-assembled data. Where capital costs were not confirmed by the institutions, we used 50% of the combined purchase and operations and maintenance ("O&M") awards made by NSF for the system as a proxy, recognizing that other types of costs are eligible for reimbursement for both types of award. All the institutions regard their cost information as confidential, so we do not include this information here.

Computational capacity ranged across 3 orders of magnitude, from the small, data-analysis-oriented Wrangler (92 Tflops/s Rpeak), to the NSF's current leadership computing system Frontera (41 Pflops/s Rpeak). Power required by this compute capability ranges across 2 orders of magnitude, from 29 kW (for Wrangler) to 3.4 MW (for Frontera). Physical size spans 2 orders of magnitude, from 2 compute racks for Wrangler to 107 racks for Frontera. Compute costs span almost 2 orders of magnitude, from roughly $2 million to roughly $50 million.

Data was also collected on the four institutions' hosting facilities, including data center power utilization efficiency (PUE), average power costs ($/kWh), and nominal space costs ($/rack-year).

Electricity costs per kWh range from 4 cents (TACC) to 13 cents (SDSC). Power utilization efficiency (PUE) at the facilities ranged from a low of 1.2 (TACC) up to 1.58 (Indiana). For each site, the average electricity cost and average PUE were converted to an annual electricity cost per compute kW, ranging from $336.38 to $1,229.90 ($/kWyear)[3] – a factor of 4 difference. An average annual PUE-adjusted energy cost of $764.32/kWyear was used in the analyses.

Hosting space costs were reported by two facilities: $3,132/rack-year in Bloomington, Indiana and $7,000 in San Diego, California – a factor of 2 difference. These figures in turn were used to estimate costs at the two other facilities based on their location. An average space cost of $5,533 per 19" rack was used in the analyses.

Carbon dioxide (CO2) emissions associated with the power required by each system are estimated based on state-wide statistics[4] as of 2019 for the data center locations in question. Carbon emissions (in metric tonnes (MT) CO2 equivalent per kWyear, adjusted for PUE and utilization levels) range from 1.774 MT in climate-conscious California to 6.646 MT in Indiana, where coal-fired power plants provide a significant proportion of the state's power. As with electricity costs, there is almost a factor of 4 difference between

[3] $/kWyear = $/kWh X PUE X 8760 hrs X 80% average utilization.
[4] https://www.eia.gov/electricity/state/

Table A.1: Summary of Key System Data Collected for the 10 Sample Systems

System	Site	NSF Funded	In Service (Mo/Yr)	Rpeak (Gflops/s)	Compute Power (kW)	Rpeak/kW	Compute Racks	Rpeak/ rack
Jetstream	1	Y	1/2016	516,096	140.0	3,686.4	10	51,610
Big Red II+	1	N	8/2016	286,157	90.0	3,179.5	3	95,386
Big Red 3	1	N	8/2019	928,512	179.0	5,187.2	5	185,702
Comet	2	Y	7/2015	2,831,699	550.9	5,140.3	29	97,645
Expanse	2	Y	7/2020	5,078,656	407.4	12,466.0	13	390,666
Stampede 2	3	Y	9/2017	18,394,522	2,200.0	8,361.1	106	173,533
Frontera	3	Y	7/2019	40,977,504	3,407.1	12,027.1	107	382,967
Wrangler	3	Y	7/2015	92,160	28.8	3,200.0	2	46,080
Chameleon	3	Y	6/2018	370,944	100.8	3,680.0	10	37,094
Bridges	4	Y	1/2016	2,131,341	283.4	7,519.8	20	106,567
Bridges 2	4	Y	10/2020	2,957,069	257.0	11,505.2	9	328,563

Table A.2: Key Data Collected for the 4 Hosting Facilities, along with averages for campus facilities

Site	Institution/ Center name	$/kWh	PUE	$/kWyear	$/rackyear	Availability	MTCO2/kWyear
1	Indiana University	0.078	1.58	860.34	3,132	100%	6.646
2	UCSD/SDSC	0.130	1.35	1,229.90	7,000	100%	1.774
3	UT Austin/TACC	0.040	1.2	336.38	5,000 (est.)	100%	3.941
4	CMU/PSC	0.060	1.5 (est.)	630.72	7,000 (est.)	100%	2.955
	Average of Campus	0.077	1.41	764.34	5,533	100%	3.829

Table A.3: Key Data for the 3 Green Hosting Facilities, along with averages used in the analysis

Site	Green Hosting Option	$/kWh	PUE	$/kWyear	$/rackyear	Availability	MTCO2/ kWyear
5	Reliable Green Co-location	0.06767	1.0	474.24	0	100%	0
6	Intermittent Green Co-location	0.06849	1.0	480.00	0	95%	0
7	Intermittent Green Computing Services (not included in average)	-0-	1.0	0	0	30%	0
Average green data center		0.06808	1.0	477.12	0	97.5%	0

Note that green data center power costs are slightly higher than the lowest cost institutional data center (TACC), but they also provide the benefit of zero space costs and zero (Scope 2) carbon emissions.

the highest and lowest figures, but California's CO2 emissions (per kWyear) are the lowest of the four locations, while UCSD's power costs are the highest of the four. Average carbon emissions of 3.829 MT CO2/kWyear were used in the analysis and were applied throughout the forecast period. Carbon emissions associated with power generation are expected to change over time, and hopefully to fall, as the power system moves away from fossil fuels and toward renewable sources of energy, but it is unclear what scenarios might realistically be considered.

The authors have explored the possibility of hosting SCE at new "green" data centers that take advantage of low-cost excess power generated by renewable power sources [5]. Table 3 summarizes the three commercial offers described, introducing the concept of reduced power availability (for Site 6, a contractually guaranteed 95% vs. the 100% typical of a research data center) as well as the concept of "bartering" compute capacity in exchange for free hosting (yielding an effective compute availability of 30% for Site 7). None of these offers charges separately for space. For this analysis we use the average of $477.12/kWyear for the two more traditional offers from Sites 5 and 6. The reduced power availability at Site 6 is reflected in the TCO calculations presented below by reducing power costs and delivered compute capacity by 2.5%.

ACKNOWLEDGMENTS

Work supported in part by NSF Grants CMMI-1832230 and OAC-2019506. Thanks to the Zero-carbon Cloud group!

REFERENCES

[1] B. Runesha, M. Dietrich, and A. Chien. Scientific Computing Equipment (SCE) Decommissioning in the University Community, 2017-2023: A Survey of the Coalition for Advanced Scientific Computing (CASC), UChicago CS TR 2020-05
[2] Frontier DOE supercomputer. See https://en.wikipedia.org/wiki/Frontier_(supercomputer)
[3] Chien and Dietrich, Navigating Dennard, Carbon and Moore: Technology Trends Shaping Future Investment in Scientific Computing Equipment. University of Chicago Technical Report TR-2022-04, May 2022. https://newtraell.cs.uchicago.edu/research/publications/techreports/TR-2022-04
[4] Chien and Dietrich, Navigating Dennard, Carbon and Moore: Scenarios for the Future of NSF Advanced Computational Infrastructure. University of Chicago Technical Report TR-2022-03, April 2022. https://newtraell.cs.uchicago.edu/research/publications/techreports/TR-2022-03
[5] Dietrich and Chien, Options for Extending the Life of Scientific Computing Equipment. University of Chicago Technical Report TR-2020-08, October 2020. https://newtraell.cs.uchicago.edu/research/publications/techreports/TR-2020-08
[6] Dietrich, Gardner, Chien. PEARC-20 BOF on How to Extend the Productive Lifetime of Scientific Computing Equipment, see https://pearc20.sched.com/event/cnWG