Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

SUSTAINABLE COMPUTING AND SIMULATION: A LITERATURE SURVEY

Suzanne DeLong

Army Modernization
The MITRE Corporation
903 Enterprise Pkwy Ste 200
Hampton, VA 23666, USA

Andreas Tolk

Modeling, Simulation, Experimentation & Analytics
The MITRE Corporation
1001 Research Park Blvd Ste 220
Charlottesville, VA 22911, USA

ABSTRACT

Smart technologies are everywhere and the creation of a smart world, from smart devices to smart cities is rapidly growing to potentially improve quality of life. Businesses, governments, and individual users of smart technology expect a level of service and access to data that is achieved through data and supercomputing centers. These centers potentially consume vast amounts of power and their continued growth may be unsustainable and contribute to greenhouse gasses. As smart technologies rely heavily on such computational capabilities their sustainability is pivotal for a smart future. This paper explores the literature to: identify the problems; categorize the challenges as well as possible solutions; explore how simulation and machine learning can improve computational sustainability; and consider the need to conduct trade-off analysis to determine when to apply simulation and machine learning benefits. A taxonomy for sustainable computing is presented for future research.

1 INTRODUCTION

The modern world is increasingly facing a dilemma. On the one side, we are relying more and more on information technology needed for our infrastructure. Cyber-physical systems utilize computational functions to allow for smart technology, from smart devices to smart cities, increasing safety and comfort and allowing for optimization on a new scale. Cloud computing and related concepts enable ubiquitous access to computation. We take streaming services for education, entertainment and on demand use of them for granted. Access to the ever-growing amount of data provides insights that help policy-makers as well as industry. However, this all comes with a price. Data centers and computation facilities are using a huge amount of energy and are quickly reaching the upper boundaries with respect to sustainability. A recent Forbes article observed that *"U.S. data centers use more than 90 billion kilowatt-hours of electricity a year, requiring roughly 34 giant (500-megawatt) coal-powered plants. Global data centers used roughly 416 terawatts (4.16 x 1014 watts) (or about 3% of the total electricity) last year, nearly 40% more than the entire United Kingdom. And this consumption will double every four years"* (Danilak 2017). Sustainable computing is a necessity and may even become one of the big factors when contributions to possible global warming are discussed.

The amount of energy consumed by data and supercomputing centers is great and projections of future energy consumption may be untenable. Action must be taken now to stem the ever-rising demand for power in order to continue the level of service provided through the use of data centers, supercomputing and distributed computing infrastructures. If we do not want to sacrifice the use of smart solutions on all levels, from smart appliances to smart cars and buildings and ultimately smart cities, computational support must become sustainable or we may find that the outcome is out-of-control global warming and its consequences. Our literature survey reveals that ideas about sustainable computing span many aspects of computing including green initiatives; distributed computing through virtualization, cloud computing,

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

and other variants; energy efficiency through hardware and software innovations and power optimization; and data center efficiencies and business management. This paper highlights the many initiatives being undertaken to make computing sustainable for the future.

There are already positive developments. Masanet et al. (2020) completed a study that shows energy usage has not grown as previously predicted. This development appears to be due in large part to the sustainable practices such as those described in this paper that have been increasingly applied over the last five years. Of particular interest, Masanet et al. (2020) cites the rise in cloud computing and machine virtualization as key contributors but also mentions improvements in hardware, such as improved cooling systems and power supply. Nonetheless, while efficiencies have been gained, the use of computing infrastructure continues to increase and we need to stay ahead and continue to look for ways to gain efficiencies. Masanet et al. (2020) acknowledge that the current efficiencies gained should allow for growth over the next three to four years but after that time frame new efficiencies will be needed to continue along the pathway of sustainable computing. Now is not the time to rest. A call to continuous and increased action is in order!

In an effort to better understand these challenges and possible solutions, a literature survey was conducted on sustainable computing, funded through MITRE's Innovation Program with the intention to help better define and identify the problem and to understand how simulation can contribute to the solution. The research started with general data center efficiencies ranging from green initiatives to hardware and software innovations. We focused on peer-reviewed publications as our main source. This survey is not meant to be complete and is intended to be an initial start for future research in this domain. This paper also includes our initial research findings on how simulation can play a role in gaining efficiencies.

This paper reflects the multiple facets revealed by our research and is organized as follows. First, green computing and sustainability is defined and the role of green policy, regulations and legislation is addressed. Second, the use of distributed computing in its various forms are identified as well as what the literature revealed in terms of sustainability for each of these. Third, ways to increase energy efficiency in computing, to include through algorithmic approaches for modeling and prediction, are discussed. Fourth, how the use of simulation and machine learning can further contribute to sustainable computing is discussed, which also includes our own research findings. Finally, all of the ways computing can be made more sustainable is summarized and a sustainable computing taxonomy is presented. While this paper only scratches the surface, the intent is to start a discussion on how we can apply simulation to contribute to sustainable computing as the foundation for smart solutions on all levels, from smart devices to smart cities.

2 GREEN COMPUTING

Sustainable technologies are often referred to as green, as they take the short- and long-term effects on the environment into account. This section defines some of the green computing and sustainability terminology, the equipment and infrastructure surrounding computing resources, the issues specifically found within data centers and the status of green computing policies, regulations and legislation.

2.1 Definitions for Green Computing and Sustainability

There is a growing interest in being more environmentally friendly, but as in many other domains, computing has not yet established common definitions or guidelines to help consumers on a big scale. However, looking at the use of the term, common themes become apparent.

According to `Techopedia` (2017), green computing is the environmentally responsible and eco-friendly use of computers and their resources. In broader terms, it is also defined as the study of green use and disposal of computer equipment, green engineering design, and green manufacturing. These four categories are described by Shukla et al. (2015) as follows:

- *Green use:* Minimizing the electricity consumption of computers and their peripheral devices and using them in an eco-friendly manner

- *Green disposal:* Re-purposing existing equipment or appropriately disposing of, or recycling, unwanted electronic equipment
- *Green engineering design:* Designing energy-efficient computers, servers, printers, projectors and other digital devices
- *Green manufacturing:* Minimizing waste during the manufacturing of computers and other subsystems to reduce the environmental impact of these activities. Increasing the biodegradability of computing hardware and manufacturing green PCs that are affordable, non-toxic, ultra low wattage but also offering a PC recycling service.

The office supply company Staples provides a similar understanding on their organization profile. Regarding their definition, green computing is an umbrella term, referring to an ecology-conscious way of developing, using and recycling technology, as well as utilizing resources in a more planet-friendly manner (Staples 2020). It is noteworthy that more and more organizations are featuring green technology and sustainability on their profile, as this increasingly becomes a recruiting aid for job seekers who want to be environmentally responsible.

2.2 Green Computing Equipment, Infrastructure and Centers

Some of the environmental hazards of computing are: coolants (especially in older systems); backup batteries used in data centers; cleaning materials used to maintain data center environments; diesel fuel generating backup generators to data centers; the electronic waste when hardware reaches its end of life cycle; chemicals used in fire suppression systems required at data centers due to the high voltage; the packaging from information technology (IT) equipment; office generated waste such as lighting, paper products, cleaning supplies; and carpet materials (Pramanil et al. 2019).

Active energy efficiency initiatives from Microsoft, IBM, CISCO and HP include eco-friendly booting and reduced energy consumption resulting in less heat production therefore reducing the need for cooling. Green computing must incorporate the economical use of computing resources for the home, industry and businesses for the benefit of the environment. Shukla et al. (2015) identify common green computing categories: making data centers more green, increase the use of virtualization, cloud computing, and grid computing, and improve the energy efficiency.

In researching green data centers, two topics show up repeatedly in the literature being of special importance, namely energy consumption and heat production. This does not exclude the methods described in the next section and in section 4 we will look at them in more detail.

Green data centers can redirect the generated heat into another facility such as a greenhouse, swimming pool, nearby buildings, district heating and so forth. When IT is combined with agriculture it is termed an organic data center. Data center waste heat is considered low grade energy at 35 degrees Celsius and below. As a heating source it will need to be heated further with a heat pump to about 85 degrees Celsius when in a temperate climate. Colder climates will require additional heating. Greenhouses however may not require additional heating and therefore co-locating data centers with greenhouses may be an ideal situation especially as data centers are not typically found in high population density areas. In addition to temperature adjustments, humidity adjustments may be required as well, especially when in a cold climate due to the cold-dry air. The best climate for optimizing reuse of data center excess heat is in climates with variable humidity, warmer air climates and in marine environments (Karnama et al. 2019).

One incentive to encourage green computing within the supercomputing community is the Green 500 or the TOP500, a list of publicly ranked organizations based on their supercomputer performance of ”being green” in terms of power efficiency measured in flops per watt, the TOP500 using slightly different metrics. Although this measure has been the standard, new measures should be determined due to the breadth of activities that organizations can take to meet green standards. Feng and Cameron (2007) describe the pros and cons of different measures that could help influence builders and users to design more efficient supercomputers.

The key elements of green computing to make data centers greener were observed in several of the publications we reviewed in our literature survey and are listed below:

- Using *energy-efficient hardware* i.e. computers, displays, servers, and printers.
- Better use of *resources*, such as reduced paper consumption and lower energy utilization.
- Greater awareness so that technology components don't end up in the *waste* stream.
- Stronger environmental controls for technology production, leading to fewer *toxic chemicals* in the finished products.
- Using power consumption and cumulative energy consumption as a control parameter as well as measure of merit.

In general, green computing works at all levels — development, manufacturing, usage and disposal — to minimize technology's impact on the global environment. However, there is little agreement between the various organizations that allow for comparisons. This includes the lack of common terms as well as common metrics.

2.3 Green Computing Policies, Regulations and Legislation

The increase in computing technology has naturally led to increases in power consumption and therefore higher greenhouse gas emissions as well as the host of other issues discussed in this section. As a result, the need for policy in this sector to regulate and legislate for the common good has become a global priority. As an example, some organizations such as the Association of Computing Machinery (ACM) recommend good practices to help achieve carbon neutral data centers. Recommended courses of action can include a carbon tax, accounting oversight, legislation for greenhouse gas reductions and the means for enforcement as well as strategic regulations. The support of the Europe Technology Policy Committee (TPC) – comprised of computer science experts from 10 countries – with the Green Deal climate, targets and encourages the establishment of "specific and mandatory emissions targets and legal limitations for the Information and Communication Technology sector" (ACM Europe Technology Policy Committee 2020). When looking at these legal limitations, it comes to mind that these limits can be used in simulation as constraints in order to optimize the data center simulation components and configurations so that these limits/constraints can be achieved.

Gilligan and Vandenbergh (2020) show how communities are helping to lead the change for green legislation by demanding the businesses they interact with hold themselves to a higher, greener standard. This effect has been termed "private governance" and is making a difference in energy conservation. They admit however that to be completely successful there needs to be a partnership between public and private governance. The effects of private governance however is an area that is in need of further research and applying Gilligan and Vandenbergh's framework to assess the impact of private governance on sustainable computing is also an area for future research (Gilligan and Vandenbergh 2020).

One place to start with sustainable computing policy could be in encouraging the setting of "efficiency standards such as energy star ratings for servers, storage and network devices while requiring such certifications in public IT procurement programs" (Masanet et al. 2020). They advocate incentives for the use of cloud services, utility rebates and Power Usage Effectiveness (PUE) reductions.

Current policy, regulation and legislation seems to be behind the curve. Although this was not a focus of the literature search, only a few publications did address the issue. Some papers, such as Kumar, et al. (2017), are addressing recycling issues, but we did not find support for legislation on the limitation of power consumption and heat emission. The reason may be that the effects of toxic waste are obvious to everybody, but the effects of high power consumption and immense heat emissions are not immediately clear. This may be the first part where computer simulation can be of help, as they can immerse decision makers into an interactive visualization of the problem domain, in which they can experience the effect of their policies, as also envisioned in Rouse (2021).

It is our interpretation that the already mentioned lack of common terms and metrics is standing in the way of policy, regulations, and legislation. We cannot unambiguously describe *what* to measure or *how* to measure. Our survey suggests that only in the case of data centers has the community accepted the use of the reduction of power as an accepted metric, although this is more a measure of individual data center success rather than a metric used to regulate data center power consumption. Metrics are key enablers to establish and agree upon appropriate policy, regulation and legislation.

3 DISTRIBUTED COMPUTING

One reoccurring idea in our literature search was to outsource computational needs to distributed computing capabilities that are more environmentally friendly than local solutions. This requires the virtualization of applications and the sustainable implementation of computer server nodes using cloud computing and related technologies. This section highlights the benefits of virtualization, cloud computing and other variants, as identified to be a main contributor to better sustainability in Masanet et al. (2020), as well as parallel and grid computing.

3.1 Virtualization

Virtualization is the creation of a simulated computing environment to include computer hardware platforms, storage devices, and computer network resources resulting in efficiencies with actual computer hardware. Virtualization is the basis for cloud computing and makes use of *"software to create an abstract layer... [allowing] a single computer to be divided into multiple virtual machines...contributing to greater efficiencies by making optimal use of the actual computer hardware...by purchasing computing resources when they need it"* (IBM Cloud Education 2019). Virtualization increases scalability while also reducing costs, and if the supporting infrastructure is more ecology-friendly than the current solutions, it also increases sustainability. Five benefits of virtualization are:

- Resource efficiency with less spending on servers as a result of consolidating applications.
- Faster provisioning which increases resiliency and decreases server downtime by moving the affected virtual machine (VM) rather than have to wait to set up a new physical server.
- Easier management with reduced maintenance time therefore increasing efficiency and productivity.
- Greater control and independence especially for development and operations (DevOps).
- Green benefits by cutting down on the number of physical servers (reducing power consumption) and reducing the carbon footprint.

Virtualization can be accomplished by various technologies, starting with alternative data storage concepts or the increased use of virtual machines reducing the number of computers necessary. Many of these methods described in detail by the various providers are targeting the migration towards a cloud or a related concept, which provides infrastructure as a service (IaaS), platform as a service (PaaS) or software as a service (SaaS) (Kavis 2014). They help the user by providing multiple machines at the same time while sharing a single physical instance of resources or applications to multiple users transforming traditional computing into a more scalable, economical and efficient form of computing.

These approaches require computing server nodes that can reside in cloud/edge or fog/mist computing. They can be categorized as a grid or parallel computing. Each of these are explored below to highlight their contributions to computing efficiency and their challenges.

3.2 Cloud Computing and Variants

Cloud computing is the use of available computer system resources on-demand for data storage and/or computing power without the user needing to actively manage the resources. Edge computing is an extension of these cloud networks and are a collection of servers in a distributed network. If the services are provided at the access points to the cloud, they also are referenced as edge services. If the computational functions are provided close to the user, even decoupled from the cloud itself and using the cloud more as background support, we enter the domain of fog computing services. Fog computing services are also understood as a mediator between the system and the edge of the cloud. Cyber-physical systems often use them to access the cloud. Finally, if the components and their various hardware-components use cloud principles on the platform, such as using them to integrate various sensors on one system, we are reaching the mist computing services. Examples for these technologies and how they support cyber–physical systems are given in Tolk (2020).

Efforts to make cloud/edge computing more efficient include the creation of cloud federations allowing differing clouds to join the cloud network dynamically allowing the entire network to flex as computing demands dictate. VM disk images are distributed across the network.

Bruneo (2013) builds a stochastic queuing model to investigate the data center performance and the quality of the service with IaaS in a cloud federation with future work to apply to flexible configuration of systems and PaaS and SaaS cloud federations. Alshahrani and Peyravi (2014) also advocate for data center networks to optimize data centers as a larger system of systems and proposes a new analytical model that is tested with modeling and simulation to explore the impact on power consumption.

Pramanil et al. (2019) suggests the use of smart phones as a crowd sourced, fog/mist, computing capability. Of course some of the challenges in adopting smartphone crowd computing are motivation of smartphone owners to participate, finding incentives and policies that would be attractive, security and privacy concerns, variable wireless network performance, middle ware that can support the disparate smartphone models, maintaining quality of service, heat generated by the phone in constant use, limited memory with no expansion capability, battery power given tasks that could drain the battery rapidly among a host of other issues identified. The future of battery life is discussed along with faster charging as well as alternative energy sources (Pramanil et al. 2019). Due to this decentralized infrastructure, fog/mist computing created as a result of crowd sourced smart phones will be more difficult to manage job scheduling, utilizing local and renewable sources of energy, and improving energy consumption without contributing to latency issues (Gougeon et al. 2020).

A domain of similar interest is grid computing. A grid is understood as a distributed collection of computing nodes where each node is set to perform a specific, assigned task. Computing nodes are often heterogeneous and can be geographically dispersed. Parallel computing on grids involves breaking down a problem into parts and simultaneously solving the smaller parts. The forms of parallel computing consist of bit-level, instruction-level, data and task parallelism. The combination of cloud and grid concepts to increase their sustainability is the topic of research in Marinos and Briscoe (2009).

All these concepts require access to data and computational power, which usually are provided by supercomputing and data centers. Besides preparing the original systems to use virtualization to migrate away from unsustainable solutions, the focus must shift in parallel to greener supercomputing and data centers, otherwise we gain nothing regarding increasing the overall sustainability.

4 ENERGY EFFICIENCY

In this section, we are focusing on increasing the sustainability of data and supercomputing centers by improving their energy efficiency. Many of the techniques described here can be applied to all computing capabilities. Ways to improve hardware energy efficiency are addressed as well as how modeling and predicting energy consumption can be beneficial. We then focus on data centers as a system of systems and how efficiencies can be gained through green practices and business management.

4.1 Increasing Energy Efficiency at the Hardware Level

Supercomputing takes place when many processors are connected by a local high speed connection. Supercomputers are being installed world wide at a record pace and are contributing to high power demand.

As documented in Borghesi et al. (2018), power capping is one technique that has been applied in the past, however without too much regard for quality of service. Constraint programming techniques can aid in lowering power consumption while maintaining some degree of quality of service with the use of energy aware schedulers. These proposed methodologies were compared to current methodologies and demonstrated an 8.5% improvement after experimentation with "state-of-the-art" methods on "historical traces of jobs" to show the impact on power reduction with Running Average Power Limit (RAPL) or frequency scaling. Power capping and power prediction methods along with energy aware schedulers can help reduce supercomputer power consumption.

A major breakthrough in computer architecture is needed as the more obvious methods to gain efficiencies have been exhausted and continued improvements in these areas will most likely be minimal (Danilak 2017). Mondal et al. (2017) examine leakage power in routers and show that power and performance aware routing with Network-on-chip (NoC) architectures can save leakage power consumption by up to 92% in base routers and 68% in hybrid routers with a 49% savings in total packet energy consumption. This is another area that could benefit from modeling and simulation where the leakage power consumption savings found by Mondal et al. (2017) can be used as input parameters in simulation modeling.

Danilak (2017) identifies that forty years ago, wires were faster than transistors, but today, transistors are much faster than wires. However, despite this development, we *"still build and use processors for when wires were faster than transistors"* (Danilak 2017). A major breakthrough in computer architecture must be forthcoming to overcome this gap. What this breakthrough will look like is still to be determined.

4.2 Energy Consumption Modeling and Prediction

Methods proposed to improve energy consumption modeling and prediction include mathematical modeling, regression modeling, artificial neural networks (ANN) (Lin et al. 2020), K-means based rand variable learning rate backpropogation neural network (K-RVLBPNN) (Lu et al. 2020), simulation with a data center's digital twin (Ikemoto 2020) , hardware refresh (Bashroush 2018), and combinatorial optimization on heterogeneous system architecture (Li 2016). The use of digital twins, i.e., the digital equivalent representation of the system of interest, will be addressed in a later section in more detail.

ANN was cited for cloud data center server power consumption prediction while modeling performance of CPU, memory and disk sub-components of a cloud data center's servers (Lin et al. 2020). K-RVLBPNN helps predict the trend in workload arrival. Energy is wasted at data centers when servers are running under low workload conditions but are obligated for unachieved anticipated demand. Predicting these workloads and forecasting server workload to optimize workload demand is needed to reduce wasted energy consumption (Lu et al. 2020). The downside to these types of models is highlighted in Strubell et al. (2019) where deep learning algorithms may expend more energy as they are *"costly to train and develop"*. This paper gives examples that detail the *"financial and environmental costs of training ... successful neural network models"* (Strubell et al. 2019). Sustainable computing should be balanced with simulation and machine learning methods with a systems thinking approach so that the methods employed to discover how to conserve energy don't inadvertently contribute to exorbitant energy consumption due to the high demand of these models for energy.

In determining when hardware refresh should occur, cumulative energy consumption is used as a metric to determine at what point in time current hardware should be upgraded to refreshed hardware so that energy savings can be realized. The relationship between hardware utilization and PUE is considered and helps drive efficiency while Moore's and Koomey's Law predict performance evolution of servers (Bashroush 2018).

4.3 Application for Data Centers

Besides supercomputing centers, data centers are drivers of energy consumption. However, they are needed to make smart cities, buildings, cars, and supporting concepts a reality as the vast amount of data needed to drive such concepts cannot be achieved otherwise. Ways of modeling data center energy consumption to make data centers more efficient is summarized below along with how data centers are now managed so that their operations are optimized.

4.3.1 Modeling Data Center Energy Consumption for Efficiency

Miyuru et al. (2015) completed a data center energy consumption modeling survey up through 2016 which also supports that data center power consumption is growing and is a chief concern. Large savings can be had when under-utilized servers can be turned off and workload consolidated. These techniques reduce system performance so a balance must be struck between energy savings and high performance (Miyuru et al. 2015).

Miyuru et al. (2015) also concludes that modeling techniques for *"temperature or energy aware scheduling, dynamic voltage frequency scaling, resource virtualization, improving algorithms used by applications, switching to low-power states, power capping, and shutting down unused servers can contribute to data center efficiency."* Power models exist in the design of data center systems, forecasting energy efficiency trends and energy consumption optimization. Power modeling research is being done on both linear and non-linear correlations between system utilization and power consumption. Data center energy consumption depends on hardware specs, workload, cooling requirements, and types of applications used. Energy consumption prediction techniques were developed to help forecast energy used based on these variables (Miyuru et al. 2015). This work organizes existing data center power models into a framework to allow multiple models to be combined and describes a more detailed data center taxonomy as a valuable resource and is extensive in its research and organization. This paper is good one stop shopping to learn about data center energy consumption as well as hardware and software implications on power consumption. With over 400 references, this paper can quickly bring you up to speed on the status of data center energy consumption modeling up until 2016.

Simple and affordable techniques to make a current data center more power efficient by enhancing the current infrastructure according to Pickering (2008) are:

- Measure where your power is going by category (IT systems, UPS, chillers and lighting)
- Virtualize and consolidate IT Systems. "Server virtualization" takes advantage of pooled network storage resulting in "savings in space, power and cooling"
- Reduce data by eliminating "duplicate data objects"
- Eliminate overcooling by monitoring temperatures and make use of "variable frequency drives on air handlers"
- Take advantage of physical properties of hot and cold air. The creation of "hot/cold aisles" segregates the air and allows for efficient cooling
- Isolate the heat with vinyl curtains to quickly dissipate heat to the outdoors
- Draw in outdoor cold air when temperatures allow to save on cooling
- Use of kinetic UPS resulting in less energy loss than batteries
- Make use of cogeneration systems to make use of heat byproduct to power adsorption chillers
- Monitor the environment routinely by measuring not only at the building perimeter but at the rack level to adjust energy consumption on an as needed basis rather than continuous basis

While the techniques described above can help enhance current data center structures, the construction of new data centers should consider attaining Leadership in Energy and Environmental Design (LEED) building certification.

In order to quantify how efficient your data center is, Reddy et al. (2017) developed a taxonomy of data center metrics in order to track data center efficiencies as upgrades are made to the data center such as managing power consumption and turning off servers when not in use as detailed in Miyuru et al. (2015) or any of the techniques described above.

4.3.2 Supporting Data Center Business Management

From a business standpoint, sustainable computing is fiscal responsibility for the environment and to the businesses bottom line. To enable entire data center management, Data Center Infrastructure Management (DCIM) software is available to optimize every physical aspect of your data center from the operating costs, electrical power consumption, cooling, optimizing floor space and information technology through the use of instrumentation and software (Taylor 2018). Before companies are able to efficiently use DCIM they need to find out some characteristic attributes for their data center:

- Data, cooling, and power information,
- Consumption of space resources,
- IT resource growth patterns,
- Capacity planning information,
- Resource utilization.

This information is then used to develop the best engagement and support plan, as DCIM is not a one-size-fits-all solution. Once configured and adapted, it can become very powerful. Among others, companies like Google, IBM, and Cisco are using DCIM. There is a lot of support for the use of DCIM software currently. The vendors of this software are offering webinars and conferences around its concepts and use. It is on the way to becoming a possible de-facto standard, but has not yet been considered a de-jure standard as well. In any case, it seems to be worthwhile to observe how the data center community is going to embrace these solutions.

5 RESEARCH FINDINGS: SIMULATION AND MACHINE LEARNING

In this section, we synthesize how simulation and machine learning can be used to improve computational sustainability in the future. Simulation and machine learning can potentially make further improvements for sustainable computing by: defining parameters and examining the effects of implementing policy, regulations and legislation; exploring the various forms of distributed computing; exploring and optimizing data center algorithms for modeling and predicting, power consumption, and performance; and data center digital twin applications. The value of simulation as an enabler for increased sustainability has been recognized for several years now and should be employed to help further gain efficiencies where practical (Tolk 2010).

When introducing policy, regulation and legislation on sustainable computing, it would be beneficial to conduct simulation studies to help define the parameters and examine the effects of implementing these policies, regulations and legislation. Being able to study the effects these might have on the environment, the economy and the public's day-to-day life might help formulate amenable policies to all stakeholders and eliminate the fear of not knowing what some of the second and third order effects of these policies might cause. While simulation may not be able to uncover all of the effects, it will be a start in the right direction and may uncover some of the issues necessary to learn about the impact of the policies, regulations and legislation.

Simulation can be used to explore the various forms of distributed computing. Examining virtualization, cloud/edge, fog/mist, parallel and grid computing in a simulated environment might reveal insights to sustainable computing without the expenditure and effort of actually building these environments. Simulation can help optimize and maybe even uncover novel approaches that have not yet been thought of.

A system that can be modeled with an algorithmic solution is ripe for a use case in simulation as these algorithmic solutions may use current information to model or forecast future resource power consumption and power leakage. Simulation allows one to experiment with techniques such as power capping, energy aware schedulers, system performance, hardware refresh and different computer architectures. In discussing algorithms such as neural networks, Strubell et al. (2019) points out that the configuration of a simulation requires significantly less power than the relearning of a neural network. More research is needed in this area to quantify the difference and learn if it is significant.

Simulation has already been utilized in data center facility layout planning (Nåfors et al. 2017). The software company Future Facilities that has built a physics based 3D digital twin of a data center to predict facility operations under varying conditions to optimize *"business goals with facility operations and efficiencies"* (Ikemoto 2020) is one example of a data center digital twin application. The increased demand for digital services and use of complex computer apps are reaching the boundaries of computing capabilities and the digital twin data center can be used to experiment with differing data center configurations and hardware to realize where greater efficiencies can be gained both before or after your data center has been established.

Another benefit of digital twin technology is enhanced control of smart facilities. Extending the ideas captured in Egeland and Gravdahl (2002), using simulation enhances control by integrating detailed system knowledge with the power of analytic prediction. In the case of the data centers, we can observe and predict when computational power needs to be provided to minimize heat emissions. We can also allocate data and supercomputing centers in advance for a fair share of the computational load. We can optimize the use of the data centers with respect to sustainability. A second order effect of this simulation experimentation may be the definition of metrics and data boundaries to maintain sustainable computing which in turn will help gain industry agreement and support by both private and public governance bodies. The closer the simulation community collaborates with the digital engineering community, the better we will be able to achieve success in computational sustainability for a smart world.

For the planning of future data centers, supercomputing centers, and supporting distributed computing methods, the use of simulation will generate numerical insight into the infrastructure as a complex system. It also generates multiple options and allows what-if analysis under multiple design options and use cases. This will provide understanding of not only what the information technology requirements will be when moving closer to the vision of a smart world, it will also help to plan how the data and supercomputing center will fulfill these future requirement while at the same time running more optimal and more sustainable solutions.

6 CONCLUSION

Sustainable computing has certainly come a long way. The biggest hurdle was recognizing that there was a problem with the excessive power consumption as a result of the growth of smart devices, data centers and supercomputing technology. The good news is that more people are aware of this problem and are implementing the best practices to fix it so that computing can be sustainable for future generations. The problem remains going forward into the future however as we continue to grow computational capabilities. Some of the future solutions may lie in simulation and machine learning to aid in developing policy, regulation and legislation; continue advancements in distributed computing; optimizing algorithms for improved modeling, prediction, power consumption and performance; and creating digital twins. Given all of the efforts to date, a new taxonomy for sustainable computing is proposed and figure 1 summarizes the literature contributions to sustainable computing.

While today's efficiencies were gained as a result of distributed computing and improving energy efficiency (Masanet et al. 2020), additional ways to achieve energy efficiencies in the future may be in the application of smart simulation and machine learning techniques. It is for the benefit of all including our planet that we must embrace the ideas of sustainable computing and put them into practice every day.

The research result presented in this paper can only be the beginning. If we want to reap the benefits of a smart world, we have to solve the challenge of sustainable computing. The various technologies and methods are promising. We hope that many simulationists will heed this call to action and further apply smart simulation, particularly in connection with machine learning, to contribute to solving the problem of sustainable computing.

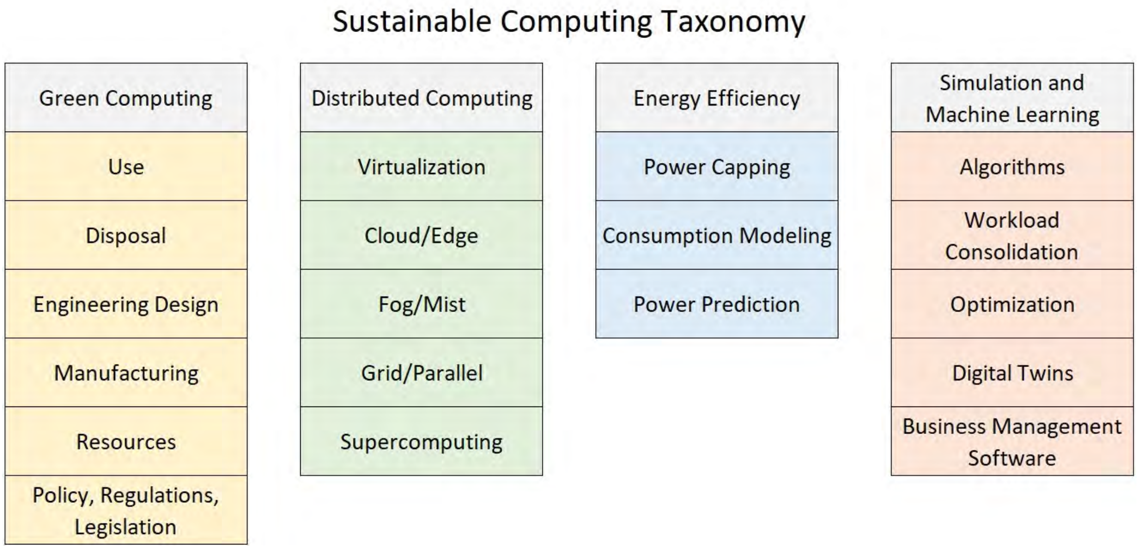

Figure 1: A Proposed Taxonomy for Sustainable Computing.

7 ACKNOWLEDGEMENT AND DISCLAIMER

The work presented in this paper was supported by the MITRE Innovation Program. The views, opinions, and/or findings contained in this paper are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation. Approved for Public Release; Distribution Unlimited. Public Release Case Number 21-1091.

REFERENCES

Alshahrani, R., and H. Peyravi. 2014. “Modeling and Simulation of Data Center Networks”. In *Proceedings of the 2nd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 75–82. Association for Computing Machinery.

Bashroush, R. 2018. “A Comprehensive Reasoning Framework for Hardware Refresh in Data Centers”. *IEEE Transactions on Sustainable Computing* 3(4):209–220.

Borghesi, A., A. Bartolini, M. Lombardi, M. Milano, and L. Benini. 2018. “Scheduling Based Power Capping in High Performance Computing Systems”. *Sustainable Computing: Informatics and Systems* 19:1–13.

Bruneo, D. 2013. “A Stochastic Model to Investigate Data Center Performance and QoS in IaaS Cloud Computing Systems”. *IEEE Transactions on Parallel and Distributed Systems* 25(3):560–569.

ACM Europe Technology Policy Committee 2020. “Association of Computing Machinery (ACM) Europe Technology Policy Committee (TPC), “RE: Comments on the EU Climate Ambition for 2030 and for the Design of Certain Climate and Energy Policies of the European Green Deal (COM/2019/640 final)”. Letter to Director General for Climate Action, Raffaele Mauro Petriccione.

Danilak, R. 2017. “Why energy is a big and rapidly growing problem for data centers”. *Forbes* 15:12–17.

Egeland, O., and J. T. Gravdahl. 2002. *Modeling and simulation for automatic control*, Volume 76. Marine Cybernetics Trondheim, Norway.

Feng, W.-c., and K. Cameron. 2007. “The Green500 list: Encouraging sustainable supercomputing”. *Computer* 40(12):50–55.

Gilligan, J., and M. Vandenbergh. 2020. “A framework for assessing the impact of private climate governance”. *Energy Research and Social Science* 60:101400.

Gougeon, A., B. Camus, and A.-C. Orgerie. 2020. “Optimizing Green Energy Consumption of Fog Computing Architectures”. In *2020 IEEE 32nd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. September 9th-11th, Porto, Portugal, 75–82.

IBM Cloud Education 2019. “Virtualization: A Complete Guide”. https://www.ibm.com/cloud/learn/virtualization-a-complete-guide, accessed July 26th 2021.

Sherman Ikemoto 2020. “Future Facilities Interview with Sherman Ikemoto”. Data Center POST.

Karnama, A., E. Bitaraf, and H. R. Vinuesa. 2019. “Organic Data Centers: A Sustainable Solution for Computing Facilities”. *Results in Engineering* 4:100063.

Kavis, M. J. 2014. *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)*. John Wiley & Sons.

Kumar, A., M. Holuszko, and D. C. R. Espinosa. 2017. "E-waste: An overview on generation, collection, legislation and recycling practices". *Resources, Conservation and Recycling* 122:32–42.

Li, K. 2016. "Energy-efficient task scheduling on multiple heterogeneous computers: Algorithms, analysis, and performance evaluation". *IEEE Transactions on Sustainable Computing* 1(1):7–19.

Lin, W., G. Wu, X. Wang, and K. Li. 2020. "An Artificial Neural Network Approach to Power Consumption Model Construction for Servers in Cloud Data Centers". *IEEE Transactions on Sustainable Computing* 5(3):329–340.

Lu, Y., L. Liu, J. Panneerselvam, X. Zhai, X. Sun, and N. Antonopoulos. 2020. "Latency-Based Analytic Approach to Forecast Cloud Workload Trend for Sustainable Data Centers". *IEEE Transactions on Sustainable Computing* 5(3):308–318.

Marinos, A., and G. Briscoe. 2009. "Community cloud computing". In *IEEE international conference on cloud computing*, 472–484. Springer.

Masanet, E., A. Shehabi, N. Lei, S. Smith, and J. Koomey. 2020. "Recalibrating Global Data Center Energy-Use estimates". *Science Magazine* 368(6481):984–986.

Miyuru, D., Y. Wen, and R. Fan. 2015. "Data Center Energy Consumption Modeling-A Survey". *IEEE Communications Surveys and Tutorials* 18(1):732–794.

Mondal, H. K., S. H. Gade, R. Kishore, and S. Deb. 2017. "P2NoC: Power- and Performance-aware NoC Architectures for Sustainable Computing". *Sustainable Computing: Informatics and Systems* 16:25–37.

Nåfors, D., E. Lindskog, J. Berglund, L. Gong, B. Johansson, and J. Vallhagen. 2017. "Realistic virtual models for factory layout planning". In *2017 Winter Simulation Conference (WSC)*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 3976–3987. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Pickering, L. 2008. "10 Ways to Make Your Data Center More Efficient". https://www.networkworld.com/article/2283606/10-ways-to-make-your-data-center-more-efficient.html, accessed July 26th 2021.

Pramanil, P., D. Kanti, S. Pal, and P. Choudhury. 2019. "Green and Sustainable High-Performance Computing with Smartphone Crowd Computing: Benefits, Enablers and Challenges". *Scalable Computing: Practice and Experience* 20(3):259–283.

Reddy, V. D., B. Setz, G. Subrahmanya, V. R. K. Rao, S. G. R. Gangadharan, and M. Aiello. 2017. "Metrics for Sustainable Data Centers". *IEEE Transactions on Sustainable Computing* 2(3):290–303.

Rouse, W. B. 2021. "Understanding the complexity of health". *Systems Research and Behavioral Science* 38(2):197–203.

Shukla, V., D. Singh, S. Kumar, and R. Singh. 2015. "Sustainable Computing – A New Pathway to Go with Green Computing". *International Journal of Engineering Research and Technology* 4(2):241–242.

Staples 2020. "Organizational Profile - Corporate Responsibility - Environment". https://www.staples.com/sbd/cre/noheader/about_us/corporate-responsibility/environment/index.html, accessed July 26th 2021.

Strubell, E., A. Ganesh, and A. McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP". *CoRR* abs/1906.02243:1–6.

Taylor, A. G. 2018. *DCIM For Dummies*. 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Techopedia 2017. "Green Computing". https://www.techopedia.com/definition/14753/green-computing, accessed July 26th 2021.

Tolk, A. 2010. "Engineering management challenges for applying simulation as a green technology". In *Proceedings of the American Society for Engineering Management Symposium*. October 13rd-16th, Fayetteville, Arkansas, USA, 137–147.

Tolk, A. 2020. "Composability Challenges for Effective Cyber Physical Systems Applications in the Domain of Cloud, Edge, and Fog Computing". In *Simulation for Cyber-Physical Systems Engineering: A Cloud-Based Context*, 25–42. Cham, Switzerland: Springer.

AUTHOR BIOGRAPHIES

SUZANNE DeLONG is a Lead Data Science Systems Engineer at The MITRE Corporation. She holds a PhD and M.Sc. in Systems Engineering from the University of Virginia and a B.Sc in Aerospace Engineering from the United States Military Academy, West Point, New York. Her email address is sdelong@mitre.org.

ANDREAS TOLK is Senior Principal Computer Scientist at The MITRE Corporation. He is an adjunct faculty member of Old Dominion University. He holds a PhD and M.Sc. in Computer Science from the University of the Federal Armed Forces, Germany. He is a Fellow of SCS and a senior member of ACM and IEEE. His email address is atolk@mitre.org.