Virtual energy storage through decentralized load control with quality of service bounds

Jonathan Brooks and Prabir Barooah
University of Florida
Gainesville, Florida USA

***Abstract*— We propose a decentralized algorithm to help reduce demand-supply imbalance in a power grid by varying the demand from loads, just like charging and discharging a battery. The algorithm ensures strict bounds on the consumers' quality of service (QoS) by constraining the bandwidth of demand variation. A model-predictive-control formulation is adopted to compute local decisions at the loads. The algorithm is decentralized in the sense that loads do not communicate with one another. Instead, loads coordinate using local measurements of the grid frequency, which provide information about global demand-supply imbalance. It is envisioned that consumers will be recruited through long-term contracts, aided by the QoS guarantees provided by the proposed scheme. Simulation results show that loads are able to reduce frequency deviations while maintaining QoS constraints and that the performance of the algorithm scales well with the number of loads. Closed-loop stability is established under some assumptions.**

I. Introduction

Reliable operation of the power grid requires generation (supply) to equal demand at all timescales. Intermittent renewable energy sources (solar and wind) exacerbate demand-supply imbalance, requiring additional balancing resources, but building additional conventional generators or installing large batteries is expensive. An alternative is to vary the demand to help restore demand-supply balance. Many loads have some flexibility in demand, e.g., air conditioners [1]. Their demand can be varied around a nominal demand profile (baseline) to absorb the volatility of renewable generation—like charging and discharging a battery. In effect, controllable loads provide *Virtual Energy Storage* (VES). The key challenge is to ensure that any loss of consumers' quality of service (QoS) is limited and predictable.

Most works on demand-side management involve use of real-time electricity prices to encourage loads to shift or reduce demand [2, 3]. An advantage of a price-based scheme is that it enables decentralized coordination where market clearing price (computed and broadcast by the balancing authority (BA)) provides a feedback signal all loads can use [4]. A price-based scheme suffers from several difficulties. First, it assumes that consumers are willing to endure some loss of QoS in return for a payment [3, 5]. There has been some work on minimizing the QoS loss experienced by consumers when deciding demand changes [6, 7]. However, modeling the loss of QoS as a function of demand variation and computing the dollar value of that loss is challenging. Second, consumers face large uncertainty in how much financial reward they may receive since the reward depends on real-time prices that historically show spiky behavior [8]. In fact, a survey of demand response found price volatility to be a significant limiting factor in consumer participation [9]. The third problem is uncertainty faced by the grid operator. Currently, generating units are committed through a bidding process in day-ahead and hour-ahead markets, whose clearing requires solving an optimization problem. The information needed to solve this problem involves capacity, ramp rate, and marginal cost of all generating units, along with forecast of demand-supply imbalance in the grid. Including a large number of consumer loads—with an uncertain cost of demand variation—into this optimization is infeasible.

In this paper, we propose an alternative approach for control of flexible loads to reduce demand-supply imbalance. The proposed approach does not involve an aggregator or real-time prices and reduces uncertainty to both consumers and grid operators. It consists of two main ideas: (1) constraining the *bandwidth* of demand variation (i.e., magnitude of the Fourier transform of demand variation) to ensure strict, predetermined bounds on the QoS, and (2) decentralized coordination based on local grid-frequency measurements.

Constraining bandwidth is inspired by prior work that argues that the change in a consumer's QoS is a function of the frequency characteristics of the demand variation [10]. E.g., variations in the air flow rate of an HVAC system in a large building will have a negligible effect on the indoor climate if the variation is of sufficiently high frequency [1, 11]. Demand variation at lower frequencies will lead to loss of QoS; e.g., $\pm 2^{-}$ °F variation in indoor temperature [12]. Similarly, pool pumps can provide service to the grid without affecting pool owners' QoS as long as the demand variations are limited to certain timescales (frequencies) [13], and this may apply to aluminum-smelting plants as well [14].

In the proposed approach, each load employs a local model-predictive-control (MPC) scheme to compute the control (demand variation) using prediction of demand-supply imbalance from the grid operator. It is assumed that the BA broadcasts these predictions over the Internet, which it obtains from forecasts of demand and renewable generation. The controller solves an optimization problem that does the best it can to reject the disturbance faced by the grid subject to local bandwidth constraints. Coordination is performed

This research was partially supported by NSF grants 1646229 and 1463316.

by using measurements of the grid frequency, which can be measured by each load locally using frequency disturbance recorders [15] and is related to the grid-level demand-supply imbalance [16, 17]. Thus, every load is able to infer global information from local measurements without requiring inter-load communication. We call the proposed control scheme the bandwidth-limited, disturbance-rejecting, decentralized model-predictive control (BaLDuR-DMPC).

It is envisioned that each load signs a contract with the BA that specifies the bandwidth of demand variation the consumer signs up for. That guarantees the QoS loss the consumer may experience for all time. For example, if a consumer is willing to tolerate a maximum temperature deviation (from the nominal value) of $2\,^{\circ}$F, then the bandwidth constraint will be tighter and monthly payment will be smaller compared to those for a consumer who is willing to tolerate a $4\,^{\circ}$F change. This contract-based mechanism makes it easy for the grid operator to estimate how much VES capacity it can count on at every instant. By the same reason, the consumer knows ahead of time how much money she will make and maximum discomfort she may experience.

One class of loads will not have sufficient flexibility to meet all the needs of the grid. It is argued in [10] that a spectral decomposition of the demand-supply imbalance can be used to assign the appropriate part thereof to the appropriate resource. For instance, the slowly varying part of the net demand (demand minus renewable generation) of a future grid can be handled by traditional generators. A "mid-pass" component can be provided by pool and irrigation pumping, a "high-pass" component can be provided, HVAC systems in buildings, etc. This is illustrated in Figure 1. Since the BA signs long-term contracts with various flexible loads, it knows ahead of time whether it has adequate capacity at all frequencies (timescales). Each class of resources therefore only focuses on the task of rejecting the part of the imbalance that its bandwidth constraints allow.

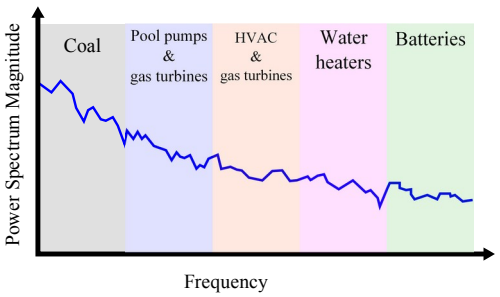

Fig. 1: Illustration of the grid's regulation needs [8].

Our work is also related to [18–20]. These works use strict output constraints to ensure QoS while optimizing a market-based objective function. This approach, however, requires an accurate model of each load's dynamics. In contrast, the proposed approach requires an imprecise characterization of a load: the local controller only needs a conservative estimate of the region in the magnitude-frequency plot that the demand variation needs to stay within to satisfy the QoS requirements of the consumer. Such a characterization can be done at the factory for consumer loads like water heaters.

Compared to prior work on demand control for grid balancing, this paper makes three contributions. Though the idea of limiting the bandwidth of demand variation to ensure QoS bounds was espoused in our earlier work [10], to the best of our knowledge, this paper is the first to use explicit bandwidth constraints in an MPC formulation for demand control. The second contribution is a novel use of locally obtained grid-frequency measurements for distributed coordination without using communication, in which disturbance predictions are scaled based on local frequency measurements. The third contribution is analysis of the closed-loop system under the proposed BaLDuR-DMPC scheme. There have been a number of papers on applying distributed MPC schemes for load control in the power grid (e.g., [18–20]). However, these references do not provide any analysis of closed-loop behavior. Although the analysis presented in this paper is preliminary and makes strong assumptions, results from dynamic simulations suggest that the algorithm performs well even when the assumptions are violated.

Our focus is on loads providing grid support in the intermediate time-scale of a few minutes to a few hours. Although in principle it can be extended to faster time-scales, some of the assumptions we have made—such as availability of disturbance predictions through periodic broadcasts from the BA—may not hold in those timescales.

The rest of the paper is organized as follows. Section II introduces the problem to be solved. The proposed BaLDuR-DMPC algorithm is described in Section III. Preliminary analytical results are presented in Section IV. Dynamic-simulation results are shown and discussed in Section V. Finally, Section VI concludes this work.

II. Problem Formulation

Let time be measured by the discrete iteration counter: $k = 0, 1, \cdots$, and let T be the discretization interval. The control action at load i at time k is the demand variation from the nominal value, denoted by $v_k^{(i)}$.

The grid is modeled as a plant with input equal to the real power injected and output equal to the grid frequency. We only consider the linearized grid dynamics around the nominal system frequency, 60 Hz, with the nominal input being the net real power (generation minus nominal demand) injected to maintain the nominal frequency. The linearized plant, discretized with sampling period T, is denoted by $G(z)$. The input to $G(z)$ is the sum of disturbance, d_k, and the control actions of the loads; denote this input by μ_k. Let ϑ_k be the total control action by the loads: $\vartheta_k = \sum_{i=1}^{n} v_k^{(i)}$. Then, $\mu_k := d_k + \vartheta_k$. The output of $G(z)$ is the deviation of the system frequency from 60 Hz and is denoted by ω_k. In the Z-transform domain, $\omega(z) = G(z)\mu(z)$.

The goal of the loads is to arrest frequency deviations from 60 Hz while ensuring no loss in QoS for consumers. We use MPC because it is readily able to enforce constraints. Let N be the prediction horizon for the MPC at every load.

The control at load i—its demand variation—must satisfy the bandwidth constraints of that load. These are specified in terms of the DFT of the control signal. The L-point DFT of the control inputs, $v_k^{(i)}$, at load i is

$$V_m^{(i)} := \sum_{k=1}^{L} v_k^{(i)} e^{-j\frac{2\pi}{N}mk}, \quad m = 0, \ldots, L-1$$

Bandwidth constraints on the control actions are enforced by specifying upper bounds, $\alpha_m^{(i)}$, on the magnitude of the control signal's L-point DFT: $|V_m^{(i)}| \le \alpha_m^{(i)}$, $m = 0, \ldots, L-1$, for an appropriately chosen L. The control actions must also satisfy upper and lower bound constraints, which are denoted by $\bar{u}^{(i)}$ and $\underline{u}^{(i)}$, respectively, for load i.

We assume that a prediction of disturbances affecting the grid are available to all loads (through periodic communication from the BA). Let $\hat{\boldsymbol{d}}_k^{(BA)} := [\hat{d}_{k|k}^{(BA)}, \ldots, \hat{d}_{k+N-1|k}^{(BA)}]^T$ be the prediction of the disturbance, $[d_k, \ldots, d_{k+N-1}]^T$, affecting the grid that is available to all loads at time k.

III. Proposed Method

A. Grid model used by loads

Since the proposed control scheme is MPC-based, each load needs a model of the plant, $G(z)$. In the proposed method, loads use an extremely simple model of $G(z)$ for control computations: a constant gain, which we call g. The rationale for choosing such a simple model is twofold. First, a simpler model aids in keeping control computations simple. Second, since the focus of this work is demand-supply balance at the slower timescale of a few minutes to a few hours, i.e., in the frequency range $f \in [\frac{1}{5\text{ min}}, \frac{1}{120\text{ min}}]$, the linearized model of a power grid at these frequencies, $G(e^{jf})$, is a nearly constant gain. This leads to the following model used by every load:

$$\omega_k = g\mu_k = g(d_k + \vartheta_k) = g(d_k + \sum_{i=1}^{n} v_k^{(i)}). \qquad (1)$$

B. MPC formulation for load i

Let $\boldsymbol{u}_k^{(i)}$ denote the decision variables of load i at time k:

$$\boldsymbol{u}_k^{(i)} := [u_{k|k}^{(i)}, \ldots, u_{k+N-1|k}^{(i)}]^T.$$

The first entry of the optimal value, $\boldsymbol{u}_k^{(i)*}$, which is denoted by $u_{k|k}^{(i)*}$, is implemented as the control, i.e., $v_k^{(i)} = u_{k|k}^{(i)*}$. The optimal value $\boldsymbol{u}_k^{(i)*}$ is obtained by solving the following optimization problem at time k by load i:

$$\min_{\boldsymbol{u}_k^{(i)}} \sum_{\ell=k}^{k+N-1} \omega_\ell^2 \qquad (2)$$

subject to the following constraints over the horizon $k \le \ell \le k+N-1$:

$$\omega_\ell = g(u_{\ell|k}^{(i)} + \hat{d}_{\ell|k}^{(i)}), \qquad (3)$$

$$\underline{u}^{(i)} \le u_{\ell|k}^{(i)} \le \bar{u}^{(i)}, \qquad (4)$$

$$|V_{m|k}^{(i)}| \le \alpha_m^{(i)}, \quad 0 \le m \le 2N-1, \qquad (5)$$

where $\hat{d}_{\ell|k}^{(i)}$) is the disturbance prediction used by load i and

$$V_{m|k}^{(i)} := \sum_{\ell=k-N}^{k-1} v^{(i)} e^{-j\frac{2\pi}{2N}m\ell} + \sum_{\ell=k}^{k+N-1} u_{\ell|k}^{(i)} e^{-j\frac{2\pi}{2N}m\ell} \qquad (6)$$

is the $2N$-point DFT of the array, $[v_{k-N}^{(i)}, \ldots, v_{k-1}^{(i)}, u_{k|k}^{(i)}, \ldots, u_{k+N-1|k}^{(i)}]$; recall that $v^{(i)}$ is the control action previously implemented by load i at time ℓ.

The reason for including past control inputs in the bandwidth constraint—through (6)—is the following. If only predicted values are used, even if the solution to the optimization problem satisfies the bandwidth constraint, because MPC only implements the first entry of the solution, the closed-loop control action may not satisfy the bandwidth constraint. This issue is faced in MPC schemes that limit the rate of change in control, where past data are used to enforce the constraint in the closed-loop [19]. The use of past data in computing the DFT in (6) was found to help maintain the bandwidth constraint in the closed-loop.

Obtaining necessary conditions for feasibility of the optimization problem (2)-(5) is left for future work. However, if (2)-(5) is feasible, then it can be shown that the problem is convex by convexity of the cost and constraints.

C. Avoiding high-gain feedback

Note that the system-model constraint (3) indicates that every load thinks it is the only load in the grid. The MPC scheme encourages each load to reject the predicted disturbance through its local control. The combined actions of all loads may lead to high-gain feedback and instability. To avoid this, load i does not use the grid-level disturbance prediction, $\hat{\boldsymbol{d}}_k^{(BA)}$, for $\hat{\boldsymbol{d}}_k^{(i)}$ in solving problem (2)-(5). Rather, the disturbance prediction is scaled according to the following: $\hat{\boldsymbol{d}}_k^{(i)} := \rho_k^{(i)} \hat{\boldsymbol{d}}_k^{(BA)}$, where $\rho_k^{(i)} > 0$ is a time-varying gain that is updated via

$$\rho_k^{(i)} = r_k^{(i)} \rho_{k-1}^{(i)}; \qquad (7)$$

the ratio, $r_k^{(i)}$, is computed from

$$r_k^{(i)} = \min\left(\max\left(\frac{\hat{d}_k^{(BA)}}{\vartheta_{k-}^{(i)}}, \underline{r}\right), \bar{r}\right), \qquad (8)$$

where $0 < \underline{r} \le 1 \le \bar{r} < \infty$ are predetermined bounds on $r_k^{(i)}$ (they ensure that $0 \neq r_k^{(i)} \neq \infty$) and $\vartheta_{k-}^{(i)}$ is an estimate obtained by load i at time k of the total control effort by all loads at time k *before* the new values of $v_k^{(i)}$ are implemented.

At time k, load i has access to a noisy measurement, $\tilde{\omega}_{k-}^{(i)}$, of the grid frequency, ω_{k-}, where $\tilde{\omega}_{k-}^{(i)} = \omega_{k-} + \eta_k^{(i)}$ with measurement noise $\eta_k^{(i)}$ and k^- retains its meaning from above. Recall from (1) that loads use the model $\omega_k = g\mu_k = g(d_k + \vartheta_k)$. Thus, load i estimates the input from the output measurements as $\mu_{k-}^{(i)} := \tilde{\omega}_{k-}^{(i)}/g$. From that, it obtains an estimate, $\vartheta_{k-}^{(i)}$, of the total demand variation, ϑ_{k-}, as

$$\vartheta_{k-}^{(i)} := \mu_{k-}^{(i)} - \hat{d}_k^{BA} = \frac{\tilde{\omega}_{k-}^{(i)}}{g} - \hat{d}_k^{BA}. \qquad (9)$$

With the update law (7)-(8), if there is too much demand variation at the previous time-step, the loads will reduce the magnitude of disturbance predictions to use less control in the next step, which will presumably bring the frequency deviation closer to 0. Conversely, if there was too little demand variation, they will increase the size of their disturbance predictions. In this way, the loads scale their disturbance predictions until the scaled predictions cause the loads to collectively execute the right amount of demand variation.

In summary, for load i at time k, the BaLDuR-DMPC algorithm is the following, with initial condition $v_0^{(i)} = 0$.

BaLDuR-DMPC Algorithm At every k, load i does:

1) Obtain the estimate, $\vartheta_{k-}^{(i)}$, via (9) from the local noisy measurement of grid frequency, $\omega_{k-}^{(i)}$.
2) Compute $r_k^{(i)}$ from (8) and $\rho_k^{(i)}$ from (7).
3) Compute $\hat{\boldsymbol{d}}_k^{(i)} = \rho_k^{(i)} \hat{\boldsymbol{d}}_k^{(BA)}$.
4) Solve problem (2)-(5) to obtain $\boldsymbol{u}_k^{(i)*}$.
5) Implement the control, $v_k^{(i)} = u_{k|k}^{(i)*}$.

The inputs to the algorithm for load i at time k are: the N past (executed) inputs $v_{\grave{}}^{(i)}$ ($\grave{} = k-N, \ldots, k-1$), the noisy frequency measurement, $\omega_{k-}^{(i)}$, and the grid-level disturbance prediction $\hat{\boldsymbol{d}}_k^{(BA)}$. The output is the control input, $v_k^{(i)}$. No inter-load communication is needed.

IV. Closed-Loop Stability

We now show that the proposed algorithm drives the output, ω_k, to 0 under some assumptions. Simulations show the controller is robust to those assumptions (see Section V).

Assumption 1.

1) *(Perfect disturbance prediction)* $\hat{d}_k^{(BA)} = d_k$ for all k.
2) *(No "plant-model mismatch")* $G(z) = g$.
3) *(Perfect frequency measurements)* $\omega_{k-} = \omega_{k-}$ for all k.
4) *(Loose bandwidth constraints)* α_m is sufficiently large so that $V_{m|k}^{(i)} < \alpha_m$ for all i, m, and k.
5) *(Sufficient actuation)* $\sum_{i=1}^{n} \bar{u}^{(i)} \geq \max_k |\hat{d}_k^{(i)}|$, and $\sum_{i=1}^{n} \underline{u}^{(i)} | \geq -\max_k |\hat{d}_k^{(i)}|$.

Assumptions 1(2),(3) imply $\vartheta_{k-}^{(i)} = \vartheta_{k-}$ for all i, k; Assumptions 1(4),(5) imply strict feasibility of problem (2)-(5).

Since ω is the deviation from the nominal system frequency, 60 Hz, the desired equilibrium of the closed-loop corresponds to $\omega_k = 0$ with $\sum_i v_k^{(i)} = -d_k$.

Theorem 1. *Suppose the disturbance affecting the grid is constant: $d_k = \delta$ for all k and some fixed δ. If Assumption 1 holds, $\underline{r} = 0$, and $\bar{r} = +\infty$, then $\omega_k \to 0$ as $k \to \infty$.*

It is straightforward to see from $\omega_k = g(d_k + \vartheta_k)$ that $\vartheta = -\delta$ and $\omega = 0$ is an equilibrium: once the control reaches this point at some k, $r_{k+1}^{(i)} = 1$ from (8), and $\rho_{k+1}^{(i)} = \rho_k^{(i)}$ by (7), meaning each load will use the same disturbance prediction at $k+1$ as at k, which will lead to the same control action at $k+1$ as at k; hence, $\vartheta_{k+1} = -\delta$ and $\omega_{k+1} = 0$. The assumptions of an accurate grid model and perfect frequency measurements and disturbance predictions are utilized in this argument. The theorem states that the closed-loop trajectories converge to this equilibrium.

Proof of Theorem 1. Without loss of generality, suppose $\delta > 0$; symmetric arguments apply for $\delta < 0$. The solution to problem (2)-(5) is the one that exactly cancels out the scaled predicted disturbance if actuator constraints allow. That is:

$$u_{k|k}^{(i)*} = -\gamma_k^{(i)} \rho_k^{(i)} \delta, \tag{10}$$

where

$$\gamma_k^{(i)} = \min\left\{ \frac{\bar{u}^{(i)}}{\rho_k^{(i)} \delta}, 1 \right\}. \tag{11}$$

Now, we have

$$\vartheta_k = -\delta \sum_{i=1}^{n} \rho_k^{(i)} \gamma_k^{(i)}. \tag{12}$$

The proof will proceed by showing $\rho_k^{(i)}$ converges, which implies $|\vartheta_k| \to \delta$ by (7) and (8). Since ϑ and δ are opposite signs by (10), $\vartheta_k \to -\delta$, which implies $\omega_k \to 0$ by (1).

From (8), using (7), (12), and $\vartheta_{k-}^{(i)} = \vartheta_{k-}$, we obtain

$$\rho_{k+1}^{(i)} = \frac{\delta}{\delta \sum_{j=1}^{n} \rho_k^{(j)} \gamma_k^{(j)}} \rho_k^{(i)} = \frac{\rho_k^{(i)}}{\sum_{j=1}^{n} \rho_k^{(j)} \gamma_k^{(j)}}. \tag{13}$$

Now, suppose $|\vartheta_{k_0}| > \delta$ for some k_0; symmetric arguments may be made if $|\vartheta_{k_0}| < \delta$. Then there is too much total demand variation at that instant. From (12), we then have $\sum_{i=1}^{n} \rho_{k_0}^{(i)} \gamma_{k_0}^{(i)} > 1$. By (13), we then have $\rho_{k_0+1}^{(i)} < \rho_{k_0}^{(i)}$, which implies $\gamma_{k_0+1}^{(i)} \geq \gamma_{k_0}^{(i)}$ by (11). Now, by iterating (13), we may observe

$$\rho_{k_0+2}^{(i)} = \frac{\rho_{k_0+1}^{(i)}}{\sum_{j=1} \rho_{k_0+1}^{(j)} \gamma_{k_0+1}^{(j)}} \frac{\frac{\rho_{k_0}^{(i)}}{\sum_{\grave{}=1} \rho_{k_0}^{(\grave{})} \gamma_{k_0}^{(\grave{})}}}{\sum_{j=1} \frac{\rho_{k_0}^{(j)}}{\sum_{\grave{}=1} \rho_{k_0}^{(\grave{})} \gamma_{k_0}^{(\grave{})}} \gamma_{k_0+1}^{(j)}}$$

$$= \frac{\rho_{k_0}^{(i)}}{\sum_{j=1} \rho_{k_0}^{(j)} \gamma_{k_0+1}^{(j)}} \leq \frac{\rho_{k_0}^{(i)}}{\sum_{j=1} \rho_{k_0}^{(j)} \gamma_{k_0}^{(j)}} = \rho_{k_0+1}^{(i)}.$$

Now, $\rho_{k_0+2}^{(i)} \leq \rho_{k_0+1}^{(i)}$ implies $r_{k_0+1}^{(i)} \leq 1$ by (7). Therefore, $|\vartheta_{k_0+1}| \geq \delta$ by (8). Hence, $|\vartheta_{k_0}| > \delta$ implies $\rho_{k_0+2}^{(i)} \leq \rho_{k_0+1}^{(i)}$, which implies $|\vartheta_{k_0+1}| \geq \delta$. By induction, it follows that $|\vartheta_k| \geq \delta$ for all $k \geq k_0$. Therefore, $\rho_{k+1}^{(i)} \leq \rho_k^{(i)}$ for all i and all $k \geq k_0$. Hence, if $|\vartheta_k| > \delta$ for any k, $\rho_k^{(i)}$ converges for all i because it is bounded below by 0.

Now, suppose $|\vartheta_k| < \delta$. Then, we have $\rho_{k_0}^{(i)} < \rho_{k_0+1}^{(i)}$. By symmetric arguments, $\rho_{k+1}^{(i)} \geq \rho_k^{(i)}$ for all $k \geq k_0$. Hence, $\rho_k^{(i)}$ either converges or increases without bound. However, if $\rho_k^{(i)}$ increases without bound, then it follows from (11) that $\sum_{i=1}^{n} \gamma_k^{(i)} \to 0$, which implies that $\rho_k^{(i)}$ converges—a contradiction because it was supposed that $\rho_k^{(i)}$ increases

without bound. Therefore, $\rho_k^{(i)}$ converges for all i, and the proof is complete.

V. Dynamic-Simulation Results

A. Simulation Setup

We use the linearized grid model,

$$G(s) = \frac{0.644s + 0.147}{s^2 + 0.4797s + 0.147}, \tag{14}$$

which was based on a study of the Texas grid in [21]. For simulations, we use a discretization interval of 1 second.

Control actions are discretized into 5-minute intervals; i.e., $T = 300$ seconds. Loads use $g = 1$ for their constant-gain model, with input in GW and output in Hz. Loads use a 2-hour prediction horizon, so $N = 24$; this means loads also use a 2-hour past horizon for bandwidth constraints.

The continuous-time disturbance to the grid is

$$d(t) = 0.05 \sin\left(\frac{2\pi}{3600}t\right) + 0.1 \sin\left(\frac{2\pi}{2 \times 3600}t\right) + \theta(t),$$

where t is in seconds and $\theta(t)$ is chosen from $N(0, 0.05^2)$. The disturbance prediction supplied to the loads from the BA, $\hat{\boldsymbol{d}}^{(BA)}$, is the discrete form of $\hat{d}^{(BA)}(t) = d(t) - \theta(t)$ with a sampling interval of $T = 300$ seconds.

The measurement noise (for grid-frequency measurements) is modeled as a zero-mean normal random variable with variance of 10^{-4}. For each i, we use $\overline{u}^{(i)} = -\underline{u}^{(i)} = \infty$ so that the effects of bandwidth constraints are clearly seen (rather than those of saturation). Finally, we use $\underline{r} = 1/\overline{r} = 0.95$ and $\rho_0^{(i)} = 0.01$ for all i.

To solve problem (2)-(5), each load uses a sequential-quadratic-programming (SQP) formulation. Any method may be used, but SQP is useful in the presence of nonlinear constraints, such as (5), and has robust theoretical foundations.

In the sequel, we present simulation results of the BaLDuR-DMPC algorithm for $n = 10$ loads and $n = 100$ loads with different choices for α_m, $m = 0, \ldots, 2N-1$.

B. Discussion

When $n = 10$ loads, the total objective value was $\sum_k \omega_k^2 = 84.3$, compared to 135 when $n = 0$ (no intelligent loads). Increasing the number of loads significantly improved performance because, with more loads, each load is responsible for a smaller portion of the disturbance, d. Results for $n = 100$ loads are presented in more detail below.

Since the bandwidth constraint is posed in terms of the DFT of a finite-length signal, the closed-loop control signal may not satisfy the bandwidth constraint even if the solution to problem (2)-(5) at every k satisfies it. This is because the bandwidth constraint (5) is enforced for each individual time-step using the past N control actions, but for the overall DFT, as the number of data points used to compute the overall DFT tends to infinity, the magnitude of the DFT also tends to infinity. At each time k, no $\boldsymbol{u}_k^{(i)*}$ violated (5).

Figure 2 shows the disturbance, d, total demand variation, ϑ, and grid frequency, ω, for $n = 100$ loads. The bottom plot shows the magnitude of the DFT of the overall control actions for load 1 as well as bandwidth constraints, α_m, for each frequency. There was a 75.6% reduction in the objective value compared to the case without intelligent loads.

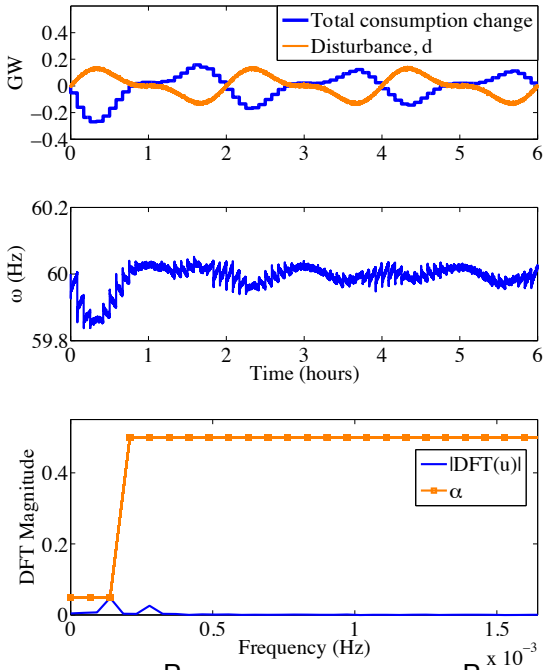

Fig. 2: 100 loads: $\sum_k \omega_k^2 = 32.6$; cf. 0 loads: $\sum_k \omega_k^2 = 135$.

In Figure 3, $n = 100$, and the values of α_m, $m = 0, 1, 2$ were ten times smaller than those in Figure 2, while α_m was the same as in that figure for all other m; i.e., the lowest three frequencies in the DFT were subjected to stricter constraints. The result of stricter constraints was a higher objective value relative to that for Figure 2. Even so, Figure 3 still shows a 58.8% reduction in objective value compared to the baseline with no load control. Additionally, the energy of the control actions in the more strictly constrained frequency was significantly attenuated compared to that in Figure 2, while the energy in the less constrained frequency remained nearly the same—effectively shaping the frequency content of the loads' actions to maintain QoS.

VI. Conclusion

We proposed a bandwidth-limited, disturbance-rejecting decentralized MPC (BaLDuR-DMPC) algorithm that enables loads to reduce demand-supply imbalance in a power grid while maintaining predetermined bounds on the loads' quality of service (QoS). Each load solves an MPC problem with constraints on the demand bandwidth (magnitude of the discrete Fourier transform of demand variation). The bandwidth constraint is the key in maintaining QoS. Decentralized coordination is achieved by scaling disturbance predictions based on local grid-frequency measurements. It is envisioned that each load will sign a long-term contract with

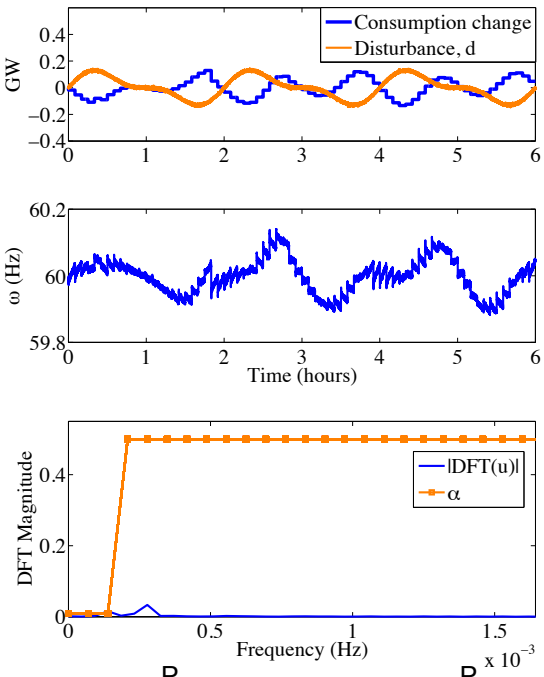

Fig. 3: 100 loads: $\sum_k \omega_k^2 = 55.6$; cf. 0 loads: $\sum_k \omega_k^2 = 135$; stricter bandwidth constraints than in Figure 2.

the balancing authority, during which its payment structure is determined based on the bandwidth it promises to make available.

In this work, we presented preliminary analytical convergence results under some idealized assumptions. Results from simulations of more realistic scenarios indicate that the algorithm performs well even in situations where those assumptions do not hold. Several aspects have been ignored in this preliminary study, especially the effect of the differential algebraic nature of power grid dynamics and its effect on closed-loop behavior [17]. A more thorough analysis of closed-loop stability is a subject of future study.

Acknowledgment

Thanks to Sean Meyn for suggesting the "constant-gain" grid model for local computations and James Rawlings for useful ideas about enforcing bandwidth constraints.

References

[1] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental evaluation of frequency regulation from commercial building HVAC systems," *IEEE Transactions on Smart Grid*, vol. 6, pp. 776 – 783, 2015.

[2] P. Siano, "Demand response and smart grids: a survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[3] DoE, "Benefits of demand response in electricity markets and recommendations for achieving them," *US Dept. Energy, Washington, DC, USA, Tech. Rep*, 2006.

[4] S. Lu, N. Samaan, R. Diao, M. Elizondo, C. Jin, E. Mayhorn, Y. Zhang, and H. Kirkham, "Centralized and decentralized control for demand response," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*. IEEE, 2011, pp. 1–8.

[5] M. Albadi and E. El-Saadany, "A summary of demand response in electricity markets," *Electric Power Systems Research*, vol. 78, no. 11, pp. 1989 – 1996, 2008.

[6] J. Brooks and P. Barooah, "Consumer-aware load control to provide contingency reserves using frequency measurements and inter-load communication," in *American Control Conference*, July 2016, pp. 5008 – 5013.

[7] C. Zhao, U. Topcu, and S. H. Low, "Optimal load control via frequency measurement and neighborhood area communication," *Power Systems, IEEE Transactions on*, pp. 3576–3587, 2013.

[8] M. Negrete-Pincetic, "Intelligence by design in an entropic power grid," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2013.

[9] G. Barbose, C. Goldman, and B. Neenan, "A survey of utility experience with real time pricing," *Lawrence Berkeley National Laboratory*, 2004.

[10] P. Barooah, A. Bušić, and S. Meyn, "Spectral decomposition of demand side flexibility for reliable ancillary service in a smart grid," in *48th Hawaii International Conference on Systems Science*, January 2015, invited paper.

[11] H. Hao, T. Middelkoop, P. Barooah, and S. Meyn, "How demand response from commercial buildings will provide the regulation needs of the grid," in *50th Annual Allerton Conference on Communication, Control and Computing*, October 2012, invited paper.

[12] Y. Lin, P. Barooah, and J. Mathieu, "Ancillary services through demand scheduling and control of commercial buildings," *IEEE Transactions on Power Systems*, vol. PP, 2016.

[13] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren, "Ancillary service to the grid from intelligent deferrable loads," *IEEE Transactions on Automatic Control*, vol. 60, pp. 2847 – 2862, March 2015.

[14] D. Todd, M. Caufield, B. Helms, A. P. Generating, I. M. Starke, B. Kirby, and J. Kueck, "Providing reliability services through demand response: A preliminary evaluation of the demand response capabilities of ALCOA INC." *ORNL/TM*, vol. 233, 2008.

[15] Y. Liu, "A US-wide power systems frequency monitoring network," in *Power Engineering Society General Meeting*, 2006.

[16] Prepared by NERC RS Committee, "Balancing and Frequency Control: A Technical Document Prepared by the NERC Resources Subcommittee," NERC Technical Report, pp. 1–53, January 26 2011.

[17] F. Dörfler and S. Grammatico, "Amidst centralized and distributed frequency control in power systems," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 5905–5914.

[18] M. Alizadeh, T.-H. Chang, and A. Scaglione, "Grid integration of distributed renewables through coordinated demand response," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 3666–3671.

[19] T. G. Hovgaard, L. F. Larsen, and J. B. Jørgensen, "Robust economic MPC for a power management scenario with uncertainties," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*. IEEE, 2011, pp. 1515–1520.

[20] G. Bianchini, M. Casini, A. Vicino, and D. Zarrilli, "Receding horizon control for demand-response operation of building heating systems," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 4862–4867.

[21] H. Chavez, R. Baldick, and S. Sharma, "Regulation adequacy analysis under high wind penetration scenarios in ercot nodal," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 4, pp. 743–750, 2012.