

## Distributed (green) data centers: A new concept for energy, computing, and telecommunications<sup>☆</sup>



Stephen Bird<sup>a,\*</sup>, Ajit Achuthan<sup>a</sup>, Othman Ait Maatallah<sup>a</sup>, Wenjin Hu<sup>a</sup>, Kerop Janoyan<sup>a</sup>, Alexis Kwasinski<sup>b</sup>, Jeanna Matthews<sup>a</sup>, David Mayhew<sup>c</sup>, Jay Owen<sup>d</sup>, Pier Marzocca<sup>a</sup>

<sup>a</sup> Clarkson University, USA

<sup>b</sup> UT Austin, USA

<sup>c</sup> Arizona State University, USA

<sup>d</sup> AMD Inc., USA

### ARTICLE INFO

#### Article history:

Received 19 June 2013

Revised 20 December 2013

Accepted 20 December 2013

Available online 13 January 2014

#### Keywords:

Renewable energy

Sustainable development

Green data centers

### ABSTRACT

We introduce a new conceptual approach to green data centers through the use of multiple, distributed networked data centers that are co-located with renewable energy. While green data centers are a well-established objective, conventional approaches have focused on either high efficiencies and/or the use of renewable energy. This model extends and innovates both approaches with three key design aspects. First, it emphasizes the use of many small, geographically distributed mini data centers. Second, such data centers can operate with the option of being dislocated from the electrical grid via a mini-grid, or behind the meter. Third, the approach enhances the use of renewable energy via the ability to transfer computing load from one mini data center to another as renewable supply or market contexts dictate. This paper introduces the conceptual approach, outlines the research broadly, and examines potential technical, economic, and policy constraints for the establishment of this approach on a broad scale. Finally, we discuss the implications of this model for the developing world, particularly in areas without substantial electricity transmission infrastructure.

© 2013 International Energy Initiative. Published by Elsevier Inc. All rights reserved.

### Introduction

This paper describes a new conceptual approach to the use of renewable, constrained, or stranded energy sources combined with a network of distributed, small, cheap, and highly efficient mini data centers that are co-located at the generation source. This allows the optimization of the use of renewable energy at its source, reducing or eliminating the need for costly transmission infrastructure (and associated line losses and other inefficiencies). This model of Distributed Green Data Centers (DGDCs) depends in part on the strategic shifting of computational load from one mini data center to the next to adjust for the non-availability of intermittent renewable power, and/or to adjust to economic conditions; particularly spot energy pricing and demand/response obligations. This concept is designed to be duplicative and resilient in its computing ability. Some mini data centers may continue to draw power off the grid to augment the intermittent renewable sources; others may shift their computational load to other locations and power-down some or all of their resources when renewable power is unavailable; and some will do a mixture of both. For many

renewable power production sites, this approach will use all power produced locally for on-site computation, avoiding the cost of expensive, inefficient conversion stages.

Both the information and electricity sectors are multi-hundred billion dollar market segments that are rapidly expanding. They are key ingredients for economic development, and critical enablers for a high quality of life. Further, each is combined with the other, with electricity requiring and adding more information than ever before, and information/computation requiring more electricity than ever before. Generally, information and electricity are not integrative in nature, because electric power and information are treated as separate layers within a single technological system.

Further, both sectors are also confronting new challenges: electricity is increasing its use of distributed, intermittent renewable sources of power such as wind and solar, and information is moving to a broadly pervasive cloud computing model. The expectation is that the need for a variety of different types of green data centers with expectations for \$17 billion in spending on the “greening” of data centers in 2012 and more than doubling to \$45 billion by 2016 (Clancy, 2012). This interconnection introduces cross-sector constraints: moving forward, the existing grid will be hobbled if it doesn't become “smart” and information will be hobbled if it doesn't become more energy-efficient. Left alone, the trends of the IT sector will exacerbate the challenges of the electric sector, and vice versa.

<sup>☆</sup> This research is funded in part by a New York State Energy Research Development Agency Program Opportunity Notice (PON) 1772. Any errors or omissions are the responsibility of the authors.

\* Corresponding author.

However, the energy and computing sectors can be integrated in a redesign in which they reduce many of these concerns in a cost-effective manner, while improving upon important national goals for economic development, national security, energy, and the environment. In essence, the DGDC concept uses photons as a proxy for dispatchable electrons. This approach merges the power and information sectors in a way that has the potential to synergistically improve the economics, resilience, performance, and environmental impacts of both.

Renewable energy sources, especially wind and solar, employ highly distributed generation formats and are generally dislocated from urban centers where their power is needed. In an analogous trend, the information sector is moving towards a distributed cloud model that shifts computing from desktops and privately owned machine rooms to third party owned and managed, shared remote servers. In many cases, remote servers in large centralized server farms are energy intensive and inefficient because transmission costs, conversion losses, and congestion create costs that may not be supplanted by their economies of scale. The location of remote servers is generally not significant to the cloud computing model. Thus they could also be geographically distributed and dislocated from urban centers. In particular, they could be collocated with renewable power plants.

### Project description

As shown in Fig. 1, many small mini data center sites are spread over an extensive geographical area, taking advantage of indigenous green energy sources and interconnected with a high-speed data network. The extensive geographically-dispersed approach provides a way to

distribute server utilization according to each site's availability of instantaneous generated local power. Although the primary focus of DGDC is on solar and wind energy resources, other renewable, low-carbon, and/or alternative sources of energy may include bio-energy, hydro-electric (particularly small scale or run-of-river hydro), geothermal, tidal, and other options. Other uses may also be developed to take advantage of currently unusable or non-economic remote energy sources. For instance, geographic locations that are inherently difficult to access for electricity transmission, such as remote natural gas pipeline substation locations (which must vent or flare gas at times) would also be optimal sites for this scheme. These could be combined and co-located with micro-turbines or fuel cells as another approach that demonstrates the modularity and flexibility of this concept.

Since much of the power generated by these local energy resources is variable or intermittent, both on a micro (expected energy availability for the next 3 min) and macro scale (expected energy availability for the next 3 h), the system needs to dynamically adjust applications and databases so that they can be shifted within the geographically dispersed system to where power is available at the time. Because there is a stochastic nature to the intermittency—as clouds block solar panels in one area, wind power picks up in another—there is generally power available *somewhere* all the time: for example, as solar power in the southeast dies down in the evening, wind picks up in the plains. This has been described as “follow the sun, follow the wind” (Van Heddeghem et al., 2012, p. 2). In essence, using geographic and source diversity allows an operator to “firm up” the variability of intermittent resources. Distributed computing can be dispatched geographically to match the availability of that power, solving the need for more

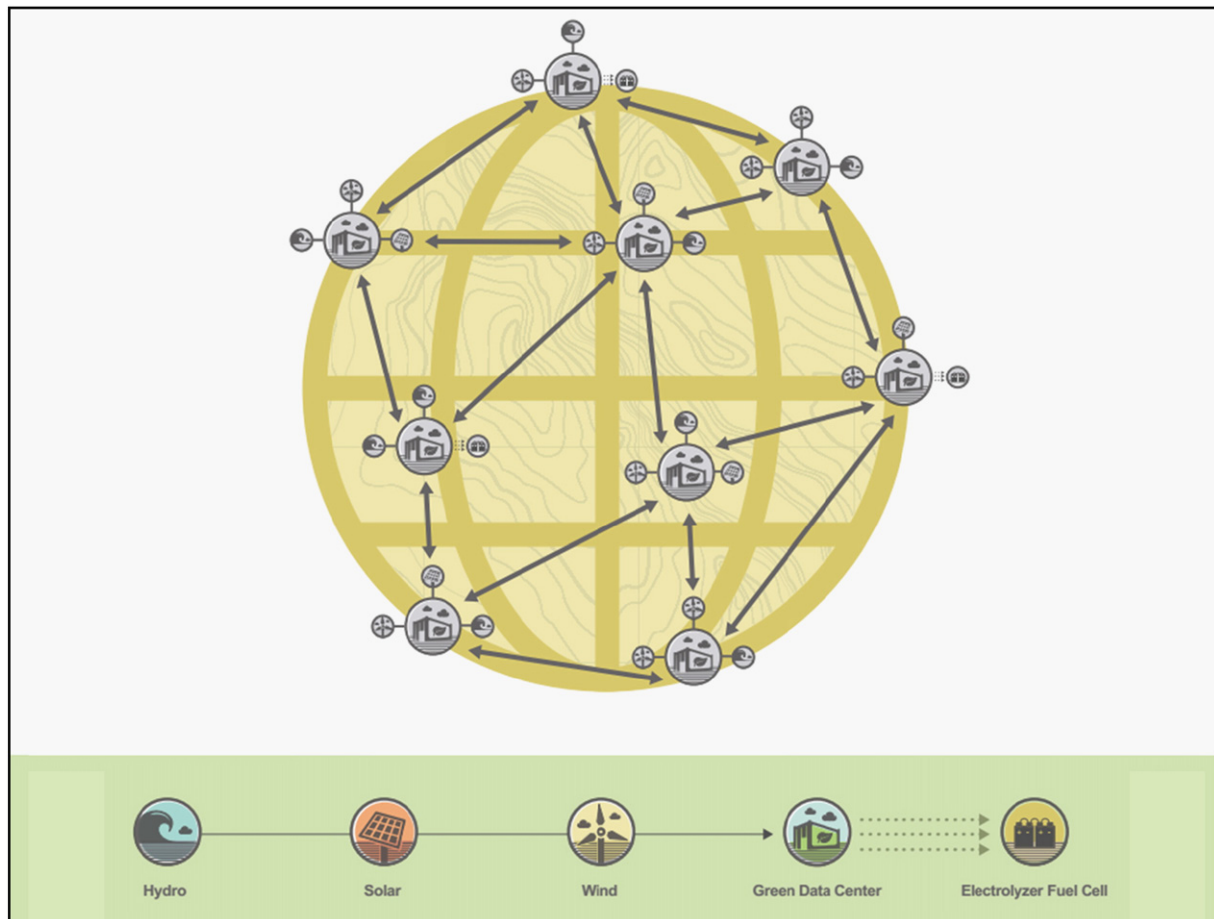


**Fig. 1.** Proposed system scheme. Distributed data management and storage resources are powered by locally available renewable and alternative energy sources. A data “cloud” environment allows running applications continuously even when locally generated power varies. Image© AMD, 2013; used with permission.

affordable power, while also more efficiently using electricity that is intermittent.

Currently, two other projects are underway that have similarities to the DGDC model: a small project at University of Cambridge described as “Free Lunch” is just beginning (Akoush et al., 2011); and the GreenStar Network in Canada (GÉANT, 2012; GreenStar Network, n.d.).

The GreenStar project has demonstrated the underlying feasibility of transferring a computing load for the purpose of optimizing the use of renewable energy resources. There are several key differences between the DGDC approach and the GreenStar Network. The distributed green data center (DGDC) model addresses the potential use of off-grid locations, with significantly more flexibility, granularity, and proposed distribution. It also involves smaller computing sites with the need for higher redundancy and coordination of computing workloads and more complex integration and dispatch modeling and coordination of power supply and server needs.

## Rationale

### Energy versus computing: Changing paradigms

Energy and computing contrast each other considerably in that they have divergent trends in cost structure. The cost of computing and computing storage has generally become significantly more efficient and less costly over a relatively short period of time. Computing storage costs have actually reduced exponentially. Alternately, energy prices, despite the introduction of efficiencies over the past 20 years, have generally increased in price, and all indicators point to increased speed of increases and higher levels of inflation in the future (this is despite a current price “flattening” from 2009 to 2012 because of the world recession and low prices associated with shale gas extraction). Fig. 2 demonstrates the contrasting trends in these prices.

As of 2010, energy costs are at least 12% of the average data center, and are the fastest rising cost of data centers (Gartner Research, 2010). Ad hoc discussions with industry indicate that these numbers are likely closer to 20% with some specific examples of data centers with energy costs over 80%. Future economic and policy analyses will address the nature and distribution of these costs in greater detail.

Currently, conservative estimation puts that at least 5% of electricity in the United States flows through information and communication technologies (including those at communications network sites used to transmit data instead of processing or storing data as occurs in data centers), and this amount is expected to continue to grow (Baer et al., 2002). As of 2005, data centers accounted for 1% of worldwide

electricity use (including cooling and distribution), and had doubled from 0.5% in 2000 in only five years (Koomey, 2008). In 2006, data centers accounted for 1.6% of all the U.S. electricity use, and these percentages are all likely higher now, given the trends (Masanet et al., 2011).

Environmentally, carbon emissions from data centers are expected to increase by more than 11% per year to 340 metric megatons by 2020 (Harmon and Auseklis, 2009). Further, “IT manufacture and use is responsible for 2% of global carbon emissions – the same amount as the airline industry – and is heading for 3% by 2020, when it will be responsible for the same amount of carbon as the United Kingdom produced in 2008” (Cubitt et al., 2011, p. 154).

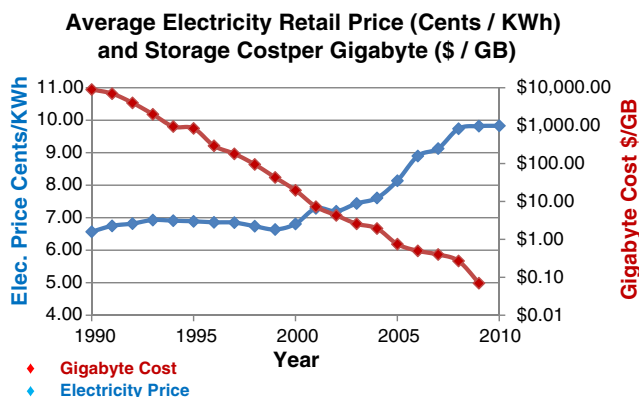
Even when there is data center duplication (because renewable resources go offline), there are still reductions in the overall carbon footprint of the system, depending on the circumstances and configuration of the network (Van Heddeghem et al., 2012). What is particularly interesting about this analysis is that it does not account for the reduced energy and carbon savings due to the lack of transmission infrastructure that can be accounted for in the DGDC approach. Other research associated with the GreenStar Network provided a protocol for assessing carbon reductions in the overall network that will be used for the future analysis of specific DGDC scenarios and configurations (CSA, 2012; Steenhof et al., 2012).

### Fiber versus electricity: Siting and cost

Electricity transmission is a major contributor to electricity costs, and a critical problem for the grid in general (Cotton and Devine-Wright, 2010; Sowinski, 2011; Woolf, 2003). First, siting processes can take very long periods of time. For instance, the 270 mile long Arrowhead line from Minnesota to Wisconsin ultimately took over 30 years to site and build. The average siting and construction timelines for large scale transmission lines are 7–9 years (Woolf, 2003).

Alternately, siting time for fiber deployments depends on the type of installation, the area that the installation is occurring, and the entities that will be approving the installation. Nonetheless, for a small build, the process is relatively simple and approvals can be obtained in 3–6 months, for a larger build, 6 months–1 year. If an environmental impact statement is required then 1–2 years is more likely. Construction times can also vary but are still significantly less than electricity transmission. These can occur in months for rural installs or along existing rights of way, or in 1–2 years for new installations, or complicated urban installations. The average siting and construction period will likely be from 1 to 3 years, with variations possible (Tuttle, 2012).

Second, costs of transmission are extremely high, generally averaging \$1.5 to \$2.0 million per mile (see Fig. 3). Alternately, siting fiber optic cable for computing is considerably cheaper, with estimates typically at \$70–150K (Lehpamer, 2004) or \$70–127K per mile (Korotky, 2012).



Sources: EIA, Table 8. Retail Sales, Revenue, and Average Retail Price by Sector, 1990 through 2010 (Energy Information Agency, n.d.; Komorowski, n.d.).

Fig. 2. Electricity price and storage cost comparison. EIA, Table 8. Retail sales, revenue, and average retail price by sector, 1990 through 2010; (Energy Information Agency, 2010; Komorowski, n.d.).

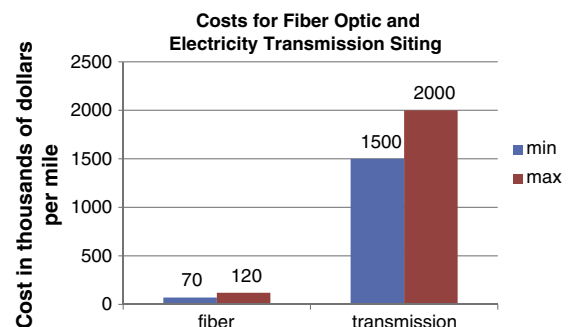


Fig. 3. Fiber and transmission siting cost comparison. Fox-Penner et al. (2008), Korotky (2012), and Lehpamer (2004).

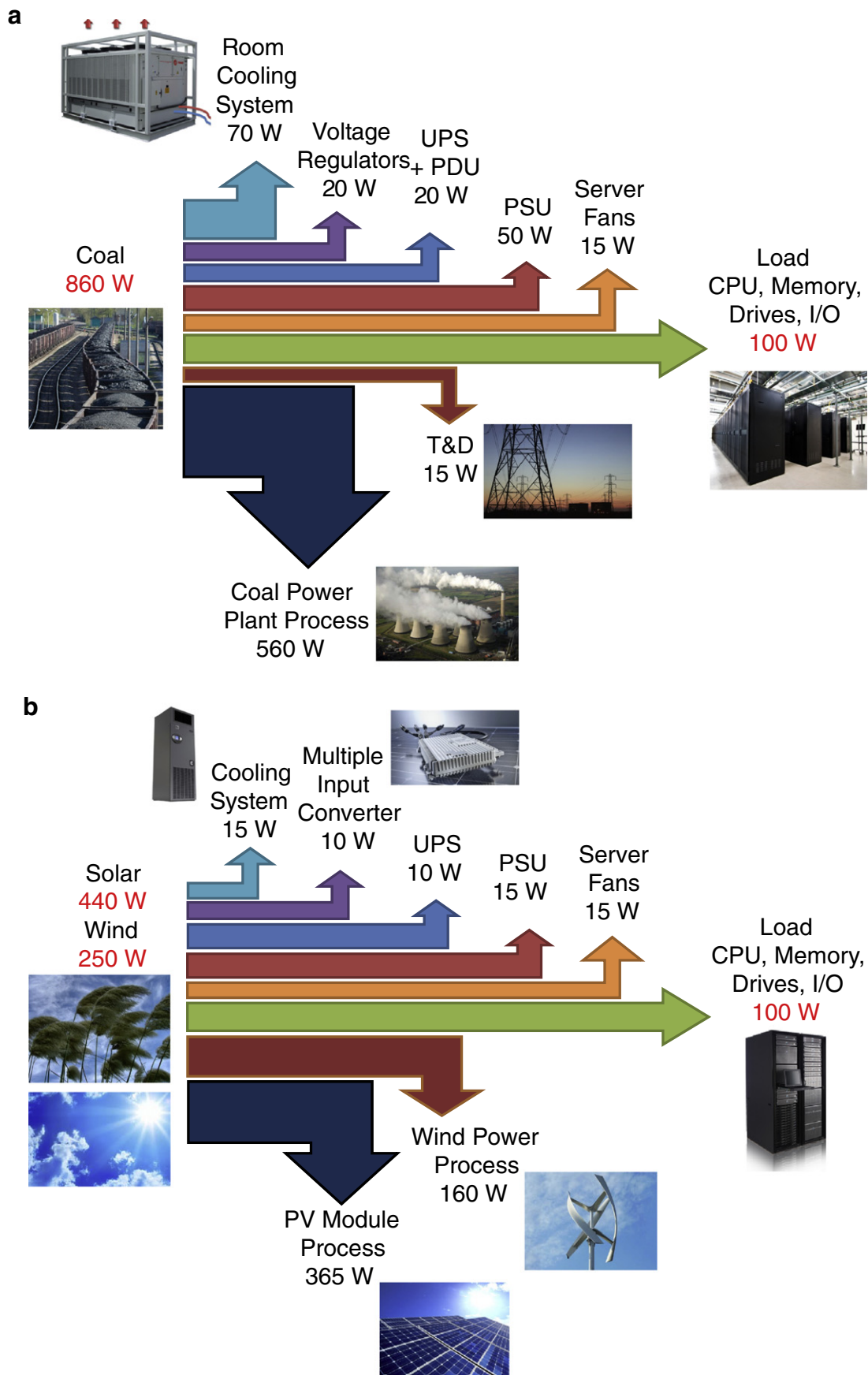


**Fig. 4.** Conventional versus distributed data center electricity utilization; power and efficiency losses. (a) Top: Energy utilization chain in a conventional data center approach; 860 W of equivalent coal power is required for a 100 W server load. (b) Bottom: Proposed approach based on distributed mini data centers; a combined 675 W of equivalent power from renewable sources is needed for the same 100 W server load.

The siting process for fiber optic networks is also considerably easier. First, large percentages of fiber optic siting are installed using extant rights of way (highways, train, electricity distribution, cable, or other pre-existing networks). Second, even if a fiber installation requires an entirely new siting and concurrent environmental impact statement (EIS), its footprint is considerably smaller, thus easing its impact on landowners, users, and the environment. The simple fact that fiber does not generate electrical fields (no electro-magnetic interference (EMI)) makes its presence in any given environment almost purely benign.

#### *Transmission line loss*

The movement of electricity to the computer can have many stages. A common scenario starts with the generation of an AC source. The AC is transformed by stepping-up its voltage, transmitted over a long distance, stepped down to usable voltages, and locally distributed to the end users. The end user then inverts the power back to DC and then back to AC again within an uninterruptable power supply and sends it on to the computer. Since computer chips require DC power, the computer's power supply converts the signal back to DC one last time. As Fig. 4 demonstrates, every step in this process steals 3–10% of the available power, causing significant losses and additional hardware expense at every stage. When inefficiencies on both sides of the electrical wall socket are taken into consideration, roughly 2 out of every 3 W of the computational power are spent outside the computer chips. When taking into account cooling needs, only 12% of the energy in the coal at a coal fired power plant is used in the final computer processing.

Alternately, the concept based on a modular distributed mini data center is relatively straightforward. Fig. 4.b shows the installation of rack servers into robust, modular, transportable cabinets or containers that can be placed next to the distributed electricity generation site. The container can be sealed to permit outdoor installation. It is completely self-contained with power input, network connections and, potentially, a connection for cooling and/or waste heat recovery. Depending on the site planning and container design, external cooling may not be necessary. The technical design team is examining a variety of designs and schematics that are expected to significantly reduce cooling needs and costs. Component failure within the container is tolerated until overall mini data center performance degrades to a predetermined level (a “fail-in-place” approach). When the failure threshold is reached, the entire container is replaced.

Thus, energy used at a source directly co-located with the data center has considerably fewer losses. Energy losses from transmission in the conventional approach are wholly useable unconverted renewable resources. If cooling is integrated into the mini data center model, its efficiency can potentially double that of the conventional approach. When cooling is also integrated into the model shown in Fig. 4.b, we expect that cooling can be as low as 20% of the mini data center's entire power load. Hence, cooling wasted energy is reduced by half compared to the conventional case (Fig. 4.a). The distributed green data center approach will reduce cost, energy consumption and negative environmental impacts. Such sites will rely on diverse forms of renewable energy sources to maximize computational up-time. A small proportion of mini data centers may remain connected to the grid to increase reliability in cases where transmission is extant. Small scale battery storage, high-tech flywheels, and/or ultra-capacitors will be used to smooth power delivery and extend the computational day of the site.

Overall, significant improvements are expected in necessary equivalent energy; in this case from renewable sources with no fuel costs and the added advantage that no greenhouse gases are generated and long transmission and distribution lines are avoided. In fact, most of the energy requirements do not translate in losses as in Fig. 4.a, but rather in unused renewable resources that are not converted in the PV modules or wind generators. Infrastructure investment is scaled down

as no bulk investments are required to transmit power over long distances. However, low efficiencies in renewable energy conversion at solar panels and in wind generators imply large footprints. These large footprints prevent extensive renewable energy use in medium to large size data centers, but the small distributed approach of the mini data center allows for much more extensive use of renewables.

We do not estimate specific energy or carbon reductions for the overall network here because the proposed DGDC networks will have extensive variability. They could be as small as four units, and as many as thousands, with many permeable variations. In current analytical work beyond the scope of this paper, we are analyzing energy and carbon use as applied to specific model variations.

#### *Security and challenges for conventional “mega-server” farms*

Current trends in the computing and telecommunications sector will exacerbate the challenges of the electric sector, and vice versa. Information technology will continue to demand more and more electricity. Electricity is lost as power is moved from wind and solar farms through an already congested and stressed electricity transmission grid. Centralized server farms will continue to grow in size, presenting extensive challenges for power, delivery, and cooling. Indeed, the possibility exists that micro-climates could emerge from the operation of large-scale data centers that will exacerbate the server farms' need to remove heat.

Security is also a concern for the deployment of large-scale conventional data-centers. The consolidation of all the data in a single location can provide a tempting target for disruption externally (infrastructure attack) or digitally (hack). A distributed model is useful from the security perspective because data has been separated, quite literally, and no single mini data center will have an extensive accumulation of potential data “riches” that conventional data-centers would contain, thus considerably reducing the potential cost of data disruption or theft at any one mini data center.

Further, the distributed model is specifically designed for outages (in some cases) or for a “disrupted” approach which migrates usage from one mini data center to the next as necessary (with data duplication). The specifics of this approach are discussed in the next section, but the critical concern is that the distributed mini data center approach is designed to migrate, or to be offline at certain times, and thus is much less vulnerable to concerns for high level loss of service that impacts an extensive swath of server function, and to the associated economic value and social functions inherent in data-center service.

#### *Regulatory environment and the impact of incentives*

For each site, a decision will need to be made about whether to rely exclusively on locally generated power in isolation or whether to invest in the infrastructure to sell power into the grid. In part this decision will be impacted by incentives, market policy, and the regulatory environment. Electrically-isolated sites may be exempt from some regulatory constraints such as NERC reliability requirements, but might also be ineligible for financial and monetary incentives that apply to grid-tied electricity. They will have lower infrastructure costs, but will need to create infrastructure for over-capacity mitigation.

Alternately, electricity sites that are grid-tied might have far more regulatory constraints and associated costs, but might gain benefits from financial and monetary incentives in electricity markets. The higher infrastructure costs might be offset by the ability to sell excess electricity to power markets when not being used for servers; or at times when the power price in electricity markets is more profitable than for server use. At times of peak demand in competitive markets, electricity prices can be exponentially higher than average, and so the ability to optimize profits by switching between computing and power generation would be advantageous.

### Grid-connected benefits: Ancillary services

When mini data centers are connected to a broader electricity grid, the flexibility of a data center's ability to come on and off-line easily will make it possible to use the power sources associated with mini data centers as dispatchable load. This capability will address some grid stability problems by giving grid operators an opportunity to shift computational load in order to meet electrical power, performance, and price specifications. These benefits and services, known as ancillary services, have specific value for grid operation and can provide another revenue stream for this model. Analysis has shown that ancillary service provision in high performance computing workloads (Aikema et al., 2011), and in traditional data center operations (Aikema, 2013; Aikema et al., 2012b) can provide up to 25–50% and 12% energy savings respectively. Our future analysis will model ancillary services in specific configurations of the DGDC model to determine the degree to which ancillary savings will be possible in grid-connected mini data centers.

### Technical concerns and research trajectory

#### Forecasting, unit commitment, and migration challenges

The greatest challenge for the distributed mini data center model is to determine the appropriate unit commitment algorithms for migration of computer service when renewable energy is unavailable at a given mini data center site. This can occur in a variety of circumstances or configurations, and is dependent in part on which particular context is in place for any given mini data center.

In general, this process involves forecasting data (potentially three different types) in various configurations, and then determining what kind of outcome is optimal depending on the information and the configuration via a unit commitment algorithm. The migration of computing load will occur (or not) depending on this output. The three different factors needed for forecasting are renewable energy forecasts (wind, hydro, or solar forecasts), computing load forecasts, and electricity market price forecasts (only in situations where a mini data center is connected to the grid).

Forecasting of renewable power availability has recently become a well-established and dependable process. A variety of forecasting companies now operate in electricity markets to provide specific geographic, climatic, and experiential models that provide 24 h, 1 h, and 5 min forecasts of expected renewable energy electricity output. These forecasts are used by grid operators for reliability and integration, and are also used by renewable energy operators to determine their output, and their market behavior.

In the context of data center operation, forecasting computer "load" or job times, scheduling, and server usage is also a common process. Forecasting in this context involves determining job size, job time, and calculating algorithms for server operation dependent on operations of multiple jobs at any given time within a given quality of service (QoS) level. It is a common operational component of data center operation. There is an extensive variation in VM live migration needs, and also performance variation across systems types. The research team has already published work examining some of the variations that will dictate constraints and provide guidance for the emergence of the DGDC model (Hu et al., 2013).

Finally, the potential market position in terms of market electricity prices for mini data centers will need to be understood. Forecasting for electricity market prices is also relatively common, and most competitive electricity markets provide expected day-ahead prices. This aspect of operation will only be relevant for a very small minority of mini data centers that satisfy three conditions. First, they are part of the small proportion of a distributed green data center network that are tied to the electricity grid; second, they operate in competitive regional electricity markets; and third, only during times of very high electricity prices (e.g. during an extensive heat wave). In this context,

it may be advantageous from a fiscal perspective to sell the renewable power to the grid market, and operate another mini data center in a cheaper market context.

The proportion of mini data centers that are connected to the power grid will vary considerably depending on the amount of mini data centers overall, the particular market niche that they supply, the profile of power supply variation, and whether or not the renewable facilities are newly developed or legacy.

Unit commitment algorithms will determine the availability of any given mini data center in the distributed green data center network as data center job requests are routed to the data center provider. Forecasting – primarily of the renewable energy availability and the current and upcoming server jobs – will create the data used for the predictive modeling. The development of integrated forecasting along with nuanced unit commitment algorithms is arguably the linchpin of the distributed green data center model. Substantial improvements have been made in renewable forecasting over the last five years (Monteiro et al., 2009; Potter et al., 2009). In combination with AWS Truepower, a renewable energy forecasting firm, the team has been incorporating algorithms for renewable energy supply into the operational models.

Migration of computing load has several technical solutions including data replication, load-balancing or fail-over and virtualization techniques such as VM migration. While the ability to migrate data and computation is well-established, specific aspects of migration over a wide area, especially in the presence of regular data center failure, are not well developed and these variations are under active analysis by the research group to determine specific opportunities and appropriate approaches for the mini data center approach. Further, characterizing the power usage of migration is also an important aspect of modeling overall power usage in the model (Aikema et al., 2012a). Different solutions for unit commitment and strategies for the operation of the DGDC network are possible.

The determination of different migration and workload scenarios is constrained by the Consistency Availability Partition (CAP) Theorem, also known as Brewer's Conjecture (Brewer, 2000; Gilbert and Lynch, 2002). This principle, shown in Fig. 5, demonstrates the impossibility of achieving strong consistency, high availability, and tolerance of network partition simultaneously in the general case. Gilbert and Lynch also discuss various strategies for addressing tradeoffs and the various applications that can be applied in different situations or needs (Gilbert and Lynch, 2002). All website systems, even large, highly duplicated systems like Amazon or Microsoft, face tradeoffs between availability and consistency and techniques to offer a range of different consistency guarantees to application developers are an active area of research. For instance, when data center services such as Google or Amazon have hardware failure, they compromise by providing weak consistency but high availability to achieve a user-acceptable service. No approach can handle data center failure seamlessly. This issue has been discussed by Browne (2009) (and others) extensively. The DGDC network is an especially good match for applications with write once (or even write rarely)/read many data access patterns. We are also investigating a wide range of strategies to make DGDC an attractive platform for a wider range of applications.

A wide variety of different applications are appropriate for the DGDC approach. In part this is due to variation in the possible configuration of a mini data center network. We show just four of the simplest possible variations below.

#### DGDC configuration options

1. Renewable only – single mini data center
2. Renewable only – small mini data center network
3. Renewable only – small-medium mini data center network w/out grid connection but extensive data duplication
4. Medium – large network with multiple mini data centers; some w/ electricity grid connection

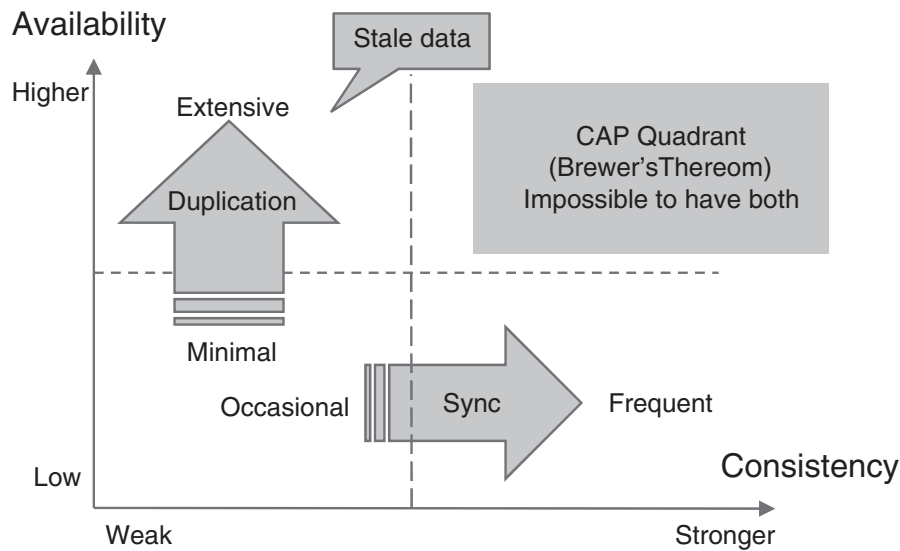


Fig. 5. Availability versus consistency (CAP Theorem).

Even under conditions of intermittency and high distribution, a variety of web services and applications have serious potential. These are generally simpler, less demanding, and non-mission-critical server functions that can be hosted in DGDC (which emphasizes availability over consistency). For instance, read-only data applications that involve archived data such as bank statements, past weather data, news or email archives, or even Netflix movie archives (with some duplication) are all appropriate for this approach. Stateless or Scientific Computing such as that used by PlanetQuest or SETI@home is also possible. Market analysis for this project will analyze these issues in much greater detail going forward. Further, a non-migration approach may be possible in some configurations for live constantly changing data needing consistency. While data redundancy would be minimal, configurations with some grid linkup, multiple energy sources, or battery storage availability could make these sorts of applications possible.

#### State of current research

In order to address the development needs involved in developing the distributed green mini data center concept, three research domains are considered in this project (see Fig. 6). A foundational layer with

focus on physical—i.e., hardware—problems, such as those related to power and cooling infrastructures, a middle layer targeting data management—i.e., software—issues, such as handover control among different sites running the same application, and a top layer addressing system environmental problems, such as evaluating the local economic impact of the proposed approach.

An initial deployment for proof of concept is being implemented with current off-the-shelf components. A scalable and reliable solution requires the deployment of an end-to-end infrastructure, including management software and will likely be implemented in the next 1–2 years (2013–14). The current research will be focused on a systems-level, integrated thermo-economic analysis that critically examines the economic, energetic and environmental balances and tradeoffs of the conventional approach versus the new approach. The research team is also focused on the identification of technological, economic, and policy barriers to implementation that can be used to specify research directions and site parameters.

A key concern is for hardware and software, including management software that provides failover capabilities among containers located throughout the grid. Initial work on specialized components that enable sealed, ultra-dense, highly reliable computation containers which will

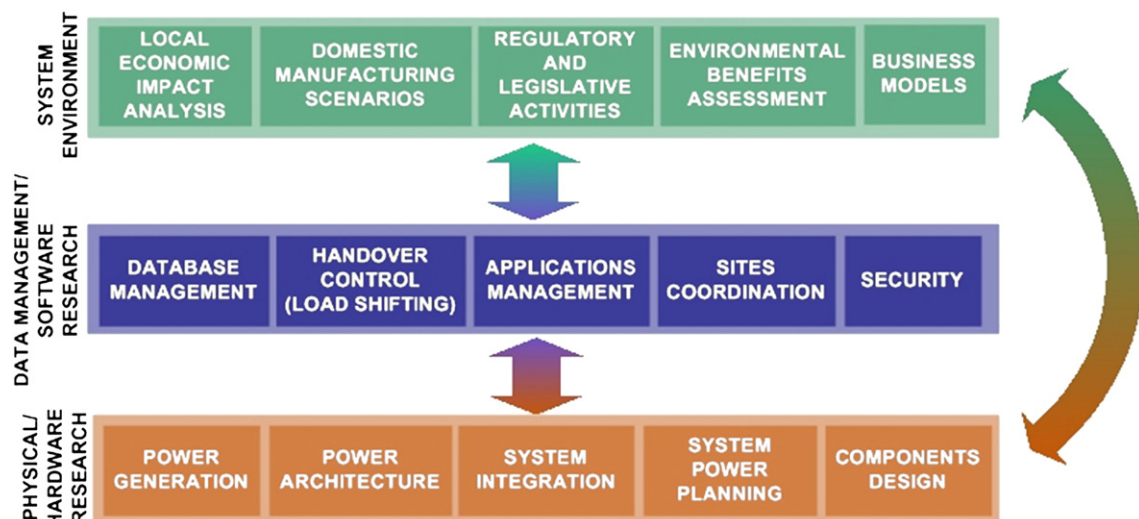


Fig. 6. Proposed research scheme.

hold advanced power infrastructure coupled with environmental and/or cooling technologies. These will integrate with ultra-light client devices that allow anyone with credentials to access the computing resources on the grid. Critical to all of this will be the design, operation, and stabilization of power electronic interfaces and an overall modular, optimized, and integrated power chain.

Finally, software will have to incorporate models to accurately predict power and energy output from renewable or other power sources. For example with wind power, two types of forecasting may be needed: very short-term forecasts (from milliseconds up to a few minutes) for the turbine active control; and longer-term (48–72 h) forecasts for power system management or energy trading.

## Societal benefits

### *Enabling renewable energy by addressing intermittency*

Many renewable energy sources, such as windmills and solar panels, do not generate continuous and reliable electricity. However a push for domestic, low-carbon, and sustainable sources of power, coupled with a surging demand for electricity, partly from the growing computing infrastructure, combines to create a rapidly growing market for these low-carbon sources of power (IEA, 2012). In parallel, a server-type computation (e.g. for on-line storage and data gathering devices) is undergoing similarly rapid growth rates, with commensurate demand for electricity. With the exception of large scale hydro-electric power, it is difficult to integrate renewable sources at the site of conventional, large-scale data centers. This is because renewable sources need an extremely large footprint to power a sizable portion of a data center's load (the density of demand for power at server farms is several times higher than photovoltaic and wind power generation density). Further, there are limitations in energy storage that can power a load for several hours at the scale of a large server farm, with affordable and scalable solutions still in the future.

Renewable power sources, including wind and solar, have been difficult to integrate into the power grid for many reasons including their intermittent production of power, the separation between the best energy creation locations and the existing electrical grid infrastructure and/or population centers, inherent losses in electric power transmission and distribution, power transmission constraints, prohibitive infrastructure costs, environmental impacts of transmission systems, and barriers to infrastructure development (permitting, siting, and regulatory constraints).

### *Congestion*

Transmission congestion in the United States is a serious issue. For instance, in New York State, most of the electricity demand is in New York City but much of the electricity production (especially for renewable energy) occurs via hydro and wind sources located in the north and north-west portions of the state. The problem is that the transmission lines delivering that power to New York City are congested, and limit the amount of power that can be delivered. The New York ISO assessed the expected congestion revenue losses at \$ 407 million for 2011, and in 2009, 1.2% of New York's wind power was not deliverable. In 2018, the NYISO predicts that amount to be 9% (NYISO, 2010). Similar issues with congestion occur in Texas, California, New England, and the Mid-Atlantic.

One can imagine a scenario in which many distributed green data centers are co-located close to renewable power in northern New York, connected by a network of fiber optic cables. There is no need to deliver power to server centers located in New York City. Congestion is avoided, and power is produced with much less efficiency loss. This novel approach changes the paradigm: Data centers and renewables will no longer exacerbate grid problems because of their inherent variability, but instead use their integration and co-location as a tool for solving their inherent problems.

### *Potential for the developing world*

The distributed mini data center concept has a particular promise for the developing world. Innovations in fast and affordable communication of information have repeatedly triggered episodes of positive, social transformation – starting with the telegraph in the late 1800 s and continuing with modern handheld mobile devices today. As the marginal cost for access to information declines, it becomes affordable to a larger percentage of the world's inhabitants, bringing economic opportunity even in areas where the communications infrastructure was previously inadequate or nonexistent.

Access is important because of the Earth's approximately 7 billion plus population, there are over 1 billion people without electricity or telephones, and almost 4.7 billion people without computers and the Internet (Dutta et al., 2011; Sharma, 2012). Obviously energy, computation, and telecommunications are all considered to be key enablers for a higher quality of life, demand for the generation and transfer of both electricity and information is growing rapidly. The novel concept proposed here enables widespread access to information and distributed computing as an affordable engine for economic growth by reducing its costs and overcoming energy constraints.

### *Potential economic and societal benefits*

Eventually, a variety of benefits for industry and society can be anticipated from the implementation of this model at scale. Economic benefits include the reduction in cost of laying transmission lines because of the substitution of optical fibers. Further the costs of line loss and other electricity losses are enormously reduced. Rural development and job creation prospects are improved because many renewable resources are located in rural areas. Growth in renewable power plants and distributed computing centers will lead to the creation of jobs in rural areas.

Environmental protection improves because this approach employs greater and more efficient use of renewable energy. It reduces energy consumption via more efficient computing systems and a more efficient computational model, while concurrently reducing the footprint of the computational infrastructure and its associated electricity supply component (fiber optic transmission has less environmental impact than electric transmission). It also allows for the creation of much smaller energy production fields. For example, a wind farm with only four turbines might become economically viable because of the dual-use with computation. There is the potential for security improvements because the resiliency and availability of a highly distributed network to extreme events such as natural disaster and/or intentional attacks is much higher.

In addition, there will be important economic benefits for existing and emerging companies to develop new equipment and service businesses. Broadly, these include the integration of renewable energy sources, telecommunications and computation and would include innovative storage technologies, power generation implementations, power conversion interfaces, power systems modeling (AC vs. DC), and thermo-economic models for effective deployment. We would also expect needs for improved semiconductor design for multiple generations of computing and communication components and client processors to enable dense computing. Finally, this approach would create incentives for improvements in management software, including performance prediction modeling and simulation.

## Conclusion

The distributed green data center model provides an entirely new approach to both the implementation of data center service, and also the use of renewable energy sites. This model improves efficiency and facilitation of renewable energy output, and has the potential to reduce costs and time delays in electricity transmission siting (and associated

congestion costs). It addresses the concerns for intermittency, and in a grid-connected context may have additional ancillary service solutions. On the computer services side, it improves data center security concerns by distributing data across multiple sites and has an application for a wide range of server workloads that require less consistency. The model also has the potential for developing nations with a minimal electricity grid infrastructure. Ongoing research will continue to address the development of efficient unit-commitment algorithms and the cost-effectiveness of the overall business model.

## References

- Aikema D. *Optimizing Data Centre Energy and Environmental Costs* [Thesis] University of Calgary; 2013.
- Aikema D, Kiddle C, Simmonds R. Energy-cost-aware scheduling of HPC workloads. World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a. Presented at the World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a; 2011. p. 1–7. <http://dx.doi.org/10.1109/WoWMoM.2011.5986476>.
- Aikema D, Mirtchovski A, Kiddle C, Simmonds R. Green cloud VM migration: power use analysis. In: Green Computing Conference (IGCC), 2012 International. Presented at the Green Computing Conference (IGCC), 2012 International; 2012a. p. 1–6.
- Aikema D, Simmonds R, Zareipour H. Data centres in the ancillary services market. In: Green Computing Conference (IGCC), 2012 International. Presented at the Green Computing Conference (IGCC), 2012 International; 2012b. p. 1–10.
- Akoush S, Sohan R, Rice A, Moore AW, Hopper A. Free lunch: exploiting renewable energy for computing. Presented at the 13th Workshop on Hot Topics in Operating Systems (HOTOS'XIII); 2011.
- Baer WS, Hassell S, Vollaard BA. Electricity Requirements for a Digital Society. RAND Corporation; 2002.
- Brewer EA. Towards robust distributed systems (Abstract). Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing, PODC'00. New York, NY, USA: ACM; 2000. p. 7.
- Browne J. Brewer's CAP Theorem: the kool aid Amazon and Ebay have been drinking [WWW Document]. <http://www.julianbrowne.com/article/viewer/browsers-cap-theorem>, 2009. [(accessed 3.1.13)].
- Clancy H. Green Data Center Market Seen Doubling by 2016 [WWW Document]. ZDNet. <http://www.zdnet.com/green-data-center-market-seen-doubling-by-2016-7000004324/>, 2012. [(accessed 9.20.12)].
- Cotton M, Devine-Wright P. NIMBYism and community consultation in electricity transmission network planning. Renewable Energy and the Public: From NIMBY to Participation 2010;115–130.
- CSA. ICT greenhouse gas reduction project protocol: quantification and reporting – Version 1 [WWW Document]. ShopCSA. <http://shop.csa.ca/en/canada/climate-change/ict-protocol-version-1/invt/27033222012>, 2012. (accessed 12.5.13).
- Cubitt S, Hassan R, Volkmer I. Does cloud computing have a silver lining? Media Culture Society 2011;33:149–58.
- Dutta S, Dutton WH, Law G. The new Internet world: a global perspective on freedom of expression, privacy, trust and security online. World Economic Forum; 2011.
- Energy Information Agency. Table 8. Retail sales, revenue, and average retail price by sector, 1990 through 2010, state electricity profiles; 2010.
- Fox-Penner PS, Chubka MW, Earle RL. Transforming America's power industry: the investment challenge. Report Forthcoming for The Edison Electric Foundation; 2008.
- Gartner Research. Energy-related Costs Account for Approximately 12 Percent of Overall Data Center Expenditures [WWW Document]. <http://www.gartner.com/it/page.jsp?id=1442113>, 2010. [(accessed 9.12.12)].
- GEANT. GEANT and GreenStar Network: Powering Networks Through Renewable Energy; 2012.
- Gilbert S, Lynch N. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. ACM SIGACT News 2002;33:51–9.
- GreenStar Network [WWW Document], n.d. URL <http://www.greenstarnetwork.com/> (accessed 9.14.12).
- Harmon RR, Auseklis N. Sustainable IT services: assessing the impact of green computing practices. Management of Engineering & Technology, 2009. PICMET 2009. IEEE; 2009. p. 1707–17.
- Hu W, Hicks A, Zhang L, Dow EM, Soni V, Jiang H, et al. A quantitative study of virtual machine live migration. Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference, CAC'13. New York, NY, USA: ACM; 2013. p. 11:1–11:10.
- IEA/Medium Term Renewable Energy Market Report 2012. International Energy Agency (IEA); 2012.
- Komorowski, M., n.d. A History of Storage Cost [WWW Document]. URL <http://www.mkomo.com/cost-per-gigabyte> (accessed 9.12.12).
- Koomey JG. Worldwide electricity used in data centers. Environ Res Lett 2008;3: 034008.
- Korotky SK. Price-points for components of multi-core fiber communication systems in backbone optical networks. IEEE/OSA J Opt Commun Networking 2012;4: 426–35.
- Lehpamer H. Microwave Transmission Networks: Planning, Design, and Deployment. McGraw-Hill Professional; 2004.
- Masanet ER, Brown RE, Shehabi A, Koomey JG, Nordman B. Estimating the energy use and efficiency potential of U.S. data centers. Proc IEEE 2011;99:1440–53.
- Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. Wind Power Forecasting: State-of-the-art 2009. Argonne National Laboratory (ANL); 2009.
- NYISO. NYISO 2010 Wind Generation Study; 2010.
- Potter CW, Archambault A, Westrick K. Building a smarter smart grid through better renewable energy information. Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES; 2009. p. 1–5.
- Sharma C. Global Mobile Market Update [WWW Document]. <http://www.chetanisharma.com/GlobalMobileMarketUpdate2012.htm>, 2012. [(accessed 9.12.12)].
- Sowinski M. Practical, legal, and economic barriers to optimization in energy transmission and distribution. J. Land Use & Envtl. Law 2011;26:503.
- Steenhof P, Weber C, Brooks M, Spence J, Robinson R, Simmonds R, et al. A protocol for quantifying the carbon reductions achieved through the provision of low or zero carbon ICT services. Sustainable Computing: Informatics and Systems 2012;2:23–32.
- Tuttle C. Personal Communication with Carrie Tuttle, Director of Engineering. Development Authority of the North Country; 2012.
- Van Heddeghem W, Vereecken W, Colle D, Pickavet M, Demeester P. Distributed computing for carbon footprint reduction by exploiting low-footprint energy availability. Futur Gener Comput Syst 2012;28:405–14.
- Woolf F. Global Transmission Expansion: Recipes for Success. Tulsa, OK: PennWell; 2003.