

Improved Battery Models of an Aggregation of Thermostatically Controlled Loads for Frequency Regulation^π

Borhan M. Sanandaji^{b,*}, He Hao^b, Kameshwar Poolla^b, and Tyrone L. Vincent^c

Abstract—Recently it has been shown that an aggregation of Thermostatically Controlled Loads (TCLs) can be utilized to provide fast regulating reserve service for power grids and the behavior of the aggregation can be captured by a stochastic battery with dissipation. In this paper, we address two practical issues associated with the proposed battery model. First, we address clustering of a heterogeneous collection and show that by finding the optimal dissipation parameter for a given collection, one can divide these units into few clusters and improve the overall battery model. Second, we analytically characterize the impact of imposing a no-short-cycling requirement on TCLs as constraints on the ramping rate of the regulation signal. We support our theorems by providing simulation results.

I. INTRODUCTION

A. Renewable Integration and Regulating Reserve Service

Vast and deep integration of renewable energy resources into the existing power grid is essential in achieving the envisioned sustainable energy future. Environmental, economical, and geopolitical concerns associated with the current power grid have motivated many countries around the globe as well as many states in the U.S. to setup aggressive Renewable Portfolio Standards (RPSs). The state of California, as an example, has targeted a 33% RPS by 2020 [1]. Volatility, stochasticity, and intermittency characteristics of renewable energies, however, present challenges for integrating these resources into the existing grid in a large scale as the proper functioning of an electric grid requires a continuous power balance between supply and demand.

Ancillary services such as regulating reserve (or frequency regulation) and load following play an important role in maintaining a functional and reliable grid under normal conditions [2]–[4]. While load following handles more predictable and slower changes in load, regulating reserve handles imbalances at faster time scales [5]. On the other hand, an increased penetration of renewable energies results in higher regulation requirements on the grid [2]–[4]. For instance, it has been shown that if California adopts its 33% RPS by 2020, the regulation procurement is anticipated to increase from 0.6 GW to 1.4 GW [6], [7]. Such

requirements can be lowered if *faster* responding resources are available [8]. For instance, it has been shown that if California Independent System Operator (CAISO) dispatches fast responding regulation resources, it would reduce its regulation procurement by 40% [9].

B. Demand-Side Flexibility for Frequency Regulation

Frequency regulation is one of the most important ancillary services for maintaining the power balance in normal conditions [5]. It is deployed in seconds (up to one minute) time scales to compensate for short term fluctuations in the *net load*.¹ This service has been traditionally provided by either fast responding generators or grid-scale energy storage units. However, the current storage technologies such as batteries have high cost while generation has both cost and an environmental footprint. Moreover, traditional generators have slow ramping rates and cannot track the fast changing regulation signal very well. These factors coupled with the search for cleaner sources of flexibility as well as regulatory developments such as Federal Energy Regulatory Commission (FERC) order 755 (2011) [10] and 784 (2013) [11] have motivated a growing interest in tapping fast responding demand-side resources for enabling deep renewable integration.

C. Aggregation of Flexible Loads

Flexible loads such as Electric Vehicles (EVs) and residential and commercial buildings have been recently considered as good candidates for providing ancillary services to the grid [12]–[26]. Residential Thermostatically Controlled Loads (TCLs) such as air conditioners, heat pumps, water heaters, and refrigerators, represent about 20% of the total electricity consumption in the United States [27], [28], and thus present a large potential for providing various ancillary services to the grid. Leveraging the inherent thermal slackness of TCLs, their electricity consumption can be varied while still meeting the desired comfort level and temperature requirements of the end user.

D. Related Work

TCLs have been recently considered for providing load following and regulation services to the grid [12]–[15], [17], [29]. In particular, it has been recently shown that the aggregate flexibility offered by a collection of TCLs can be succinctly modeled as a *stochastic battery* with dissipation [30], [31]. The power limits and energy capacity of this battery model can be calculated in terms of TCL

^bBorhan M. Sanandaji, He Hao, and Kameshwar Poolla are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

*Corresponding author. Email: sanandaji@berkeley.edu.

^cTyrone L. Vincent is with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401.

^πSupported in part by EPRI and CERTS under sub-award 09-206; PSERC S-52; NSF under Grants CNS-0931748, EECS-1129061, CPS-1239178, and CNS-1239274; the Republic of Singapore National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore for the SinBerBEST Program; Robert Bosch LLC through its Bosch Energy Research Network funding program.

¹Net load is defined as forecasted load minus predicted variable generation.

model parameters and exogenous variables such as ambient temperature and user-specified set-points. Simple battery models are also considered in [13], [17]. Clustering and no-short-cycling of TCLs have been reported in [18], [19].

E. Main Contributions

In this paper, we address some practical aspects associated with our earlier proposed battery model [30], [31]. First, we consider the impact of dividing a heterogeneous collection of TCLs into clusters and show that by finding the optimal dissipation parameter for a given collection of TCLs, one can divide these units into a few stochastic batteries and improve the battery model. Second, we consider the effect of enforcing a requirement of no-short-cycling. In order to avoid damages, TCLs manufacturers require a minimum duration of time between any two switches of ON/OFF state. If this minimum time is not met, the unit is said to be short-cycled. In particular, we show that the no-short-cycling constraint can be expressed as constraints on the ramping rate (first difference) of the Automatic Generation Control (AGC) signal. Consequently, a characterization of regulation signals that can be feasibly met by a TCL aggregation is the intersection of signals feasible for the stochastic battery model, and this new constraint on the first difference of the regulation signal. To the best of our knowledge, this work is the first which explicitly represents a ramping rate constraint on the regulation signal as a consequence of units no-short-cycling requirements.

F. Paper Organization

The remainder of the paper is organized as follows. Section II describes preliminaries on individual TCL models. In Section III, we summarize the stochastic battery model. Optimal dissipation and clustering of a collection of TCLs are presented in Section IV. We address the no-short-cycling of TCLs in Section V. Whenever needed and within each section, we provide simulation results to support our theorems.

II. THERMOSTATICALLY CONTROLLED LOADS

The temperature evolution of the k th TCL can be described by a standard dead-band model as

$$\dot{\theta}^k(t) = \begin{cases} -a^k(\theta^k(t) - \theta_a) - b^k P_m^k + w^k(t), & \text{ON state,} \\ -a^k(\theta^k(t) - \theta_a) + w^k(t), & \text{OFF state,} \end{cases} \quad (1)$$

where $\theta^k(t)$ is the internal temperature of the k th TCL at time t , θ_a is the ambient temperature, P_m^k is the rated electrical power, $a^k := 1/C^k R^k$, $b^k := \eta^k/C^k$, and R^k , C^k , and η^k are model parameters as described in Table I. For more details on the TCL model, please see [12], [14], [30].² Each TCL has a temperature set-point θ_r^k with a hysteretic

²Four types of TCLs are: (i) air conditioners, (ii) heat pumps, (iii) water heaters, and (iv) refrigerators.

TABLE I

TYPICAL PARAMETER VALUES FOR A RESIDENTIAL AIR CONDITIONER.

Parameter	Description	Value	Unit
C	thermal capacitance	2	kWh/°C
R	thermal resistance	2	°C/kW
P_m	rated electrical power	5.6	kW
η	coefficient of performance	2.5	
θ_r	temperature set-point	22.5	°C
Δ	temperature deadband	0.3125	°C
θ_a	ambient temperature	32	°C

ON/OFF local control within a dead-band $[\theta_r^k - \Delta^k, \theta_r^k + \Delta^k]$. The operating state of the k th TCL, $q^k(t)$, evolves as

$$\lim_{\epsilon \rightarrow 0} q^k(t + \epsilon) = \begin{cases} q^k(t), & |\theta^k(t) - \theta_r^k| < \Delta^k, \\ 1 - q^k(t), & |\theta^k(t) - \theta_r^k| = \Delta^k, \end{cases}$$

where $q^k(t) = 1$ when the TCL is ON and $q^k(t) = 0$ when it is OFF. The average power consumed by the k th TCL over a cycle is

$$P_a^k = \frac{P_m^k T_{\text{ON}}^k}{T_{\text{ON}}^k + T_{\text{OFF}}^k},$$

where T_{ON}^k and T_{OFF}^k are given by

$$T_{\text{ON}}^k = R^k C^k \ln \frac{\theta_r^k + \Delta^k - \theta_a + R^k P_m^k \eta^k}{\theta_r^k - \Delta^k - \theta_a + R^k P_m^k \eta^k},$$

$$T_{\text{OFF}}^k = R^k C^k \ln \frac{\theta_r^k - \Delta^k - \theta_a}{\theta_r^k + \Delta^k - \theta_a},$$

and represent the ON and OFF state durations per cycle, respectively. For a large collection of TCLs that is uncoordinated, the instantaneous power drawn by this collection will be very close to the combined average power requirement due to the fact that any specific TCL will be at a uniformly random point along its operating cycle. For a heterogeneous collection of TCLs indexed by k , the baseline power is

$$n(t) := \sum_k P_a^k.$$

The aggregated instantaneous power consumption is

$$P_{\text{agg}}(t) := \sum_k q^k(t) P_m^k.$$

As an approximation to the dead-band model, we consider a continuous-power model. Here, a TCL accepts any continuous power input $p^k(t) \in [0, P_m^k]$ and the dynamics are

$$\dot{\theta}^k(t) = -a^k(\theta^k(t) - \theta_a) - b^k p^k(t).$$

As common in the literature, the disturbance $w^k(t)$ in Model (1) is assumed to be Gaussian with zero mean and small variance [12], [14], [32], and can be neglected. Maintaining the temperature $\theta^k(t)$ within the user-specified dead-band $\theta_r^k \pm \Delta^k$ is treated implicitly as a *constraint* on the power signal $p^k(t)$. When evaluating the trajectory $\theta^k(t)$, it is assumed that $\theta^k(0) = \theta_r^k$. The parameters that specify this continuous-power model are $\chi^k = (a^k, b^k, \theta_r^k, \theta_a, \Delta^k, P_m^k)$.

The nominal power required to keep the k th TCL at its set-point is

$$P_o^k = \frac{a^k(\theta_a - \theta_r^k)}{b^k} = \frac{\theta_a - \theta_r^k}{\eta^k R^k}.$$

We note that P_o^k is a random process as it depends on the ambient temperature and the user-defined set-point. Simple calculations with typical parameters reveal that the nominal power P_o^k under the continuous-power model closely follows the average power P_a^k under the dead-band model for a wide range of operating conditions. In [23], [30], we showed that the aggregate behavior of a population of TCLs with the dead-band model could be accurately approximated by those using the continuous-power model. The continuous-power model is used for analysis, and the dead-band model is used for simulations.

III. STOCHASTIC BATTERY MODEL

Each TCL can accept perturbations around its nominal power consumption ($p^k(t) = P_o^k + e^k(t)$) that will meet user-specified comfort bounds. Define

$$\mathbb{E}^k := \left\{ e^k(t) \mid \begin{array}{l} 0 \leq P_o^k + e^k(t) \leq P_m^k, \\ P_o^k + e^k(t) \text{ maintains } |\theta^k(t) - \theta_r^k| \leq \Delta^k \end{array} \right\}.$$

This set of power signals represents the flexibility of the k th TCL with respect to its nominal. The *aggregate flexibility* of the collection of TCLs is defined as the Minkowski sum

$$\mathbb{U} = \sum_k \mathbb{E}^k.$$

The difficulty is in evaluating \mathbb{U} as its geometry is, in general, bulky. In [30], [31], we showed that the set \mathbb{U} can be nested within two generalized battery models.

Definition 1: Let $\phi = (\mathcal{C}, n_-, n_+, \alpha)$ be non-negative parameters. A *Generalized Battery Model*, $\mathbb{B}(\phi)$, is a set of signals $u(t)$ that satisfy

$$\begin{aligned} -n_- \leq u(t) \leq n_+, \quad \forall t > 0, \\ \dot{x}(t) = -\alpha x(t) - u(t), \quad x(0) = 0 \Rightarrow |x(t)| \leq \mathcal{C}, \quad \forall t > 0. \end{aligned}$$

One can regard $u(t)$ as the power drawn from or supplied to a battery and $x(t)$ as its State of Charge (SoC). One should note that the parameters ϕ are random and depend on ambient temperature and participation rates. As a result, we regard $\mathbb{B}(\phi)$ as a stochastic battery. This battery model provides a compact framework to characterize the aggregate power limits and energy capacity of a population of TCLs.

Theorem 1 ([30]): Consider a heterogeneous collection of TCLs modeled by the continuous-power model with parameters χ^k . Let $\alpha > 0$ be the dissipation parameter. Let

$$f^k := \Delta^k / (b^k(1 + |1 - \alpha/a^k|)).$$

The aggregate flexibility \mathbb{U} of the collection satisfies

$$\mathbb{B}(\phi_s) \subseteq \mathbb{U} \subseteq \mathbb{B}(\phi_n),$$

where the necessary battery model parameters are given by

$$\phi_n : \begin{cases} \mathcal{C} = \sum_k \left(1 + \left|1 - \frac{a^k}{\alpha}\right|\right) \frac{\Delta^k}{b^k}, \\ n_- = \sum_k P_o^k, \\ n_+ = \sum_k (P_m^k - P_o^k), \end{cases} \quad (2)$$

For a given α , the sufficient battery model parameters are any triple (\mathcal{C}, n_-, n_+) that satisfies

$$\phi_s : \begin{cases} \beta^k n_- \leq P_o^k, & (\forall k) \\ \beta^k n_+ \leq P_m^k - P_o^k, & (\forall k) \\ \beta^k \mathcal{C} \leq f^k, & (\forall k) \end{cases} \quad (3)$$

where $\beta^k \geq 0$ satisfies $\sum_k \beta^k = 1$. Further, if $u(t) \in \mathbb{B}(\phi_s)$, the causal power allocation strategy

$$e^k(t) = \beta^k u(t)$$

satisfies the deadband constraints $|\theta^k(t) - \theta_r^k| \leq \Delta^k$.

One should note that the gap between the proposed battery models $\mathbb{B}(\phi_s)$ and $\mathbb{B}(\phi_n)$ in Theorem 1 depends on the choice of allocation β^k , the dissipation α , and heterogeneity level of the collection of TCLs. In the next section, we explain how we can obtain an optimal dissipation for a given collection of TCLs. Moreover, we show how one can improve the battery models by means of clustering of units.

IV. OPTIMAL DISSIPATION AND CLUSTERING OF TCLS

As mentioned earlier, there exist different choices of β_k that satisfy (3). For each choice, a different battery model $\mathbb{B}(\phi_s)$ will be obtained that assures feasibility. One choice is

$$\beta_k = \frac{P_m^k - P_o^k}{\sum_k (P_m^k - P_o^k)}, \quad (4)$$

which yields the smallest gap between the necessary and sufficient battery models for n_+ as compared to other choices of β^k . However, it results in larger gaps for n_- and \mathcal{C} .³ Based on this particular choice of β_k ,

$$\phi_s : \begin{cases} \mathcal{C} = \sum_k (P_m^k - P_o^k) \min_k \frac{f^k}{P_m^k - P_o^k}, \\ n_- = \sum_k (P_m^k - P_o^k) \min_k \frac{P_o^k}{P_m^k - P_o^k}, \\ n_+ = \sum_k (P_m^k - P_o^k). \end{cases} \quad (5)$$

This result is valid for any given dissipation parameter α .

A. Optimal Dissipation Parameter

While the bound on n_+ is the tightest possible based on the allocation (4), one would like to tighten the bound on \mathcal{C} as well. This can be done easily by considering the following optimization problem for a given heterogeneous collection:

$$\alpha^* := \arg \max_{\alpha} \min_k \frac{f^k}{P_m^k - P_o^k} \quad (6)$$

to find the *optimal dissipation* parameter α^* . While solving (6) for the general case may require a numerical solution, for some heterogeneity scenarios we get an analytical solution.

³We can maximize the bounds on \mathcal{C} and n_- by choosing different allocations β_k . Please refer to [31] where we discuss how different choices of β_k would affect the bounds on power limits and energy capacity.

1) *Thermal Capacity*: Consider the case where all of the parameters are homogenous and the only heterogeneity is in C^k .⁴ Then, the battery capacity can be written as

$$\mathcal{C}(\alpha) = N\Delta(\min_k g^k)/\eta,$$

where $g^k := C^k/(1+|1-\alpha RC^k|)$. Note that the dependance of the capacity on α has been made explicit.

Lemma 1: Consider a heterogeneous collection of TCLs where the heterogeneity is only in C^k . Then

$$\mathcal{C}^* := \max_{\alpha} \mathcal{C}(\alpha) = N\Delta C_{\min}/\eta \quad \text{and} \quad \alpha^* = 1/RC_{\min},$$

where $C_{\min} := \min_k C^k$.

Proof: See Appendix. ■

2) *Dead-band*: Consider the case where all of the parameters are homogenous and the only heterogeneity is in Δ^k . Then, the battery capacity can be written as

$$\mathcal{C}(\alpha) = NC(\min_k g^k)/(\eta(1+|1-\alpha RC|)),$$

where $g^k := \Delta^k$.

Lemma 2: Consider a heterogeneous collection of TCLs where the heterogeneity is only in Δ^k . Then

$$\mathcal{C}^* = NC\Delta_{\min}/\eta \quad \text{and} \quad \alpha^* = 1/RC,$$

where $\Delta_{\min} := \min_k \Delta^k$.

Proof: See Appendix. ■

One can derive similar results for cases where more parameters contain heterogeneity. For example, when the heterogeneity is in both C^k and Δ^k , then $\mathcal{C}(\alpha) = N(\min_k g^k)/\eta$ where $g^k := \frac{C^k \Delta^k}{1+|1-\alpha RC^k|}$. One can show that under this assumption,

$$\mathcal{C}^* \gtrsim NC_{\min}\Delta_{\min}/\eta \quad \text{and} \quad \alpha^* \lesssim 1/RC_{\min}.$$

B. Optimal Clustering

As the diversity of the TCL model parameters increases, the gap between $\mathbb{B}(\phi_s)$ and $\mathbb{B}(\phi_n)$ increases. Fig. 1 illustrates how the sufficient and necessary capacity values provided by $\mathbb{B}(\phi_s)$ and $\mathbb{B}(\phi_n)$, respectively, increases as the heterogeneity level increases. In order to improve the battery models, one can divide a heterogenous collection TCLs into a few clusters and derive battery models for each of those clusters. The number of clusters should be small in order to keep the benefits of aggregation within each battery model, while limiting the complexity of the overall model. We assume the number of clusters m is small and given.

Consider the case where only C^k contains heterogeneity. We know that from our previous analysis, for a given collection of N TCLs with heterogeneous C^k 's, $N\Delta C_{\min}/\eta$ is the maximum energy capacity (Lemma 1). Let's assume we want to divide N units into m clusters where N_i is the size of cluster i such that $\sum_{i=1}^m N_i = N$. The goal is to find optimal cluster sizes such that the overall energy capacity of the collection is maximized. The following theorem provides the

optimal cluster sizes and optimal capacity under clustering when there is a uniform distribution on C^k 's.

Theorem 2: Consider a heterogeneous collection of N TCLs with model (1) where C^k has a uniform distribution as $C^k \sim U(C_{\min}, C_{\max})$ but all other parameters are identical between units. Then, an optimal clustering is achieved by sorting the units based on their C^k value and then putting the first N/m units in the first cluster, the second N/m units in the second cluster, etc, with an optimal cluster size of

$$N_i^* = N/m, \quad i = 1, 2, \dots, m, \quad (7)$$

where m is the number of clusters. The optimal capacity is

$$\mathcal{C}_m^* = \left(C_{\min} + \frac{(C_{\max} - C_{\min})}{2} \frac{N}{(N-1)} \frac{(m-1)}{m} \right) N\Delta/\eta, \quad (8)$$

where \mathcal{C}_m^* is the optimal capacity with m clusters.

Proof: See Appendix. ■

Remark 1: Similar results can be achieved when more parameters contain heterogeneity (and even with different heterogeneity distributions) following the steps explained in the proof of Theorem 2 and based on the optimal dissipation parameters as discussed in Section IV-A.

Fig. 1 illustrates how the sufficient and necessary capacity bounds change over different heterogeneity levels and under different scenarios on the dissipation parameter and clustering. As can be seen, when a nominal (an average of the time constants $1/RC^k$) dissipation is considered with no clustering, the gap between the sufficient and necessary capacity bounds increases as the heterogeneity level increases. This gap can be decreases when an optimal dissipation parameter is used for the collection. Moreover, the gap can be further tightened when we divide the collection into a few clusters (in this example 3 clusters). Apparently when $m = 1$, the optimal capacity is the same as provided by Lemma 1. A keen reader would note that when $m = N$, the optimal capacity (8) is

$$\mathcal{C}_N^* = \frac{N\Delta C}{\eta},$$

which is the capacity bound of a homogenous collection [30], [31]. As mentioned earlier, we keep m small such that the complexity of the overall battery model is as low as possible.

V. NO-SHORT-CYCLING AND RAMPING RATE CONSTRAINTS

In this section, we first present our priority-stack-based control framework for manipulating the power consumption of a population of TCLs and for providing regulation service to the grid. We then augment our control structure with a no-short-cycling constraint. Moreover, we analytically characterize the no-short-cycling constraint in terms of bounds on the ramping rate of the regulation signal.

A. Priority-Stack-Based Control

We adopt a centralized control architecture. This choice is dictated by the stringent power quality, auditing and telemetry requirements necessary to participate in regulation service

⁴In total there are 6 parameters whose heterogeneity can affect \mathcal{C} in (5): C^k , R^k , η^k , Δ^k , P_m^k , and θ_r^k .

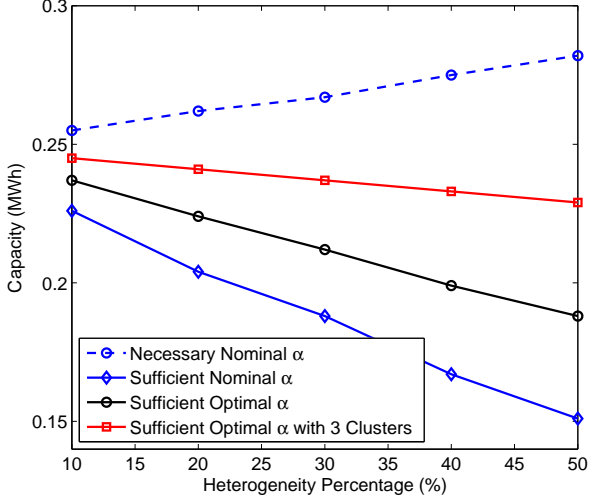


Fig. 1. The effect of optimal dissipation parameter and clustering on the battery model. A collection of 1000 heterogenous TCLs are considered whose nominal parameter values are given in Table I. A uniform distribution is considered as the heterogeneity pattern of C^k .

market [33]. At each sample time, the aggregator compares the regulation signal $r(t)$ with the aggregate power deviation $\delta(t) = P_{\text{agg}}(t) - n(t)$, where $P_{\text{agg}}(t)$ is the instantaneous power drawn by TCLs and $n(t)$ is their baseline power.

If $r(t) < \delta(t)$, the population of TCLs needs to “discharge” power to the grid which requires turning OFF some of the ON units. Conversely, if $r(t) > \delta(t)$, then the population of TCLs must consume more power. This requires turning ON some of the OFF units. To track a regulation signal $r(t)$, the system operator needs to determine appropriate switching actions for each TCL so that the power deviation of TCLs, $\delta(t)$, follows the regulation signal $r(t)$.

In practice, it is more favorable to turn ON (or OFF) the units which are going to be turned ON (or OFF) by their local hysteretic control law. To this end, we propose a priority-stack-based control method. The unit with the highest priority will be turned ON (or OFF) first, and then units with lower priorities will be considered in sequence until the desired regulation is achieved. This priority-stack-based control strategy minimizes the ON/OFF switching action for each unit, reducing wear and tear of TCLs. Priority stacks are illustrated in Fig. 2. A normalized *temperature distance* is considered as the sorting criterion.

B. No-Short-Cycling Constraint

The proposed priority-stack-based control scheme attempts to reduce the consecutive switching times of each TCL. However, it can not guarantee that none of the units will not be switched quicker than allowed. To this end, one should *explicitly* impose such no-short-cycling constraints on the priority stacks. As shown in Fig. 2, the ON and OFF priority stacks can be modified to account for such no-short-cycling constraints. Once a unit is turned ON or OFF, it must remain in that state for at least a certain amount of time (that

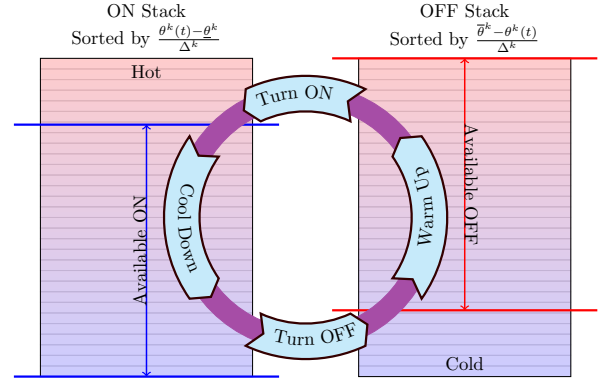


Fig. 2. The ON and OFF priority stacks with explicit no-short-cycling constraints. A unit that is hotter has a higher priority to be switched ON and a unit that is cooler has a higher priority to be turned OFF. However, when no-short-cycling constraints are imposed, we are only allowed to manipulate units that are Available ON or Available OFF. The lower and upper temperature bounds are given by $\underline{\theta}^k = \theta_r^k - \Delta^k$ and $\bar{\theta}^k = \theta_r^k + \Delta^k$.

is specified by the manufacture) before it is switched again.

For clarity of presentation, we list some of the terms that we will frequently use in this section in Table II. When the controller must satisfy the no-short-cycling constraints, a certain percentage of TCLs will be unavailable to be switched from ON to OFF or OFF to ON. The effect of this loss of use is to create an additional constraint on *changes* in feasible regulation signals $r(t)$. Quite simply, if there is no available ON unit to be switched OFF, the regulation signal cannot request decreased power draw (and similarly for increased power draw). To determine feasible regulation signals, the battery model must be augmented with the constraints

$$-\mu_-(t) \leq \Delta r(t) \leq \mu_+(t), \quad (9)$$

where $\Delta r(t) = r(t) - r(t-1)$, and $\mu_-(t)$ and $\mu_+(t)$ are time varying constraints.⁵ However, we will show that μ_- and μ_+ are easily estimated if the following information is available: (i) power draw of each unit (when ON); (ii) average power draw of each unit P_o^k ; and (iii) the total rated power for units that are about to be turned ON or OFF due to their temperature limits, denoted by $P_{OFF \rightarrow ON}^{lim}(t)$ or $P_{ON \rightarrow OFF}^{lim}(t)$, respectively.

Theorem 3: Assume a collection of TCLs defined by P_m^k, P_o^k , and a minimum short cycle time of τ (samples). If the regulation signal $r(t)$ has been met through sample time t , then the total power of units available at t is

$$P_{OFF}^{avail}(t) = P_{tot} - P_{ave} - r(t) - \sum_{k=t-\tau}^t (P_{ON \rightarrow OFF}^{lim}(k) + [-D(k)]_+)$$

and

⁵There might exist other known constraints/bounds on changes $\Delta r(t)$ specified by the system operator which we are not considering here.

Term	Description
P_m^k	Power draw of unit k when ON
P_{tot}	$\sum_k P_m^k$
P_o^k	Average power draw of unit k
P_{ave}	$\sum_k P_o^k$
$r(t)$	Regulation signal requesting power draw of $P_{ave} + r(t)$
Available ON	Units that have been ON for more than a certain amount of time
Unavailable ON	Units that have been ON for less than a certain amount of time
Available OFF	Units that have been OFF for more than a certain amount of time
Unavailable OFF	Units that have been OFF for less than a certain amount of time
$P_{ON \rightarrow OFF}^{lim}(t)$	Total power of units switched from ON to OFF at time t due to temperature bound
$P_{OFF \rightarrow ON}^{lim}(t)$	Total power of units switched from OFF to ON at time t due to temperature bound
$P_{ON}(t)$	Total power of ON units
$P_{OFF}(t)$	Total power of OFF units
$P_{ON}^{avail}(t)$	Total power of units that are available ON
$P_{ON}^{unavail}(t)$	Total power of units that are unavailable ON
$P_{OFF}^{avail}(t)$	Total power of units that are available OFF
$P_{OFF}^{unavail}(t)$	Total power of units that are unavailable OFF

TABLE II
NOMENCLATURE OF SOME OF THE FREQUENTLY-USED TERMS.

$$P_{ON}^{avail}(t) = P_{ave} + r(t) - \sum_{k=t-\tau}^t (P_{OFF \rightarrow ON}^{lim}(k) + [D(k)]_+),$$

where $D(t) := \Delta r(t) - (P_{OFF \rightarrow ON}^{lim}(t) - P_{ON \rightarrow OFF}^{lim}(t))$ and $[x]_+ := \max(x, 0)$. In addition, feasible $\Delta r(t)$ satisfies (9) with

$$\begin{aligned} \mu_+(t) &= P_{OFF}^{avail}(t-1) - \max(P_{ON \rightarrow OFF}^{lim}(t), P_{OFF \rightarrow ON}^{lim}(t)), \\ \mu_-(t) &= P_{ON}^{avail}(t-1) - \max(P_{ON \rightarrow OFF}^{lim}(t), P_{OFF \rightarrow ON}^{lim}(t)). \end{aligned}$$

Proof: See Appendix. ■

C. Simulation Results

We run the priority-stack-based controller with the reference command shown as the solid line in Fig. 3(a). For comparison, the power and capacity limits found using the battery model are also shown in Fig. 3(a) and Fig. 3(b), respectively. As can be seen, the power and capacity limits are not violated by this regulation signal. We take $P_{OFF \rightarrow ON}^{lim}(t)$ and $P_{ON \rightarrow OFF}^{lim}(t)$ as that reported by the local unit controllers, and use that information along with $r(t)$ to calculate $\mu_+(t)$ and $\mu_-(t)$. In Fig. 3(c) these are plotted along with $\Delta r(t)$. Note that at time 150 (s), the lower bound approaches zero, meaning that negative $\Delta r(t)$ is no longer feasible. Fig. 3(d) depicts the difference between the desired regulation signal and the actual power draw $P_{agg}(t) - P_{ave}$, confirming that regulation signal is not well followed downward during the time that μ_- is close or equal to zero.

VI. CONCLUSIONS AND FUTURE WORK

An aggregation of Thermostatically Controlled Loads (TCLs) can be utilized to provide fast regulating reserve service for power grids and the behavior of the aggregation can be captured by a stochastic battery with dissipation. In this paper, we addressed two practical issues associated with the proposed battery model. First, we addressed clustering of a heterogeneous collection of TCLs and showed that

by finding the optimal dissipation parameter for a given collection, one can divide these units into few clusters and improve the overall battery model. Second, we analytically characterized the impact of imposing a no-short-cycling requirement on TCLs in terms of constraints on the ramping rate of the AGC signal.

One of the future directions of this work is to better understand the thermal characteristics and dynamics of an individual TCL. We should also better understand the actual heterogeneity pattern of a collection of TCLs. These can be done by examining the proposed models against experimental data captured from installed TCLs.

APPENDIX

Proof of Lemma 1 Let $C_{\min} := \min_k C^k$ and $C_{\max} := \max_k C^k$. If $\alpha > 1/RC_{\min}$, then $\alpha RC^k > 1, \forall k$, and $g^k = 1/\alpha R$. If $\alpha < 1/RC_{\max}$, then $\alpha RC^k < 1, \forall k$, and $g^k = \frac{C^k}{2 - \alpha RC^k}$. With a change of variable $x := 1/RC^k$, $g^k(x) = 1/(2Rx - \alpha R)$. If $1/RC_{\max} < \alpha < 1/RC_{\min}$, for C^k 's such that $C^k > 1/\alpha R$, $g^k = 1/\alpha R$, and for C^k 's such that $C^k < 1/\alpha R$, $g^k = \frac{C^k}{2 - \alpha RC^k}$. At $x = \alpha$, $1/(2Rx - \alpha R) = 1/\alpha R$. Consequently, when $\alpha < 1/RC_{\min}$, $\min_k g^k = 1/(2/C_{\min} - \alpha R)$ and when $\alpha > 1/RC_{\min}$, $\min_k g^k = 1/\alpha R$. Thus, when the heterogeneity is only in C^k ,

$$\max_{\alpha} \mathcal{C}(\alpha) = N \Delta C_{\min} / \eta, \quad \alpha^* = 1/RC_{\min}.$$

Proof of Lemma 2 If $\alpha > 1/RC$, then $\mathcal{C}(\alpha) = N \Delta_{\min} / \alpha \eta R$. If $\alpha < 1/RC$, then $\mathcal{C}(\alpha) = N \Delta_{\min} C / \eta (2 - \alpha RC)$. The breakpoint is at $\alpha = 1/RC$. Thus, when the heterogeneity is only in Δ^k ,

$$\max_{\alpha} \mathcal{C}(\alpha) = N C \Delta_{\min} / \eta \quad \text{and} \quad \alpha^* = 1/RC.$$

Proof of Theorem 2 When the heterogeneity is only in C^k , the optimal cluster sizes can be found by solving the

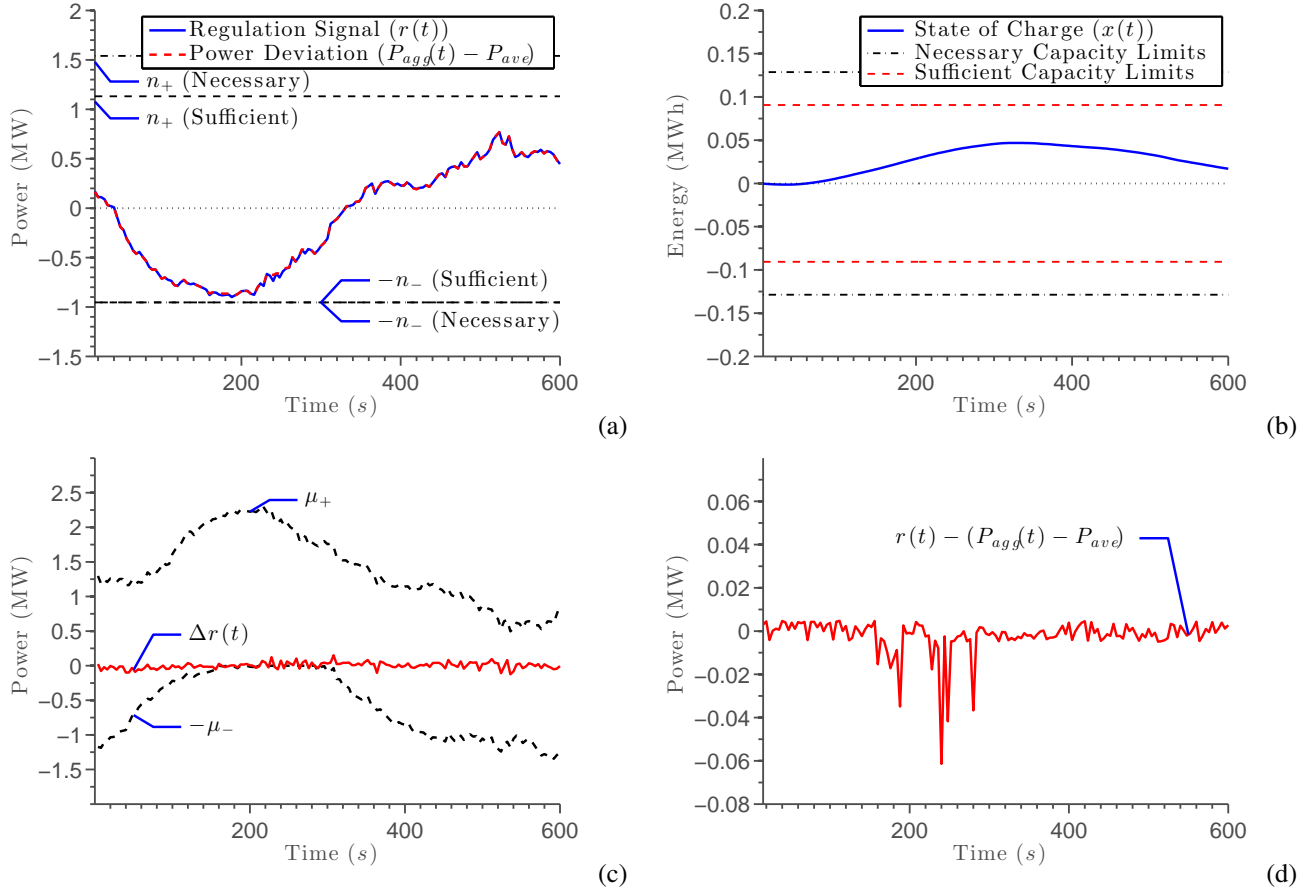


Fig. 3. Illustration of the effect of short cycling and ramping rate constraints. (a) The regulation signal and battery model bounds on power. (b) The SoC and the capacity limits. (c) $\Delta r(t)$ and its bounds given in Theorem 3. (d) The difference between the desired regulation signal $r(t)$ and the actual power draw $P_{agg}(t) - P_{ave}$. At time 150 (s), $-\mu_-(t)$ approaches zero, meaning that negative $\Delta r(t)$ is no longer feasible. Fig. 3(d) confirms that the regulation signal is not well followed downward during the time that $-\mu_-(t)$ is close to zero.

following optimization problem:

$$\begin{aligned}
 & \underset{N_1, \dots, N_m}{\text{maximize}} && f(1)N_1 + \sum_{i=2}^m N_i f\left(1 + \sum_{j=1}^{i-1} N_j\right) \\
 & \text{subject to} && \sum_{i=1}^m N_i = N,
 \end{aligned} \tag{10}$$

where $f(\cdot)$ is a function that represents the sorted C^k values in an ascending order. In the case where a uniform distribution is assumed as the heterogeneity of C^k 's, the sorted C^k values construct an affine function f between C_{\min} and C_{\max} as

$$f(x) = C_{\min} + \frac{C_{\max} - C_{\min}}{N - 1}(x - 1),$$

where x only takes integer values between 1 and N . It can be shown that under linearity assumption on $f(\cdot)$, the optimal solution to (10) is

$$N_1^* = \dots = N_m^* = N/m.$$

The proof is not presented here for the sake of saving space. Consequently, the optimal capacity is derived by using the

optimal cluster sizes in the objective function of (10) as

$$C_m^* = \left(C_{\min} + \frac{C_{\max} - C_{\min}}{2} \frac{N}{(N-1)} \frac{(m-1)}{m}\right) N \Delta / \eta.$$

Proof of Theorem 3 Let $P_{ON}(t)$ and $P_{OFF}(t)$ denote the total power of units ON and OFF, respectively. If $r(t)$ is satisfied, then by definition

$$\begin{aligned}
 P_{ON}(t) &= P_{ave} + r(t), \\
 P_{OFF}(t) &= P_{tot} - P_{ave} - r(t).
 \end{aligned}$$

Note that $P_{ON}(t) + P_{OFF}(t) = P_{tot}$. Let $P_{ON}^{unavail}(t)$ and $P_{OFF}^{unavail}(t)$ be the total power of units that are unavailable and ON or OFF, respectively. Clearly

$$\begin{aligned}
 P_{ON}^{avail}(t) &= P_{ON}(t) - P_{ON}^{unavail}(t), \\
 P_{OFF}^{avail}(t) &= P_{OFF}(t) - P_{OFF}^{unavail}(t).
 \end{aligned}$$

Now, $P_{ON}^{unavail}$ is given by the sum of the power of units that have been turned ON in the last τ seconds. The first result follows by noting that if $r(t)$ is satisfied, then the power of units turned ON at time t must balance the difference between units turned ON and OFF due to local controllers,

along with the change in $r(t)$. For example, the units turned from OFF to ON must be given by

$$P_{OFF \rightarrow ON}(t) = P_{OFF \rightarrow ON}^{lim}(t) + [D(k)]_+,$$

where the second term represents the potential imbalance due to a change in $\Delta r(t)$ plus a difference between $P_{OFF \rightarrow ON}^{lim}(t)$ and $P_{ON \rightarrow OFF}^{lim}(t)$. Finally, the limits $\mu_+(t)$ and $\mu_-(t)$ are achieved using the worst case assumption that all units that hit the temperature limits are currently unavailable and no units unavailable at time $t - 1$ become available.

REFERENCES

- [1] California Energy Commission, "California renewable energy overview and programs," 2013. [Online]. Available: <http://www.energy.ca.gov/renewables/index.html>
- [2] J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons, "Utility wind integration and operating impact state of the art," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 900–908, 2007.
- [3] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational impacts of wind generation on California power systems," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039–1050, 2009.
- [4] S. Meyn, M. Negrete-Pincetic, G. Wang, A. Kowli, and E. Shafiepoorfar, "The value of volatile resources in electricity markets," in *CDC2010*, 2010, pp. 1029–1036, and submitted to IEEE TAC, 2012.
- [5] B. Kirby, "Ancillary services: Technical and commercial insights," Report prepared for Wartsila, Tech. Rep., July 2007.
- [6] U. Helman, "Resource and transmission planning to achieve a 33% RPS in California—ISO modeling tools and planning framework," in *FERC Technical Conference on Planning Models and Software*, 2010.
- [7] Market and Infrastructure Policy, "2013 flexible capacity procurement requirement," California Independent System Operator – CAISO, Tech. Rep., March 2012. [Online]. Available: <http://www.CAISO.com/>
- [8] K. Vu, R. Masiello, and R. Fioravanti, "Benefits of fast-response storage devices for system regulation in ISO markets," *IEEE Power Energy Society General Meeting, 2009*, pp. 1–8, 2009.
- [9] Y. V. Makarov, L. S., J. Ma, and T. B. Nguyen, "Assessing the value of regulation resources based on their time response characteristics," Pacific Northwest National Laboratory, Tech. Rep. PNNL-17632, 2008.
- [10] Market and Infrastructure Policy, "Frequency regulation compensation - FERC order no. 755," Federal Energy Regulatory Commission – FERC, Tech. Rep., 2011. [Online]. Available: <https://www.ferc.gov/whats-new/comm-meet/2011/102011/E-28.pdf>
- [11] —, "Frequency regulation compensation - FERC order no. 784," Federal Energy Regulatory Commission – FERC, Tech. Rep., 2013. [Online]. Available: <https://www.ferc.gov/whats-new/comm-meet/2013/071813/E-22.pdf>
- [12] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, 2009.
- [13] S. Koch, M. Zima, and G. Andersson, "Active coordination of thermal household appliances for load management purposes," *Proceedings of the IFAC Symposium on Power Plants and Power Systems Control*, pp. 149–154, 2009.
- [14] S. Koch, J. Mathieu, and D. Callaway, "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services," *Proceedings of the 17th Power Systems Computation Conference – PSCC*, 2011.
- [15] J. Mathieu and D. Callaway, "State estimation and control of heterogeneous thermostatically controlled loads for load following," *Proceedings of the 45th Hawaii International Conference on System Sciences – HICCS45*, pp. 2002–2011, 2012.
- [16] J. L. Mathieu, S. Koch, and D. S. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 430–440, February 2013.
- [17] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway, "Energy arbitrage with thermostatically controlled loads," in *European Control Conference (ECC)*, 2013.
- [18] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, , and D. Chassin, "Aggregate model for heterogeneous thermostatically controlled loads with demand response," *Proceedings of the 2012 IEEE PES General Meeting*, 2012.
- [19] C.-Y. Chang, W. Zhang, J. Lian, and K. Kalsi, "Modeling and control of aggregated air conditioning loads under realistic conditions," *Proceedings of the IEEE PES Innovative Smart Grid Technologies Conference – ISGT*, 2013.
- [20] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi, "Aggregated modeling and control of air conditioning loads for demand response," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4655–4664, 2013.
- [21] S. Bashash and H. K. Fathy, "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads," *Proceedings of the 2011 American Control Conference – ACC*, pp. 4546–4553, 2011.
- [22] —, "Modeling and control of aggregate air conditioning loads for robust renewable power management," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1318–1327, 2013.
- [23] B. M. Sanandaji, H. Hao, and K. Poolla, "Fast regulation service provision via aggregation of thermostatically controlled loads," *Proceedings of the 47th Hawaii International Conference on System Sciences – HICSS47*, pp. 2388–2397, 2014.
- [24] S. Bashash and H. K. Fathy, "Robust demand-side plug-in electric vehicle load control for renewable energy management," *Proceedings of the 2011 American Control Conference – ACC*, pp. 929–934, 2011.
- [25] A. Nayyar, J. Taylor, A. Subramanian, K. Poolla, and P. Varaiya, "Aggregate flexibility of a collection loads," *Proceedings of the 52th IEEE Conference on Decision and Control – CDC*, 2013.
- [26] M. Maasoumy, B. M. Sanandaji, K. Poolla, and A. Sangiovanni-Vincentelli, "Model predictive control of regulation services from commercial buildings to the smart grid," *Proceedings of the 2014 American Control Conference – ACC*, 2014.
- [27] "Buildings energy data book." [Online]. Available: <http://buildingsdatabook.eren.doe.gov/default.aspx>
- [28] "U.S. Energy Information Administration, annual energy review," 2010. [Online]. Available: <http://www.eia.gov/totalenergy/data/annual/#consumption>
- [29] S. Kundu, N. Sinityn, S. Backhaus, and I. Hiskens, "Modeling and control of thermostatically controlled loads," *Proceedings of the 17th Power Systems Computation Conference – PSCC*, 2011.
- [30] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "A Generalized Battery Model of a Collection of Thermostatically Controlled Loads for Providing Ancillary Service," *Proceedings of the 51th Annual Allerton Conference on Communication, Control and Computing*, pp. 551–558, 2013.
- [31] —, "Aggregate flexibility of thermostatically controlled loads," *IEEE Transactions on Power Systems, under review*, 2013. [Online]. Available: http://www.eecs.berkeley.edu/~sanandaji/my_papers/TPS-TCL.pdf
- [32] R. Malhame and C.-Y. Chong, "Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system," *IEEE Transactions on Automatic Control*, vol. 30, no. 9, pp. 854–860, 1985.
- [33] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Frequency Regulation from Flexible Loads: Potential, Economics, and Implementation," *Proceedings of the 2014 American Control Conference – ACC*, 2014.