[Mahnoosh Alizadeh, Xiao Li, Zhifang Wang, Anna Scaglione, and Ronald Melton]

Demand-Side Management in the Smart Grid

[Information processing for the power switch]

ISTOCKPHOTO.COM© SIGAL SUHLER MORAN

Technical Challenges of the Smart Grid

Over the course of several decades after their introduction, power systems merged into large interconnected grids to introduce redundancy and to leverage on a wider pool of generation resources and reserves. As the system grew in size and complexity, a cyberphysical infrastructure was progressively developed to manage it. Traditionally, general-purpose computing and communication resources have been used in power systems, specifically to serve two needs: 1) that of monitoring the safe operation of the grid and logistics of power delivery, and 2) that of gathering information required to dispatch the generation optimally and, later on, to operate the energy market.

Enhanced sensory measurement systems under development today can help utilize the existing transmission infrastructure more efficiently, thereby reducing the current wide margins of operations. This is clearly of great interest to the present stakeholders: the utilities, who may spend less in upgrading infrastructures as well as the generation providers, whose competitive prices are affected by congestion costs. However, investing in fewer transmission lines and transformers is not going to produce a fundamental shift in the portfolio of energy resources used to meet the demand of electricity. Hence, ultimately, advances in measurement systems will just help contain costs, without paving the way for greener energy use.

A GREEN FUTURE FOR THE GRID

Wind and solar power cannot be easily thrown into the mix of generation resources due to their limited dispatchability and intermittent nature when compared to fuel combustion. Two essential factors make delivering green electrons especially complex: 1) the need for the demand and generation to be continuously balanced with limited energy reserve and storage; 2) the fact that in the retail energy market the tariffs are fixed, and the demand of energy is not responsive to grid conditions or, in other words, is treated as inelastic. This state of things favors fossil fuel generation, which can be dispatched at will to follow the demand.

Digital Object Identifier 10.1109/MSP.2012.2192951
Date of publication: 20 August 2012

Fossil fuels not only support the major part of our electricity generation needs but also provide energy for our transportation systems. This may change in the near future, as electric vehicles (EVs) enter the market. EVs have higher energy conversion efficiency, produce little or no CO_2 emissions, and allow, in principle, to reduce the dependency on foreign oil imports. While this is an enticing future to look forward to, the main technical concern is whether the grid can support a widespread adoption of EVs, as these vehicles will act as a great number of high-wattage, long-duration loads in the system [1]. Even with full observability of this volatile load, there is the increasing consensus that EV demand should be managed differently, or else it will require increasing energy generation, reserves, and substantial spending on distribution infrastructure, with adverse effects on energy bills. Fortunately, unlike many appliances that are already manufactured on a commercial scale today, EVs are still in their introductory stage and do not have many fixed standards. In the past few years, many researchers have focused on studying possible ways of predicting and smartly managing the battery charge of EVs. The mobility of EVs adds another level of complexity to these management techniques but offers new research opportunities as well. For example, some researchers are working on helping mobile customers find appropriate charging spots or help the grid forecast demand via traffic information posted online.

WHAT MAKES THE TIME RIPE NOW FOR INTRODUCING DSM MORE BROADLY IS THE MATURITY OF IT AND THE RISING COST OF MEETING THE DEMAND OF ELECTRICITY.

WHY GREEN ENERGY REQUIRES INFORMATION PROCESSING

There are two opposite approaches to integrate more intermittent resources into wholesale energy markets while ensuring that the generation and demand balance holds at all times: 1) generation reserves and 2) demand-side management (DSM) and demand response (DR).

The first approach requires volatile and nondispatchable plants like wind and solar to be backed up by clean and controllable resources, like hydro or natural gas units, which can start up whenever the intermittent resources fail to meet their scheduled generation dispatch, or by large scale storage. A serious challenge is to be able to find reserves with rapid ramping rates to overcome frequency variations that threaten the system stability. Hence, the main problem in this approach lies in the limited availability of clean and dispatchable plants and the high cost of storage. One view is that more of these resources can be provided by merging independently operating markets of electricity to cover larger areas (like the common markets in Europe), thus providing larger-scale balancing options.

In the second approach, the volatility introduced by intermittent resources on the generation side will be compensated for by a more responsive and controllable consumption behavior. DSM and DR programs, among other options, can make market rules for generators more flexible so that more renewables can participate in the wholesale market.

The first approach is more traditional while the second one may prevail in this century as automated solutions for the EV charging problem advance and become increasingly ubiquitous. In fact, it is very likely that other appliances will follow suit, and provide, as an aggregate, the large-scale inertia needed to cope with volatile generation. Specifically, applications and interfaces to thermostatically controlled (TC) appliances that already are in the DSM domain are likely to spark new inventions, software, and protocols for their control. These protocols will fuse specifications that are at the intersection of power, communication, signal processing, computation, and control.

In fact, load management through DSM and DR can be realized only if electrical appliances and systems become responsive, which in turn makes the inclusion of communication and embedded intelligence in electrical devices essential. Ubiquitous network connectivity and embedded processing is becoming commonplace; in fact, several sensor networking communication protocols were developed in the past decade, promising to give machines a voice. We are also witnessing an increasing number of complex human transactions that are carried out in real time over wide areas. Can inelastic end-use loads become responsive to the cost and impact of the energy they get? In other words, can cheap bits and flops make the demand become elastic, so that greener and cheaper watts can flow in the electric grid?

The ideas of DSM and of DR are not new. DSM and DR systems emerged in the 1970s and have evolved over the past three decades through systematic activities of researchers, power utilities, and government policies, designed to change the amount and the timing of electricity consumption. Such DSM measures have been implemented for load management, enhancing energy efficiency, and electrification (i.e., the strategic increase of electricity use).

The application of DSM and DR today is still quite limited in scope, and it is mostly applied to large industrial and commercial customers. However, experiments like the GridWise Olympic Peninsula project [2] have shown that, intrinsically, households demand is far more elastic than previously thought. What makes the time ripe now for introducing DSM more broadly is the maturity of IT and the rising cost of meeting the demand of electricity. However, DSM techniques themselves require further research to be the cornerstone for the integration of more renewable energy.

The goal of this article is to first survey the evolution of DSM and DR approaches over the years and the current research trends. We also showcase a new paradigm for holding back demand, mimicking the effect of managing fuel reserves by queuing load requests, which we coin as digital direct load scheduling (DDLS). DDLS provides a first concrete mapping

between switching jobs and the physical shape of the load injection, which allows us to view the access to power as a network service model. Interestingly, quantization principles and filterbank reconstruction are the key ingredients of this mapping; in this sense DDLS is a clear example of how old signal processing tools can be used to create new concepts and technologies for optimizing the power switching of networked smart devices.

TWO-WAY COMMUNICATIONS AND EMBEDDED INTELLIGENCE ARE ESSENTIAL INGREDIENTS FOR DSM AND DR.

Efficient demand management techniques will be essential for the grid to integrate considerable amounts of renewables, however, they are only one side of the coin. The other side, which is equally important, is the development of architectures that accrue information about the availability of these resources and manage their physical and financial interaction with the generation and transmission assets. These two sides cannot be treated separately. Motivated by the need to define scalable information processing architectures through which these two types of resources (renewable energy and volatile demand) can be appropriately managed, this article is going to cover not only the most recent developments in the area of demand management, but it will also consider possible applications of DSM techniques by entities that own distributed energy resources (DER), e.g., microgrids, aggregators, and commercial and residential buildings.

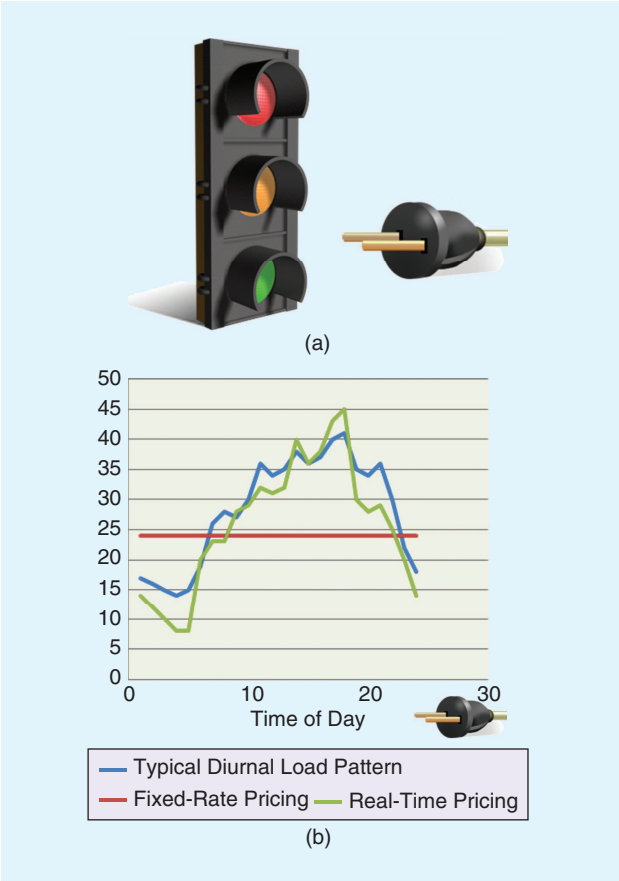
Finally, as we mentioned, two-way communications and embedded intelligence are essential ingredients for DSM and DR. The application layers that describe them should gain research prominence compared to other media applications, as our energy supply infrastructure will increasingly need to shift some level of control from the generation side to the demand side. The interface between the physical world and the application layers offer great opportunities for research in signal processing, as the analog world of the machine actions, and the service and satisfaction they provide to the end user, will need to be described in a digitized form, to be ingested by software that determines the optimum action to take, or predicts parameters to make the future decisions. The thread that links signal processing to DSM is quite clear in DDLS, but the DSM and DR ideas that we generally described along with the DER architectures can considerably benefit from efforts in digitally describing the end-use service DSM and DR need to deliver and the volatile resources available from DER. These are discussed as future research directions in the "Conclusions" section.

LOAD MANAGEMENT MODELS AND INFRASTRUCTURES

Initially, DSM programs mainly comprised reliability-driven load management measures, used occasionally to manage emergency situations. Over time, more sophisticated and rapid forms of DSM emerged, extending the level of consumer interaction for this services through appropriate incentives [3]. The most popular topic of research today is the inclusion of programs that enable direct price responsiveness, even of individual loads.

One of the main concerns in DSM and DR is the overall value of the service provided, which is largely dependent on the deployment cost, a lot of which is in the sensory and telemetry costs necessary to realize a responsive load behavior. Today, demonstrations and pilot programs are retrofitting systems that do not have the built-in intelligence to become a smart load. However, the vision of the Internet of things that emerged in the last decade of sensor networking may find in the smart grid its killer application.

In this section, we will discuss in detail the two most popular trends of DSM, which lie at the two opposite sides of the control spectrum, illustrated in Figure 1. At one end are dynamic and real-time pricing (RTP) strategies, discussed in the section "Price-Based Load Control Strategies," while at the other end are direct load control (DLC) strategies, presented in the section "Load Control Through Curtailment."



[FIG1] Contending DSM strategies: (a) DLC and (b) RTP. RTPs will have an average value similar to fixed rate tariffs but exhibit a diurnal pattern that is correlated with the daily load pattern and incentivizes load shifting from peak to off-peak periods.

DSM advances, and dynamic pricing in particular, are possible thanks to technological progress in data processing and communications technology. But the network and processing infrastructure for DSM is clearly best defined by knowing what the right DSM application is. As this article outlines, this is still a subject of heated debate and research. As we will see, DLC programs lack a sufficiently good interface to the consumer that can gauge the impact of interruptions on the overall service while RTP programs, although very appealing since consumers can define their own preferences, face the very challenging problem of determining what these price signals should be to have efficient and secure grid operation. One invention that can help strike the right balance between these two ends of the design spectrum is the digital direct load scheduling (DDLS) scheme we propose, for its ability to categorize and place the jobs in queues, which can be served using basic quality of service (QoS) requirements.

DSM ADVANCES, AND DYNAMIC PRICING IN PARTICULAR, ARE POSSIBLE THANKS TO TECHNOLOGICAL PROGRESS IN DATA PROCESSING AND COMMUNICATIONS TECHNOLOGY.

PRICE-BASED LOAD CONTROL STRATEGIES

One of the most fundamental operational problems in power systems is keeping the produced and consumed power balanced at all times. In the bulk power system, multisettlement markets are used as an integral element of balancing the system—consisting of at least a day-ahead and a spot market. Today, electricity is traded in a central wholesale market through which utilities purchase electricity from generators and sell it to end-use customers, who are shielded from making short-term market decisions. There has been, in fact, a barrier between wholesale (bulk power) and retail. Retail electric service providers in most areas operate as natural monopolies and thus lack the basis for free market trading of electricity as an element of providing that commodity to their customers.

From time to time, several operational issues, such as lack of storage options, generator and transmission line failures, and the ignorance of customers about the real cost of the electricity they consume, all contribute to cause the wholesale price of electricity experience spikes. The purpose of price-based load control strategies is to introduce economic elasticity to consumption, with a side effect of transferring part of the risks of buying electricity through this volatile market from the utility company to the customers. However, as will be discussed next, dynamic pricing tariffs have the challenge of identifying a basis for further refining the dynamic prices from the wholesale level, to be supportive of the distribution utilities need to recover their costs, provide reliable service, and make a profit.

Different techniques, like time of use (TOU), critical peak pricing (CPP), or RTP all expose end-use customers to the risks of managing wholesale prices to different extents.

In TOU pricing strategies, the price is usually decided months or years before the actual TOU. There have been several studies on determining these rates, which requires a dynamic model for the price response of customers, based on experiments [4]. Due to their slow update rate, TOU rates do not require substantial communication with customers. One of the problems with TOU rates is that they are not dynamic, so they only enable a response to the gross diurnal differences in peak load and not to specific circumstances occurring in real time. Recently, more complex nonflat billing rates have been made possible through the deployment of the so-called advanced metering infrastructure (AMI), with smart meters in the homes. An example of this includes CPP programs, which use TOU rates except for the duration of a number of emergency and peak events.

One of the most serious contenders and a popular research topic in the DSM research arena is RTP. The concept of real-time prices has been around for about three decades now: instead of shielding the customer completely from the real fluctuations of energy costs in the spot market (as it is done when using flat rates and TOU tariffs), in RTP, price signals delivered to the customers will act as an economic incentive to modify their demand and alleviate the pressure on the grid, with the reward of lowering their bill. The key difference in the RTP model, as the name suggests, is that the price is updated and provided only hours/minutes before consumption, so that the signal can reflect actual grid congestion. Clearly, under RTP, the decisions on how to consume electricity should be mostly automated. There is, in fact, an extensive literature emerging on home energy management systems (HEMS) (e.g., [5] and [6]), which encompass the software and hardware at home that would respond to the price signal (see Figure 2). The idea is that the HEMS software will run a control program that optimally plans the use of appliances, based on their power consumption, price data, job deadlines, and customer preferences.

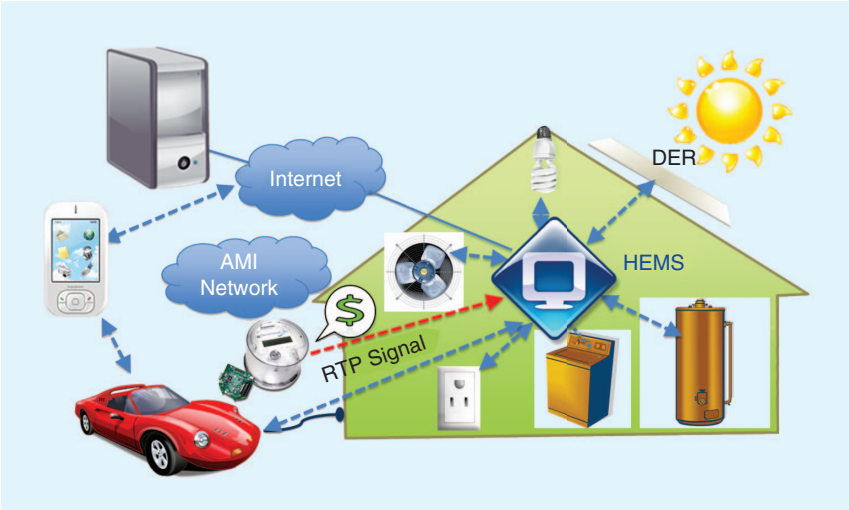
The HEMS control objective is, typically, to minimize the total cost of energy consumption by all the controllable appliances, given certain constraints on each appliance that describes job deadlines, required energy profile, and consumption limits. Generally, HEMS optimizations over a horizon T can be cast in the following form:

$$\min_{P_{n,t}} \sum_{t \in T} C_t \left(\sum_{n \in \mathcal{A}} p_{n,t} \right)$$
$$\text{s.t. } \forall n \in \mathcal{A}, \ \{P_{n,t}\}_{t \in T} \in \mathcal{L}_n, \quad t_n^f \le d_n, \tag{1}$$

where the cost function $C_t(.)$ describes the cost of consuming electricity at time t and is a function of the real-time prices, received from the price setting entity. Real-time prices after the announced period can be predicted using price predictor units by applying a weighted averaging filter to past prices [5]. $\mathcal{A}$ represents the set of appliances that can be managed. $P_{n,t}$ represents the power consumed by appliance n at time t.

$\mathcal{L}_n$ represents the set of all the acceptable consumption profiles of appliance n, and t_n^f represents the time at which appliance n finishes its job, required to be before its deadline, d_n.

What is not obvious is how the utility should calculate and post price signals to attain a stable operating point for the system, balancing generation and demand and avoiding transients that can cause grid instability. To discuss this issue, we first present a simplified model for how electricity is priced in today's market. The power grid, through which electricity is traded, is operated under several reliability and operational constraints. Thus, the price of electricity at any time instant is a function of the congestion status of this interconnected system. Theoretically, the price of electricity at each node in the power grid is the shadow price obtained from solving the optimization for the spot market security constrained generation dispatch [7], [8], which minimizes the cost that retailers pay to generators under operational constraints. More specifically, in a system with N generator and L load buses, each generating G_i or consuming D_i units of power respectively, and W transmission lines, the dispatch optimization in one of its simplest forms is given by



[FIG2] HEMS control based on RTP data received from smart meters, through the AMI network. The computer represents the local utility's demand control server.

$$
\begin{aligned}
G_i^* &= \arg\min_{G_i} \sum_{i=1}^{N} C_i(G_i) \\
&\text{s.t.} \sum_{i=1}^{N} G_i - \sum_{i=1}^{L} D_i - \text{Loss} = 0 \\
&|H[G_1 \ldots G_N\, D_1 \ldots D_L]^T| \le \bar{F}^{\max} \\
&G_i^{\min} \le G_i \le G_i^{\max}\,,
\end{aligned} \tag{2}
$$

where $C_i(\cdot)$ is the generation cost function and H is a matrix that relates power flow on transmission lines to nodal power inputs. The first constraint ensures power balance; the second ensures that flows on all the transmission lines lie within their limit, given by the vector $\bar{F}^{\max}$; and the third one defines generation capacity limits. The Lagrangian of (2) is

$$
\begin{aligned}
\mathcal{L} = &\sum_{i=1}^{N} C_i(G_i) - \theta\left(\sum_{i=1}^{N} G_i - \sum_{i=1}^{L} D_i - \text{Loss}\right) \\
&- \bar{\mu}^T (H[G_1 \ldots G_N\, D_1 \ldots D_L]^T - \bar{F}^{\max}) \\
&+ \sum_{i=1}^{N} [\nu_i^{\max}(G_i - G_i^{\max}) - \nu_i^{\min}(G_i - G_i^{\min})],
\end{aligned} \tag{3}
$$

where θ, $\bar{\mu} = [\mu_1, \ldots, \mu_W]^T$, $\nu_i^{\max}$, and $\nu_i^{\min}$ are the Lagrange multipliers at the optimal solution. The locational marginal price (LMP) for load bus i is

$$
\lambda_i = \frac{\partial \mathcal{L}}{\partial D_i}\Big|_{G_i = G_i^*}, \tag{4}
$$

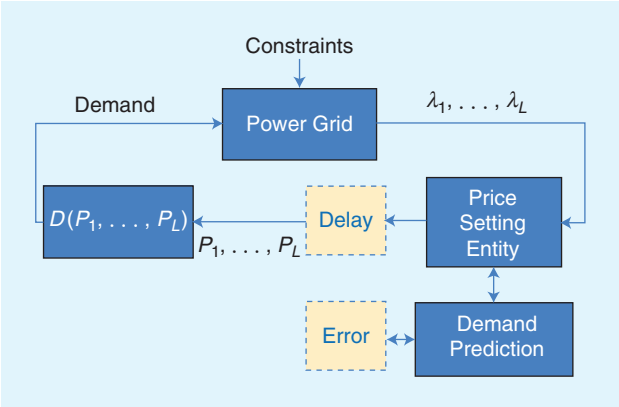
which represents the marginal cost of providing one additional unit of power at that bus under the optimal generation dispatch. In real scenarios, the optimization in (2) usually includes many more constraints to take into account as many known facts as possible when dispatching the generation and to ensure grid reliability and security. Examples include reliability constraints like $(N-1)$ contingency analysis and temporally variable generation capacities. Also, potential transient stability issues are not accounted for in this model [9]. LMPs mirror the true cost of serving the next increment of load, which includes the cost of generation, grid losses, and congestion. Theoretically, one could calculate these prices before real-time operation and by releasing them the system should converge to its optimal operation point. But this will only happen if the following takes place:

- Perfect forecasts of demand values D_is are available.
- The customers are shielded from this true cost and do not react to fluctuations of the their associated LMP (like what happens in most of the power grid today).
- Most importantly, all generators should be price taking rational agents, so that posting the calculated LMP prices will make the generators choose the dispatched value obtained from (2).

If, however, the customers are exposed to even limited information about these true costs, like what all RTP techniques aim to do, there will be an extra feedback loop added to the equation, i.e., the demand at the ith load bus, D_i will be a function of λ_i and thus, (4) has to be modified to

$$
\lambda_i = \frac{\partial \mathcal{L}(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_L)}{\partial D_i}. \tag{5}
$$

A diagram of the system is shown in Figure 3. Calculating the true cost of serving the demand if the customers react to real-time market fluctuations will require perfect knowledge of customer behavior in reacting to price signals for every different load bus in the system at the time the LMPs are being



[FIG3] Model of the generation market and the demand interaction with RTP.

calculated. Reference [10] provides a dynamic model to analyze the dynamics of supply, demand, and clearing prices in a power grid with real-time retail pricing and information asymmetry. It shows that the power market under RTP may possibly experience volatile prices and loads, or even lose its stability. Although proven with major simplifying assumptions, the important point made is that stability should be taken into account when designing a load control system with an unknown and highly variable element in the feedback loop. The authors in [10] conclude that more intricate models for demand in each area, knowledge about consumer behavior in response to dynamic prices, and a thorough understanding of the implications of different market mechanisms and system architectures are needed before RTP techniques can be implemented in large-scale networks.

Currently, power system operators have two major approaches to calculate LMPs for the real-time market. Ex-anté prices indicate the value of the LMPs before the true value of the demand is released, using predicted values of the stochastic variables like the demand and intermittent renewable resources. Ex-post values, on the other hand, are calculated after the load is served and with deterministic knowledge of all the values for demand and generation.

Since real-time price signals need to be delivered to customers beforehand to allow some planning time, they should be of the same nature as ex-anté LMPs. Also, it is very unlikely that ex-post adjustments will be allowed to affect how customers are billed, since this would expose the public to unacceptable risks. The same approach just described is adopted by most RTP researchers in the literature. The real-time price sent to customers is derived either from a direct ex-anté analysis of wholesale market prices, or by adding some modifications to account for consumer satisfaction. To do so, either a term representing the benefit of customers from consuming electricity is added to the cost function in (2) or, a cap on the variations of the price signal in enforced. For example, in [11], the authors maximize the social benefit, which includes known cost functions $C_i(\cdot)$ representing the cost of production of energy and known benefit functions $B_i(\cdot)$representing the consumers

$$\max \sum_{i=1}^{L} B_i(D_i) - \sum_{i=1}^{N} C_i(G_i). \tag{6}$$

To calculate the price, the authors assume that both generation units and customers are price taking agents with known limits on their consumption and generation values and declare the price at each bus as the marginal value of the objective function in (6) under various capacity and transmission constraints. References [12]–[14] follow a similar approach. The gap caused by the difference between ex-anté and ex-post prices should be compensated by the utility, similar to what is done today for the gap between flat or TOU rates and the LMPs. Some of the problems that can arise from this approach are as follows:

- Perfect knowledge of the utility of the customers is assumed at every load bus (usually assumed to be simple analytic functions), which is unrealistic, at least in the current situation.
- If the price is set using an incorrect prediction of the customers usage behavior, prices that are posted may lead to system instability.
- Demand is only assumed to be dependent on the current price. This assumption is not valid because energy is not delivered instantly in packets and appliances need time to finish their jobs.
- Generation owners may try to arbitrage the market.
- For a market that incorporates RTP techniques to be safely operated, at least the major part of generation assets should have a fully deterministic and controllable nature. This assumption no longer holds with the addition of a considerable amount of intermittent resources to the grid as their capacities are highly stochastic and thus, they add more uncertainties to the real-time price calculations.

This is why the socioeconomic feedback of the energy market, with all the uncertainty it carries, cannot be closed in actual real time and some degree of direct control may prove necessary, blurring the boundary between RTP and DLC. Unlike pricing techniques, which require an active participation from users to make price-aware decisions, direct control strategies assume a more passive behavior on the user side. For example, a fleet of EVs will be available to an entity (aggregator/utility) that will be in charge of controlling the charging of vehicles, as long as they will receive full charge in a reasonable time (specified by the user or based on a QoS agreement). Next, we will discuss how these DLC programs traditionally work.

LOAD CONTROL THROUGH CURTAILMENT

Unlike RTP, DLC, or interruptible load programs, have been widely and successfully practiced for over two decades now. During peak load hours, the ISO or the utilities have the option to curtail the load due to certain appliances like air conditioners or water heaters, belonging to participating customers, for a predetermined duration of time (usually, 15–30 min).

This is done by sending a curtailment signal to the target appliance from a dispatch center.

Since the 1970s, when DLC was first conceived, several researchers have worked on finding an optimum curtailment schedule that will maximize the benefits of the ISO or the utility while avoiding unacceptable dissatisfaction for participating customers. Typically, an optimization problem is defined to minimize a certain cost function of the load during a look-ahead horizon. Common constraints are maximum curtailment time and a minimum payback period between two successive curtailments, during which the appliance is allowed to function without interruptions, to catch up on its duty. This type of control is effective only for appliances that can be interrupted. The typical formulation of the optimal DLC is based on stochastic dynamic programming methods [15], [16]. Alternatively, the optimal DLC can be relaxed into a linear program, which is less computationally intensive.

DLC decisions are usually made considering sufficiently long look-ahead horizons, to avoid rebound peaks due to payback periods, since the modified load $L'(t)$ at time t is given by

$$L'(t) = L(t) - \text{Curtailed Load from DLC} + \text{Payback Load from Previous Curtailments}.$$

Note that DLC schedules are sometimes solved for in coordination with the unit commitment problem (2) since they will affect the demand values D_i [17].

Most DLC studies are mainly focused on commercial and industrial buildings and business infrastructures because of the economy of scales provided by controlling their energy consumption that justifies the value of the sensing and telemetry network to perform the control. However, there are several exceptions to this rule, and AMI networks deployed in the last several years enable DLC for residential customers. Figure 4 is a schematic representation of the way DLC can be implemented in homes. One of the issues of advancing DLC is the fact that most of the AMI networks currently being deployed are extremely slow, with downlink speeds of one minute (though with the ability of multicasting), but with a staggering four hours delay in the uplink. This is why it is possible that Internet connections, as well as other communication links, may be used in conjunction with the AMI network to further advance DLC methods.

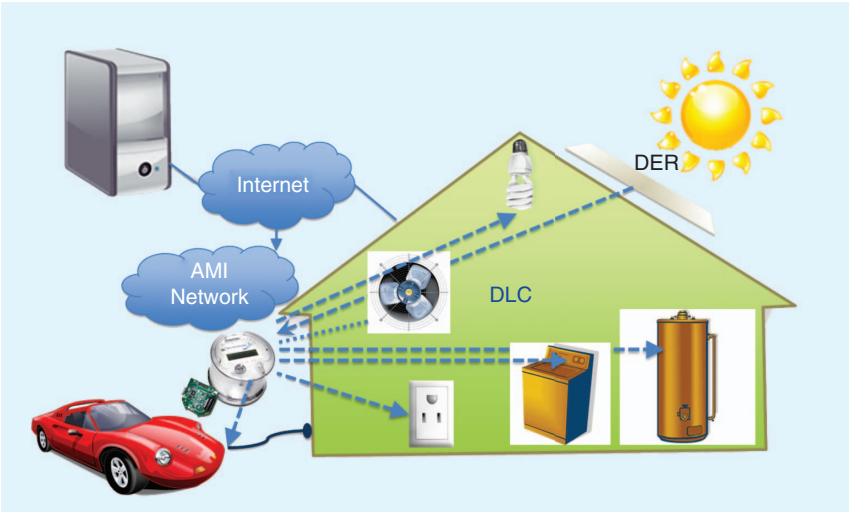
AN IMPORTANT AREA OF RESEARCH RIGHT NOW IS HOW TO MARKET DLC, DEFINING WHAT IS CALLED AN ANCILLARY SERVICE MODEL THAT CAN BE INCORPORATED IN THE ENERGY MARKET.

An important area of research right now is how to market DLC, defining what is called an ancillary service model that can be incorporated in the energy market [18]. Utilities/aggregators examine the degree at which they are able to shed load in case they are called up on the next day and announce this capacity to the wholesale electricity market.

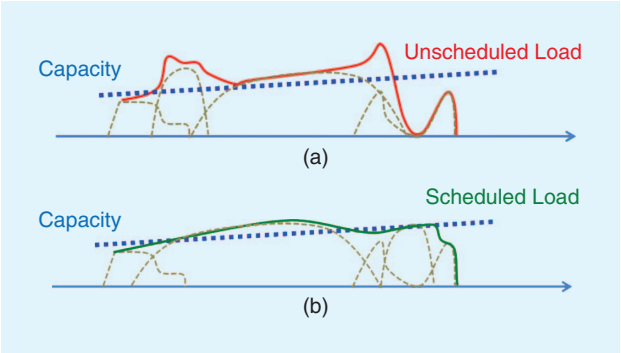
While interrupting certain loads can help alleviate the problems with high peak demand when facing generation shortage, the drawback of most DLC solutions is that they are designed for emergency situations. With the addition of a sizable amount of unpredictable renewable resources, like wind and solar energy to the grid, the frequency of these emergency situations will increase substantially and forced curtailment of the load, even if it is backed by customer participation, will no longer be a sufficient measure to match volatile and unpredictable generation with mostly inelastic demand. Also, deciding how the customers should be paid for these interruptions will become increasingly unclear, bringing about similar complexities as RTP. However, we anticipate that this is an area where new applications may emerge as a result of capillary information and control of events on the grid, as discussed in the section "Information Domains."

DEMAND BIDDING PROGRAMS

Another option that has recently regained popularity is demand bidding (sometimes referred to as *negawatts*) programs, where market participants managing a certain DLC customer base, a large industrial load or even single residential customers directly make an offer to the wholesale market to use less electricity during peak hours on the next day. According to the Federal Energy Regulatory Commission (FERC), demand bidding "will be less costly than a program



[FIG4] Example of DLC architecture for residential users. The computer represents the DLC dispatch center of the regional utility.

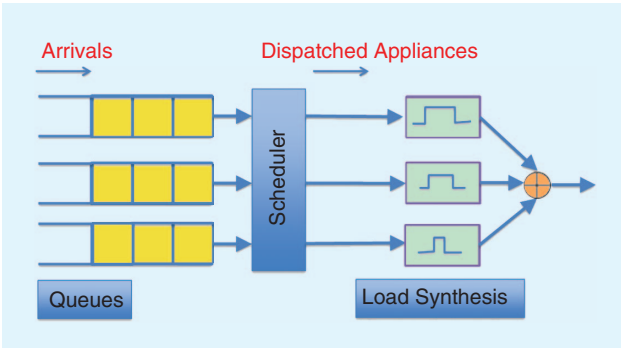


[FIG5] Load scheduling: (a) unscheduled and (b) scheduled load profiles.

where an end user receives payments greater than the market-clearing price to reduce its demand." This compensation by actual market price can happen because, unlike conventional DLC, which is mostly regulatory driven, these programs are market driven as negative watts are dispatched in a manner similar to those offered by conventional generation units. Whether this idea is acceptable is still under debate, since the amount of energy reduction can only be measured based on the consumption history of the bidder and there are market design problems that may arise due to this fact.

EMERGING COORDINATED DR TECHNOLOGIES

Recently, we have observed an emerging body of work that propose to use the famous concept of network utility maximization (NUM) to optimally manage electricity demand under the control of an aggregator/utility company in a distributed fashion (e.g., [18]–[20]). Though there are a few different versions of the utility function used to characterize the social benefit of the entire population, these approaches share a common merit: they can address the previously discussed problems caused by unknown customer behavior since in this case, similar to RTP, customers can individually tailor their demand based on their preferences. However, under NUM, the prices are determined by iterating back and forth between the control center and the customers, thus allowing the control center to take into account individual customer preferences before announcing a final price.



[FIG6] Appliances are put in different queues based on their quantized request code $\hat{C}$, and they are switched on when they are allowed to depart their associated queue by the scheduler. The multichannel structure shows how the load injection is associated with the queue departure processes.

While developing techniques that coordinate how controllable loads owned by different customers respond to fluctuations in the price of their shared resource is a crucial step in making the demand flexible, optimization models that incorporate individual customer preferences (including deadlines) cannot be solved centrally due to their inherent computational complexity and massive communication overhead. Consequently, they cannot be used to model the aggregate service that a population of controllable loads can offer to the grid, either as a price taker or as a DR service provider. Thus, currently, optimizations that need such a model assume that providing electricity to controllable loads is analogous to filling a tank by a certain deadline, which can result in many difficulties and miscalculations. To address this issue, we present a method that can provide a concrete mapping between switching on individual energy requests and the aggregate demand imposed on the grid by the entire population of controllable appliances. This method follows the path of many scalable computing and communication systems existing today that provide best-effort scheduling services, optimizing average service metrics rather than meeting individual customer constraints.

FROM PACKET TO POWER SWITCHING: DIGITAL DIRECT LOAD SCHEDULING

In the section "Price-Based Load Control Strategies," we saw that the problems with determining appropriate price signals to control the load in a decentralized and efficient manner are mainly due to the uncertainty in how customers respond to variable real-time prices, i.e., an unknown feedback behavior. If, however, this feedback loop is somehow opened or its responses are based on settings that can be learned by a control center, the stated problems will no longer exist (or will be mitigated). What we propose is to add a certain intermediate mechanism to help reshape the original demand subject to a mutually agreed-upon level of QoS.

Holding fuel reserves is much more efficient than storing electrical energy. But what would it take to hold back the demand, considering its intrinsic elasticity? The key idea we proposed in [21] is to unbundle the load from certain appliances and view each contribution like a set of LEGO pieces that can be reassembled to follow a desired load profile, by delaying appropriately the power delivery to each individual appliance (see Figure 5).

In our model, each arriving smart appliance has an associated parameter vector C_i that determines uniquely the time evolution of the load contribution when that appliance is turned on. In fact, we assume that, if turned on at time zero, the load is the complex phasor signal $g(t;C_i)$, one to one with C_i. In general, $C_i \in \mathbb{C}^N$ can be the Nyquist samples, or the Fourier or Wavelet coefficients of the known load evolution after activation. A simple example is that of EVs, for which C_i is a two-dimensional vector, representing the charging rate and the fraction of battery charge needed by the car upon its

arrival. What makes our model scalable and practical though, is the quantization step: the continuous load injection parameters C_i are mapped onto a quantized request $\hat{C}_i$, with Q quantization levels. These quantizations will bundle different requests by appliances into a set of discrete load classes. Hence, the scheduling design consists of managing the departures from a set of Q queues, with a FIFO discipline, as shown in Figure 6. For example, with EVs, the charge duration has a maximum of eight hours and can be quantized into 15-min intervals, thus requiring 5 B to communicate and resulting in 32 different queues in which EVs can be bundled to wait for energy. Solving for an optimal departure process from these queues has a considerably lower computational complexity and communication overhead than solving for deadline-constrained schedules, precisely because of the described discretization and bundling technique.

PROSPECTIVE DSM STRATEGIES WILL BE OF SUBSTANTIAL IMPORTANCE IN THE FUTURE GRID FOR BALANCING DEMAND AND GENERATION.

DDLS can be implemented as a voluntary program, similar to DLC, through which customers can receive cheap energy in return for their patience, since their demand is being modulated to consume green energy opportunistically. The neighborhood scheduler will operate as a local energy retailer, which determines the activation time of the requesting smart loads such that it can shape the load profile to be as close as possible to the available generation supply (the day-ahead bid + the available local intermittent resources), avoiding the rebound peak problems that arise in DLC and also accounting for the customer service level through its queue management. We call this scheme DDLS.

As previously discussed, prospective DSM strategies (i.e., load control via price-induced customer response or load curtailment etc.) will be of substantial importance in the future grid for balancing demand and generation. By providing DR services to the grid using these strategies, end-use customers can get cheaper energy as a reward. However, if these services are offered independently by each individual customer, they will come in small amounts and have little degrees of freedom for transactions in the wholesale electricity market. Therefore, to have more bargaining power, these services may be pooled together by intermediary businesses responsible for planning and scheduling these distributed DR offers, usually referred to as energy aggregators (which can be a utility) [22]. Furthermore, the idea of "aggregation" can also be employed on the generation side to incorporate renewable energy resources. Some popular aggregation schemes, such as virtual power plant (VPP) and microgrid, are discussed in the following section.

RENEWABLE INTEGRATION FROM THE DISTRIBUTION SIDE

The increasing penetration of renewables introduces another level of difficulty before the benefits of DSM technologies can be fully exploited. As mentioned earlier, electricity generation from renewable energy sources (RES), which is a fast growing segment in the smart grid implementation, adds to the difficulty of ensuring power grid stability and reliability. It is projected that by the year 2020, 33% of the U.S. energy supply will come from renewable resources. Wind power and solar generation can be expected to gain essential penetration in the coming years. In particular, RES will be integrated in the grid in a dispersed fashion because this strategy has many potential benefits, including lower transmission losses, providing reliable power supplies during emergencies, reduction in peak power requirement, provision of ancillary services and so on [23].

The increasing interest in distributed generation (e.g., microgeneration) is due to the convergence of two main factors: the many recent developments in technologies to generate and control a variety of microscale DER and the fact that the electricity business is restructuring. Typical microgeneration systems include, e.g., windmills, photovoltaic (PV) panels, fuel cells, microturbines that use fuel such as gas or biofuel, and community energy storage (CES) devices (flywheels or batteries). With widespread deployment of these resources, the cost of expanding and updating the transmission capacity can be partly offset by placing assets near the consumer. Unfortunately, the limited dispatchability and intermittent nature of wind or solar RES may pose severe operational challenges to the grids, in terms of market performance, transmission congestion, and system reliability, if the volatility in their power outputs is not monitored, managed, or coordinated appropriately.

As explained later, electricity storage incurs prohibitive costs and options. For instance, curtailing wind or solar generation or further adding additional gas turbines to compensate for the variability of renewables are imperfect solutions, since they essentially counteract the benefits and purpose of deploying renewables. Therefore, a scalable and robust integration coordination architecture, in addition to the intelligence of DSM technologies, is vital for connecting these distributed resources to the power grid.

INTEGRATION ARCHITECTURES FOR DISTRIBUTED RESOURCES

To integrate renewable resources from the distribution side into the power grid, three main concepts have been proposed, centering on the idea of aggregation of either the DER, the loads, or both.

1) ***VPP***: A VPP [24] is an aggregated group of distributed generation units. The concept of VPP is mainly proposed to allow DER to bid more effectively in the market, because distributed sources come in small quantities and experience large fluctuations and uncertainty, which makes participation of any single source unit in the wholesale market impossible. VPP is comparable to a power plant connected

to the transmission grid, which pools together available distributed resources. With this intermediary entity, DER participation in the energy market is facilitated and the balance between distributed generation and demand is better managed and coordinated.

2) ***Energy Hub***: The energy hub is a market-driven concept [25] of input/output relationship of energy, where multiple energy carriers/resources can be converted, conditioned, and stored as one comprehensive unit. Energy hubs consume power at their input ports connecting to, e.g., electricity and natural gas infrastructures, and provide certain required energy services such as electricity, heating, cooling, and compressed air at the output ports. Within the hub, energy is converted and conditioned using, e.g., CHP technology, transformers, power-electronic devices, compressors, heat exchangers, and other equipment. Real facilities that can be considered as energy hubs are, for example, industrial plants (steel works, paper mills, data centers), big building complexes (airports, hospitals, shopping malls), rural and urban districts, and small isolated systems (trains, ships, aircrafts).

THE MAIN CHALLENGE IN USING STORAGE TODAY IS THE HINDERING COST OF PRODUCTION AND MAINTENANCE OF STORAGE DEVICES AND THEIR LIMITED CONVERSION EFFICIENCY.

3) ***Microgrid***: A microgrid [26] is a coordinated cluster of DER units, loads and storage with appropriately designed intelligence for management and control. The microgrid can cover an urban area, or serve the subnetwork connected to a transformer, which can be a conglomerate of industry, retail, or other and it operates in essentially in two modes: the grid-connected and autonomous island modes. The first is the normal interconnected mode, in which the load is served by both drawing power from the medium voltage (MV) grid, as well as from the microgrid generation resources, which can inject power back in the MV network. The second one is usually only adopted in emergencies, when the microgrid does not receive power from the MV network and relies only on local generation to meet the local demand.

It is noteworthy to remark that microgrid sources cannot be interconnected and provide energy in the MV pool, because they are not synchronous and sometimes they do not produce energy in the form of ac voltage. In fact, fuel cells and PV panels produce dc current. The control of the microgrid needs therefore to include an electronic interface (dc/ac or ac/dc/ac) to operate. At a small scale the microgrid is a microcosm that replicates, and somewhat exacerbates, the communication and control needs of the power grid. To be able to operate in the emergency model, typically, microgrid load control (e.g., load curtailment mechanisms) and operational strategies are vital to provide sufficient generation capacity, and maintain the possible level of service in the emergency mode. On the other hand, with the increasing penetration of DER units, sufficient provisions for both island and grid-connected modes of operations and smooth transition between the two (i.e., fast response) are also necessary to enable exploiting microgrid resources. Apart from its resilience and flexibility, microgrid technology is considered as a promising candidate in future power grid to use distributed resources [27] because it serves as a comprehensive aggregator of both distributed generation and loads to optimally implement DSM technologies, balance demand with generation, and operate as an intelligent entity in parallel with or connected to the transmission grid.

PHYSICAL INTERFACE FOR RENEWABLE INTEGRATION

With all the above concepts proposed for renewable integration (e.g., VPP and microgrid), a fundamental issue is how the distributed units (including microgeneration, loads, and storage) interact with the current grid. The interfaces for such interactions span from maker participation to physical interfaces that coordinate the generation sources, information processing and control units as well as electrical configuration.

The key for incorporating renewable generation is to have intelligent controllers inside each bundled structure to gather and process local information (e.g., DER bids, local market prices for the renewables, demand/renewable forecast, network security constraints [26]) to optimize the DER unit productions, determine the set points of loads to be served or shed, and issue market prices for the next local bidding process. This can be implemented in both centralized and decentralized fashions [24].

In a centralized scheme, the DER units are monitored and coordinated by a central controller (CC). Local demand information is collected by the CC, where the distributed generation and demand scheduling are optimized under network and resource constraints. Then the decisions are fed to the DER units for outputs. The centralized deployment is able to incorporate both technical and economical functions in the optimization, to gain benefit of aggregating distributed resources. For decentralized systems, each DER unit or cluster has its own local controller (LC). The communication topology between the LCs is similar to a mesh network, where the optimization is carried out at the LCs jointly via communicating with one and another in a multiagent fashion. Different than the centralized scheme, the decentralized scheme is more resilient to communication failures and more flexible for reconfiguration and integration.

EMERGING STORAGE OPTIONS

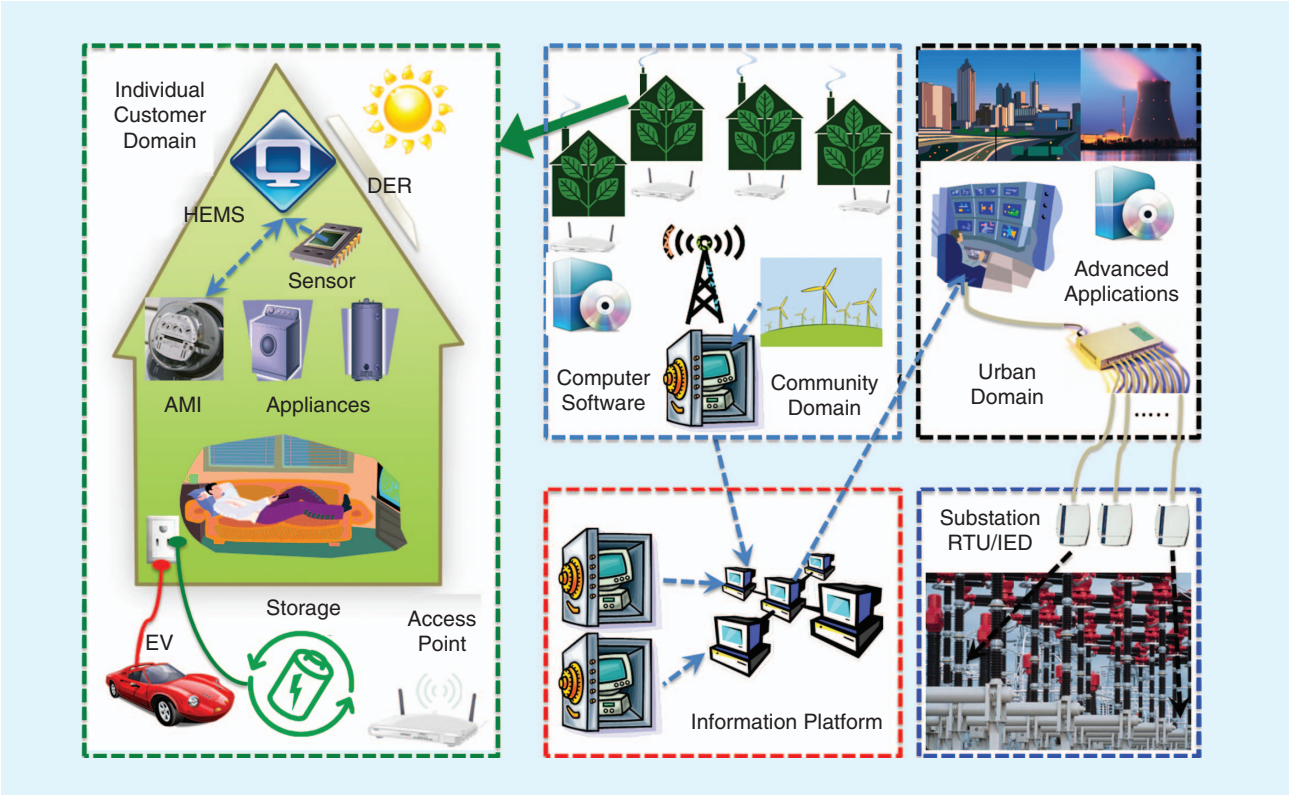
With the introduction of battery-based loads such as EVs or industrial and CES devices, serving as a buffer or a source of energy, the possibility and benefits of storing energy in power

systems have been investigated in recent years [28]. Access to advanced energy storage will increase the reliability and security of power systems. In the short term, fast-response storage has already been demonstrated as a viable means of managing grid imbalances and volatility through the regulation service. In the longer term, storage can be deployed to shift energy in time to smooth the output of renewable generation or reduce the peak load on constrained transmission assets. However, the option of energy storage should not be viewed as a panacea to all these issues. The main challenge in using storage today is the hindering cost of production and maintenance of storage devices and their limited conversion efficiency. If these problems are solved, storage will be a viable option to modify the demand for electricity, with much less consumer interaction.

INFORMATION DOMAINS

Although there are several models for integrating renewables, microgrid solutions are very expensive, and what is still lacking is an abstract model for sensing, communication, service and actuation (decision making) that would allow a seamless integration of microgeneration and local RES technologies with DSM and DR of loads. In the past, the Manufacturing Message Specification (MMS) protocol of ISO9506 provided a generic description for networked automation. Today, OpenADR (http://openadr.lbl.gov/) is emerging as a specification protocol to ease the development of DR applications. What are the information domains around which these new applications will gravitate? Due to their considerable share of the total electricity consumption, in addition to the transportation system, buildings will play a vital role in allowing to use local renewables opportunistically. Within a microgrid, DSM and DR are essential tasks to support. However, outside these settings, it is only natural to combine RES assets with local DSM and DR, since the idea of using the grid as a stable and cheap supply cannot scale. During the past decade, many researchers have worked on developing new sensing, networking, and control architectures specifically for demand management in commercial and residential buildings, creating a new domain where DSM and DR can be implanted. If networked together, communities including such buildings can actively participate in the energy market by owning DERs and managing their energy use smartly. Heating, ventilation, and air conditioning (HVAC) systems and smart lighting are considered as the most important venues for demand management on this scale.

Figure 7 is an abstraction of the network architecture and highlights the different information domains. Contrary to the traditional model for the grid, to ensure scalability in intelligently dispatching the resources, it is critical to perform a number of decisions at the edge of the network. In this vision AMI and submetering systems provide information to the local HEMS for optimization and feedback control, including load consumption patterns, renewable generation, storage status and other measurements obtained from distributed sensors. At



[FIG7] Communications and information processing in smart distribution networks.

the same time, the HEMS from different households perform DR and DSM control actions issued by the local demand control server. In this vision, machine-to-machine communications will become the prevalent source of traffic for the information infrastructure that will support them. This traffic is likely to have very peculiar features compared to traditional communication services, supporting voice, or traditional data transmission, since the messages will report individual events from heterogeneous user devices, with sensitive time requirements and with correlated traffic patterns.

THERE ARE MANY AREAS WHERE SIGNAL PROCESSING RESEARCH CAN CONTRIBUTE TO PAVE THE WAY FOR A GREENER FUTURE FOR THE GRID.

We can roughly summarize the information [29] that needs to be gathered locally for effective planning and design as well as efficient operation and management as follows:

1) Equipment should be capable of communicating context information describing the generation of the technology and its power and energy capacities, such as the type of technology, network location, ownership and responsibility, range of operations; it should also provide a service description in a standardized form, which would enable a mapping into standardized actions.

2) Accurate sensor data, with context information on time, location, type of measurement, which will be needed for real-time load forecasting, based on proper consumer behavior analysis.

3) Pricing-incentive descriptions should be easily accessible for the development of HEMS software.

Responsive loads will have to be able to communicate their class: for example, if they have large consumption power, long job duration, and predictable operation cycles (e.g., EV or PHEV, washing machines, dishwasher) if they are interruptible (e.g., home cooling/heating/ventilation); if they are delay intolerant loads, that exhibit random arrivals and durations with some predictable pattern (e.g., TV, computer).

Signal processing is likely to operate between the physical world, to ease the descriptions, or extract them from data, as well as in the application layer, where algorithms will use information to produce valuable analytics for optimization.

CONCLUSIONS

In this article, we discussed the most recent developments in the area of load management, and considered possible interaction schemes of novel architectures with DER. To handle the challenges faced by tomorrow's smart grid, which are caused by volatile load and generation profiles (from the large number of plug-in EVs and from renewable integration), the conventional grid operation principle of load following needs to be changed into load shaping or generation following. DSM will be a promising and powerful solution to the above challenges. However, a lot of other issues such as load forecasting, pricing structure, market policy, and renewable integration interface need to be taken into the design to search for the most effective and applicable solution.

A similar fertile ground for new inventions happened when communication networks changed their access model from circuit switched networks, designed to follow peak requirements of analog transmission, to packet switched network that constantly negotiated how to stream the next block of data efficiently. Soon, digital standards emerged that could transmit the same information content, albeit with best effort services.

There are many areas where signal processing research can contribute to pave the way for a greener future for the grid. At the high level, one is in defining the DSM and DR application layer, finding new abstractions and digital descriptions of the end use of electricity, that can potentially incorporate a customer tailored response, much like video compression, which uses perception to push the limit of scalable video coding. The other is in supporting the wealth of sensing, telemetry, and signal processing that is needed to follow uncertain generation, in particular if the grid will increasingly rely on DER to bypass transmission bottlenecks. Rapid sensing and automated reconnaissance will be of paramount importance, to serve the flexible and responsive demand with greener energy production. Last but not least, a number of classic statistical signal processing tools can be used to model, learn, and predict sensors data, future demand, and financial signals that will permeate the architecture of the future grid.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability through the Consortium for Electric Reliability Technology Solutions, Trustworthy Cyber Infrastructure for the Power Grid program and Pacific Northwest National Laboratory (operated by Battelle under contract DE-AC05-76RL01830).

AUTHORS

Mahnoosh Alizadeh (malizadeh@ucdavis.edu) has been a Ph.D. student at the University of California, Davis, since 2009. She received the B.S. degree in electrical engineering from Sharif University of Technology. Her research interests mainly lie in the area of the smart grid, in particular DSM and DR.

Xiao Li (eceli@ucdavis.edu) has been pursuing the Ph.D. degree at the University of California, Davis, since 2009. He received the B.Eng. degree (2007) from Sun Yat-Sen (Zhongshan) University, China, and the M.Phil. degree (2009) from the University of Hong Kong. His research interests are in signal processing and optimization algorithms, with applications in communications, networking, and power systems.

Zhifang Wang (zfwang@ucdavis.edu) received the B.S. and M.S. degrees in electrical engineering in 1995 and 1998, respectively, from Tsinghua University Beijing, China, and the Ph.D. degree in electrical and computer engineering in 2005

from Cornell University, Ithaca, New York. She is currently an associate specialist in the Department of Electrical and Computer Engineering at the University of California, Davis. She was previously a postdoctoral research associate at the Information Trust Institute of the University of Illinois at Urbana-Champaign. Her research interests include smart grid communications, large-scale system analysis, cascading failures, and real-time controls of electric power grids. She is a Senior Member of the IEEE.

Anna Scaglione (ascaglione@ucdavis.edu) received the Laurea (M.Sc. degree) in 1995 and the Ph.D. degree in 1999 from the University of Rome, "La Sapienza." In 2001, she was at Cornell University, Ithaca, New York, where she became an associate professor in 2006; prior to joining Cornell she was an assistant professor at the University of New Mexico from 2000 to 2001. She is currently a professor in the Department of Electrical and Computer Engineering at the University of California, Davis, where she joined in 2008. She has held several editorial posts and is currently the editor in chief of *IEEE Signal Processing Letters* and guest editor of *IEEE Journal on Selected Areas in Communications*, for a series of special issues on the smart grid. She received the 2000 IEEE Signal Processing Transactions Best Paper Award, and the 2002 NSF Career Award, and she was also the corecipient of the Ellersick Best Paper Award (MILCOM 2005). Her expertise is in the broad area of signal processing for communication systems and networks. Her current research focuses on communication and wireless networks, sensors' systems for monitoring, control and energy management, and network science. She is a Fellow of the IEEE.

Ronald Melton (ron.melton@pnnl.gov) received his B.S.E.E. degree from the University of Washington and his M.S. and Ph.D. degrees in engineering science from the California Institute of Technology. He is a senior power systems engineer at Battelle/Pacific Northwest National Laboratory and is involved in a variety of smart grid projects with an emphasis on interoperability standards and methods. He is currently the administrator of the GridWise Architecture Council and project director for the Pacific Northwest Smart Grid Demonstration. He has over 30 years of experience in systems engineering applied to interdisciplinary problems. In recent years, he has been a technical leader in cybersecurity for industrial control systems used in critical infrastructures. He is a member of the ACM and a Senior Member of the IEEE.

REFERENCES

[1] J. Lopes, F. Soares, and P. Almeida, "Integration of electric vehicles in the electric power system," *Proc. IEEE*, vol. 99, no. 1, pp. 168–183, Jan. 2011.

[2] D. Hammerstrom, R. Ambrosio, J. Brous, T. A. Carlon, D. P. Chassin, J. G. DeSteese, R. T. Guttromson, G. R. Horst, O. M. Järvegren, R. Kajfasz, S. Katipamula, L. Kiesling, N. T. Le, P. Michie, T. V. Oliver, R. G. Pratt, S. E. Thompson, and M. Yao, "Pacific northwest gridwise testbed demonstration projects: Olympic peninsula project final report," Pacific Northwest Nat. Lab., Richland, WA, Rep. NNL-17167, 2007.

[3] J. Eto, C. Goldman, B. Kirby, J. Kueck, M. Kintner-Meyer, J. Dagle, T. Mount, W. Schultze, R. Thomas, and R. Zimmerman, "Innovative developments in load as a reliability resource," in *Proc. IEEE Power Engineering Society Winter Meeting*, 2002, vol. 2, pp. 1002–1004.

[4] Y. Jia-Hai, "Customer response under time-of-use electricity pricing policy based on multi-agent system simulation," in *Proc. PSCE*, 2006, pp. 814–818.

[5] A. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 120–133, Sept. 2010.

[6] D.-M. Han and J.-H. Lim, "Design and implementation of smart home energy management systems based on zigbee," *IEEE Trans. Consumer Electron.*, vol. 56, no. 3, pp. 1417–1425, 2010.

[7] W. Hogan, E. Read, and B. Ring, "Using mathematical programming for electricity spot pricing," *Int. Trans. Oper. Res.*, vol. 3, no. 3–4, pp. 209–221, 1996.

[8] E. Litvinov, "Design and operation of the locational marginal prices-based electricity markets," *IET Gen. Trans. Distrib.*, vol. 4, no. 2, pp. 315–323, Feb. 2010.

[9] M. Huneault and F. D. Galiana, "A survey of the optimal power flow literature," *IEEE Trans. Power Syst.*, vol. 6, pp. 762–770, May 1991.

[10] M. Roozbehani, M. A. Dahleh, and S. K. Mitter. (2011, June). Volatility of power grids under real-time pricing, Quantitative Finance Papers 1106.1401 [Online]. Available: http://ideas.repec.org/p/arx/papers/1106.1401.html

[11] J. Y. Choi, S.-H. Rim, and J.-K. Park, "Optimal real time pricing of real and reactive powers," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1226–1231, Nov. 1998.

[12] M. Roozbehani, M. Dahleh, and S. Mitter, "Dynamic pricing and stabilization of supply and demand in modern electric power grids," in *Proc. IEEE SmartGridComm*, Oct. 2010, pp. 543–548.

[13] R. Sioshansi, "Evaluating the impacts of real-time pricing on the cost and value of wind generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 741–748, May 2010.

[14] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. W. S. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *Proc. IEEE SmartGridComm*, Oct. 2010, pp. 415–420.

[15] A. Cohen and C. Wang, "An optimization method for load management scheduling," *IEEE Trans. Power Syst.*, vol. 3, no. 2, pp. 612–618, May 1988.

[16] K.-Y. Huang and Y.-C. Huang, "Integrating direct load control with interruptible load management to provide instantaneous reserves for ancillary services," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1626–1634, Aug. 2004.

[17] Y.-Y. Hsu and C.-C. Su, "Dispatch of direct load control using dynamic programming," *IEEE Trans. Power Syst.*, vol. 6, no. 3, pp. 1056–1061, Aug. 1991.

[18] S. Caron and G. Kesidis, "Incentive-based energy consumption scheduling algorithms for the smart grid," in *Proc. IEEE SmartGrid-Comm*, Oct. 2010, pp. 391–396.

[19] M. Pedrasa, T. Spooner, and I. MacGill, "Coordinated scheduling of residential distributed energy resources to optimize smart home energy services," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 134–143, Sept. 2010.

[20] N. Li, L. Chen, and S. Low, "Optimal demand response based on utility maximization in power networks," in *Proc. IEEE PES General Meeting*, July 2011, pp. 1–8.

[21] M. Alizadeh, A. Scaglione, and R. Thomas, "From packet to power switching: Digital direct load scheduling," *IEEE J. Select. Areas Commun.*, to be published.

[22] E. Koch and M. Piette, "Scenarios for consuming standardized automated demand response signals," Lawrence Berkeley Nat. Lab., Berkeley, CA, LBNL Paper LBNL-1362E, 2009.

[23] G. Joos, B. Ooi, D. McGillis, F. Galiana, and R. Marceau, "The potential of distributed generation to provide ancillary services," in *Proc. PES Summer Meeting*, 2000, vol. 3, pp. 1762–1767.

[24] E. A. Setiawan, "Concept and controllability of virtual power plant," Ph.D. dissertation, Kassell Univ., Germany, 2007.

[25] M. Geidl and G. Andersson, "Optimal coupling of energy infrastructures," in *Proc. IEEE Power Tech*, 2007, pp. 1398–1403.

[26] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management," *IEEE Power Energy Mag.*, vol. 6, no. 3, pp. 54–65, May–June 2008.

[27] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Apr. 2009.

[28] W. Lachs and H. Tabatabaei-Yazdi, "Energy storage in power systems," in *Proc. IEEE Int. Conf. Power Electronics and Drive Systems (PEDS 1999).*, vol. 2, pp. 843–848.

[29] C. Foote, P. Espie, G. Ault, G. Burt, J. McDonald, and F. Silvestro. Distributed generation information management: Requirements and methods [Online]. Available: http://www.iset.uni-kassel.de

[SP]