

Flexibility capacity of thermostatically controlled loads with cycling/lock-out constraints

Austin R. Coffman^{*,†}, Neil Cammardella^{*}, Prabir Barooah^{*}, Sean Meyn^{*}

Abstract—Thermostatically Controlled Loads (TCLs), e.g., A/Cs and water heaters, are a source of flexible power demand for the power grid: many different power consumption trajectories exist that can maintain consumers’ quality of service (QoS). Extensive research has shown that flexible loads can provide valuable grid services. Quantifying the flexibility capacity of a collection of TCLs is a well-studied problem. However, many studies consider temperature constraints alone, while most TCLs are on/off loads that have cycling (or lock-out) constraints. Studies that have considered lock-out constraints have proposed quantifications that depend on the control algorithm used to coordinate loads to provide grid services.

In this work, we present a characterization of the capacity of a collection of TCLs that considers not only temperature, but also cycling and total energy constraints. Our characterization is independent of the algorithm used to control the TCLs; it depends only on the QoS constraints on the individual TCLs. The proposed characterization can be used for planning a feasible power deviation trajectory for a collection of TCLs by solving a convex optimization problem.

I. INTRODUCTION

Currently, power balance in power grids is maintained mostly through supply-side actions, i.e., generators are ramped up and down to meet demand, resulting in negative economic and environmental impacts. These negative impacts motivate an active area of research: controlling flexible loads to provide grid support.

Flexible loads can alter their power consumption without violating consumers’ quality of service (QoS) constraints. A grid operator or balancing authority (BA) can utilize flexible loads by requesting they consume more or less power, with respect to a baseline. Baseline refers to the power consumption that would have occurred without the BA interfering. From the perspective of the BA, an increase (or decrease) of power consumption is identical to the charging (or discharging) of a battery. Due to this similarity, these resources are often termed *Virtual Energy Storage* (VES) [1]. In fact, VES can be cheaper than grid-scale batteries [2]. Examples of flexible loads that are suitable for VES are Thermostatically Controlled Loads (TCLs) [3]–[7], HVAC systems in commercial buildings [8], and electric pumps for irrigation [9] and pool cleaning [10].

To provide useful VES to the grid, a collection of TCLs have to change their power consumption so that the deviation from the baseline tracks a reference signal that is provided by a BA. Coordination of TCLs so that their aggregate power deviation tracks a reference has been studied extensively [3]–[6], [11]. An important related question is that of capacity of the collective. In the literature, no formal definition of capacity exists. A conceptual definition is that capacity represents limitations in aggregate behavior due to QoS constraints at the individual. Knowledge of capacity is essential to design a feasible reference signal for a collection of TCLs, so that the collection can track the reference without any TCL having to violate its local QoS.

A characterization of capacity must account for *all* QoS constraints. For TCLs, there are at least three QoS: (i) temperature, (ii) cycling rate (or, lock-out period), and (iii) total energy (used in providing VES over a certain time period). Since TCLs are on/off devices and frequent cycling between on and off states reduces lifetime and performance, excessive cycling must be avoided. Similarly, if the total additional energy use is not tightly constrained, consumers will have to pay extra for helping the grid. If a BA designs a reference signal with an incomplete notion of capacity, the BA must accept poor tracking or the TCL users must accept QoS violations. In both scenarios the long-term outlook is grim: either the BA views TCLs as an unreliable resource, or the TCL users view the BA as an authoritative monarch with unrealistic expectations.

A significant amount of research has focused on characterizing TCL capacity [12]–[17]. A subset of these work have only account for temperature constraints [12], [13]. The works that do include cycling rate constraints, provide answers that depend not only on the properties of the TCLs but also on the coordination algorithm used to control the TCLs [14]–[17]. Additionally, the methods in [14], [15] are not suitable for the grid to plan a reference signal for a collection of TCLs that is within the capacity of the collection.

In this work we characterize the capacity of a collection of TCLs as constraints on the aggregate power deviation. Our work is novel in three regards. First, our characterization accounts for all three QoS measures: temperature, cycling, and additional energy use. Second, the characterization is independent of the algorithm used to control the ensemble of TCLs. Third, the capacity characterization can be used by a BA to compute a reference for a collection of TCLs that is within their capacity. This is done by solving an optimization problem that is always *feasible* and *convex*. Together, these

^{*} University of Florida.

[†] corresponding author, email: bubbaroney@ufl.edu.

AC and PB are with the Dept. of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32601, USA. NC and SM are with the department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32601, USA. The research reported here has been partially supported by the NSF through award 1646229 (CPS-ECCS).

facets ensure that the reference signal so planned can be tracked with any well-designed algorithm that respects the QoS constraints of each TCL.

The effectiveness of our capacity characterization is investigated in simulation experiments. Given a desired power deviation signal from a BA in the Pacific Northeast, we compute a reference that is closest to it while within capacity of an ensemble of TCLs. The TCLs are then coordinated to track the computed reference using a priority stack controller that is a modified version of the one developed in [12]. The modification is to enforce cycling constraints strictly at each TCL. We offer a comparison of reference planning and tracking when the reference signal is planned using the aggregate capacity constraints of [12], which do not include information on individual TCLs cycling and energy QoS. The results of the comparison confirm the need to include all relevant individual TCL QoS requirements in reference planning.

The paper proceeds as follows: Section II contains descriptions of individual TCL behavior, Section III contains descriptions of aggregate TCL behavior, and Section IV contains the derived aggregate capacity constraints. In Section V, the proposed reference planning method is described. Lastly, Section VI reports the results of numerical experiments.

II. THE INDIVIDUAL TCL

A. QoS constraints

An on/off TCL is any device that turns on or off to maintain a temperature within a preset deadband. Here, we denote the state space of a TCL as X , and elements of X as the couple $x = \{m \in \{0, 1\}, \theta \in \mathbb{R}\}$, that consists of the off (0) and on (1) status (mode) and temperature of the TCL. We denote the electrical power consumption as P , which is the power consumed by the TCL when it is on. Time is discrete and denoted by the index k , with total time horizon N_t , and the TCL index is j .

The quality of service (QoS) constraints for the j^{th} TCL are that (i) the temperature must remain within $\pm\delta$ of the setpoint θ_{set} , and (ii) it can only switch once within a specified period τ_{tcl} :

$$\text{QoS 1: } |\theta_k^j - \theta_{set}| \leq \delta, \quad \forall k, \quad (1)$$

$$\text{QoS 2: } \sum_{i=0}^{\tau_{tcl}-1} |m_{k-i}^j - m_{k-1-i}^j| \leq 1, \quad \forall k. \quad (2)$$

We denote the set $Q_s^j \triangleq \{\theta_{set}, \delta, \tau_{tcl}\}^j$ as the ‘‘QoS set,’’ which contains the user defined parameters that appear in (1)-(2).

We represent ‘‘switch on and off’’ as the variables, $S_{k-1}^{\text{on},j}$ and $S_{k-1}^{\text{off},j}$, respectively. An on or off switch can occur because of two events: (i) the TCL switches to maintain the temperature QoS (1), or (ii) the TCL switches for the purpose of providing VES:

$$S_{k-1}^{\text{on},j} = F_{k-1}^{\text{on},j} + D_{k-1}^{\text{on},j}, \quad (3)$$

$$S_{k-1}^{\text{off},j} = F_{k-1}^{\text{off},j} + D_{k-1}^{\text{off},j}. \quad (4)$$

The quantity $F_{k-1}^{\text{on},j}$ (respectively, $F_{k-1}^{\text{off},j}$) represents the on switch to provide VES (respectively, off switch). The quantity $D_{k-1}^{\text{on},j}$ (respectively, $D_{k-1}^{\text{off},j}$) represents a switch to maintain the temperature QoS (1).

A TCL is *stuck off* (respectively, *stuck on*) if it is off (respectively, on) and has changed mode once in the past τ_{tcl} times. We represent stuck on and off as the variables, $\gamma_k^{\text{on},j}$ and $\gamma_k^{\text{off},j}$.

B. Modeling and Control of the individual TCL

As in much of prior work [4], [5], [12], evolution of the temperature θ_k^j is modeled in discrete time as a linear difference equation

$$\theta_{k+1}^j = \bar{a}\theta_k^j + (1 - \bar{a}) \left(\theta_k^a - R_{th}m_k^j Q_{ac} \right), \quad (5)$$

$$\bar{a} = \exp \left(\frac{-T_s}{R_{th}C_{th}} \right), \quad (6)$$

where R_{th} and C_{th} represent the thermal resistance to ambient temperature θ_k^a and thermal capacitance, respectively. The thermal power consumption Q_{ac} is related to the electrical power consumption by $Q_{ac} = \eta P$, where η is the Coefficient of Performance (COP). The thermal energy deviation quantity [12] of the j^{th} TCL is denoted

$$z_k^j \triangleq \frac{C_{th}}{\eta} (\theta_k^j - \theta_{set}). \quad (7)$$

The dynamics for thermal energy are obtained by substituting the definitions for z_k^j and z_{k+1}^j into (5),

$$z_{k+1}^j = \bar{a}z_k^j - b \left(m_k^j P - \frac{\theta_k^a - \theta_{set}}{\eta R_{th}} \right), \quad (8)$$

$$b = (1 - \bar{a})C_{th}R_{th}. \quad (9)$$

We identify the RHS term in parenthesis in (8) as the power deviation for the j^{th} TCL, so that the baseline power consumption for the j^{th} TCL is,

$$\bar{P}_k^j = \frac{\theta_k^a - \theta_{set}}{\eta R_{th}}. \quad (10)$$

This form of the baseline power consumption is a consequence of the equation (5) used to model the TCL.

III. AGGREGATE QUANTITIES

Section II was devoted to the individual TCL; we now define variables for a collection of N TCLs. Two quantities are of interest at the aggregate level: (i) quantities in units of power (Watt) and (ii) fractional quantities normalized by N , the number of TCLs. Furthermore, a homogeneous collection is defined as an ensemble of TCLs for which the parameters (C_{th}, R_{th}, η, P) that appear in (5) and the QoS set Q_s^j are uniform over the population. A homogeneous collection is considered in the following.

The total and maximum power consumption of the collection at time k are, respectively,

$$Y_k \triangleq N_k^{\text{on}}P = P \sum_{j=1}^N m_k^j, \quad \text{and,} \quad P_{\text{aggmax}} \triangleq NP, \quad (11)$$

where N_k^{on} is the number of TCLs on at time k ; the number of TCLs off at time k is $N_k^{\text{off}} = N - N_k^{\text{on}}$. The aggregate thermal energy deviation is denoted

$$z_k \triangleq \sum_{j=1}^N z_k^j, \quad (12)$$

which has dynamics

$$z_{k+1} = \bar{a}z_k - br_k. \quad (13)$$

Another important aggregate quantity is the baseline power consumption for the ensemble, denoted

$$\bar{P}_k \triangleq \sum_{j=1}^N \bar{P}_k^j = N \left(\frac{\theta_k^a - \theta_{\text{set}}}{\eta R_{th}} \right). \quad (14)$$

The fractional quantities are denoted

$$s_k^{\text{on}} \triangleq \frac{\sum_{j=1}^N S_k^{\text{on},j}}{N}, \quad s_k^{\text{off}} \triangleq \frac{\sum_{j=1}^N S_k^{\text{off},j}}{N}, \quad (15)$$

$$n_k^{\text{on}} \triangleq \frac{N_k^{\text{on}}}{N}, \quad n_k^{\text{off}} \triangleq \frac{N_k^{\text{off}}}{N}, \quad (16)$$

$$d_k^{\text{on}} \triangleq \frac{\sum_{j=1}^N D_k^{\text{on},j}}{N}, \quad d_k^{\text{off}} \triangleq \frac{\sum_{j=1}^N D_k^{\text{off},j}}{N}, \quad (17)$$

$$\gamma_k^{\text{on}} \triangleq \frac{\sum_{j=1}^N \gamma_k^{\text{on},j}}{N}, \quad \gamma_k^{\text{off}} \triangleq \frac{\sum_{j=1}^N \gamma_k^{\text{off},j}}{N}. \quad (18)$$

The aggregate power deviation, over baseline power consumption, is denoted

$$y_k \triangleq n_k^{\text{on}} P_{\text{aggmax}} - \bar{P}_k. \quad (19)$$

The power deviation reference signal is

$$r_k \triangleq \text{Desired value of } y_k \text{ at time } k. \quad (20)$$

Comment 1: For a homogeneous collection of TCLs, the fraction of loads on and the total power consumption are proportional. Thus in the developments to follow, “fraction of loads on” and “total power consumption” can be freely interchanged, modulo a scaling factor.

IV. CONSTRAINTS FOR REFERENCE PLANNING

Aggregate capacity constraints refers to constraints on aggregate quantities due to constraints at the individual TCL, e.g. (1)-(2). We formulate constraints on aggregate quantities of the two individual TCL states (i) power deviation (19) and (ii) thermal energy deviation (12). Our constraints on aggregate power and thermal energy deviation account for the temperature and cycling constraints at the individual, specified by (1)-(2). That is, these aggregate constraints ensure that if a power and thermal energy deviation trajectory were to satisfy them, then a collection of TCLs could track the power deviation signal while enforcing (1)-(2). Conversely, if the aggregate constraints are violated, then there would exist at least a single TCL that violates its individual QoS constraints.

A. Fraction Stuck

The fraction of TCL's stuck on, or off, can be represented as an inventory model with deterministic demand,

$$\gamma_k^{\text{on}} = \gamma_{k-1}^{\text{on}} + s_{k-1}^{\text{on}} - s_{k-1-\tau_{BA}}^{\text{on}}, \quad (21)$$

where τ_{BA} is a design parameter that will be discussed shortly. In words, (21) means the fraction that is stuck on, γ_{k-1}^{on} (respectively, off), is increased by the fraction that switch on s_{k-1}^{on} (respectively, off) from $k-1$ to k and decreased by the fraction that had switched on (respectively, off) $k-1-\tau_{BA}$ sample times in the past. We define an input for stuck on (respectively, off) as the following column vector,

$$u_{k-1}^{\text{on}} \triangleq [s_{k-1}^{\text{on}}, \dots, s_{k-1-\tau_{BA}}^{\text{on}}]^T, \quad (22)$$

the T superscript denotes matrix transpose. Eq. (21) can now be represented as follows, which is a linear state space model,

$$\gamma_k^{\text{on}} = \gamma_{k-1}^{\text{on}} + B(\tau_{BA}) u_{k-1}^{\text{on}}, \quad \gamma_0^{\text{on}} = 0, \quad (23)$$

$$\gamma_k^{\text{off}} = \gamma_{k-1}^{\text{off}} + B(\tau_{BA}) u_{k-1}^{\text{off}}, \quad \gamma_0^{\text{off}} = 0. \quad (24)$$

For both systems, the matrix $B(\tau_{BA})$ is

$$B(\tau_{BA}) \triangleq [1, \mathbf{0}_{\tau_{BA}-2}, -1], \quad (25)$$

where $\mathbf{0}_\tau$ is a row vector of zeros of length τ . The quantity τ_{BA} is elected as $\tau_{BA} > \tau_{tcl}$, and is the cycling QoS parameter the BA uses for reference planning.

1) $\tau_{BA} > \tau_{tcl}$: While TCLs may have lockout times as short as 5 minutes [18], this does not mean it is desirable for a TCL to switch every 5 minutes. So, using $\tau_{BA} > \tau_{tcl}$ will allow the BA to plan a reference signal that would require TCLs to switch less over a given time horizon.

B. Power Deviation Limits

We start by considering how much the fraction of on devices could be changed in a given sample time, relative to the current fraction of on devices. To obtain an upper bound on the change $n_k^{\text{on}} - n_{k-1}^{\text{on}}$, assume that $n_k^{\text{on}} \geq n_{k-1}^{\text{on}}$. The quantity $n_{k-1}^{\text{off}} - \gamma_{k-1}^{\text{off}}$ represents the current fraction of TCLs that are off and can switch on, so that the upper bound on $n_k^{\text{on}} - n_{k-1}^{\text{on}}$ should include at least:

$$n_{k-1}^{\text{off}} - \gamma_{k-1}^{\text{off}}.$$

However, this is not complete as some TCLs may be forced to switch due to the temperature constraint (1). The upper bound should then be increased by d_{k-1}^{on} and decreased by d_{k-1}^{off} . Letting $\Delta d_{k-1} := d_{k-1}^{\text{on}} - d_{k-1}^{\text{off}}$, an upper bound is

$$n_k^{\text{on}} \leq 1 - \gamma_{k-1}^{\text{off}} + \Delta d_{k-1}, \quad (26)$$

where n_{k-1}^{on} is eliminated through the relation $n_{k-1}^{\text{off}} = 1 - n_{k-1}^{\text{on}}$. The steps necessary to obtain the lower bound are symmetric, and the result is

$$\gamma_{k-1}^{\text{on}} + \Delta d_{k-1} \leq n_k^{\text{on}} \leq 1 - \gamma_{k-1}^{\text{off}} + \Delta d_{k-1}. \quad (27)$$

Neglecting the quantity Δd_k results in,

$$\gamma_{k-1}^{\text{on}} \leq n_k^{\text{on}} \leq 1 - \gamma_{k-1}^{\text{off}}. \quad (28)$$

With Δd_k neglected, measures must be taken so that reference signals designed with (28) do not cause significant tracking errors. We believe the following will help mitigate tracking errors: $\tau_{BA} > \tau_{tcl}$ where τ_{tcl} and τ_{BA} are the cycling QoS parameters for individual TCLs and the BA, respectively (described near (23)-(24)).

When a BA designs a reference signal with $\tau_{BA} > \tau_{tcl}$, it is underestimating the capacity of the collection. That is, the BA is assuming that TCLs can switch less than they actually can. The hope is that when Δd_k contributes in a non-conservative manner to (27), the extra capacity available due to $\tau_{BA} > \tau_{tcl}$ will enable the collection to counteract the effect of Δd_k and continue to track the reference signal designed with (28). This hypothesis is numerically tested in [19], where the action $\tau_{BA} > \tau_{tcl}$ reduces the tracking error in the performed numerical experiments.

C. Thermal Energy Limits

From (1) and (7), it follows that $|z_k^i| \leq \frac{C_{th}\delta}{\eta} =: \bar{C}$. The bounds on the aggregate thermal energy are obtained by applying the triangle inequality,

$$|z_k| = \left| \sum_{i=1}^N z_k^i \right| \leq \sum_{i=1}^N |z_k^i| \leq N\bar{C} \quad (29)$$

This is not a new result, and often credited to the work of Hao et al. [12].

D. Relation to “fraction on”

The fraction stuck on γ_k^{on} (respectively, off γ_k^{off}) is related to the fraction of on (respectively, off) switches through the dynamics (23) - (24). Another “inventory equation” couples switching and power model dynamics:

$$n_k^{\text{on}} = n_{k-1}^{\text{on}} + s_{k-1}^{\text{on}} - s_{k-1}^{\text{off}}. \quad (30)$$

In words, the fraction of on devices at time k is the fraction on at $k-1$, plus the fraction that switch on (s_{k-1}^{on}) and minus the fraction that switch off (s_{k-1}^{off}) from time step $k-1$ to time step k . For notational consistency, we re-write (30) as,

$$n_k^{\text{on}} = n_{k-1}^{\text{on}} + u_{k-1}^{\text{on}}[1] - u_{k-1}^{\text{off}}[1], \quad (31)$$

where $u_{k-1}^{\text{off}}[1]$ represents the first element of the vector u_{k-1}^{off} , i.e. $u_{k-1}^{\text{off}}[1] = s_{k-1}^{\text{off}}$.

E. Additional energy use

The total additional energy used by the TCLs in providing VES should be 0 over a long term so that the loads are in fact providing a battery-like service and not a generator-like or load-like service. Fixing a planning time horizon N_t , this requirement becomes

$$\sum_{k=0}^{N_t-1} y_k = 0. \quad (32)$$

F. Capacity of the collective

Suppose a BA is interested in computing a reference $\{r_k\}_{k=1}^{N_t}$ over a time interval N_t such that the resulting reference is within the capacity of a collection of N TCLs. We characterize the capacity as a set of constraints by collecting the “battery model” (13) with the aggregate power deviation constraint (28) and thermal energy deviation constraint (29). The aggregate power (28) and thermal energy (29) deviation constraints are coupled to the variables of the battery model (13) through (30). Due to this coupling, several related signals must also be constrained, not simply $\{r_k\}_{k=1}^{N_t}$. We define the vector

$$\psi \triangleq [\{z_k\}_1^{N_t}, \{y_k\}_0^{N_t-1}, \{u_k^{\text{on}}\}_0^{N_t-1}, \dots, \{u_k^{\text{off}}\}_0^{N_t-1}, \{\gamma_k^{\text{on}}\}_1^{N_t}, \{\gamma_k^{\text{off}}\}_1^{N_t}]. \quad (33)$$

The constraint set that defines the capacity of the TCL ensemble, based on the aggregate constraints developed in Section IV, is:

$$\Omega \triangleq \left\{ \psi \text{ s.t. } (29), (28), (31), (23), (24), (13) \text{ and } (32) \text{ hold for } k = 0, 1, \dots, N_t \right\}. \quad (34)$$

V. REFERENCE PLANNING

The *reference planning* problem utilizes the capacity characterization developed in Section IV to plan an optimal power deviation trajectory (reference signal) for an ensemble of TCLs. The problem data is the BA’s *total desired demand deviation*, r_k^{BA} . The signal r_k^{BA} can be a regulation signal, or it can be obtained by filtering the net load [1], though a discussion is outside the scope of this work. The BA would ideally like $y_k = r_k^{BA}$ but r_k^{BA} can be outside the capacity of the collection, so the goal is to find a reference r_k that is closest to r_k^{BA} but within the capacity of the TCLs.

Reference planning is achieved by projecting the signal r_k^{BA} onto the aggregate constraint set, Ω . We first define the full length projection vector as,

$$\psi^{BA} \triangleq [\{0\}_1^{N_t}, \{r_k^{BA}\}_1^{N_t}, \{0\}_1^{N_t}, \{0\}_1^{N_t}, \{0\}_1^{N_t}, \{0\}_1^{N_t}],$$

so that the optimization problem to be solved is,

$$\begin{aligned} \min_{\psi} J(\psi) &= \sum_{k=0}^{N_t-1} \|\psi_k^{BA} - \psi_k\|_2^2 \xi_k \\ \text{s.t. } &\psi \in \Omega \end{aligned} \quad (35)$$

where N_t is the planning horizon and $\xi_k > 0$ are weights. The component $\{y_k^*\}$ of the optimal solution ψ^* is the optimal reference r_k for the TCLs.

VI. NUMERICAL EXPERIMENTS

We survey here numerical experiments conducted to compare our proposed reference planning method to the capacity constraints of [12], which we term the alternative method. For clarity, the simulated TCLs are residential air conditioner units (ACs). Additionally, all scenarios involve the solution

TABLE I
SIMULATION PARAMETERS

Par.	Unit	value	Par.	Unit	value
N	thousand	60	η	N/A	2.5
$\bar{C}$	MWh	50	θ^a	$^{\circ}C$	30
τ_{BA}	Mins.	20	θ_{set}	$^{\circ}C$	21
τ_{tcl}	Mins.	10	Δ	$^{\circ}C$	2
R_{th}	$KW/^{\circ}C$	2.5	T_s	Mins.	2
C_{th}	$^{\circ}C/KWh$	2.5	P	KW	2.24
$\bar{P}$	MW	86	P_{aggmax}	MW	134.4

of a convex optimization problem, which is performed using CVX [20]. For all experiments the sampling time is $T_s = 2$ minutes.

In the method comparison scenario, the purpose is to illustrate that all individual TCL QoS *must* be accounted for in reference planning. The alternative method does *not* account for the individuals' cycling constraint. The proposed and alternative methods are used to plan two reference signals. We then use a priority stack controller to coordinate an ensemble of TCLs to track the planned reference signals. Under the priority stack control, we present two tracking scenarios: (t-i) tracking the reference from the proposed method and (t-ii) tracking the reference from the alternative method. We find that only in scenario (t-i) will the ensemble of TCLs be able to track the planned reference while each individual enforces its own QoS.

The priority stack controller used in these experiments is a modified version of the one presented in [12], so to also enforce the individual TCLs cycling QoS (2); it by default enforces the temperature QoS (1).

A. Method Comparison: Reference Planning and Tracking

For both reference planning methods the BA supplied reference, r_k^{BA} , is obtained from BPA, a Balancing Authority in the Pacific Northwest of the United States, and is shown in Figure 1. The parameters for the individual loads are elected based on the values provided in [21] and are shown in Table I, along with other simulation parameters.

Figure 1 shows the planned reference signals for both methods. The reference signal planned with the proposed method is noticeably less aggressive than the reference signal planned with the alternative method. That is, when cycling constraints are not taken into account higher ramp rates are asked from the collection of TCLs. As we will see briefly, this leads to either poor reference tracking, violation of individual TCLs QoS, or both.

In Figure 2 (top), the reference tracking results are shown for our proposed method that includes cycling information in reference planning. The priority stack controller is able to coordinate the collection of AC units to track the planned reference signal with minimal tracking error (see Table II). For verification, the individual cycling QoS results are shown in Figure 2 (bottom). Every AC unit maintains to the preset level, as no units cycle faster than $\tau_{tcl} = 10$ minutes.

In Figure 3 (top), the reference tracking results are shown for the alternative method that does not include cycling

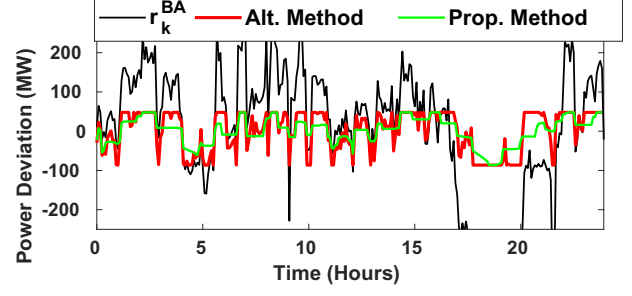


Fig. 1. BA signal (r_k^{BA}) and the reference trajectories (r_k) for a collection of 60,000 TCLs.

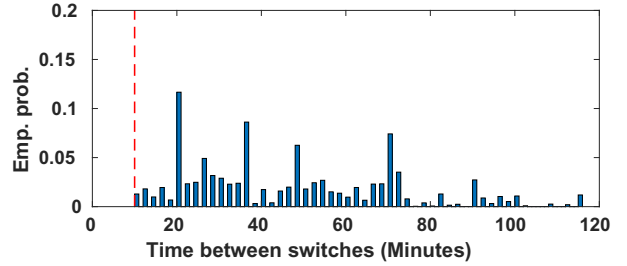
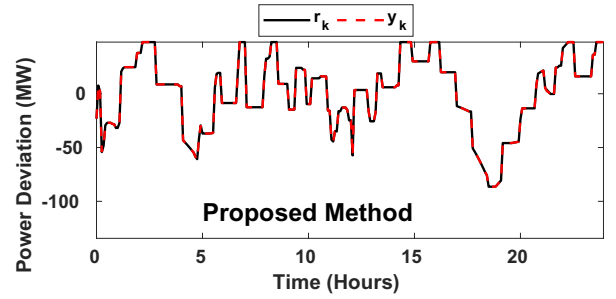


Fig. 2. Results of tracking the reference planned from the proposed method. (Top): reference tracking results, (Bottom): individual TCL cycling QoS results. The dashed red line indicates τ_{tcl} .

information in reference planning. Since this reference is beyond the capacity of the TCLs, and the priority stack controller enforces cycling QoS, it is unable to coordinate the collection of AC units to track the planned reference. For comparison, the reference tracking error reported in Table II is *two orders of magnitude* higher than the error with our proposed method. This illustrates that TCLs cycling constraints should be incorporated in reference planning.

Another consequence of the reference from the alternative method neglecting the capacity is that this actually *prevents* the priority stack controller from enforcing the cycling QoS, Figure 3 (bottom). The reference signal is requiring TCLs to switch on or off too close to the deadband, so that when a TCL switches to enforce (1) it will have switched in a time less than τ_{tcl} from its previous switch.

TABLE II
REFERENCE TRACKING ERRORS

Reference planning method	Tracking Error
Proposed method	0.06 %
Alternative method	21 %

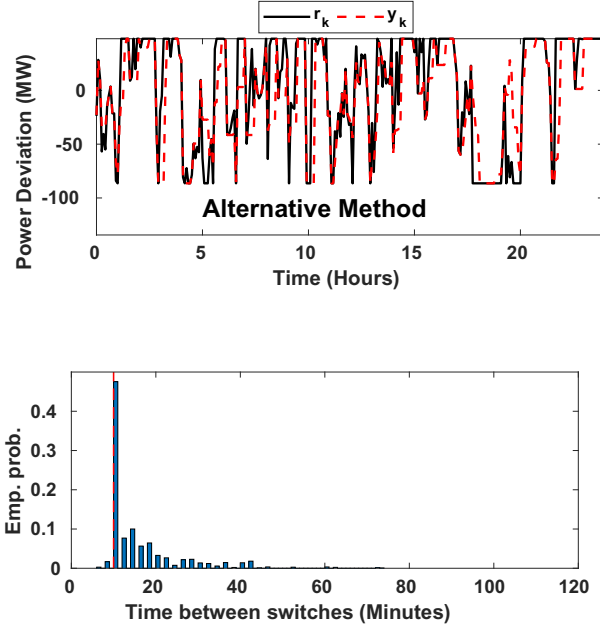


Fig. 3. Results of tracking the reference planned from the alternative method. (Top): reference tracking results, (Bottom): individual TCL cycling QoS results. The dashed red line indicates τ_{tcl} .

VII. CONCLUSION

We present an aggregate capacity characterization for collections of TCLs that takes into account constraints on temperature, cycling rate and energy use of the TCLs. We then use this characterization to pose the BA's reference planning problem as an optimization problem, in which the power deviation desired by the BA is projected onto the set of signals that are within the ensemble's capacity. As updated predictions for the BA's needs become available it can plan a new reference for the TCL collection. The proposed method ensures the reference so computed is always within capacity of the collection. All the developments in the paper are applicable to time varying ambient conditions.

There are many avenues for extensions. One is to heterogeneous loads. Another is a method for capacity characterization of a collection that does not need information about every single load. Yet another is a capacity characterization that is suitable for long-term planning, such as, what fraction of a BA's long term storage needs can be satisfied by 1000 or 10,000 TCLs? Finally, a question that is as important as it has been neglected in flexible load research: what is the kW/kWh equivalent of a virtual battery made of a collection of flexible loads?

REFERENCES

- [1] P. Barooah, *Smart Grid Control: An Overview and Research Opportunities*. Springer Verlag, 2019, ch. Virtual energy storage from flexible loads: distributed control with QoS constraints, pp. 99–115.
- [2] N. J. Cammardella, R. W. Moye, Y. Chen, and S. P. Meyn, "An energy storage cost comparison: Li-ion batteries vs Distributed load control," in *2018 Clemson University Power Systems Conference (PSC)*, Sep. 2018, pp. 1–6.
- [3] Y. Chen, M. U. Hashmi, J. Mathias, A. Bušić, and S. Meyn, "Distributed control design for balancing the grid using flexible loads," in *IMA Volume on the Control of Energy Markets and Grids*, 2017, pp. 1–26.
- [4] A. Coffman, A. Bušić, and P. Barooah, "A study of virtual energy storage from thermostatically controlled loads under time-varying weather conditions," in *5th International Conference on High Performance Buildings*, July 2018, pp. 1–10.
- [5] 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, pp. 430–440, 2013.
- [6] 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.
- [7] M. Liu, S. Peeters, D. S. Callaway, and B. J. Claessens, "Trajectory tracking with an aggregation of domestic hot water heaters: Combining model-based and model-free control in a commercial deployment," *IEEE Transactions on Smart Grid*, 2019.
- [8] H. Hao, A. Kowli, Y. Lin, P. Barooah, and S. Meyn, "Ancillary service for the grid via control of commercial building HVAC systems," in *American Control Conference*, June 2013, pp. 467–472.
- [9] A. Aghajanzadeh and P. Therkelsen, "Agricultural demand response for decarbonizing the electricity grid," *Journal of Cleaner Production*, vol. 220, pp. 827 – 835, 2019.
- [10] Y. Chen, A. Bušić, and S. Meyn, "State estimation for the individual and the population in mean field control with application to demand dispatch," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1138–1149, March 2017.
- [11] M. Liu and Y. Shi, "Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1963–1971, May 2016.
- [12] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 189–198, Jan 2015.
- [13] L. Zhao, W. Zhang, H. Hao, and K. Kalsi, "A geometric approach to aggregate flexibility modeling of thermostatically controlled loads," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4721–4731, Nov 2017.
- [14] C. Ziras, S. You, H. W. Bindner, and E. Vrettos, "A new method for handling lockout constraints on controlled TCL aggregations," in *2018 Power Systems Computation Conference (PSCC)*, June 2018, pp. 1–7.
- [15] B. M. Sanandaji, T. L. Vincent, and K. Poolla, "Ramping rate flexibility of residential hvac loads," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 865–874, April 2016.
- [16] D. Cheng, W. Zhang, and K. Wang, "Hierarchical reserve allocation with air conditioning loads considering lock time using benders decomposition," *International Journal of Electrical Power & Energy Systems*, vol. 110, pp. 293 – 308, 2019.
- [17] A. R. Coffman, A. Bušić, and P. Barooah, "Aggregate capacity for TCLs providing virtual energy storage with cycling constraints," in *IEEE Conference on Decision and Control*, December 2019.
- [18] T. C. Chang, "Time interlock for air conditioning compressor and the like," Nov 1973, US Patent 3,774,082.
- [19] A. Coffman, N. Cammardella, P. Barooah, and S. Meyn, "Aggregate capacity of TCLs with cycling constraints," *arXiv preprint arXiv:1909.11497*, 2019.
- [20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," <http://cvxr.com/cvx>, Feb. 2011.
- [21] 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.