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

Borhan M. Sanandaji[a,⋆], He Hao[a], Kameshwar Poolla[a], and Tyrone L. Vincent[b]

***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 a few clusters and improve the overall battery power limit and energy capacity. 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 (RESs) into the existing power grid is essential in achieving the envisioned sustainable energy future. Environmental, economical, and political issues 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 portfolios. The state of California, as an example, has targeted a 33% RES portfolio by 2020 [1]. Volatility, stochasticity, and intermittency characteristics of renewable energies, however, present a challenge 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 RES results in higher regulation requirements on the grid [2]–[4]. For instance, it has been shown that if California adopts its 33% renewable penetration target by 2020, the regulation procurement is anticipated to increase from 0.6 GW to 1.3 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 the short term fluctuations in the total system load and uncontrolled generation. 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 have motivated a growing interest in tapping fast responding demand-side resources for enabling deep penetration of RESs into the grid.

C. Aggregation of Flexible Loads Modeled as a Battery

Flexible loads such as Electric Vehicles (EVs) and Thermostatically Controlled Loads (TCLs) have been recently considered as good candidates for providing ancillary services to the grid [10]–[14]. Residential TCLs such as air conditioners, heat pumps, water heaters, and refrigerators, represent about 20% of the total electricity consumption in the United States [15], [16], and thus present a large potential for providing various ancillary services to the grid. TCLs have inherent thermal storage, so that electricity consumption can be varied while still meeting the desired comfort level and temperature requirements of the end user. In fact, 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 [17], [18]. The power limits and energy capacity of this battery model can be calculated in terms of TCL parameters and exogenous variables such as ambient temperature and user-specified set-points.

D. Main Contributions

In this paper, we address some practical aspects associated with the proposed battery model. First, we consider the

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

[a]Borhan M. Sanandaji, He Hao and Kameshwar Poolla are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

[b]Tyrone L. Vincent is with the Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401.

[π]Supported 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.

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 increase the net capacity. Second, we consider the effect of enforcing a requirement of no-short-cycling. In order to avoid damage, TCL manufacturers will state a minimum time that a unit must remain ON or OFF after a switch between the two states. If this minimum time is not met, the unit is said to be short-cycled. We show that the no-short-cycling constraint can be expressed simply as constraints on the first difference of the regulation or Automatic Generation Control (AGC) signal. Thus, a characterization of regulation signals that can be feasibly met by a TCL aggregation is simply the intersection of signals feasible for the stochastic battery, and this new constraint on successive changes of the regulation signal.

E. Related Work

Flexible loads in general and TCLs in particular have been recently considered for providing load following and regulation services to the grid [10], [11], [13], [19], [20]. A battery model with no dissipation is considered in [13]. Clustering and no-short-cycling of TCLs have been reported in [21], [22]. However, to the best of our knowledge, our work is the first to analytically characterize the aggregate flexibility of TCLs as a stochastic battery, their optimal clustering, and no-short-cycling constraints.

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. The paper ends with Conclusions given in Section VI.

II. Thermostatically Controlled Loads

The temperature evolution of a TCL can be described by a standard hybrid-system model

$$\dot{\theta}(t) = \begin{cases} -a(\theta(t)-\theta_a) - bP_m + w(t), & \text{ON state,} \\ -a(\theta(t)-\theta_a) + w(t), & \text{OFF state,} \end{cases} \tag{1}$$

where $\theta(t)$ is the internal temperature of the TCL at time t, θ_a is the ambient temperature, P_m is the rated electrical power, $a := \frac{1}{CR}, b := \frac{\eta}{C}$, and R, C, and η are model parameters as described in Table I. For more details on the TCL model, please see [10], [17], [19].[1] Each TCL has a temperature set-point θ_r with a hysteretic ON/OFF local control within a deadband $[\theta_r - \Delta, \theta_r + \Delta]$. The operating state $q(t)$ evolves as

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

[1]Four types of TCLs are: (i) air conditioners, (ii) heat pumps, (iii) water heaters, and (iv) refrigerators. See [18], [23] for more details.

TABLE I
Typical parameter values for a residential Air Conditioner.

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

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

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

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

$$T_{\text{ON}} = RC \ \ln \frac{\theta_r + \Delta - \theta_a + RP_m\eta}{\theta_r - \Delta - \theta_a + RP_m\eta},$$
$$T_{\text{OFF}} = RC \ \ln \frac{\theta_r - \Delta - \theta_a}{\theta_r + \Delta - \theta_a},$$

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 given by

$$P_{\text{ave}} := \sum_k P_a^k.$$

The aggregated instantaneous power consumption is

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

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

$$\dot{\theta}(t) = a(\theta_a - \theta(t)) - bp(t).$$

Note that in this model, as common in the literature, the disturbance w in Model (1) is assumed to be Gaussian distributed with zero mean and small variance [10], [19], [24], and thus neglected. Maintaining the temperature $\theta(t)$ within the user-specified deadband $\theta_r \pm \Delta$ is treated implicitly as a *constraint* on the power signal $p(t)$. When evaluating the trajectory $\theta(t)$, it is assumed that $\theta(0) = \theta_r$. The parameters that specify this continuous power model are $\chi = (a, b, \theta_a, \theta_r, \Delta, P_m)$. The nominal power required to keep a TCL at its set-point is

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

We note that P_o 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 under the continuous power model closely follows the average power P_a under the hybrid model for a wide range of operating conditions. In [17], we showed the aggregate behavior of a population of TCLs with the hybrid model could be accurately approximated by the those using the continuous power model. The continuous model was used for analysis, and the hybrid model was used in numerical experiments.

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

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

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 geometry of the set $\mathbb{U}$ is, in general, unwieldy. In [17], however, we showed that the aggregate flexibility $\mathbb{U}$ can be captured by two generalized battery models.

Definition 1: A *Generalized Battery Model* $\mathbb{B}$ is a set of signals $u(t)$ that satisfy

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

The model is specified by non-negative parameters $\phi = (\mathcal{C}, n_-, n_+, \alpha)$, and we write this compactly as $\mathbb{B}(\phi)$.

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 an succinct and compact framework to characterize the aggregate power limits and energy capacity of a population of TCLs. The following theorem is derived in [17].

Theorem 1 ([17]): Consider a heterogeneous collection of TCLs modeled by the continuous-power model with parameters χ^k. Let $\alpha > 0$ be the dissipation parameter and define

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

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

$$\mathbb{B}(\phi_1) \subseteq \mathbb{U} \subseteq \mathbb{B}(\phi_2),$$

where the necessary battery model parameters are given by

$$\phi_2 : \begin{cases} \mathcal{C} &= \sum_k \left(1 + \left|1 - \frac{a^k}{\alpha}\right|\right) \frac{\Delta^k}{b^k}, \\ n_- &= \sum_k P_o^k, \\ n_+ &= \sum_k (P_m^k - P_o^k), \end{cases} \tag{2}$$

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

$$\phi_1 : \begin{cases} \beta^k n_- &\le P_o^k, \\ \beta^k n_+ &\le P_m^k - P_o^k, \\ \beta^k \mathcal{C} &\le f^k, \end{cases} \tag{3}$$

where $\beta^k \ge 0$ satisfies $\sum_k \beta^k = 1$. Further, if $u(t) \in \mathbb{B}(\phi_1)$, the causal power allocation strategy

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

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

One should note that the gap between the proposed battery models $\mathbb{B}(\phi_1)$ and $\mathbb{B}(\phi_2)$ in Theorem 1 depend on the choice of allocation β^k, the dissipation α, and heterogeneity level of the considered 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 power limit and energy capacity 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 of these choices, a different battery model $\mathbb{B}(\phi_1)$ will be obtained that assures feasibility. One choice is

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

which yields the smallest gap between the necessary and sufficient battery models on n^+ as compared to other choices of β^k. However, it results in larger gaps for n^- and $\mathcal{C}$.[2] Based on this particular choice of β_k,

$$\phi_1 : \begin{cases} \mathcal{C} &= \sum_k (P_m^k - P_o^k) \min_k \frac{f^k}{P_m^k - P_o^k}, \\ n_- &= \sum_k (P_m^k - P_o^k) \min_k \frac{P_o^k}{P_m^k - P_o^k}, \\ n_+ &= \sum_k (P_m^k - P_o^k). \end{cases} \tag{5}$$

Note this result is valid for any fixed 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:

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

to find the *optimal dissipation* parameter α^*, for a given heterogeneous collection of TCLs. While the general case would require a numerical solution, there are analytical results in the following specific heterogeneity scenarios.

[2] We can maximize the bounds on $\mathcal{C}$ and n^- by choosing different allocations β_k. Please refer to [18] 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.[3] Under this assumption the battery capacity is

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

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

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

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

Proof: See Appendix. ■

2) Deadband: Consider the case where all of the parameters are homogenous and the only heterogeneity is in Δ^k's. Under this assumption,

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

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

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

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

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,

$$\mathcal{C}^* \gtrapprox NC_{\min}\Delta_{\min}/\eta \text{ and } \alpha^* \lessapprox 1/RC_{\min}.$$

B. Optimal Clustering

As the diversity of the TCL model parameters increases, the gap between $\mathbb{B}(\phi_1)$ and $\mathbb{B}(\phi_2)$ increases. Fig. 1 illustrates how the sufficient and necessary capacity values provided by $\mathbb{B}(\phi_1)$ and $\mathbb{B}(\phi_2)$, 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. In the following analysis, 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 heterogenous 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

$$N_i^* = N/m, \qquad i = 1, 2, \ldots, m, \tag{7}$$

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

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

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 [17], [18]. 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.

[3] In 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.

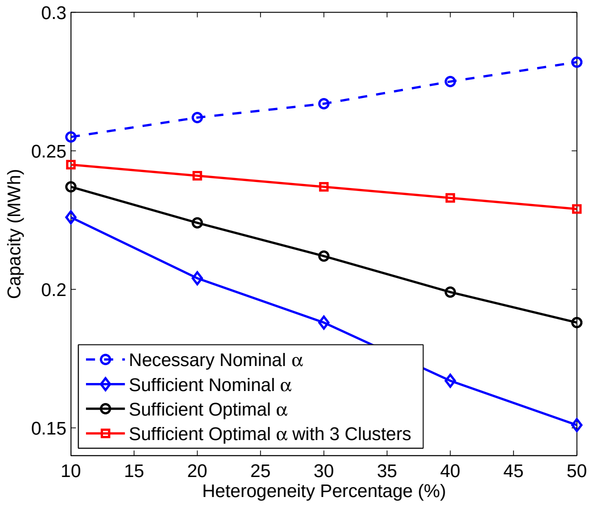

Fig. 1. The effect of optimal dissipation parameter and clustering on the battery model. A collection of 1000 heterogenous TCLs are considered whose parameter values are given in Table I. A uniform distribution is considered as the heterogeneity pattern of C^k.

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 market [25]. At each sample time, the aggregator compares the regulation signal $r(t)$ with the aggregate power deviation $\delta(t) = P_{\text{agg}}(t) - P_{\text{ave}}$, where $P_{\text{agg}}(t)$ is the instantaneous power drawn by TCLs and P_{ave} 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. The *temperature distance* is considered as the sorting criteria, i.e., $\pi^k(t) = (\theta^k(t) - \underline{\theta}^k)/\Delta^k$, where $\underline{\theta}^k = \theta_r^k - \Delta^k$. The temperature distance is normalized to account for heterogeneity. The priority-stack-based control algorithm is summarized in Algorithm 1.

B. No-Short-Cycling Constraint

The proposed priority-stack-based control scheme attempts to reduce the consecutive switching times of each

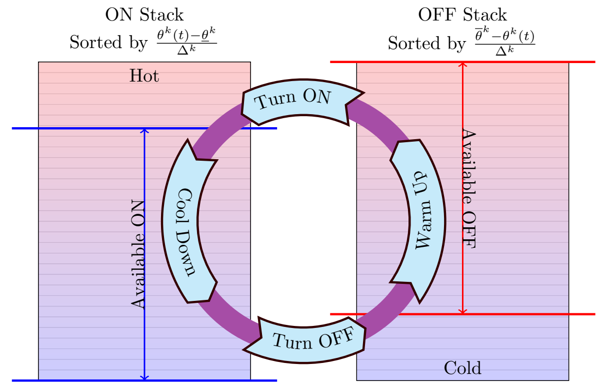

Fig. 2. The ON and OFF priority stacks with explicit no-short-cycling constraints. In general, a unit that is hotter has a higher priority to be switched ON and a unit that is cooler has a higher priority to be turned OFF. However, when no-short-cycling constraints are imposed, we are only allowed to 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 $\overline{\theta}^k = \theta_r^k + \Delta^k$.

Algorithm 1 Priority-stack-based control algorithm

loop
  receive $\pi^k(t)$, $P^k(t)$, and availability of unit k;
  construct priority stacks with no-short-cycling constraint;
  read $r(t)$;
  compute $\delta(t) = P_{\text{agg}}(t) - P_{\text{ave}}$;
  if $\delta(t) < r(t)$ **then**
    find $j^* = \min\left\{j \mid j \le N_1,\ \sum_{i=1}^{j} P^i(t) \ge r(t) - \delta(t)\right\}$;
    turn ON units indexed by $\{1, 2, \cdots, j^*\}$.
  else if $\delta(t) > r(t)$ **then**
    find $j^* = \min\left\{j \mid j \le N_0,\ \sum_{i=1}^{j} P^i(t) \ge \delta(t) - r(t)\right\}$;
    turn OFF units indexed by $\{1, 2, \cdots, j^*\}$.
  end if
end loop

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 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 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). This means that to determine feasible regulation signals, the battery model must be augmented with the constraints

$$-\eta_-(t) \le \Delta r(t) \le \eta_+(t), \tag{9}$$

where $\Delta r(t) = r(t) - r(t-1)$, and $\eta_-(t)$ and $\eta_+(t)$ are time

TABLE II

NOMENCLATURE OF SOME OF THE FREQUENTLY-USED TERMS.

Term	Description
P^k	Power draw of unit k when ON
P_{tot}	$\sum_k P^k$
P_o^k	Average power draw of unit k
P_{ave}	$\sum_k P_o^k$
$r(t)$	Regulation signal requesting power draw of $P_{\text{ave}} + r(t)$
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
$P^{\text{lim}}_{\text{ON}\to\text{OFF}}(t)$	Total power of units switched from ON to OFF at time t due to temperature bound
$P^{\text{lim}}_{\text{OFF}\to\text{ON}}(t)$	Total power of units switched from OFF to ON at time t due to temperature bound
$P_{\text{ON}}(t)$	Total power of ON units
$P_{\text{OFF}}(t)$	Total power of OFF units
$P^{\text{avail}}_{\text{ON}}(t)$	Total power of units that are available ON
$P^{\text{unavail}}_{\text{ON}}(t)$	Total power of units that are unavailable ON
$P^{\text{avail}}_{\text{OFF}}(t)$	Total power of units that are available OFF
$P^{\text{unavail}}_{\text{OFF}}(t)$	Total power of units that are unavailable OFF

varying constraints.[4] However, we will show that η_- and η_+ 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 $P^{\text{lim}}_{\text{OFF}\to\text{ON}}(k)$ or $P^{\text{lim}}_{\text{ON}\to\text{OFF}}(k)$, respectively.

Theorem 3: Assume a collection of TCLs defined by P^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 given by

$$P^{\text{avail}}_{\text{ON}}(t) = P_{\text{ave}} + r(t) - \sum_{k=t-\tau}^{t} \left(P^{\text{lim}}_{\text{OFF}\to\text{ON}}(k) + [D(k)]_+\right)$$

and

$$P^{\text{avail}}_{\text{OFF}}(t) = P_{\text{tot}} - P_{\text{ave}} - r(t) - \sum_{k=t-\tau}^{t} \left(P^{\text{lim}}_{\text{ON}\to\text{OFF}}(k) + [-D(k)]_+\right),$$

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

$$\eta_+(t) = P^{\text{avail}}_{\text{OFF}}(t-1) - \max(P^{\text{lim}}_{\text{ON}\to\text{OFF}}(t), P^{\text{lim}}_{\text{OFF}\to\text{ON}}(t)),$$
$$\eta_-(t) = P^{\text{avail}}_{\text{ON}}(t-1) - \max(P^{\text{lim}}_{\text{ON}\to\text{OFF}}(t), P^{\text{lim}}_{\text{OFF}\to\text{ON}}(t)).$$

Proof: See Appendix. ■

In the following simulation, we run the controller with the reference command shown as the solid line in Fig. 3 (a). For comparison, the power limits found using the battery model are also shown. We take $P^{\text{lim}}_{\text{OFF}\to\text{ON}}(t)$ and $P^{\text{lim}}_{\text{OFF}\to\text{ON}}(t)$ as that reported by the local unit controllers, and use that information along with $r(t)$ to calculate $\eta_+(t)$ and $\eta_-(t)$. In Fig. 3 (b) 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, and this in-feasibility continues even after $r(t)$ moves up away from the lower power limit after time 200 (s). In Fig. 3 (c) the difference between the desired regulation signal and the actual power draw $P_{\text{agg}}(t) - P_{\text{ave}}$ shows that regulation signal is not well followed downward during the time that η_- is close to zero.

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

VI. CONCLUSIONS

Thermostatically Controlled Loads (TCLs) present a large potential to provide regulation service to grid. Our prior work showed that the aggregate flexibility offered by a collection of TCLs could be succinctly modeled as a stochastic battery model with dissipation. In this paper, we addressed two practical issues associated with the proposed battery model: clustering of heterogeneous units and preventing short cycling. We showed that by finding the optimal dissipation parameter and/or clustering a population of heterogeneous units, the overall battery power limit and energy capacity could be improved substantially. Moreover, we also derived an explicit characterization of the constraints on upward and downward movement of feasible regulation signals under the no-short-cycling requirement.

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 $C^k < 1/\alpha R$, $g^k = \frac{C^k}{2-\alpha RC^k}$. Figure shows these 3 different situations. Note that at $x = \alpha$ (the breakpoint), $1/(2Rx - \alpha R) = 1/\alpha R$.

Consequently, when $\alpha < 1/RC_{\min}$, $\min_k g^k = 1/(2/C_{\min} - \alpha R)$ and when $\alpha > 1/RC_{\min}$, $\min_k g^k =$

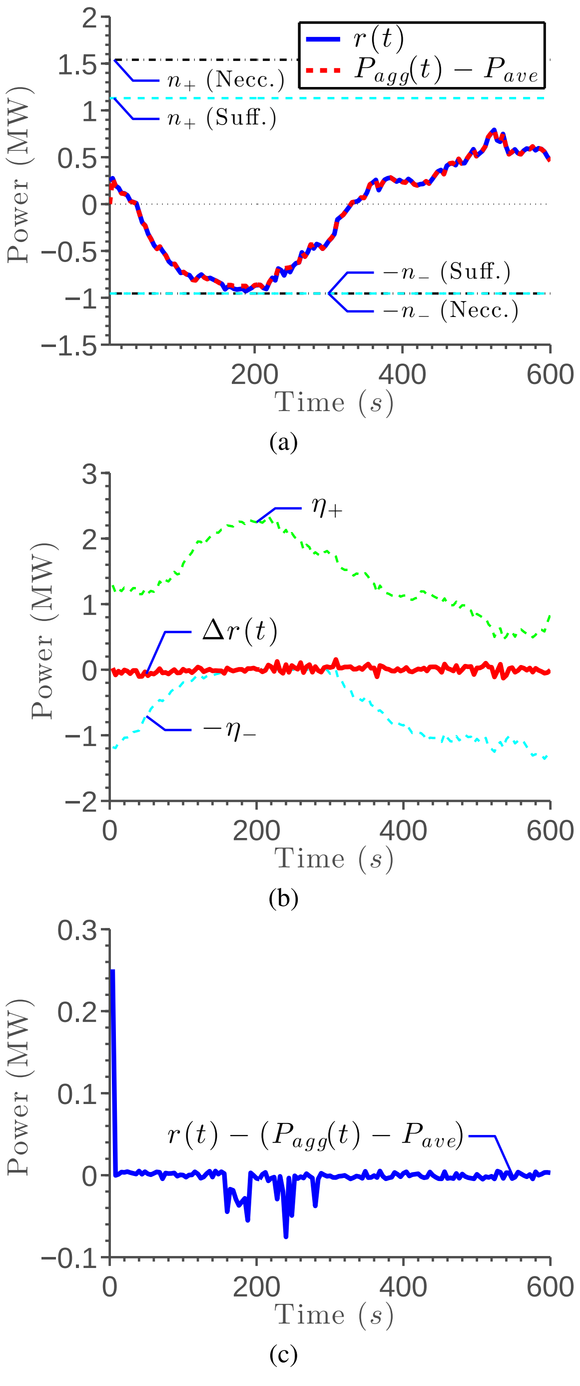

Fig. 3. Illustration of the effect of short cycling and ramping rate constraints. (a) The regulation signal and battery model bounds on power. (b) $\Delta r(t)$ and its bounds given in Theorem 3. (c) The difference between the desired regulation signal $r(t)$ and the actual power draw $P_{\mathrm{agg}}(t) - P_{\mathrm{ave}}$.

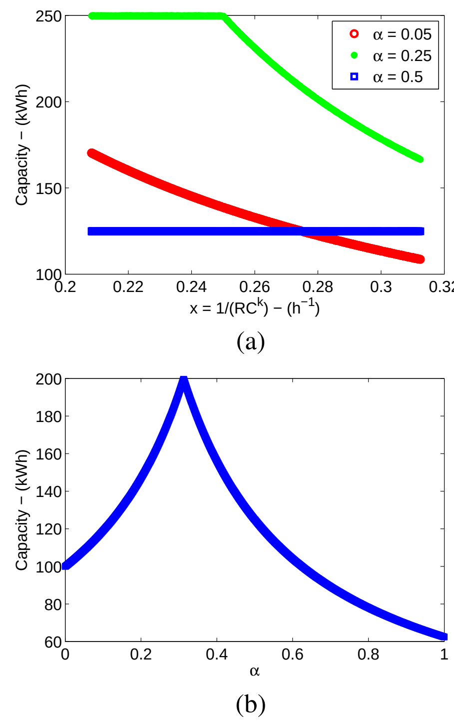

Fig. 4. A collection of 1000 heterogenous TCLs considered whose parameter values are given in Table I. The heterogeneity is in C^k. (a) Capacity over time constant $1/RC^k$ for various dissipation value. (b) Capacity $\mathcal{C}(\alpha)$ as a function of dissipation. A maximum is achieved at $\alpha^* = 1/RC_{\min}$.

$1/\alpha R$. Figure shows $\mathcal{C}(\alpha)$. The breakpoint is at $\alpha = 1/RC_{\min}$. Thus, when the heterogeneity is only in C^k we have

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

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,

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

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

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

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

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

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

$$\mathcal{C}_m^* = \big(C_{\min} + \frac{(C_{\max} - C_{\min})}{2} \frac{N}{(N-1)} \frac{(m-1)}{m}\big) N\Delta/\eta.$$

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

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

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

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

Now, $P_{\text{ON}}^{\text{unavail}}$ 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

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

where the second term represents the potential imbalance due to a change in $\Delta r(t)$ plus a difference between $P_{\text{OFF}\to\text{ON}}^{\text{lim}}(t)$ and $P_{\text{ON}\to\text{OFF}}^{\text{lim}}(t)$. Finally, the limits $\eta_+(t)$ and $\eta_-(t)$ 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. Smith, M. Milligan, E. 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, aug. 2007.

[3] Y. 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, may 2009.

[4] S. Meyn, M. Negrete-Pincetic, G. Wang, A. Kowli, and E. Shafieep-oorfard, "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," 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," in *IEEE Power Energy Society General Meeting, 2009*, july 2009, pp. 1 –8.

[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, Richland, WA, Tech. Rep. PNNL-17632, June 2008.

[10] S. Koch, J. Mathieu, and D. Callaway, "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services," in *Proc. PSCC*, 2011, pp. 1–7.

[11] J. Mathieu and D. Callaway, "State estimation and control of heterogeneous thermostatically controlled loads for load following," in *2012 45th Hawaii International Conference on System Sciences*. IEEE, 2012, pp. 2002–2011.

[12] 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, Feburary 2013.

[13] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway, "Energy arbitrage with thermostatically controlled loads," in *European Control Conference (ECC)*, 2013.

[14] A. Nayyar, J. Taylor, A. Subramanian, K. Poolla, and P. Varaiya, "Aggregate flexibility of a collection loads," *to appear in the* 52*-th IEEE Conference on Decision and Control – CDC*, 2013.

[15] "Buildings energy data book." [Online]. Available: http://buildingsdatabook.eren.doe.gov/default.aspx

[16] "U.S. Energy Information Administration, annual energy review," 2010. [Online]. Available: http://www.eia.gov/totalenergy/data/annual/#consumption

[17] 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," *to appear in proceedings of the* 51*-th Annual Allerton Conference on Communication, Control and Computing*, 2013.

[18] ——, "Aggregate flexibility of thermostatically controlled loads," *submitted to the IEEE Transactions of Power Systems*, 2013.

[19] 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.

[20] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens, "Modeling and control of thermostatically controlled loads," in *the 17-th Power Systems Computation Conference*, 2011.

[21] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, , and D. Chassin, "Aggregate model for heterogeneous thermostatically controlled loads with demand response," *in proceedings of IEEE PES General Meeting, San Diego, CA*, 2012.

[22] C. Chang, W. Zhang, J. Lian, and K. Kalsi, "Modeling and control of aggregated air conditioning loads under realistic conditions," *in proceedings of IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, 2013.

[23] J. Mathieu, M. Dyson, and D. Callaway, "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations," in *2012 ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.

[24] R. Malhame and C.-Y. Chong, "Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system," *Automatic Control, IEEE Transactions on*, vol. 30, no. 9, pp. 854–860, 1985.

[25] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Frequency regulation from flexible loads: Potential, economics, and implementation," in *American Control Conference, submitted*, September 2013. [Online]. Available: http://plaza.ufl.edu/hehao/publication.html