Virtual Energy Storage from Flexible Loads: Distributed Control with QoS Constraints

Prabir Barooah

Abstract Loads are expected to help the power grid of the future in balancing the highs and lows caused by intermittent renewables such as solar and wind. With appropriate intelligence, loads will be able manipulate demand around a nominal baseline so that the increase and decrease of demand appears like charging and discharging of a battery, thereby creating a virtual energy storage (VES) device. An important question for the control systems community is: how to control these flexible loads so that the apparently conflicting goal of maintaining consumers' quality of service (QoS) and providing reliable grid support are achieved? We advocate a frequency domain thinking of handling both of these issues, along the lines of a recent paper. In this article, we discuss some of the challenges and opportunities in designing appropriate control algorithms and coordination architectures in obtaining reliable VES from flexible loads.

1 Introduction

A future power grid is likely to experience significant intermittency in generation from renewable sources such as solar and wind. This intermittency is illustrated in Fig. 1; the data comes from BPA (http://www.bpa.org), a balancing authority (BA) in the Pacific Northwest. The net demand, which is the difference between demand for power and renewable power generated, must be supplied by controllable generation resources. The sharp ramps and fast variations in the net demand are a cause of concern for conventional generators. They are not designed to track such a fast varying signal. Inability to track the net demand can seriously degrade reliability of the power grid: if demand–supply imbalance becomes too large, the grid frequency deviates far from the nominal value of 60 Hz, and cascading blackouts can occur.

Additional resources are needed to mitigate the volatility created by solar and wind. One possibility is to employ sufficient standby generation that can ramp up and down quickly, such as hydro and gas. Hydro is limited by geography, while the

P. Barooah (✉)
University of Florida, Gainesville, FL, USA
e-mail: pbarooah@ufl.edu

J. Stoustrup et al. (eds.), *Smart Grid Control*, Power Electronics and Power Systems, https://doi.org/10.1007/978-3-319-98310-3_6

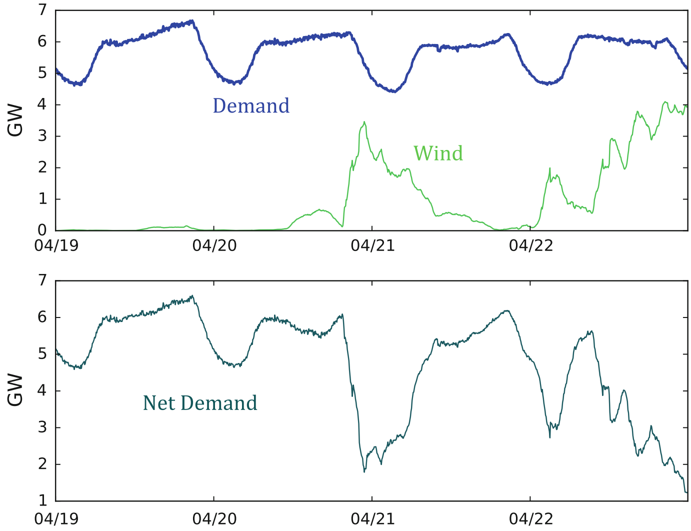

Fig. 1 (Top) Total demand and renewable generation, and (bottom) net demand in BPA (Bonneville Power Administration: http://www.bpa.gov), April 19–22, 2016

use of additional fossil plants as backup will negate the environmental benefits of renewables, apart from increasing the overall cost of energy. The business case for the power plant owners is also questionable since the plants would not sell much energy, which is already causing a few power plants to close [1]. Another possibility is to employ sufficient energy storage resources such as batteries, flywheels, pumped hydro, and compressed air systems. At present, this is a prohibitively expensive option. We discuss the cost of batteries in Sect. 4. The third possibility is to equip loads with intelligence so that their demand can be varied in such a way that mismatch between demand and generation is reduced. In fact, with the help of appropriate control algorithms, loads that have some flexibility in their power demand can be made to provide the same service as that of a battery. We call this virtual energy storage (VES) from flexible loads; see Fig. 2 for a schematic. This is to be contrasted with real energy storage (RES), which include batteries, pumped hydro, flywheels, compressed air, etc.

This paper describes some of the technological challenges and opportunities in obtaining VES from flexible loads. Any technological solution to obtaining grid support from loads must consider its effect on consumers. After all, all loads are used by consumers to provide a certain function, and they have certain expectation of the quality of service (QoS) from those loads.

There is a fast-growing literature on the control of flexible loads to provide grid support services. A dominant paradigm in this literature is control and coordination of loads through real-time prices of electricity, or some other market-based mechanism; see [2, 3] and references therein. These viewpoints have several weaknesses. One, real-time prices subject consumers to high levels of risk. Real- time prices of

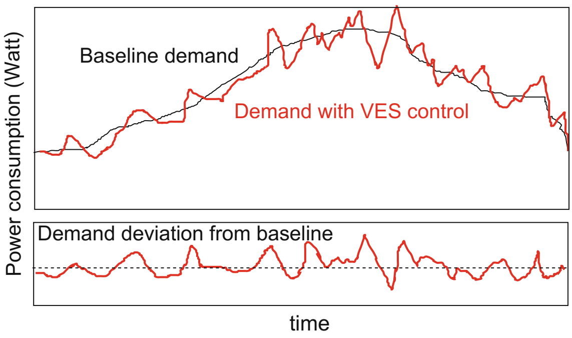

Fig. 2 Virtual energy storage (VES) from flexible loads: demand is varied around a baseline with the help of a control algorithm so that the demand deviation from the baseline is akin to the charging and discharging of a battery

electricity are volatile even without high penetration of intermittent renewables; see [4] for examples from around the world. In fact, [4] shows that these volatilities persist even in an idealized market with participants having no market power ("price takers"), but occur purely as a result of uncertainty and ramp rate constraints. Two, they require consumers to assign a dollar value to a change in consumption with an uncertain QoS loss, e.g., "how much payment is adequate to compensate for a 1 kW decrease in power consumption?", such as in [3]. However, the answer to this question is likely to change frequently for the same consumer, depending on the context (during a party, after a workout session), and also depending on how long the loss of QoS will have to be endured. More recent work on market-based "demand response" has sought to address some of these issues by moving away from real-time prices; but using price as a coordination signal meant to help reach an equilibrium; such as [5]. However, these works also require complex information, such as specification of utility functions (utility of consumers as a function of consumption). If deployed at scale, market-based mechanisms may not lead to a reliable service that grid operators can rely on.

Evidence from existing demand response programs indicate that long- term contracts reduce the risk to consumers while providing a more reliable service to the balancing authorities. Florida Power & Light has 760,000 residential consumers enrolled in their *On Call* demand response program [6]. In return for a monthly rebate, these consumers allow FP&L to turn off their pool pumps and air conditioners a few times in a year. This program has been in place for more than a decade, and has been effective since consumers are getting a reliable return for a known loss of QoS. We, therefore, argue that a control architecture based on long-term contracts between consumers and BAs, with negotiated QoS bounds, offer a reliable consumer engagement. The control system must ensure that QoS never deviates outside of the pre-negotiated bounds. Although the rest of the paper is not dependent on long-term contracts being the only form of payment, we use that assumption to remove market considerations.

It was argued in [7] that Fourier decomposition provides a convenient framework to assign grid's needs to all supply side resources, including traditional generators, loads providing VES service, and batteries providing RES services. In this paper, we further explore the frequency domain thinking. We emphasize that current grid operation and planning already is based on a similar framework, by breaking down the requirements by timescale. Base-load power generation is scheduled based on predictions of the net demand at the slowest timescale (lowest frequency), load following and frequency regulation at intermediate and fast time scales is performed by automatic generation control that adjusts generation set points [8]. However, current taxonomy of generation-side services, such as "frequency regulation" and "load following" are inadequate in a renewable-rich power grid. In the future, "renewable following" may be as important a service as load following. Therefore, we avoid using that taxonomy in the paper.

The rest of the paper is structured as follows. Section 2 describes the VES idea in detail, and summarizes the main challenges in developing local control algorithms for a load to deliver VES to the grid with guaranteed bounds on its QoS. Section 3 discusses the challenges in developing architectures for distributed coordination of millions of loads to meet the VES service needed by the grid. Section 4 discusses cost of battery-alone storage and what it means for cost targets of VES technology.

2 Virtual Energy Storage from Flexible Loads

A load's power consumption can be varied around a baseline to provide a battery-like service. Let $p_b(t)$ be the baseline power demand of a load (or a collection of loads). Suppose its (their) demand is varied through the use of appropriate control software to be $p(t)$ so that the demand deviation from the baseline:

$$p_{\text{ves}}(t) := p(t) - p_b(t) \tag{1}$$

is zero mean: $\lim_{T\to 0} \frac{1}{T}\int_0^T p_{\text{ves}}(t)dt = 0$; cf. Fig. 2. We can then say that the load is providing VES, or, that it is acting like a virtual battery. The demand deviation $p_{\text{ves}}(t)$ is the charging power consumption of the virtual battery. Positive $p_{\text{ves}}(t)$ means the load is drawing more power from the grid than what it would have under baseline conditions; so the virtual battery is charging. Conversely, negative $p_{\text{ves}}(t)$ means it is discharging. The zero-mean nature of the demand deviation means the net energy consumption/generation of the virtual battery is 0, just like a real battery.

Two questions arise:

- For a specific load and a bound on change of its QoS, what kind of demand deviation ("virtual charge/discharge signal") $p_{\text{ves}}(t)$ is allowable that ensures the QoS bound is satisfied? And, how does this vary from load to load?
- How is the net demand signal to be apportioned among the loads so that together they can supply it, while each load maintains its QoS bound?

2.1 Constraining Loss of QoS via Constraining Bandwidth

QoS measures vary depending on load type. There are a large variety of flexible loads, such as refrigeration systems, electric vehicles, pool pumps, water heaters, data centers, municipal pumping systems, HVAC systems, etc. Each has their own QoS metrics, and a distinct degree of flexibility. For HVAC, measures of QoS include indoor temperature and ventilation rate (as a surrogate for indoor air quality). Hot water heaters—and pool pumps in some areas—are also large sources of demand. A QoS measure for a pool pump is the average number of hours the pump is on (as a surrogate for water cleanliness) [9, 10]. For hot water heaters, it is the availability of hot water that is critical. For an aluminum plant, a measure of QoS is the temperature of the smelter [11]. For all loads, whether commercial, residential or industrial, QoS metrics include the cost of energy used[1] and equipment lifetime.

The diversity of QoS metrics among distinct load types is a challenge in developing control algorithms to exploit their demand flexibility. We argue that, in fact, a unifying framework can be developed based on the spectral content of the demand variation, a viewpoint first expounded in [7]. For every load type, maintaining a specific bound on the QoS can be translated to maintaining a bound on the *bandwidth* of its demand deviation. For instance, a small and fast variation of power consumption of a commercial HVAC system can be obtained by a small and fast variation of airflow. The resulting temperature deviations will be small since the large thermal inertia of the building will act like a low-pass filter to such airflow variations. However, even a small amplitude airflow variation can lead to large deviation in indoor temperature if the variation persists for a long time, i.e., the frequency is small enough. For a given amplitude, the higher the frequency of airflow variation, the smaller the effect on QoS metrics of indoor temperature and average ventilation rate. However, above a certain frequency, QoS will reduce since equipment life will degrade. Figure 3 illustrates this idea. For loads that can only be turned on or off, such as hot water heaters, again limiting the frequency of turning on and off is needed to reduce short-cycling and ensure delivery of hot water.

In essence, the VES capacity of a load can be characterized in terms of the power spectral density (PSD) $P_{ves}(\omega)$ of the demand variation, $p_{\mathrm{ves}}(t)$. The PSD must lie in a specific region to meet a given QoS constraint, which can be parameterized by, say, a scalar q. For every value of q, there is a curve $c_q(\omega)$ so that that QoS will be respected only if the PSD of p_{ves} lies under the curve $c_q(\omega)$. *The curve corresponding to the minimum acceptable QoS q^* determines the load's VES capacity. We call $c_{q^*}(\omega)$ the load's capacity curve.*

An illustration of the curve $c_q(\omega)$, for some q, is shown in Fig. 3. For a specific load, or load class, determination of the curve $c_q(\omega)$ can be determined either through modeling or experimental evaluation [12].

Challenges and opportunities A weakness of the frequency domain characterization of VES capacity is that variations over time, especially due to exogenous factors such

[1] For some large consumers, "utility bill" is a better measure since their peak demand charges may constitute a large part of the bill.

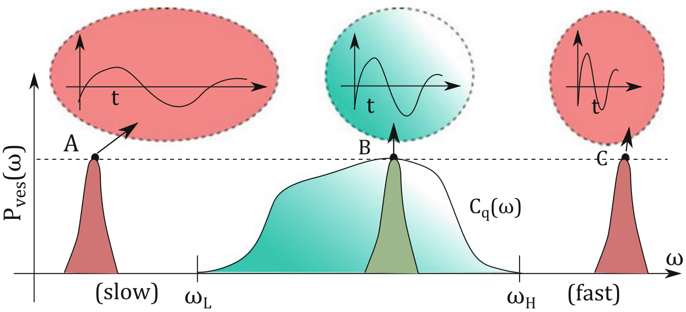

Fig. 3 Constraint on QoS is a constraint on bandwidth of demand variation. The x-axis is frequency and the y-axis is the PSD of demand variation. The PSD must lie in the region under the curve $c_q(\omega)$ to meet the QoS measure q. For a different value of q, this curve would change. The low and high limits of the frequency in which this particular load class can provide VES service are denoted as ω_L and ω_H, respectively. The three signals shown in A, B, and C, have PSDs that have the same total power (i.e., the integral of their PSDs are the same), but distinct bandwidths. The signals A and C violate the QoS metric q, because their bandwidths are too low and too high, respectively. The signal B satisfies the bandwidth requirement

as weather are not conveniently captured. For instance, during afternoon hours of very hot days, an HVAC system may have to run at peak power, and in that case a zero-mean deviation from the baseline is not possible. An alternate way of quantifying capacity that has been explored is a time-varying range (upper and lower bound) of total power consumption so that as long as power consumption stays within that bound, QoS metrics will be satisfied [13, 14]. These approaches necessarily lead to conservative estimates since a constant power deviation from a baseline that still maintains QoS constraints must be allowed in this framework. A general framework that combines the advantage of frequency-based characterization, but is capable of modeling the effect of exogenous factors on VES capacity is still lacking.

Another challenge in this approach is its dependence on baseline for its definition. The baseline is not possible to measure if a load is providing VES services, only the total power is, leading to the issues of estimating the baseline and associated estimation errors [14, 15].

2.2 Matching VES Resources to Grid's Needs

The grid needs controllable resources to meet the net demand. The net demand[2] $p_d(t)$ at time t is defined as

$$p_d(t) := p_b(t) - g_r(t) \tag{2}$$

[2]Usually called net-load, but we avoid that term since "load" in this paper refers to physical entities that consume power.

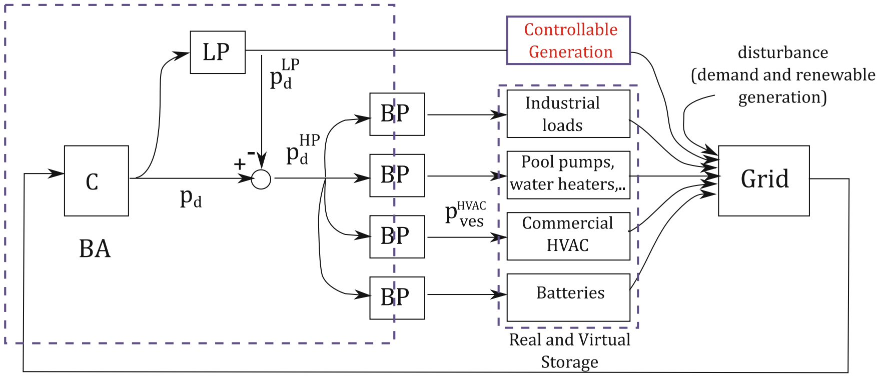

Fig. 4 A potential control architecture for the smart grid with VES based on spectral decomposition. The "Grid" block represents everything other than controllable generation and storage resources, such as loads (baseline), transmission and distribution networks, etc.

where $p_b(t)$ is the *baseline* power demand (in MW) in the grid and $g_r(t)$ is the uncontrollable renewable generation (in MW). The word baseline refers to the nominal demand from all loads, when loads are operated without employing any of the algorithms designed to extract flexibility. The net demand is the signal the grid's remaining resources will have to provide, which include traditional generators, flexible loads providing VES, and other energy storage (ES) devices such as pumped hydro, flywheels, and batteries.

How to ensure that available resources together supply the total needs of the grid, i.e., how do they together track the net demand? Our approach is based on a spectral decomposition of the net demand into distinct frequency bands, by passing it through a number of bandpass filters, as shown in Fig. 4. The "C" block at the BA computes/predicts the net demand p_d, which serves as a reference command to the aggregate controllable resources in the grid. Its low-pass component, $p_d^{LP}(t)$, is obtained by passing p_d through a low-pass filter ("LP" in Fig. 4). As long as the low-pass filter LP is designed by keeping the ramping abilities of the controllable generators in mind, the bandwidth of the signal $p_d^{LP}(t)$ will be low enough that controllable generators will be able to track it. The remaining high-pass component of the net demand is $p_d^{HP}(t) := p_d(t) - p_d^{LP}(t)$, which is *zero mean*. Because of the zero-mean property, $p_d^{HP}(t)$ can be tracked by controllable storage resources (whether real or virtual), by charging when $p_d^{HP}(t)$ is positive and discharging when $p_d^{HP}(t)$ is negative. The bandpass filters (BPs in Fig. 4) can be located either in a centralized manner at the BA, or in a distributed manner at the resources, or in some combination thereof, depending on the control architecture chosen.

To match to resources of appropriate ability, the zero-mean component of the net demand is passed through a number of bandpass filters to create reference signals for various energy storage resources: the "BP"s in Fig. 4. Each of the reference signals is band-limited to a particular frequency band that is suitable for a distinct class of resource. For instance, the highest frequency component of the net demand can be

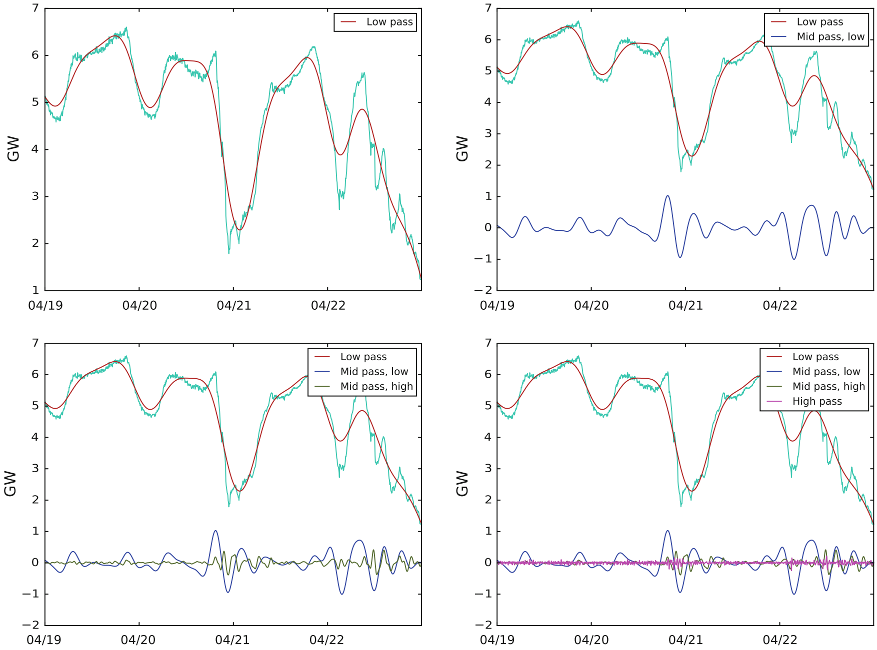

Fig. 5 Frequency decomposition of the net demand of Fig. 1: each bandpass- filtered component is a reference for a distinct class of resource that is appropriate for that frequency band

the reference signal for batteries, while the one with a slightly lower frequency can be the reference for HVAC loads providing VES. The sum of all these reference signals is the net demand. Thus, the needs of the grid are met, and yet no resource (including a conventional generator and a battery) is asked to provide a service that is not appropriate for it. Figure 5 shows an example of the frequency decomposition of the net demand based on data from BPA.

Challenges and opportunities

- *VES capacity characterization*: Based on experiments in a commercial building in the University of Florida reported in [12], we know that variable speed fans in HVAC systems can provide VES service in the frequency range of [1/(10 min) 1/(1 min)] and up to 30% of their average power without any perceptible change in indoor climate. Simulations with calibrated models show that with both chillers and fans engaged, HVAC systems can provide VES service in a slower frequency range of [1/(1 hr)]; [1/(10 min)] and up to 50% of its rated power, with an indoor temperature deviation of 2 °C [15]. Collection of pool pumps can provide VES in lower frequencies of hours [10], and so can residential air conditioners and heat pumps [16]. Industrial loads may be able to provide much lower frequency VES—than, say, HVAC—by deferring production in a timescale of days or weeks.

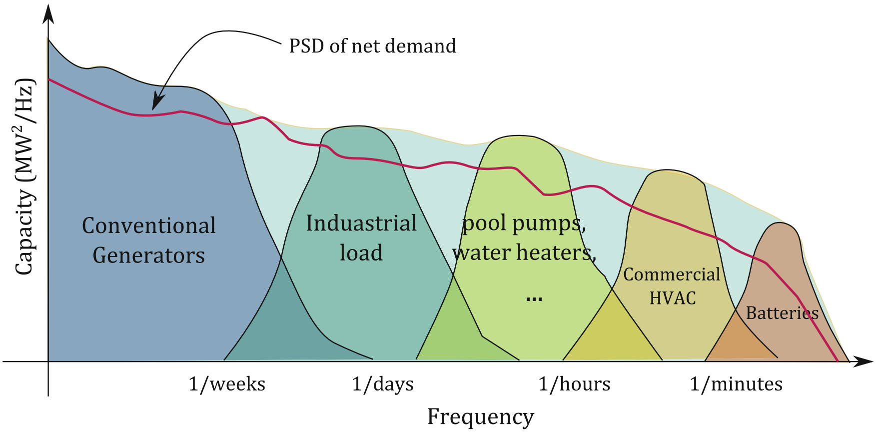

Fig. 6 Power spectral density of the net demand versus total capacity of the grid's resources

An important open question is to provide a complete characterization of the VES capacity of various classes of loads—especially industrial and residential loads—as a function of QoS within the frequency domain framework introduced here. Even for HVAC, which has been more thoroughly examined, VES capacity is likely to vary depending on the thermal load it experiences. A purely frequency domain framework may not be suited to characterize these variations [14].
Information on VES capacity as a function of QoS constraints is essential for loads to enter into contractual agreements with BAs. The appropriate payment structure is not clear yet, but at the simplest form it can be a fixed monthly payment depending on the load's QoS bound q. For more sophisticated loads such as industrial loads or large commercial HVAC, the payment can also consist of a "milage" payment depending on the actual VES service the load provided [17].

- *Ensuring resource adequacy*: The combined capacity of various resources (generation, VES, and RES resources) must be larger than the net demand. The grid's needs can again be quantified by the PSD of the net demand. Figure 6 illustrates a hypothetical scenario in which resources are adequate: the capacity curves of each category of resources—limited to various frequency bands due to QoS constraints—including conventional generators, VES resources, and RES resources, together cover the PSD at all frequencies. In this case, we can say that adequate resources exist.
- *Optimal allocation of VES and battery storage*: The cost of various types of VES resources are likely to be distinct. How much of each kind should a BA recruit to meet its requirements with sufficient margin at the minimum cost? Methodologies for answering such questions are essential to the BAs for planning purposes. Currently, a bottleneck in answering this question is the lack of estimates of VES cost. Section 4 discusses cost of battery storage that provides an upper bound for allowable cost of VES before VES becomes noncompetitive with battery-based energy storage.

3 Coordination of Loads to Obtain Required VES

To obtain VES service without violating QoS constraints, a two-tier strategy is required: local control and coordinated control. The local controller ensures that the load's QoS constraints are respected. Each load can provide only a small amount of VES, so a large number of loads need to act together to provide the desired VES service, which is ensured by the coordination algorithm.

In this paper, we consider *loads with continuously variable demand (LCVD)* such as commercial HVAC systems with variable speed drives. The demand of such a load can be varied to be any number within a range. In contrast, many residential loads can only be turned on or off; their demand cannot be continuously varied. However, if a load aggregator is used, the aggregator becomes a LCVD from the BA's point of view even if the all loads managed by the aggregator are on/off type [16].[3]

3.1 Local Intelligence

Here, the task is for the power demand deviation (from the baseline) $p_{\text{ves}}(t)$ of a load to track an external reference. The external reference must satisfy the bandwidth constraint described in Sect. 2 to maintain QoS, which can be ensured by locally bandpass filtering a grid-level reference.

Challenges and opportunities

- *Baseline uncertainty*: The challenges in designing the local intelligence to ensure tracking is measuring the output, the power deviation from the baseline, since the baseline, by definition, cannot be measured. In [12], this challenge was addressed by exploiting timescale separation between the VES reference to be tracked and the baseline. Since the baseline power consumption is dictated by the normal climate control system, it is of lower frequency than the high-frequency VES reference the system was designed to track. As a result, the baseline can be recovered by low-pass filtering the power consumption measurement.
  When the VES reference signal is of the same timescale as the baseline, the problem of separating the baseline becomes quite challenging. In our prior work [15] as well as in [13, 14], the baseline was prespecified by solving an optimization problem that ensured QoS (indoor climate) constraints were satisfied. The local controller was then tasked with tracking the total power: baseline plus VES reference.
- *Continuously variable demand from on/off actuators*: Chillers in commercial buildings are a much bigger load than fans, but they are predominantly on/off actuators, since their motors do not have variable speed drives. It is still possible to vary their power demand continuously in a range by indirect means, such as

[3] The problem of controlling an aggregate of on/offloads so that the power consumption of the collective tracks a smooth signal while respecting every load's QoS constraints has a different set of challenges that we do not go into in this paper; see [10].

airflow rate, due to the inlet guide vane controls. However, models of appropriate complexity that can be used to design and study local controllers for such equipment are lacking. Existing dynamic models of chillers are too complex for control design; e.g., [18]. A similar issue exists for packaged air conditioning units used in small commercial buildings, which may have variable speed fans but constant speed compressor motors. For chillers, especially larger ones, avoiding short-cycling is a key QoS requirement.

- *Round trip efficiency*: For thermal loads such as air conditioners, it is not clear if there is a loss of efficiency in varying their demand over a baseline instead of running them at their baseline. In other words, what is the "round trip efficiency" of the virtual battery? Work in this direction is preliminary [19].

3.2 Coordination

How does one break up the grid-level reference signal among many LCVD, each with its own QoS constraints? For the purpose of exposition, let us limit our attention to one particular frequency band, say, the component—p_{ves}^{HVAC} in Fig. 4—that will be supplied by commercial HVAC systems.

One possibility is for the grid to broadcast p_{ves}^{HVAC} and each load locally bandpass filters it to compute its own VES reference signal. This architecture is shown in Fig. 7: the goal is to ensure $y(t) = r(t)$, where $r(t)$ is the grid-supplied reference signal for demand deviation. The bandpass filter $F_i(s)$ at load i has to be designed so that load i's QoS is satisfied and the grid-level tracking goal, $y = r$, is also satisfied. Load i's QoS will be satisfied if the PSD of its local reference signal lies within its capacity curve $c_i(\omega)$. Recall that capacity curve was defined in Sect. 2.1. Note that if $P_{ves}^{HVAC}(\omega)$ is the PSD of the grid-level reference signal $p_{ves}^{HVAC}(t)$, then the PSD of the i-th load's local reference is $|F_i(j\omega)|^2 P_{ves}^{HVAC}(\omega)$. The CL_i block in Fig. 7 represents the closed-loop system consisting of a load and its local intelligence that

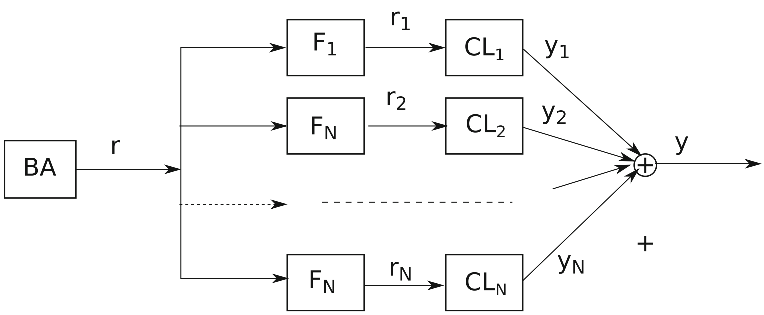

Fig. 7 Part of an open-loop coordination architecture for an aggregate of LCVDs to track a grid-level VES reference. Only the forward path between the BA and the loads are shown; the outer loop feedback between the "Grid" block of Fig. 4 and the BA is omitted

can track a reference signal for its demand deviation. The load belonging to the LCVD class is crucial; only such a load can track a reference other than a square-wave. Assuming the local intelligence at each load i is such that it tracks the local reference signal $r_i(t)$ perfectly, i.e., $y_i(t) = r_i(t)$, the equation $\sum_i r_i(t) = r(t)$ must be satisfied for the grid-level tracking goal to be satisfied. That is, if there are N loads supplying VES in the "high pass" category, then the following must hold to ensure that the loads together track the grid-level reference:

$$\sum_{i}^{N} F_i(j\omega) = 1, \quad \omega \in [\omega_L^{(HP)} \quad \omega_H^{(HP)}]. \tag{3}$$

When the grid operator enters into an agreement with a load to obtain VES resource, it obtains the load's VES capacity curve $c_i(\omega)$, either through modeling or through a system identification test. The local bandpass filter F_i is mutually agreed upon at that time. The grid operator must engage enough loads to ensure that (3) holds.

Even though this architecture satisfies the needs of both the grid and the loads, it lacks robustness to uncertainty due to its open-loop nature. There are many sources of uncertainty: the number of loads providing service at any given time, the capacity of some of the loads, etc., are all likely to vary over time in less-than predictable manner.

An alternate, more robust, architecture using feedback is proposed in [20], in which load coordinate their actions by using a global feedback signal that can be measured locally. Figure 8 shows this architecture. In particular, each load measures the grid frequency, which can be locally measured at loads [21, 22]. Since the deviation of the grid frequency from its nominal value (60 Hz) is a measure of demand–supply mismatch, it can estimate the demand–supply mismatch from this measurement. Since total supply is conventional plus renewable generation, the demand–supply mismatch—total demand minus total supply—is precisely the net demand minus conventional generation, so it is the zero-mean component of the net demand after the low-pass component is removed. The load computes the appropriate VES reference for itself by passing the estimated demand–supply imbalance with its local bandpass filter.

The control algorithm proposed in [20] goes one step further, and assumes that the BA broadcasts a prediction (for the next hour) of the demand–supply imbalance. The BA is in a unique position to predict this signal, since it has statistical models to predict grid-level baseline demand $d_b(t)$ and renewable generation $g_r(t)$, and it can predict the power generation by conventional generators $g_c(t)$ based on the contracts in place. The VES controller at each load uses an MPC scheme to compute appropriate power deviation (VES reference) subject to a QoS constraint expressed in terms of the Fourier transform of its local reference. High gain feedback due to the actions of other loads is avoided by estimating the VES supplied by other loads from the estimating the grid-level demand–supply imbalance and its own VES signal. The grid-level demand–supply imbalance is estimated from locally measured grid frequency.

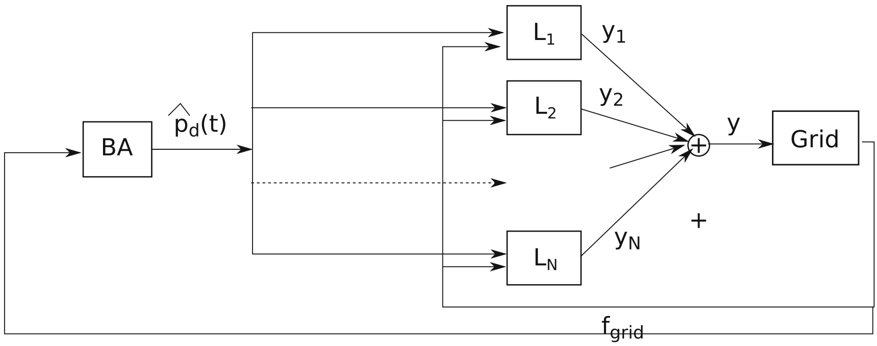

Fig. 8 A potential control architecture for coordination among VES resources by using local feedback (on grid frequency f_{grid}) and broadcast from BA on the predicted demand–supply imbalance $\hat{p}_d$

An extremely simplified schematic representation of this architecture, with N loads $L_1, L_2, \ldots, L_N$, is shown in Fig. 8. The goal is not for the aggregate response y to track some BA-supplied reference. Rather, it is to determine y_i's so that the aggregate response y minimizes demand-generation mismatch and each y_i satisfies the QoS constraint of load i.

The advantage of this architecture is that it is much more robust to uncertainty in how many loads are providing VES service at a given time and what their capacities are. In addition, distributed coordination among loads is achieved without any sort of inter-load communication. Only one-way broadcast from the BA to the loads is needed. Simulation studies reported in [20] shows the architecture is effective in providing robust tracking in presence of uncertainty.

Resource adequacy can be ensured by the BA by signing enough contracts so that the following holds:

$$|\sum_{i}^{N} c_i(j\omega)| > 1, \quad \omega \in [\omega_L^{(HP)} \;\; \omega_H^{(HP)}], \tag{4}$$

where $c_i(\omega)$ is the capacity curve of the ith load. The subscript q^* in $c_{q^*}(\omega)$, which was used in defining the capacity curve in Sect. 2.1 is suppressed here to avoid clutter. The advantage is that the inequality (4) is far easier to ensure than the equality (3), especially when a large number of loads are involved.

Challenges and opportunities

- *Communication architecture*: A large body of literature exist on distributed control, and the architectures discussed above are not the only possible ones. Most of the distributed coordination architectures proposed in the literature rely on inter-agent communication *within a neighborhood* for meeting network-wide goals. With the recent push toward an Internet of Things (IoT) paradigm, it is likely that smart loads will be part of the IoT. In that case, it is not clear what an appropriate notion of neighborhood is. All to all communication may be infeasible, but there is no rationale for limiting to a geographically defined neighborhood. Communicating with very far off (in a geographic sense) agents may be possible over the Internet. That may help with certain performance metrics, but may introduce larger delays. Determining these tradeoffs for distributed control in the age of IoT remains an important open question, one that is particularly relevant to the smart grid.

- *Contract/mechanism design*: A load may not provide the maximum capacity that was used at the time of signing contracts. That may not be malicious; if all of them provide maximum capacity at all times that may, in fact, cause demand–supply mismatch. If some loads bear a much larger share of the burden of required storage, it is reasonable they should be incentivized more than others. It is not clear what is an appropriate incentive to loads providing VES in such a scenario. Currently, generators in many ISOs are paid based on a two-part scheme based on capacity and mileage, but such a scheme may not be scalable to millions of loads.
- *Characterizing loads on-line*: The capacity of a load needs to be known to ensure that the loads together have enough bandwidth to track the reference. This can be done through a system identification experiment, as was done in [12] for the fan motor of an HVAC system. However, such a method may not be scalable to a large number of loads, and it may fail to identify slow variations in load's VES capacity over long time periods. Is it possible for the BA to be sure—without examining every single load—that the loads together have enough capacity to meets its need?

4 Cost

Without a cost advantage over real energy storage, virtual energy storage has little justification. Cost of VES is hard to estimate. On one hand, VES involves a change of software, with little change in hardware. Yet, the cost of large-scale deployment of VES may vary a lot depending on the kind of communication infrastructure and hardware retrofits needed. Cost of retrofitting existing consumer loads to make them VES-friendly is likely to be prohibitive, but it is equally likely that the additional cost of equipping loads with the required hardware and software at the factory will be negligible. However, precise estimates are lacking at this point.

Although the cost of VES may be hard to estimate at this point, we can establish an upper bound on the cost of VES beyond which VES loses its economic advantage. This upper bound is the minimum cost of the main competitor of VES, that of battery storage.

To estimate the cost of battery-based storage, we examine how the levelized cost of electricity (LCOE) will change if a battery is used to store the average daily generation of energy from an intermittent renewable source, say solar. The LCOE is the total cost incurred in the lifetime of the generator, divided by the total energy generated over the same period.

Consider a renewable generator with peak generation capacity 1 kW. Suppose the capacity factor of the generator is f, so that the average energy it produces in a day is $24f$ kWh. Let the lifespan of the generator be t_{lg} years. The total energy generated by the generator in its lifespan is $365t_{lg}24f$ kWh.

Suppose a battery is added to the generator so that it can store the average daily energy produced. That is, the energy capacity of the battery is $24f$ kWh. Let the lifetime of the battery be t_{lb} years, and its cost be c \$/kWh. Then, the cost of batteries over the life of the renewable generator is $24fct_{lg}/t_{tb}$ \$.

Since adding a battery does not change the energy generated, the *additional* LCOE due to the battery is the total cost of battery over the lifetime of the generator divided by the total energy generated during the same period:

$$\Delta\text{LCOE}_{\text{battery}} = \frac{24fc\frac{t_{lg}}{t_{tb}}}{365t_{lg}24f} = \frac{c}{365t_{lb}}. \quad (\$/\text{kWh}) \tag{5}$$

Among the myriad types of batteries, Sodium Sulfur (NaS) batteries have had a lead in terms of grid storage, but the cost of Li-ion batteries—used in mobile phones and electric cars—is decreasing the fastest: at an annual rate of approximately 14% per year during 2009–2014 [23]. The cheapest Li-ion batteries in 2015 cost about \$300/kWh (batteries used in Tesla's model S electric car [23]), and they have a lifetime of approximately 5000 charge–discharge cycles [24]. If the battery undergoes one charge–discharge cycle every day, its lifespan will be $5000/365 = 13.7$ years.

Plugging $c = 300$ and $t_{lb} = 13.7$, we see that the additional LCOE due to batteries is ≈ 6 ¢/kWh. Since several important costs are ignored here, especially the cost of balance of systems and the cost of capital, the true cost will be higher than this estimate. A more thorough cost estimate can be performed using the methodology in [24]. Even this low estimate of battery cost is quite high compared to the mean retail electricity rate in the U.S., which in December 2016 was 12.2 ¢/kWh (from https://www.eia.gov/electricity/). If we take the estimate, 6 ¢/kWh, as the true cost of battery storage, the cost of VES must be less than 6 ¢/kWh for it be competitive with battery-based energy storage.

In comparing batteries with VES, one should keep in mind that battery-based energy storage is likely to be much more reliable than VES. Availability of VES may depend on time of day, weather, etc., while batteries are a firm resource. Therefore, an optimal solution will probably consist of expensive but highly reliable batteries as well as inexpensive but less reliable VES.

5 Summary

Loads can vary their power around a baseline in a zero-mean fashion to effectively act like batteries, thereby providing virtual energy storage (VES) to help the grid. A frequency domain framework for characterizing loads flexibility vis-a-vis consumer's QoS is advocated, following [7]. The framework is powerful enough to handle not just flexible loads but also conventional generators and batteries. However, it is highly simplified: issues of transmission constraints, distribution network and voltage support, contingency reserves are not considered yet, which are worthwhile avenues for further refinement. Some results on local control and distributed coordination of loads within this framework, are mentioned. Challenges and opportunities in extending this framework to design reliable VES services, including some of the open problems, are summarized.

Acknowledgements Financial support from NSF (through grant 1646229) and DOE (through a BTO-GMLC project titled *virtual batteries*) is gratefully acknowledged.

References

1. G. Parkinson, UBS: closures coal and gas fired power plants in Europe accelerating (2015). http://www.energypost.eu/ubs-closures-coal-gas-fired-power-plants-europe-accelerating/. Accessed 4 May 2015
2. J.H. Yoon, R. Baldick, A. Novoselac, Dynamic demand response controller based on real-time retail price for residential buildings. IEEE Trans. Smart Grid **5**(1), 121–129 (2014)
3. D.J. Hammerstrom, R. Ambrosio, J. Brous, T.A. Carlon, D.P. Chassin, J.G. DeSteese, R.T. Guttromson, G.R. Horst, O.M. Jrvegren, R. Kajfasz, S. Katipamula, L. Kiesling, N.T. Le, P. Michie, T.V. Oliver, R.G. Pratt, S. Thompson, M. Yao, Pacific northwest gridwise testbed demonstration projects. Part I. Olympic peninsula project. Pacific Northwest National Laboratory, Tech. Rep. PNNL-17167, Oct 2007
4. G. Wang, M. Negrete-Pincetic, A. Kowli, E. Shafieepoorfard, S. Meyn, U. Shanbhag, Real-time prices in an entropic grid, in *IEEE PES Innovative Smart Grid Technologies (ISGT)*, Jan 2012, pp. 1–8
5. J. Knudsen, J. Hansen, A.M. Annaswamy, A dynamic market mechanism for the integration of renewables and demand response. IEEE Trans. Control Syst. Technol. **24**(3), 940–955 (2016)
6. FPL on call saving program (2013). http://tinyurl.com/k3ldwe3
7. P. Barooah, A. Bušić, S. Meyn, Spectral decomposition of demand side flexibility for reliable ancillary service in a smart grid, in *48th Hawaii International Conference on Systems Science*, Jan 2015 (invited paper)
8. B. Kirby, Ancillary services: technical and commercial insights (prepared for Wärtsilä North America Inc, 2007)
9. A. Nayyar, M. Negrete-Pincetic, K. Poolla, P. Varaiya, Duration-differentiated energy services with a continuum of loads. IEEE Trans. Control Netw. Syst. **3**(2), 182–191 (2016)
10. S. Meyn, P. Barooah, A. Bušić, Y. Chen, J. Ehren, Ancillary service to the grid from intelligent deferrable loads. IEEE Trans. Auto. Control **60**, 2847–2862 (2015)
11. D. Todd, M. Caufield, B. Helms, A. Generating, I. Starke, B. Kirby, J. Kueck, Providing reliability services through demand response: a preliminary evaluation of the demand response capabilities of alcoa inc. ORNL/TM **233** (2008)
12. Y. Lin, P. Barooah, S. Meyn, T. Middelkoop, Experimental evaluation of frequency regulation from commercial building HVAC systems. IEEE Trans. Smart Grid **6**, 776–783 (2015)
13. H. Hao, D. Wu, J. Lian, T. Yang, Optimal coordination of building loads and energy storage for power grid and end user services. IEEE Trans. Smart Grid **PP**(99), 1–1 (2017)
14. E. Vrettos, F. Oldewurtel, G. Andersson, Robust energy-constrained frequency reserves from aggregations of commercial buildings (2015). arXiv:1506.05399v1 [cs.SY]
15. Y. Lin, P. Barooah, J. Mathieu, Ancillary services through demand scheduling and control of commercial buildings. IEEE Trans. Power Syst. (2017)

16. B. Biegel, P. Andersen, J.S. Mathias, B. Madsen, L. Henrik, H. Lotte, H. Rasmussen, Aggregation and control of flexible consumers—a real life demonstration. IFAC, 1–6 (2014)
17. FERC, Frequency Regulation Compensation in the Organized Wholesale Power Markets, FERC 755, Docket Nos. RM11-7-000 and AD10-11-000 , Oct 2011, pp. 1–123
18. S. Bendapudi, J. Braun, E. Groll, A dynamic model of a vapor compression liquid chiller, in *International Refrigeration and Air Conditioning Conference* (2002)
19. I. Beil, I. Hiskens, S. Backhaus, Round-trip efficiency of fast demand response in a large commercial air conditioner. Energy Build. **97**, 47–55 (2015)
20. J. Brooks, P. Barooah, Demand control for maintaining grid frequency without affecting consumers quality of service through decentralized bandwidth-constrained MPC, in *American Control Conference* (2017, accepted)
21. Y. Liu, A US-wide power systems frequency monitoring network, in *Power Engineering Society General Meeting* (2006)
22. P.J. Douglass, R. Garcia-Valle, P. Nyeng, J. Stergaard, M. Togeby, Demand as frequency controlled reserve: implementation and practical demonstration, in *2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, Dec 2011, pp. 1–7
23. B. Nykvist, M. Nilsson, Rapidly falling costs of battery packs for electric vehicles, in *Nature Climate Change*, vol. 5 (2015), pp. 329–332
24. B. Zakeri, S. Syri, Electrical energy storage systems: a comparative life cycle cost analysis. Renew. Sustain. Energy Rev. **42**, 569–596 (2015)