Automatica 152 (2023) 111002

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# A unified framework for coordination of thermostatically controlled loads[☆]

Austin Coffman[a], Ana Bušić[b], Prabir Barooah[a,*]

[a] University of Florida, Gainesville, FL, USA
[b] Inria, Paris, France

ARTICLE INFO

*Article history:*
Received 13 August 2021
Received in revised form 13 October 2022
Accepted 23 February 2023
Available online 1 April 2023

*Keywords:*
Decentralized control
Power grid
Demand response
Randomized control
Thermostatically controlled loads

ABSTRACT

A collection of thermostatically controlled loads (TCLs) – such as air conditioners – can vary their power consumption within limits to help the balancing authority of a power grid maintain demand supply balance. Doing so requires loads to coordinate their on/off decisions so that the aggregate power consumption tracks a grid-supplied reference. At the same time, each consumer's quality of service (QoS) must be maintained. While there is a large body of work on TCL coordination, they do not provide guarantees on the reference tracking performance or QoS maintenance, and they do not provide a means to compute a suitable reference signal for power demand of a collection of TCLs. In this work we provide a framework that addresses these two weaknesses. The framework enables coordination of an arbitrary number of TCLs that: (i) is computationally efficient, (ii) is implementable at the TCLs with local feedback and low communication, and (iii) enables reference tracking by the collection while ensuring that temperature and cycling constraints are satisfied at every TCL at all times. The framework is based on a Markov model obtained by discretizing a pair of Fokker–Planck equations derived in earlier work by Malhame and Chong (1985). We then use this model to design randomized policies for TCLs. The balancing authority broadcasts the same policy to all TCLs, and each TCL implements this policy which requires only local measurement to make on/off decisions. Simulation results are provided to support these claims. Matlab implementation is made publicly available.



## 1. Introduction

Thermostatically controlled loads (TCLs) – such as residential air conditioners, heat pumps, space heaters, refrigerators and water heaters – are recognized to be valuable sources of *flexible* demand (Callaway & Hiskens, 2011; Chen, Hashmi, Mathias, Bušić and Meyn, 2017; Lee et al., 2020; Mathieu, Koch, & Callaway, 2013). They can vary their demand from the nominal without adversely affecting consumers' quality of service (QoS). The flexibility can be used by a balancing authority (BA) to balance supply and demand in a power grid. Since the rated power of each load is small, it is necessary to coordinate a collection of loads. Coordination of TCLs involves two conflicting requirements: (i) the TCLs collectively need to track a reference power demand, and (ii) every TCL's QoS need to be maintained. For air conditioners, refrigerators and heat pumps with *on/off actuation*, there are at least two QoS requirements: the space temperature must be maintained within a prespecified range and compressor short-cycling must be avoided. Meaning, once the compressor turns on it cannot turn off until a prespecified time period elapses, and vice versa. Variable speed air conditioners are common in commercial – and increasingly in residential – buildings; but they are outside the scope of this paper.

A framework for coordinating TCLs needs two parts. A coordination scheme, consisting of an algorithm and information exchange architecture is one part. The other part is reference computation: the framework must provide the BA with a method to determine a suitable reference signal for the TCLs. The reference should be feasible for the collection, meaning it should be possible for the TCLs to collectively track the reference while each TCL maintains its QoS. Otherwise, even the best coordination scheme will fail to meet either the BA's need, which is reference tracking, or the consumers' need, which is maintaining indoor temperature etc., or both.

There has been intense research on the problem of designing TCL coordination schemes to support the power grid. There has also been some work on computing appropriate references. We

[☆] The research reported here has been partially supported by the National Science Foundation through awards 2122313 (ECCS) and 1934322 (CMMI), and the French National Research Agency grant ANR-16-CE05-0008. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Subhrakanti Dey under the direction of Editor Ian R. Petersen.

[*] Corresponding author.
*E-mail addresses:* bubbaroney@ufl.edu (A. Coffman), ana.busic@inria.fr (A. Bušić), pbarooah@ufl.edu (P. Barooah).

https://doi.org/10.1016/j.automatica.2023.111002
0005-1098/

will discuss these in detail in Section 1.1. These two bodies of work have been carried out in their own silos, and it is not easy to combine them. For instance, existing work on reference computation does not guarantee that there exists a coordination scheme that can track the computed reference.

This work presents a framework for coordination of a collection of TCLs for providing demand side services to the power grid. The framework includes both of the above mentioned components, i.e. (i) planning a suitable reference for a collection of TCLs and (ii) designing an algorithm for coordinating the TCLs to track the reference, so that both the BA's requirement and consumers' QoS are satisfied. To avoid confusion, we use "controller" for the computations at the BA and "policy" for the on/off decision maker at each TCL. The proposed framework uses a decentralized information architecture adopted in many prior works, in which a controller at the BA broadcasts a signal to all the TCLs and the on/off decision making policy at each TCL is influenced by that signal. In particular, the BA computes *randomized policies* for the TCLs and broadcasts them to all the TCLs. Each TCL receives the same policy and implements it – to make on/off decisions – using locally measurable information. The framework is computationally tractable for an arbitrary number of TCLs. The communication burden is low: only a few numbers need to be broadcast by the BA at every sampling instant. Feedback from TCLs to the BA can be infrequent.

### 1.1. Literature review and contribution

Centralized control in which the BA directly commands on/off status of each TCL is not scalable to large populations. Among non-centralized coordination schemes, an idea that many works on TCL coordination use is for the BA to broadcast a low dimensional control command to all TCLs, which is translated by each TCL into its actuation command with a local policy. We classify the information architecture in such a scheme as *decentralized*. Another non-centralized information architecture is *distributed*, in which decisions at a TCL are computed based on information exchanged with a set of neighboring TCLs.

**Decentralized:** The literature on decentralized coordination of TCLs differ in their choice of the broadcast signal (i.e., BA's control command) and the policy at the TCL that translates this broadcast to on/off decisions. The literature can be divided into two broad categories based on these choices: (i) thermostat set point change and (ii) probabilistic. There are many forms of probabilistic policies, which can be roughly subdivided into two sub categories: (ii-A) bin switching and (ii-B) randomized. We discuss these in detail below.

In coordination schemes based on thermostat setpoint change, a time-varying thermostat set point is broadcast by the BA to all TCLs, and each TCL makes on/off decisions based on this new setpoint, e.g. Bashash and Fathy (2013), Callaway and Hiskens (2011), Lee and Max Zhang (2021), Mahdavi, Braslavsky, Seron, and West (2017) and Soudjani and Abate (2015). The underlying thermostat policy is not changed. This approach may ask for an extremely small change in thermostat setpoint in response to small changes in the aggregate power reference, far below the resolution of the temperature sensor at each TCL. Retrofitting thermostats to increase the temperature resolution will lead to chattering due to inevitable measurement noise. Or the method may ask for large changes which may violate consumers' QoS.

In probabilistic bin-switching coordination schemes, the TCL policy – the mapping from BA's broadcast command to a TCL's on/off decision – is a non-deterministic mapping, e.g. Chen, Hashmi et al. (2017), Coffman, Bušić, and Barooah (2018), Liu, Shi, and Liu (2016) and Mathieu et al. (2013). Works in this category typically first model the population of TCLs under thermostat control as a Markov chain. The continuous temperature range is divided into a number of discrete bins. A finite dimensional state vector, a probability mass function, is then defined. Each entry of the state vector represents "the fraction of TCLs that are on (or off) and has temperature in a certain range". The control command from the BA is chosen so as to affect the fraction of TCLs in the temperature bins directly (Liu & Shi, 2016; Liu et al., 2016; Mathieu et al., 2013; Totu, Wisniewski, & Leth, 2017). Since the basic Markov model is derived for the thermostat policy, introduction of the BA's control to manipulate TCLs' on/off state is somewhat ad-hoc. In Mathieu et al. (2013), the BA's control command is chosen to be another vector, whose $i$th entry represents "the fraction of TCLs in bin $i$ to increase/decrease". A policy is then proposed to translate this command to on/off action at each TCL, which requires knowledge of the state of the Markov model, leading to a state estimation problem. In Liu et al. (2016), BA's control command is chosen to be a scalar. The probability of a TCL turning on or off is proportional to this scalar. Subsequent works have proposed various refinements, such as BA's command affecting the rate of fractions to switch instead of fraction to switch (Totu et al., 2017). Providing performance guarantees with bin switching architecture has proved challenging, either on reference tracking or on QoS maintenance for individual TCLs.

An alternative to bin switching that still uses probabilistic on/off decision making is randomized policy, e.g. Bušić and Meyn (2016), Chen, Hashmi et al. (2017) and Coffman et al. (2018). A randomized policy is a specification of the conditional probability of turning on or off given the current state of the TCL. On/off decisions are computed with the help of a random number generator and the policy. In this architecture it is envisioned that the thermostat policy at the TCL is replaced with a randomized policy. The key advantage of doing so is that the control design problem is converted to controlling the probability of the TCL being "on". By the law of large numbers, this fraction is close to the fraction of TCLs "on" for a large number of TCLs (Meyn, Barooah, Bušić, Chen, & Ehren, 2015). For a homogeneous collection the aggregate power consumption is equal to the number of TCLs times the fraction of TLCs "on". Even for a heterogeneous collection the two are approximately equal except in case of severe heterogeneity. The coordination problem thus simplifies to the problem of designing the randomized policy that manipulates the probability of a single TCL being "on". Aggregate demand is manipulated by the BA by changing the policy, as discussed next.

In Bušić and Meyn (2016) and Chen, Hashmi et al. (2017), the randomized policy is parameterized by a scalar $\zeta(t)$. Coordination of the population is then achieved by appropriate design of $\zeta(t)$, which is computed and broadcast by the BA. This scheme also uses a Markov model of the evolution of binned temperature, but assumes a certain factorization: the next values of the temperature and mode are conditionally independent given the current joint pair of temperature and mode values under the effects of the randomized policy and exogenous disturbances, especially weather. That is, the transition matrix of the state process is a point wise product of two controlled transition matrices. In an optimal control setting, computation of the BA's control command, $\zeta(t)$, for reference tracking is a non-convex optimization problem (Coffman, Bušić, & Barooah, 2019). The probability of turning on when temperature exceeds the upper limit, or off when temperature dips below the lower limit, is set to 1 by design. This will ensure the temperature QoS constraint is maintained. Attempts have been made to maintain the cycling constraint (Coffman et al., 2018). But a formal design method to incorporate the cycling constraint has been lacking.

**Distributed:** In distributed coordination schemes, the coordination problem is typically cast as an optimization problem that is solved in a distributed fashion by using a iterative algorithm that can eventually converge to an optimal solution under certain conditions. Distributed optimization is not ideally suited for discrete decisions. Works on TCL coordination using a distributed scheme, (e.g., Burger & Moura, 2017; Kim & Giannakis, 2013 and Franceschelli, Pilloni, & Gasparri, 2021), handle this challenge through various heuristics, principally via relaxing {0, 1} to [0, 1] and then interpreting a solution in [0, 1] as a probability to make on/off decisions (Kim & Giannakis, 2013) or choose integer-valued set point change (Burger & Moura, 2017).

**Reference Computation:** As mentioned earlier, a complete framework for coordination of TCL needs not only a control algorithm to make decisions at TCLs, but also a method to compute a *feasible* reference signal for the collection's power demand. Feasible means that no TCL needs to violate local constraints in order for the collection to track the reference. Reference computation is related to the topic of "flexibility capacity", and the latter has been examined in many recent works with various definitions of flexibility (Coffman, Cammardella, Barooah, & Meyn, 2022; Coffman, Guo and Barooah, 2021; Hao, Sanandaji, Poolla, & Vincent, 2015; Paccagnan, Kamgarpour, & Lygeros, 2015). The work (Coffman et al., 2022) developed necessary conditions on the reference. Meaning, if the reference for the collection does not satisfy the developed conditions then there will be at least one TCL that will violate at least one local QoS constraint. However, they do not guarantee the existence of a coordination algorithm that can track a reference that satisfies those conditions.

#### *1.1.1. Contributions*

The previous discussion shows that existing work TCL coordination have a number of scattered disadvantages. Among decentralized coordination schemes, thermostat set-point based methods have implementation issues due to resolution of temperature sensors and measurement noise. Bin switching methods do not provide guarantees on reference tracking and often require solving a challenging state estimation problem. Prior works on randomized control require non-convex optimization and are based on an assumed conditional independence. Distributed optimization based coordination schemes are not ideal for integer-valued optimization. When it comes to reference planning, existing works do not guarantee existence of a coordination scheme – whether centralized, decentralized, or distributed – that can track that reference. In short, a unified framework that treats reference computation and coordination algorithm design simultaneously is lacking.

In this work we develop a unified framework for *decentralized* coordination of TCLs that *performs both reference computation and coordination algorithm design* simultaneously. Our major contributions are as follows.

(1) We provide a framework that allows the BA to compute (a) an optimal reference signal that is feasible for the collection and (b) optimal randomized policies for the TCLs to track the said reference. When the TCLs implement these policies, their aggregate power demand collectively tracks the reference, while the policies guarantee that temperature and cycling QoS requirements at each TCL are satisfied. Optimal reference means it is closest to what the BA wants while being feasible for the TCLs. Implementation of the policy at a TCL is easy; it requires only local measurements and a random number generator but does not require solving an optimization problem. The communication burden for coordination is also low. At each sampling time, a randomized control policy – parameterized by a few numbers – is broadcast to all TCLs. Feedback from TCLs to the BA can be infrequent.

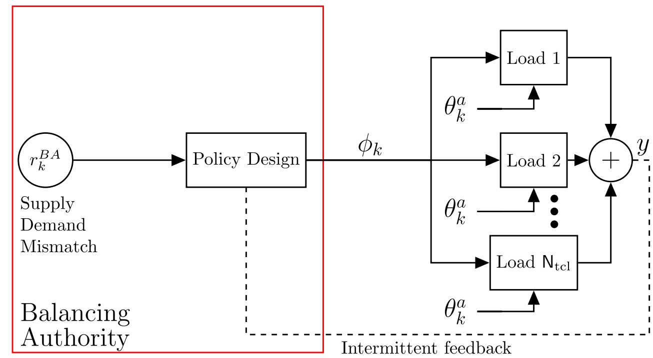


**Fig. 1.** Information architecture of the proposed framework.

(2) Our framework is based on a careful discretization of the partial differential equation (PDE) model described in Malhame and Chong (1985) which leads to a Markov chain. We show that a certain "conditional independence" that was assumed in Bušić and Meyn (2016) indeed holds. This independence separates the effects of the policy at the TCL (control) and weather (disturbance) on the transition matrix, and greatly facilitates computation of policies.

(3) The unified framework is made possible by the low dimensional parameterization of randomized policies with entries of certain condition probability mass functions. Thus the aggregate reference and policies at each TCLs are posed as decision variables in an optimization problem whose solution yields both the components of the unified framework: reference power demand of the collection and a coordination algorithm.

(4) Numerical experiments are provided to support the claims made in (i). Matlab implementation is made publicly available at Coffman (2021).

Fig. 1 illustrates the two parts of the proposed framework.

The rest of the paper proceeds as follows. In Section 2 the model of the individual TCL is introduced. In Section 3 the PDEs introduced are discretized and in Section 3.3 the structure of the discretized model is identified. Since the PDE discretization was previously reported in Coffman, Bušić, and Barooah (2021a), technical details including some of the proofs are moved to the expanded version (Coffman, Bušić, & Barooah, 2021b); only the parts necessary for completeness are described in this section. The proposed framework for reference and policy design is presented in Section 4 and numerical experiments are reported in Section 5.

### *1.2. Notation*

The symbol $\mathbb{1}$ denotes the vector of all ones, $\mathbf{e}_i$ denotes the $i$th canonical basis vector, and $\mathbf{0}$ denotes the zero matrix or vector, all of appropriate dimension. For a vector $v$, $\mathrm{diag}(v)$ denotes the diagonal matrix with entries of $v$, i.e., $\mathrm{diag}(v)\mathbb{1} = v$. Further, $\otimes$ denotes matrix Kronecker product and $\mathbf{I}_A(\cdot)$ the indicator function of the set $A$.

## 2. TCLs and the control problem

During its operation, a TCL must adhere to certain operational requirements (QoS constraints). We consider two.

- The temperature constraint: the TCL's temperature must remain within a prespecified deadband, $[\Theta^{\min}, \Theta^{\max}]$.

- The cycling constraint: The TCL can only change from "on" to "off" or vice versa once every $\tau$ (discrete) time instants, where $\tau$ is a prespecified constant. The cycling constraint is to ensure the mechanical hardware is not damaged.

### 2.1. Temperature dynamics of a TCL

The typical model for the TCL's temperature $\theta(t)$ in the literature is the following ordinary differential equation (ODE), where $\theta^a(t)$ is the ambient temperature:

$$\frac{d}{dt}\theta(t) = f_m(\theta(t), t), \quad \text{with}$$
$$f_m(\theta, t) = -\frac{1}{RC}\left(\theta(t) - \theta^a(t)\right) - m(t)\frac{\eta p^{\text{rated}}}{C}. \tag{1}$$

The rated electrical power consumption of a TCL is denoted $p^{\text{rated}}$ with coefficient of performance (COP) $\eta$. The parameters $R$ and $C$ denote thermal resistance and capacitance, respectively. The quantity $m(t) \in \{0, 1\}$ is the on/off *mode*, also called *mode state*: 1 means "on" and 0 means "off". The thermostat setpoint is denoted by $\Theta^{\text{set}}$, and $\Theta^{\text{set}} \in [\Theta^{\min}, \Theta^{\max}]$.

A model for the temperature state that accounts for modeling errors in (1) – and which will be crucial for the development in the next section – is the following Itô stochastic differential equation,

$$d\theta(t) = f_m(\theta, t)dt + \sigma dB(t). \tag{2}$$

The term $B(t)$ is Brownian motion with parameter $\sigma > 0$, and the quantity $\sigma dB(t)$ captures modeling errors in (1). In either model, the baseline power demand for the TCL, denoted by $p^{\text{BL,TCL}}$, is the value of the quantity $m(t)p^{\text{rated}}$ so that $f_1(\lambda^{\text{set}}, t) = 0$, solving which yields:

$$p^{\text{BL,TCL}}(t) = \frac{\theta^a(t) - \Theta^{\text{set}}}{\eta R}. \tag{3}$$

The mode state of a TCL evolves according to a policy. The following policy, which we denote as the *thermostat policy*, ensures the temperature constraint:

$$\lim_{\epsilon \to 0} m(t+\epsilon) = \begin{cases} 1, & \theta(t) \geq \Theta^{\max}. \\ 0, & \theta(t) \leq \Theta^{\min}. \\ m(t), & \text{o.w.} \end{cases} \tag{4}$$

We assume the following about the individual TCL discussed so far.

- **A.1** The thermostat policy does not violate the cycling constraint.
- **A.2** For all $t \geq 0$ and $\theta \in [\Theta^{\min}, \Theta^{\max}]$, $f_{\text{on}}(\theta, t) \leq 0$ and $f_{\text{off}}(\theta, t) \geq 0$.
- **A.3** The TCL's cycling and temperature constraint are both simultaneously feasible.

The sizing/design of the TCL is most likely to satisfy **A.1** and **A.3**. Assumption **A.2** states that when the TCL is on, the temperature does not increase and when the TCL is off the temperature does not decrease. All prior works focusing on cooling TCLs (e.g., air conditioners) implicitly make this assumption. Every result that is to follow is also valid for heating TCLs (e.g., a water heater or a heat pump) with a sign reversal. Assumption **A.3** is also implicit in any work that considers both the TCLs temperature and cycling constraint.

### 2.2. Problem statement

In the sequel the continuous time $t$ will be uniformly sampled with discretization interval $\Delta$. Let $k$ be the corresponding discrete time index. The total electrical power demand of the collection at time $k$, whether with thermostat policy or some other policy, is denoted by $y_k$:

$$y_k \triangleq p^{\text{rated}} \sum_{\ell=1}^{N_{\text{tcl}}} m_k^\ell \tag{5}$$

where $m_k^\ell$ is the on/off mode state of the $\ell$th TCL at $k$. The goal of coordinating TCLs is to help the BA balance supply and demand of electricity in the grid. We denote $r_k^{\text{BA}}$ as the desired demand from all flexible loads and batteries that will reduce the imbalance to 0. It is unreasonable to expect any collection of TCLs to meet the entire desired demand $r_k^{\text{BA}}$ while maintaining their QoS. Only a portion of $r_k^{\text{BA}}$ can be supplied by TCLs, and we denote this portion by $r_k$; Fig. 1 shows a discrete-time version of these quantities. The problem under study is to compute $r_k$ – given $r_k^{\text{BA}}$ as problem data – and design a decentralized coordination algorithm so that the aggregate power demand of the collection, the output $y_k$ in Fig. 1 — tracks the reference $r_k$ without any TCL having to violate its temperature and cycling QoS constraints.

### 2.3. PDE model from Malhame and Chong (1985)

We now describe a PDE model of a single TCL's temperature density with thermostat policy, with $\Theta$ for the temperature axis and $t$ for the time axis. This PDF was originally derived in Malhame and Chong (1985). Consider the following marginal pdfs $\mu_{\text{on}}, \mu_{\text{off}}$:

$$\mu_{\text{on}}(\Theta, t)d\Theta = \mathsf{P}\left((\Theta < \theta(t) \leq \Theta + d\Theta),\ m(t) = \text{on}\right), \tag{6}$$
$$\mu_{\text{off}}(\Theta, t)d\Theta = \mathsf{P}\left((\Theta < \theta(t) \leq \Theta + d\Theta),\ m(t) = \text{off}\right), \tag{7}$$

where $\mathsf{P}(\cdot)$ denotes probability, $\theta(t)$ evolves according to (2) and $m(t)$ evolves according to (4). It was shown in Malhame and Chong (1985) that the densities $\mu_{\text{on}}$ and $\mu_{\text{off}}$ satisfy the Fokker–Planck equations,

$$\frac{\partial}{\partial t}\mu_{\text{on}}(\Theta, t) = \frac{\sigma^2}{2}\nabla_\Theta^2 \mu_{\text{on}}(\lambda, t) - \nabla_\lambda\Big(f_{\text{on}}(\Theta, t)\mu_{\text{on}}(\Theta, t)\Big) \tag{8}$$
$$\frac{\partial}{\partial t}\mu_{\text{off}}(\Theta, t) = \frac{\sigma^2}{2}\nabla_\Theta^2 \mu_{\text{off}}(\Theta, t) - \nabla_\lambda\big(f_{\text{off}}(\Theta, t)\mu_{\text{off}}(\Theta, t)\big) \tag{9}$$

that are coupled through their boundary conditions. The boundary conditions are listed in Coffman et al. (2021b).

## 3. Markov model from PDE and and generalization to non-thermostat policies

We use the finite volume method (FVM) to discretize the PDEs (8) and (9) that yield a finite dimensional probabilistic model – a Markov chain – for a single TCL (Eq. (13)).

### 3.1. Spatial (temperature) discretization

The FVM bins the continuous temperature into $2N_{\text{bin}}$ control volumes (CV), half for the on modes and half for the off modes. The layout of the CVs is shown in Fig. 2. The CVs are defined through the nodal temperature values, $\Theta^i_{\text{on}}$ and $\Theta^i_{\text{off}}$, $i = 1, \ldots, N_{\text{bin}}$ and their left and right boundaries, with $\Delta\Theta$ is being the CV width.

The steps taken to obtain the spatially discretized PDEs is detailed in Coffman et al. (2021b). We describe here the end result of the derivation. First, define the following quantities

$$\nu_{\text{off}}(\Theta^i, t) \triangleq \mu_{\text{off}}(\Theta^i, t)\Delta\Theta,$$
$$\nu_{\text{on}}(\Theta^i, t) \triangleq \mu_{\text{on}}(\Theta^i, t)\Delta\Theta, \tag{10}$$

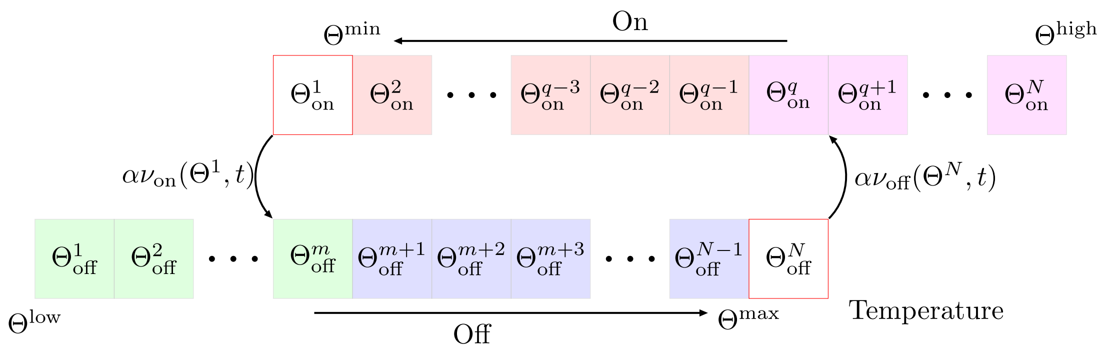


**Fig. 2.** The control volumes (CVs). The colors correspond to the colors found in Fig. 3. The values in each CV represent the nodal temperature for the CV. The arrows describe the sign of the convection of the TCL through the CVs. The values are such that $\mathsf{N}_{\text{bin}} = m + q$. The terms involving $\alpha$ model rate of transfer between the corresponding CVs due to the thermostat policy, where $\alpha = \gamma + \frac{\sigma^2}{(\Delta\lambda)^2}$. The parameter $\gamma > 0$ is a design parameter; see Remark 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

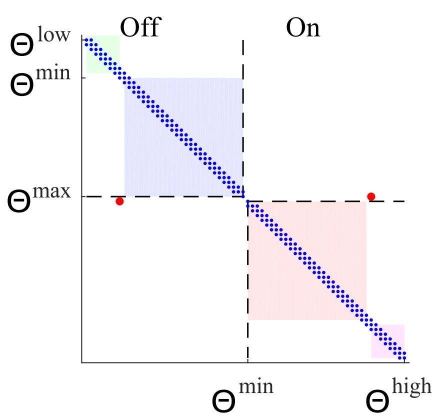


**Fig. 3.** Sparsity pattern of the matrix $A(t)$ for $\mathsf{N}_{\text{bin}} = 51$ CVs for both the on and off states. The colors correspond to the colors found in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for $i = 1, \ldots, \mathsf{N}_{\text{bin}}$, and the row vector $\nu(t) \triangleq [\nu_{\text{off}}(t), \nu_{\text{on}}(t)]$ with

$$\begin{aligned}\nu_{\text{off}}(t) &\triangleq [\nu_{\text{off}}(\Theta^1, t), \ldots, \nu_{\text{off}}(\Theta^{\mathsf{N}_{\text{bin}}}, t)],\\ \nu_{\text{on}}(t) &\triangleq [\nu_{\text{on}}(\Theta^1, t), \ldots, \nu_{\text{on}}(\Theta^{\mathsf{N}_{\text{bin}}}, t)].\end{aligned} \tag{11}$$

This spatial discretization leads to $2\mathsf{N}_{\text{bin}}$ coupled ordinary differential equations (ODEs), one for each of the $\nu_{\text{off}}(\Theta^1, t)$, $\nu_{\text{on}}(\Theta^i, t)$'s, for $i = 1, \ldots, \mathsf{N}_{\text{bin}}$. By combining these ODEs we obtain the linear time varying system

$$\frac{d}{dt}\nu(t) = \nu(t)A(t). \tag{12}$$

The sparsity pattern of $A(t)$ is shown in Fig. 3. The system (12) is the spatially discretized version of the PDEs (8)–(9). The matrix $A(t)$ also satisfies the properties of a transition rate matrix, described in the following lemma.

**Lemma 1.** *For all $t$, the matrix $A(t)$ is a transition rate matrix. That is, for all $t$*

*(i):* $A(t)\mathbb{1} = \mathbf{0}$.

*(ii):* *for all $i$, $A_{i,i}(t) \le 0$, and for all $j \neq i$ $A_{i,j}(t) \ge 0$.*

We omit the proof here since it is a standard result that has been arrived by several others independently; see Benenati, Colombino, and Dall'Anese (2019) and Paccagnan et al. (2015). The interested reader can refer to Coffman et al. (2021b) for a proof.

**Remark 1.** The choice of the FVM and how we discretize the convection and diffusion terms appearing in (8)–(9) is important for $A(t)$ to satisfy the conditions in Lemma 1. This issue is well known in the CFD literature (Versteeg & Malalasekera, 2007), and also recognized in the related work (Benenati et al., 2019). If a finite difference method had been used with central differences for both diffusion and convection terms, the resulting $A(t)$ would require restrictive conditions on both $\sigma^2$ and $\Delta\Theta$ to satisfy the properties in Lemma 1 (Versteeg & Malalasekera, 2007).

### 3.2. Temporal discretization

To temporally integrate the dynamics (12) we use a first order Euler approximation with time step $\Delta t > 0$. Making the identifications $\nu_k \triangleq \nu(t_k)$ and $A_k \triangleq A(t_k)$ we have

$$\nu_{k+1} = \nu_k P_k, \quad \text{with} \quad P_k = I + \Delta t A_k. \tag{13}$$

In the continuous time setting elements of the vector $\nu(t)$ were referred to as, for example, $\nu_{\text{on}}(\Theta^i, t)$. The counterpart to this, in the discrete time setting, is referring to elements of $\nu_k$ as, for example, $\nu_{\text{on}}[\Theta^i, k]$. We further have the following.

**Lemma 2** (Coffman et al., 2021a). *The matrix $P_k$ is a Markov transition probability matrix if*

$$\forall\ i,\ \text{and}\ \forall\ k, \quad 0 < \Delta t \le \left|[A_k]_{i,i}\right|^{-1}.$$

*where $[A_k]_{i,i}$ is the ith diagonal element of the matrix $A_k$.*

### 3.3. Markov model (with thermostat policy)

Recall that the dynamics (13) derived in the previous section was for the thermostat policy. We now delve into the structure of these dynamics so to introduce a BA control input. We first formalize a discrete state space for the dynamics (13). We will then show that the transition matrix $P_k$ in (13) can be written as a product of two matrices, one that depends solely on the thermostat policy and the other solely on weather. This conditional independence allows constructing a randomized TCL, one in which the deterministic TCL policy is replaced by a randomized policy. The BA can design this policy, which then becomes the BA's control input.

Now denote $\theta_k \triangleq \theta(t_k)$, $m_k \triangleq m(t_k)$, and

$$I_k \triangleq \sum_{i=1}^{\mathsf{N}_{\text{bin}}} i\mathbf{I}_{\text{CV}(i)}(\theta_k, m_k). \tag{14}$$

The quantity $I_k$ indicates which CV the TCL's temperature resides in at time $k$. It also is a function of $m_k$ since the CV index for the on mode is different from the index for the off mode. We then define the following discrete state space:

$$\mathsf{Z} \triangleq \{m \in \{\text{on}, \text{off}\},\ I \in \{1, \ldots, \mathsf{N}_{\text{bin}}\}\}, \tag{15}$$

with cardinality $|\mathsf{Z}| = 2\mathsf{N}_{\text{bin}}$. Using the newly defined quantity $I_k$ we rewrite the marginals $\nu_{\text{on}}[\Theta^i, k]$ and $\nu_{\text{off}}[\Theta^i, k]$ as functions on $\mathsf{Z}$,

$$\nu_{\text{on}}[\Theta^i, k] = \mathsf{P}\left(I_k = i,\ m_k = \text{on}\right), \quad \text{and} \tag{16}$$

$$\nu_{\text{off}}[\Theta^i, k] = \mathsf{P}\left(I_k = i,\ m_k = \text{off}\right). \tag{17}$$

From the above, the matrix $P_k$ (with the conditions of Lemma 2 satisfied) is the transition matrix for the joint process $(I_k, m_k)$ on the state space $\mathsf{Z}$. The dynamic equation $\nu_{k+1} = \nu_k P_k$ is then a probabilistic model for a TCL with state space $\mathsf{Z}$ and operating under the thermostat policy.

In the following, we refer to the values of $I_k$ with $i$ and $j$ and the values of $m_k$ with $u$ and $v$. We introduce the following notation to refer to the elements of the transition matrix $P_k$:

$$P_k((i, u), (j, v)) \triangleq \tag{18}$$
$$\mathsf{P}\Big(I_{k+1} = j,\ m_{k+1} = v \ \Big|\ I_k = i,\ m_k = u,\ \theta^a_k = w_k\Big).$$

We will now show that the matrix $P_k$ can be written as the product of two matrices. One depends only on the thermostat policy (control) and the other depends only on weather and TCL temperature dynamics. That is, we show that each entry of $P_k$ factors as

$$P_k((i, u), (j, v)) = \phi^{\text{TS}}_u(v \mid i) P^u_k(i, j) \tag{19}$$

where, for each given value of $\theta^a_k$, $P^u_k(i, j)$ is a *controlled transition matrix*:

$$P^u_k(i, j) \triangleq \mathsf{P}\left(I_{k+1} = j \mid I_k = i,\ m_k = u,\ \theta^a_k = w_k\right) \tag{20}$$

and $\phi^{\text{TS}}_u(v \mid i)$ is an instance of a *randomized policy* $\phi_u(v \mid i)$ on $\mathsf{Z}$:

$$\phi_u(v \mid i) \triangleq \mathsf{P}\left(m_{k+1} = v \mid I_k = i,\ m_k = u\right). \tag{21}$$

The quantity $\phi^{\text{TS}}_u(v \mid i)$ in (21) is the thermostat policy on $\mathsf{Z}$, which is formally defined as follows.

**Definition 1.** The thermostat policy on $\mathsf{Z}$ is specified by the two vectors, $\phi^{\text{TS}}_{\text{off}}, \phi^{\text{TS}}_{\text{on}} \in \mathbb{R}^{\mathsf{N}_{\text{bin}}}$, where $\phi^{\text{TS}}_{\text{off}} \triangleq \phi^{\text{TS}}_{\text{off}}(\text{on} \mid \cdot) = \mathbf{e}_{\mathsf{N}_{\text{bin}}}$, $\phi^{\text{TS}}_{\text{on}} \triangleq \phi^{\text{TS}}_{\text{on}}(\text{off} \mid \cdot) = \mathbf{e}_1$, and $\phi^{\text{TS}}_{\text{off}}(\text{off} \mid \cdot) \triangleq 1 - \phi^{\text{TS}}_{\text{off}}$, $\phi^{\text{TS}}_{\text{on}}(\text{on} \mid \cdot) \triangleq 1 - \phi^{\text{TS}}_{\text{on}}$.

The quantity $P^u_k(i, j)$ in (20) represents the open loop evolution of the TCL on $\mathsf{Z}$. That is, it describes how the TCLs temperature evolves under a fixed mode. We define matrices with entries $P^u_k(i, j)$ next.

**Definition 2.** Let $P^{\text{off}}_k, P^{\text{on}}_k \in \mathbb{R}^{\mathsf{N}_{\text{bin}} \times \mathsf{N}_{\text{bin}}}$ have $(i, j)$ entries given by,

$$P^{\text{off}}_k(i, j) = P_{w_k}((i, \text{off}), (j, \text{off})), \quad i \neq N \text{ and } j \neq \mathsf{N}_{\text{bin}},$$
$$P^{\text{on}}_k(i, j) = P_{w_k}((i, \text{on}), (j, \text{on})), \quad i \neq 1 \text{ and } j \neq 1,$$

with $P^{\text{off}}_k(\mathsf{N}_{\text{bin}}, \mathsf{N}_{\text{bin}}) = 1$ and $P^{\text{on}}_k(1, 1) = 1$.

The quantities defined in Definitions 1 and 2 correspond to entries of $P_k$. To construct the promised factorization, from these definitions, the idea is to construct its four sub-matrices that correspond to all possible combinations of $u, v \in \{\text{on}, \text{off}\}$ (see Fig. 3). For example, the $\text{off} - \text{off}$ quadrant of $P_k$ is given by the matrix product

$$\left(I - \text{diag}(\phi^{\text{TS}}_{\text{off}})\right) P^{\text{off}}_k.$$

However, since the temperature associated with the $i$th CV for the on mode is not the same temperature associated with the $i$th CV for the off state (see Fig. 2) it is *not* true that the $\text{off} - \text{on}$ quadrant of $P_k$ is given as $\text{diag}(\phi^{\text{TS}}_{\text{off}}) P^{\text{off}}_k$. The entries of the matrix $P^{\text{off}}_k$ need to be re-arranged so to correctly account for the difference in CV index between the on/off mode. We define such correctly re-arranged matrices next.

**Definition 3.** Let $I^{\text{off}} = \{m, \ldots, \mathsf{N}_{\text{bin}}\}$, $I^{\text{on}} = \{1, \ldots, q\}$, $m^- = m - 1$, and $S^{\text{off}}_k, S^{\text{on}}_k \in \mathbb{R}^{\mathsf{N}_{\text{bin}} \times \mathsf{N}_{\text{bin}}}$ with $(i, j)$ entries

$$S^{\text{off}}_k(i, j - m^-) = \begin{cases} P^{\text{off}}_k(i, j) & i, j \in I^{\text{off}} \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

$$S^{\text{on}}_k(i, j + m^-) = \begin{cases} P^{\text{on}}_k(i, j) & i, j \in I^{\text{on}} \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The above definition is based on the construction that $\mathsf{N}_{\text{bin}} = q + m$. The quantities in Definition 3 let us construct, e.g., the $\text{off} - \text{on}$ quadrant of $P_k$ as $\text{diag}(\phi^{\text{TS}}_{\text{off}}) S^{\text{off}}_k$.

The next result provides the promised factorization.

**Lemma 3.** *Let the time discretization period $\Delta t$ and the parameter $\alpha$ that appears as a design choice in discretizing the PDEs to ODEs be chosen to satisfy $\alpha = (\Delta t)^{-1}$. Let $\Phi^{TS}_{off} \triangleq diag(\phi^{TS}_{off})$ and $\Phi^{TS}_{on} \triangleq diag(\phi^{TS}_{on})$, where the vectors $\phi^{(\cdot)}_{(\cdot)}$'s are defined in Definition 1, and*

$$\Phi^{TS} \triangleq \begin{bmatrix} I - \Phi^{TS}_{off} & \Phi^{TS}_{off} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi^{TS}_{on} & I - \Phi^{TS}_{on} \end{bmatrix}, \tag{24}$$

$$G_k \triangleq \begin{bmatrix} \mathbf{0} & S^{off}_k & \mathbf{0} & P^{on}_k \\ P^{off}_k & \mathbf{0} & S^{on}_k & \mathbf{0} \end{bmatrix}^T. \tag{25}$$

*Then*

$$P_k = \Phi^{TS} G_k. \tag{26}$$

**Proof.** See Appendix in Coffman et al. (2021b). □

**Remark 2.** The condition $\alpha = 1/\Delta t$ can be satisfied as long as time and temperature discretization intervals $\Delta t, \Delta\Theta$ are chosen to satisfy $\Delta t < (\Delta\Theta)^2/\sigma^2$. To understand how, recall that in the discretizing the PDE to the coupled ODEs, a design parameter $\gamma$ appears that must satisfy $\gamma > 0$, and that $\alpha \triangleq D + \gamma$ where $D = \frac{\sigma^2}{(\Delta\Theta)^2}$. Thus, as long as $1/\Delta t > D$, which is equivalent to $\Delta t < (\Delta\Theta)^2/\sigma^2$, a positive $\gamma$ can be chosen while meeting the condition $\alpha = 1/\Delta t$. As a result, if a fine temperature resolution is chosen, the temporal resolution must also be fine enough.

**Remark 3.** The conditional independence factorization (26) has been a useful assumption in the design of algorithms in Bušić and Meyn (2016). In the present it is a byproduct of our spatial and temporal discretization of the PDEs (8)–(9). There are other works (Bashash & Fathy, 2013; Benenati et al., 2019; Paccagnan et al., 2015) that develop Markov models for TCLs through discretization of PDEs. However, to our knowledge, our work is the first to uncover this factorization.

### *3.4. Expanded state space for lock-out*

We now augment the Markov model of an individual TCL's temperature dynamics to include cycling dynamics, following Liu and Shi (2016) and Totu et al. (2017). Recall the cycling constraint that the policy needs to enforce: as soon as a TCL switches its mode, the TCL becomes stuck in that mode for $\tau$ time instances. This constraint can be represented as the evolution of a counter variable. Define a binary variable $s_k$ as $s_k = 1$ if the TCL is stuck in the current mode at time $k$ and 0 if it is not stuck. The counter variable is defined as follows

$$\mathsf{L}_{k+1} \triangleq \begin{cases} \mathsf{L}_k + 1, & s_k = 1. \\ 0, & s_k = 0. \end{cases} \tag{27}$$

This variable counts the time spent in the "stuck" mode. If $\mathsf{L}_k > 0$, it is stuck in either the on or off mode, and switching the mode by the policy will violate the cycling constraint.

We now expand the state space to consist not only on/off state and temperature but also the counter, and denote this newly expanded state space $\mathsf{X}$ as

$$\mathsf{X} \triangleq \Big\{ m \in \{0,1\}, I \in \{1,\dots,\mathsf{N}_{\text{bin}}\}, \mathsf{L} \in \{0,\dots,\tau\} \Big\}, \tag{28}$$

with cardinality $|\mathsf{X}| = 2\mathsf{N}_{\text{bin}}(\tau+1)$. The corresponding marginal pmf and transition matrix will be presented after we introduce generalized policies that go beyond thermostat logic.

### 3.5. Grid support policy (= BA's control command)

In light of Lemma 3 and the discussion preceding it, an arbitrary randomized policy can replace the thermostat policy. From the viewpoint of the BA this randomized policy *is* the control input that it must design and broadcast to a TCL. The TCL now implements this policy to make on/off decisions instead of using the thermostat policy. As we shall soon see, if the BA appropriately designs and sends the randomized policy to multiple TCLs it can achieve coordination of the TCLs for grid support.

To distinguish from thermostat policies in the prior section, we denote the newly introduced policies with the superscript 'GQ' to emphasize that these are policies introduced for providing *Grid support with QoS preservation*. We require these randomized policies to have the following structure

$$\phi^{\text{GQ}}_{\text{off}}(\text{on} \mid j, l) = \begin{cases} \kappa^{\text{on}}_j, & (m+1) \le j \le (\mathsf{N}_{\text{bin}} - 1), l = 0. \\ 1, & j = \mathsf{N}_{\text{bin}}, l = 0. \\ 0, & \text{o.w.} \end{cases} \tag{29}$$

$$\phi^{\text{GQ}}_{\text{on}}(\text{off} \mid j, l) = \begin{cases} \kappa^{\text{off}}_j, & 2 \le j \le (q-1), l = 0. \\ 1, & j = 1, l = 0. \\ 0, & \text{o.w.} \end{cases} \tag{30}$$

with $\phi^{\text{GQ}}_{\text{off}}(\text{off} \mid \cdot) = 1 - \phi^{\text{GQ}}_{\text{off}}(\text{on} \mid \cdot)$ and $\phi^{\text{GQ}}_{\text{on}}(\text{on} \mid \cdot) = 1 - \phi^{\text{GQ}}_{\text{on}}(\text{off} \mid \cdot)$ and $\kappa^{\text{on}}_j, \kappa^{\text{off}}_j \in [0,1]$ for all $j$. The policies could also be time varying, for example: $\kappa^{\text{off}}_j[k]$ and $\kappa^{\text{on}}_j[k]$. The dependence of the policies on time is denoted as $\phi^{\text{GQ}}_{\text{off}}[k]$ and $\phi^{\text{GQ}}_{\text{on}}[k]$. In the above, we have required $\phi^{\text{GQ}}_{\text{off}}(\text{on} \mid j) = 0$ for $1 \le j \le m$ since the temperatures corresponding to these indices are below the permitted deadband temperature, $\Theta^{\min}$. Hence, turning on at these temperature does not make physical sense. The arguments for the zero elements in $\phi^{\text{GQ}}_{\text{on}}$ are symmetric. This construction ensures a TCL will not violate its temperature and cycling constraints under Assumptions **A.2** and **A.3**.

*Designing such a policy is equivalent to choosing the values of* $\kappa^{on}_j[k]$ *and* $\kappa^{off}_j[k]$ *for all j and k.*

**Remark 4** (*Implementation at a TCL*)**.** For an individual TCL, implementing a randomized policy $\phi^{\text{GQ}}_{\text{on}}, \phi^{\text{GQ}}_{\text{off}}$ is straightforward: (i) the TCL measures its current temperature and on/off status, (ii) the TCL "bins" this temperature value according to (14) and (iii) the TCL flips a coin to decide its next on/off state according to the probabilities given in (29)–(30). Note that the thermostat policy is a special case of the grid support policy, and both policies enforce the temperature constraint. Only the grid support policy enforces the cycling constraint explicitly.

### 3.6. The Markov model (with grid support policy)

The Markovian model of a TCL with grid support policy on the expanded state space $\mathsf{X}$ defined in (28) is

$$\nu^{\text{GQ}}_{k+1} = \nu^{\text{GQ}}_k \Phi^{\text{GQ}}_k G^{\text{GQ}}_k. \tag{31}$$

where the marginal pmf $\nu^{\text{GQ}}$ and the matrices $\Phi^{\text{GQ}}, G^{\text{GQ}}$ are defined below.

Each entry of the grid-support policy is denoted as $\phi^{\text{GQ}}_{\text{off}}(u \mid j,\ l)$ and $\phi^{\text{GQ}}_{\text{on}}(u \mid j,\ l)$. The corresponding marginals are $\nu^{\text{GQ}}_{\text{off}}[\Theta^j, l, k]$ and $\nu^{\text{GQ}}_{\text{on}}[\Theta^j, l, k]$. We use $\nu^{\text{GQ}}_{\text{off},l}$ (resp., $\nu^{\text{GQ}}_{\text{on},l}$) as shorthand for $\nu^{\text{GQ}}_{\text{off}}[\cdot, l, k]$ (resp., $\nu^{\text{GQ}}_{\text{on}}[\cdot, l, k]$). In vectorized form, the marginal is $\nu^{\text{GQ}} = [\nu^{\text{GQ}}_{\text{off}}, \nu^{\text{GQ}}_{\text{on}}]$ where $\nu^{\text{GQ}}_{\text{off}} = [\nu^{\text{GQ}}_{\text{off},0}, \dots, \nu^{\text{GQ}}_{\text{off},\tau}]$ and $\nu^{\text{GQ}}_{\text{on}} = [\nu^{\text{GQ}}_{\text{on},0}, \dots, \nu^{\text{GQ}}_{\text{on},\tau}]$. Define

$$G^{\text{GQ}}_k \triangleq \begin{bmatrix} \mathbf{0} & D_\tau \otimes S^{\text{on}}_k & \mathbf{0} & C_\tau \otimes P^{\text{off}}_k \\ C_\tau \otimes P^{\text{on}}_k & \mathbf{0} & D_\tau \otimes S^{\text{off}}_k & \mathbf{0} \end{bmatrix}^T, \tag{32}$$

where $D_\tau \triangleq \mathbb{1}^T \otimes \mathbf{e}_2 \in \mathbb{R}^{\tau+1 \times \tau+1}$ and

$$C_\tau \triangleq \begin{bmatrix} 1 & 0 & \mathbf{0}^T_{\tau-1} \\ \mathbf{0}_{\tau-1} & \mathbf{0}_{\tau-1} & I_{\tau-1} \\ 1 & 0 & \mathbf{0}^T_{\tau-1} \end{bmatrix} \in \mathbb{R}^{(\tau+1)\times(\tau+1)}. \tag{33}$$

The matrix $\Phi^{\text{GQ}}_k$ has the same structure as $\Phi^{\text{TS}}$, but with policy $\phi^{\text{GQ}}_{\text{off}}, \phi^{\text{GQ}}_{\text{on}}$, i.e.,

$$\Phi^{\text{GQ}}_k \triangleq \begin{bmatrix} I - \Phi^{\text{GQ}}_{\text{off}}[k] & \Phi^{\text{GQ}}_{\text{off}}[k] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi^{\text{GQ}}_{\text{on}}[k] & I - \Phi^{\text{GQ}}_{\text{on}}[k] \end{bmatrix}, \tag{34}$$

where

$$\Phi^{\text{GQ}}_{\text{off}}[k] \triangleq \text{diag}(\phi^{\text{GQ}}_{\text{off}}[k]),\ \Phi^{\text{GQ}}_{\text{on}}[k] \triangleq \text{diag}(\phi^{\text{GQ}}_{\text{on}}[k]) \tag{35}$$

The structure of the transition matrix $\Phi^{\text{GQ}}_k G^{\text{GQ}}_k$ is shown in Fig. 4. For comparison, the transition matrix with policy $\phi^{\text{GQ}}$ and *without* the cycle counter variable would simply be the four red shaded blocks appearing in their respective quadrant. In the expanded system, an on to off mode switch forces probability mass from the red shaded region ($l = 0$ and $m = \text{on}$) to the green shaded region ($l = 1$ and $m = \text{off}$). Mass must then transition through the chain of $\tau$ green blocks until it reaches the red block again, so to respect the cycling constraint.

In the sequel, we will use "policy" to mean either the pair $(\phi^{\text{on}}_k, \phi^{\text{off}}_k)$ or the matrix $\Phi^{\text{GQ}}_k$.

## 4. Proposed framework

We are now ready to present a framework to solve the problem stated in Section 2.2. We first transition from the viewpoint of a single TCL to that of a collection of $\mathsf{N}_{\text{tcl}}$ TCLs: $\ell = 1, \dots, \mathsf{N}_{\text{tcl}}$. For example, $m^\ell_k$ and $I^\ell_k$ are the mode and binned temperature of the $\ell^{th}$ TCL at time $k$. Recall (5): the total power consumption $y_k$ of the collection of $\mathsf{N}_{\text{tcl}}$ TCLs is $y_k = p^{\text{rated}} \sum_{\ell=1}^{\mathsf{N}_{\text{tcl}}} m^\ell_k$. In addition, the aggregate baseline demand of a $\mathsf{N}_{\text{tcl}}$ TCLs — denoted by $p^{\text{BL,Agg}}$ – and the maximum power — denoted by $p^{\text{max,Agg}}$ are

$$p^{\text{BL,Agg}}(t) \triangleq \mathsf{N}_{\text{tcl}}\, p^{\text{BL,TCL}}(t),\ \ p^{\text{max,Agg}} \triangleq \mathsf{N}_{\text{tcl}}\, p^{\text{rated}}. \tag{36}$$

### 4.1. Aggregate model of a collection of TCLs

So far the PDEs and their discrete version (the Markov model) have been described as a model of a single TCL. They can also model a collection of TCLs due to the Law of Large Numbers (LLN), which was in fact the motivation behind the PDE derivation in Malhame and Chong (1985). See also Chen, Bušić and Meyn

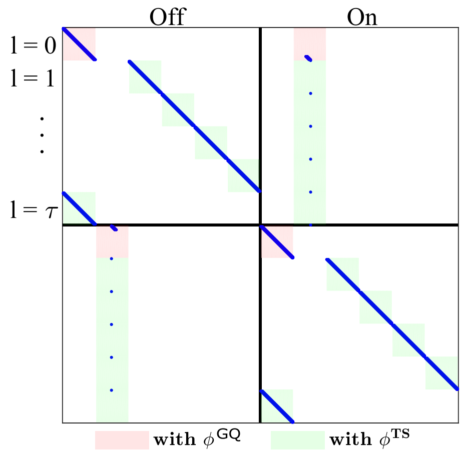


**Fig. 4.** The sparsity pattern of the expanded transition matrix (the dots represent non-zero entries in the matrix) with $\tau = 5$. Each shaded block is over the entire range of temperature values.

(2017) and Meyn et al. (2015) for a more formal treatment of the basis in LLN of Markovian models of load collections. The marginal pmf in our Markov model thus also describes sample average over a collection of TCLs. Since the state $\nu_k^{\mathrm{E}}$ is a marginal pmf for a single TCL, for a collection of TCLs we expect this pmf to approximate the histogram $h_k[\cdot]$ as $\mathsf{N}_{\mathrm{tcl}} \to \infty$, where:

$$h_k[u, i, l] \triangleq \frac{1}{\mathsf{N}_{\mathrm{tcl}}} \sum_{\ell=1}^{\mathsf{N}_{\mathrm{tcl}}} \Big( \mathbf{I}_{\{i\}}(I_k^\ell) \mathbf{I}_{\{u\}}(m_k^\ell) \mathbf{I}_{\{l\}}(\mathsf{L}_k^\ell) \Big), \tag{37}$$

for each state $(u, i, l) \in \mathsf{X}$. In the same regard, we define

$$\gamma_k^{\mathrm{GQ}} \triangleq \nu_k^{\mathrm{GQ}} C^{\mathrm{GQ}}, \text{ with } C^{\mathrm{GQ}} \triangleq [\mathbf{0}^T, p^{\max,\mathrm{Agg}} \mathbb{1}^T]^T, \tag{38}$$

where $p^{\max,\mathrm{Agg}}$ is the maximum possible power of the collection, defined in (36). The *control oriented aggregate model of a TCL collection* is the dynamics (31) together with the output (38):

$$\nu_{k+1}^{\mathrm{GQ}} = \nu_k^{\mathrm{GQ}} \Phi_k^{\mathrm{GQ}} G_k^{\mathrm{GQ}}. \text{ and } \gamma_k^{\mathrm{GQ}} = \nu_k^{\mathrm{GQ}} C^{\mathrm{GQ}}. \tag{39}$$

Due to the LLN, we expect $\gamma_k^{\mathrm{GQ}} \approx y_k$ for large $\mathsf{N}_{\mathrm{tcl}}$.

Effectiveness of (39) in modeling a population of TCLs can be seen in our prior work (Coffman et al., 2021a) and the extended version (Coffman et al., 2021b).

### 4.2. Co-design of policy and demand reference

Recall the problem statement from Section 2.2: the BA needs to determine a reference signal $r_k$ - given $r_k^{\mathrm{BA}}$ as problem data - and a policy for the TCLs so that with that policy the TCLs can collectively track $r_k$ without any TCL having to violate its temperature and cycling QoS constraints. Determining $r_k$ becomes an optimal control problem due to the time coupling produced by TCLs' dynamics. We consider a planning horizon of $T_{\mathrm{plan}}$, and define $\mathsf{T} \triangleq \{\mathsf{T}(0), \ldots, \mathsf{T}(0) + T_{\mathrm{plan}} - 1\}$, with $\mathsf{T}(0)$ denoting the initial time index.

In the proposed framework, the BA solves the following optimization problem to simultaneously design grid support policies and a feasible reference signal $r_k$ for $k \in \mathsf{T}$:

$$\min_{\nu_k^{\mathrm{GQ}}, \Phi_k^{\mathrm{GQ}}} \sum_{k \in \mathsf{T}} \Big( r_k^{\mathrm{BA}} - \gamma_k^{\mathrm{GQ}} \Big)^2 \tag{40}$$

$$\text{s.t. } \nu_{k+1}^{\mathrm{GQ}} = \nu_k^{\mathrm{GQ}} \Phi_k^{\mathrm{GQ}} G_k^{\mathrm{E}}, \gamma_k^{\mathrm{GQ}} = \nu_k^{\mathrm{GQ}} C^{\mathrm{E}}, \tag{41}$$

$$\nu_k^{\mathrm{GQ}} \in [0, 1], \quad \Phi_k^{\mathrm{GQ}} \in \Phi. \tag{42}$$

where $\nu_k^{\mathrm{GQ}} \in [0, 1]$ holds elementwise, the initial condition $\nu_{\mathsf{T}(0)}^{\mathrm{GQ}}$ is given as problem data, and $\Phi$ is the following convex set:

$$\begin{aligned} \Phi \triangleq \Big\{ & \Phi \in \mathbb{R}_{[0,1]}^{|\mathsf{X}| \times 2|\mathsf{X}|} \;\Big|\; \mathbb{1} = \Phi \mathbb{1}, \\ & \Phi \text{ satisfies (34)–(35) for some } \phi_{\mathrm{on}}^{\mathrm{GQ}}, \phi_{\mathrm{off}}^{\mathrm{GQ}}, \\ & \text{which in turn satisfy (29)–(30)} \Big\}. \end{aligned} \tag{43}$$

where $\mathbb{R}_{[0,1]}^{|\mathsf{X}| \times |\mathsf{X}|}$ is the set of $|\mathsf{X}| \times |\mathsf{X}|$ matrices with entries in [0, 1].

A solution $\nu_k^{\mathrm{E},*}, \Phi_k^{\mathrm{E},*}$ to (40) yields, for $k \in \mathsf{T}$, two things: (i) an optimal reference for the power demand of the TCL collection, defined as $r_k \triangleq \gamma_k^{\mathrm{E},*} (= \nu_k^{\mathrm{GQ},*} C^{\mathrm{GQ}})$ and (ii) optimal randomized policies $\phi_{\mathrm{off}}^{\mathrm{GQ},*}[k], \phi_{\mathrm{on}}^{\mathrm{GQ},*}[k]$, obtained from the solution $\Phi_k^{\mathrm{GQ},*}$ as follows. Given a $\Phi \in \Phi$, one recovers vectors $\phi_{\mathrm{off}}^{\mathrm{GQ}}, \phi_{\mathrm{on}}^{\mathrm{GQ}}$ from (34)–(35). Due to (29)–(30) and the [0, 1] constraint in $\Phi$, the vectors $\phi_{\mathrm{off}}^{\mathrm{GQ}}, \phi_{\mathrm{on}}^{\mathrm{GQ}}$ are valid conditional pmfs. The reference is optimal in the following sense: among all power demand signals the collection can track without requiring any TCL to violate its local QoS constraints in so doing, it is the closest to the BA's desired demand $r^{\mathrm{BA}}$ in 2-norm. The reference is also the predicted power consumption of the TCLs whilst using the policies $\phi_{\mathrm{off}}^{\mathrm{GQ},*}[k]$ and $\phi_{\mathrm{on}}^{\mathrm{GQ},*}[k]$. Recall that the equality constraints (29)–(30) ensure the temperature and cycling constraints by placing zero probability on state transitions that would violate QoS.

**Remark 5.** Since the reference $r_k (= \gamma_k^{\mathrm{GQ},*})$ from (40) is the best the TCLs can do to help the BA without any TCL having to violate its QoS, Problem (40) therefore also provides an answer to the "aggregate flexibility" question; see the discussion in Section 1.1. Unlike earlier works, proposed framework not only characterizes the aggregate flexibility and an optimal reference within the flexibility set but also provides a coordination algorithm that can track that reference.

### 4.3. Convex policy design

The problem (40) is non-convex due to the product $\nu_k^{\mathrm{E}} \Phi_k^{\mathrm{E}}$ in the constraint. A well known convexification remedy for (40) is to optimize over the marginal and joint distributions instead of the marginal and the policy (Benenati et al., 2019; Manne, 1960). The joint distribution, in matrix form, is:

$$J_k = \mathrm{diag}(\nu_k^{\mathrm{GQ}}) \Phi_k^{\mathrm{GQ}} \in \mathbb{R}^{|\mathsf{X}| \times 2|\mathsf{X}|}. \tag{44}$$

By construction, we have that $\nu_{k+1}^{\mathrm{E}} = \mathbb{1}^T J_k G_k^{\mathrm{GQ}}$ and $(\nu_k^{\mathrm{GQ}})^T = J_k \mathbb{1}$ since $\mathbb{1}^T \mathrm{diag}(\nu_k^{\mathrm{GQ}}) = \nu_k^{\mathrm{GQ}}$ and $\mathbb{1} = \Phi_k^{\mathrm{GQ}} \mathbb{1}$. It is straightforward to convert the constraint set $\Phi_k^{\mathrm{GQ}} \in \Phi$ to the new decision variables. We denote the transcription of $\Phi_k^{\mathrm{GQ}} \in \Phi$ to the new variables as $(J_k, \nu_k^{\mathrm{GQ}}) \in \bar{\Phi}$. For instance, for the equality constraints in $\Phi$, if we have $\phi_{\mathrm{off}}^{\mathrm{GQ}}(u \mid j, l) = \kappa$ for some scalar $\kappa$, then in the decision variables $J_k$ and $\nu_k^{\mathrm{GQ}}$ we will have a linear constraint of the form

$$\begin{aligned} & \mathsf{P}\,(m_{k+1} = u,\ I_k = j,\ \mathsf{L}_k = l,\ m_k = \mathrm{off}) \\ & = \kappa\, \nu_{\mathrm{off}}[\Theta^j, l, k], \end{aligned} \tag{45}$$

where the LHS of the above is some element in the matrix $J_k$. In addition, both $J_k$ and $\nu_k^{\mathrm{GQ}}$ to be within [0, 1] entrywise. Optimizing over $J_k$ and $\nu_k^{\mathrm{GQ}}$ yields the convex program:

$$\begin{aligned} \eta^* = \min_{\nu_k^{\mathrm{GQ}}, J_k} \ & \eta(\hat{\nu}) = \sum_{k \in \mathsf{T}} \Big( r_k^{\mathrm{BA}} - \gamma_k^{\mathrm{GQ}} \Big)^2 \\ \text{s.t. } \ & \nu_{k+1}^{\mathrm{GQ}} = \mathbb{1}^T J_k G_k^{\mathrm{GQ}}, \quad \nu_{\mathsf{T}(0)}^{\mathrm{GQ}} = \hat{\nu}, \quad \gamma_k^{\mathrm{GQ}} = \nu_k^{\mathrm{GQ}} C^{\mathrm{GQ}}, \\ & \nu_k^{\mathrm{GQ}}, J_k \in [0, 1],\ (\nu_k^{\mathrm{GQ}})^T = J_k \mathbb{1},\ (J_k, \nu_k^{\mathrm{GQ}}) \in \bar{\Phi}. \end{aligned} \tag{46}$$

Once the convex problem is solved, the matrices $\Phi_k^{\mathrm{GQ}}, k \in \mathsf{T}$ need to be recovered from it by using the relation (44). If the matrix

$\mathrm{diag}(\nu_k^{\mathrm{GQ}})$ is invertible, then $\Phi_k^{\mathrm{E}}$ obtained trivially from inversion of $\mathrm{diag}(\nu_k^{\mathrm{GQ}})$. If it is not invertible, more care is required. The next Algorithm describes this reconstruction.

**Algorithm 1.** For $v \in \{\mathrm{off}, \mathrm{on}\}$, if $\nu_v[\Theta^j, l, k] > 0$ then set

$$\phi_v^{\mathrm{GQ}}(u \mid j, l) := \frac{\mathsf{P}\left(m_{k+1} = u,\ I_k = j,\ \mathsf{L}_k = l,\ m_k = v\right)}{\nu_v[\Theta^j, l, k]} \tag{47}$$

where the value of the joint distribution $\mathsf{P}(m_{k+1} = u,\ I_k = j,\ \mathsf{L}_k = l,\ m_k = v)$ is obtained from the corresponding entry of the matrix $J_k$. If $\nu_v[\Theta^j, l, k] = 0$, consider two scenarios depending on the value of $v$. If $v = \mathrm{off}$, choose $\phi_v^{\mathrm{GQ}}(\mathrm{on} \mid j, l)$ to satisfy the equality constraints (29) and then elect $\phi_v^{\mathrm{GQ}}(\mathrm{off} \mid j, l) = 1 - \phi_v^{\mathrm{GQ}}(\mathrm{on} \mid j, l)$. Second, if $v = \mathrm{on}$, choose $\phi_v^{\mathrm{GQ}}(\mathrm{off} \mid j, l)$ to satisfy the equality constraints (30) and then elect $\phi_v^{\mathrm{GQ}}(\mathrm{on} \mid j, l) = 1 - \phi_v^{\mathrm{GQ}}(\mathrm{off} \mid j, l)$. □

Note that there is some design flexibility in the case a marginal entry is 0, since many $\kappa_j^{(\cdot)}$'s can be chosen to meet the equality constraints (29). Any of these choices are feasible since due to the constraints (45), the only way the solution has a zero entry in the marginal $\nu$ is if the corresponding entry in the joint density $J$ is also 0.

**Lemma 4.** *Suppose for all $k \in \mathsf{T}$ that $\nu_k^{GQ}$ and $J_k$ satisfy the constraints in problem (46). Then, the quantity $\Phi_k^{GQ}$ constructed according to Algorithm 1 satisfies (44) and it is a valid randomized policy, i.e., $\Phi_k^{GQ} \in \Phi$.*

**Proof.** See Appendix A.1. □

The nonconvex problem and its convex relaxation have a certain equivalence described in the following Theorem.

**Lemma 5.** *Let $\eta^*_{NCVX}$ and $\eta^*_{CVX}$ be the optimal costs for problems (40) and (46). Then, $\eta^*_{CVX} = \eta^*_{NCVX}$.*

**Proof.** See Appendix A.2. □

This result, for a similar problem setup, is also reported in Benenati et al. (2019). While we have no guarantee on the difference of the argument minimizers (and hence the policies obtained from both), Lemma 5 says that the policies will produce the same tracking performance. Further, from Lemma 4, the policies produced from either problem are guaranteed to ensure TCLs' QoS.

The sparse nature of the matrix $J_k$ can be exploited to reduce computational burden significantly. For instance, in the numerical examples reported later in the paper, the problem (46) has $\approx$ 500,000 decision variables. We were able to reduce the number of variables to $\approx$ 75,000 by exploiting sparsity of $J_k$. The interested reader is referred to Coffman et al. (2021b) for details of this and other computational aspects.

Matlab implementation of (46) and the algorithm to extract the policies from $J_k$ is available at Coffman (2021).

#### *4.3.1. Communication burden*

Once solved, the policies obtained from (46) need to be sent to each individual TCL. Many of the policy state values are constrained to either zero or one, which could be pre-programmed into each TCL. At each time index, $q - 2$ (for the on to off policy) plus $\mathsf{N}_{\mathrm{bin}} - m - 1$ (for the off to on policy) numbers are not constrained and need to be sent from the BA to each TCL. Recall that the numbers $m$ and $q$ are temperature bin indices (see Fig. 2) and $\mathsf{N}_{\mathrm{bin}}$ is the number of temperature bins. For illustrative purposes, consider the values used in numerical experiments reported in the sequel: $\mathsf{N}_{\mathrm{bin}} = 12$ with $q = 10$ and $m = 2$ and a time discretization $\Delta t = 1$ min. Since $\mathsf{N}_{\mathrm{bin}} = q + m$, then the BA has to broadcast $2q - 3 = 17$ numbers every 1 min to the TCLs. Each TCL receives the same 18 numbers.

Communication from TCLs to the BA – about their temperature and on/off state – is needed at the beginning of every planning period so that the BA can determine the initial condition $\hat{\nu}$ in (46). The frequency of this feedback is a design choice. In our numerical simulations reported later, a planning horizon of 6 h was used, and this feedback was necessary only once in six hours. More frequent loop closure may be needed for higher robustness to uncertainty in weather prediction etc., a topic outside the scope of this paper.

**Remark 6.** The reference tracking objective in the policy design problem (40) can be changed depending on the application, but *the constraints of Problem (40) will have to be retained so that TCLs' QoS are not violated.* For instance, a load aggregator participating in an ancillary services market seeking to maximize revenue may simply change the objective without changing the constraints to compute a reference for its TCL collection and design the corresponding policies. Depending on the application, the constraints may have to be augmented with additional ones. In additional constraints are introduced, the convex reformulation may or may not enjoy the equivalence with the non-convex problem described in Lemma 5.

## 5. Numerical experiments

Simulation involving coordination of $\mathsf{N}_{\mathrm{tcl}} = 20{,}000$ TCLs through our proposed framework is presented here. Recall the two parts of the coordination architecture shown in Fig. 1: (i) planning and (ii) real time coordination. Planning refers to the solution of the problem (46) at the BA to compute the following two things for each $k$ in the planning period $\mathsf{T}$:

(i) $r_k$: the reference power consumption of the TCL collection, given the problem data $r_k^{\mathrm{BA}}$.
(ii) $\phi_{\mathrm{off}}^{\mathrm{GQ},*}[k]$ and $\phi_{\mathrm{on}}^{\mathrm{GQ},*}[k]$: grid support control policies for each TCL.

This computation is performed at $\mathsf{T}(0)$. Real time coordination is the implementation of the grid support policies at each TCL to make on/off decisions, which is done as explained in Remark 4. We imagine the BA broadcasts the policies $\phi_{\mathrm{off}}^{\mathrm{GQ},*}[k]$ and $\phi_{\mathrm{on}}^{\mathrm{GQ},*}[k]$ at each $k$, though it can also broadcast all the policies, for all $k \in \mathsf{T}$, at $\mathsf{T}(0)$ and not broadcast again until the beginning of the next planning horizon.

The goal of the numerical simulations of real time control is to show the following.

(i) When each TCL uses the computed policies $\phi_{\mathrm{off}}^{\mathrm{GQ},*}[k]$ and $\phi_{\mathrm{on}}^{\mathrm{GQ},*}[k]$ to decide on/off actuation, the collection's power demand indeed tracks $r_k$.
(ii) Every TCL's QoS constraints – both temperature and cycling – are satisfied at all times.

There is some "plant-model mismatch" in these simulations: temperature of each TCL is computed with the ODE model (1) even though the policy design and reference computation is based on the Markov model.

### *5.1. Planning*

The demand needed for demand–supply imbalance at the BA, $r_k^{\mathrm{BA}}$, is chosen arbitrarily, and shown in Fig. 5 (top). It is infeasible for the collection: sometimes negative and sometimes far higher than the maximum power demand of the collection. This is done to simulate a realistic scenario in which many sources of demand and generation, not just TCLs, are managed by the BA.

**Table 1**
Simulation parameters.

| Par. | Unit | Value | Par. | Unit | Value |
|---|---|---|---|---|---|
| $N_{tcl}$ | N/A | $2\times10^4$ | $\eta$ | $\frac{\text{kW-e}}{\text{kW-th.}}$ | 2.5 |
| $C$ | kWh/°C | 1 | $p^{rated}$ | kW | 5.5 |
| $\Theta^{min}$ | °C | 20 | $\Theta^{max}$ | °C | 22 |
| $(\Delta t)\tau$ | Min. | 5 | $P_{agg}$ | MW | 110 |
| $R$ | °C/kW | 2 | $\Delta t$ | Min. | 1 |
| $q$ | N/A | 10 | $m$ | N/A | 2 |
| $N_{bin}$ | N/A | 12 | $T_{plan}$ | N/A | 360 |

The baseline demand trajectory is defined by Eq. (36), which is approximately the power consumption for this collection of air conditioners under thermostat control. The ambient air temperature is time varying and is obtained from wunderground.com for a typical summer day in Gainesville, Florida, USA. The other parameters relevant to the simulations are shown in Table 1.

Planning computations are done with Matlab and CVX (Grant & Boyd, 2011) using a desktop Linux machine, with $N_{bin} = 12$, and for a six hour planning horizon with 1 min discretization ($T_{plan} = 360$). The problem (46) takes about a minute to solve. The quantity $r_k^{BA}$, the baseline power $p_k^{BL,Agg}$ defined in (36), and the reference signal $r_k$ obtained from solving (46), are shown in Fig. 5 (top). Fig. 5 (bottom) shows the two grid support control polices for an arbitrarily chosen time instant.

### 5.2. Real time coordination

The power consumption of the collection making on/off decisions according to the obtained policies is shown in Fig. 6 (top). The figure shows that the TCLs are able to collectively track the reference signal $r_k$. We emphasize that the computational effort at each TCL is negligible. Recall Remark 4: once a TCL receives a grid support policy (17 floating point numbers, see Section 4.3.1) it only has to measure its current state (temperature and on/off mode) and generate a uniformly distributed random number in [0, 1] to implement the policy.

Verification of the grid support policies in ensuring QoS is shown in Fig. 6. The bottom plots show a histogram of the times between switches for 300 randomly chosen TCLs. The middle plot shows a histogram of temperature from 200 randomly chosen TCLs' temperature trajectories. The histograms show that the policies designed with (46) indeed satisfy the QoS constraints, which is specified by the vertical lines in the figures. Some TCLs do escape the temperature deadband by a little bit, which is expected and occurs also in thermostatic control: the sensor must *first* register a value outside the deadband in order decide to switch the on/off state.

## 6. Conclusion

In this work we present a framework for the decentralized control of TCLs. The framework unifies: (i) reference planning for a collection of TCLs and (ii) design of randomized control policies for individual TCLs by posing them as the solution of a single optimization problem. The resulting framework is (i) scalable to an arbitrary number of loads and is implemented through *local* feedback and minimal communication, (ii) able to guarantee both temperature and cycling constraints maintenance in each TCL, and (iii) based on convex optimization. Matlab/cvx implementation is publicly available (Coffman, 2021). The exposition is in terms of a reference demand supplied by a grid authority, but the framework is flexible enough to aid other applications, such as a load aggregator bidding in a day-ahead market.

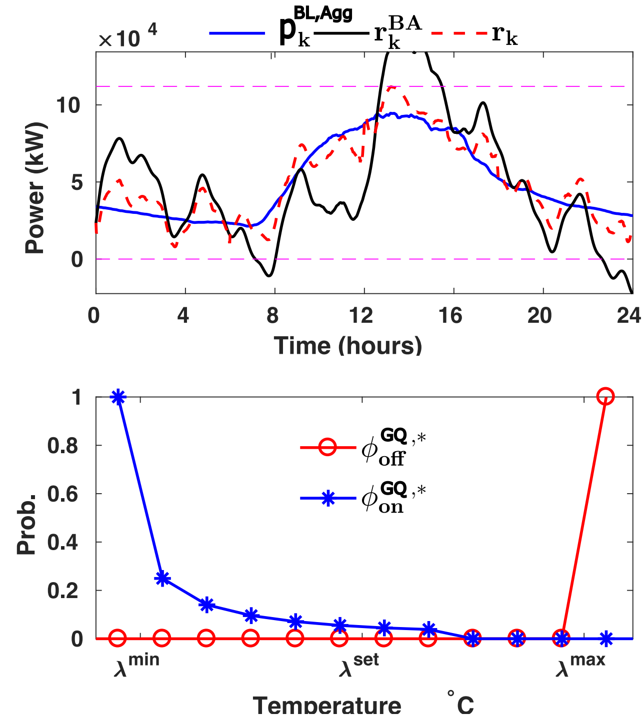


**Fig. 5.** (Top): The optimal reference $r_k$ obtained from solving (46), the dashed horizontal lines represent all of the TCLs on (top line) and off (bottom line). (Bottom): Grid support control policies, obtained from solving (46), at an arbitrary time instance.

There are several avenues for future work. The optimal control problem is solved in an open-loop fashion here. Feedback from TCLs is used only to compute an initial condition that is needed as problem data for the off-line planning problem. It is straightforward to close the loop between the TCL collection and the BA with greater frequency for robustness to uncertainty in weather forecast and TCL parameters. It will be of interest to identify scenarios where closing loop, say, by using Model Predictive Control, is (i) necessary, and (ii) at what frequency should information be communicated from the TCLs to the BA. Another avenue is to investigate how the problem (46) could be solved at each TCL, intermittently, instead of at the BA. Since the computational power of the processor at each TCL is lower than that of the processor at the BA, online distributed algorithms for convex optimization could play a role. The Fokker–Planck equations from Malhame and Chong (1985) we used here are convenient for modeling TCL populations with a small degree of heterogeneity. Distributed computation of optimal policies locally at each TCLs may help extend the method to a highly heterogeneous population of TCLs. The proposed framework performs reference computation and coordination algorithm design simultaneously by posing both as decision variables in an optimization problem. The coordination algorithm is parameterized by a randomized policy. It is possible that with appropriate parameterization of a policy, reference+algorithm design can be extended to non-randomized grid support policies, and to applications beyond coordinating TCLs.

## Appendix. Proofs

### A.1. Proof of Lemma 4

Eq. (45) is used to define the individual entries of the policy $\phi$, which then form the matrix version $\Phi$ as seen in (44). Hence

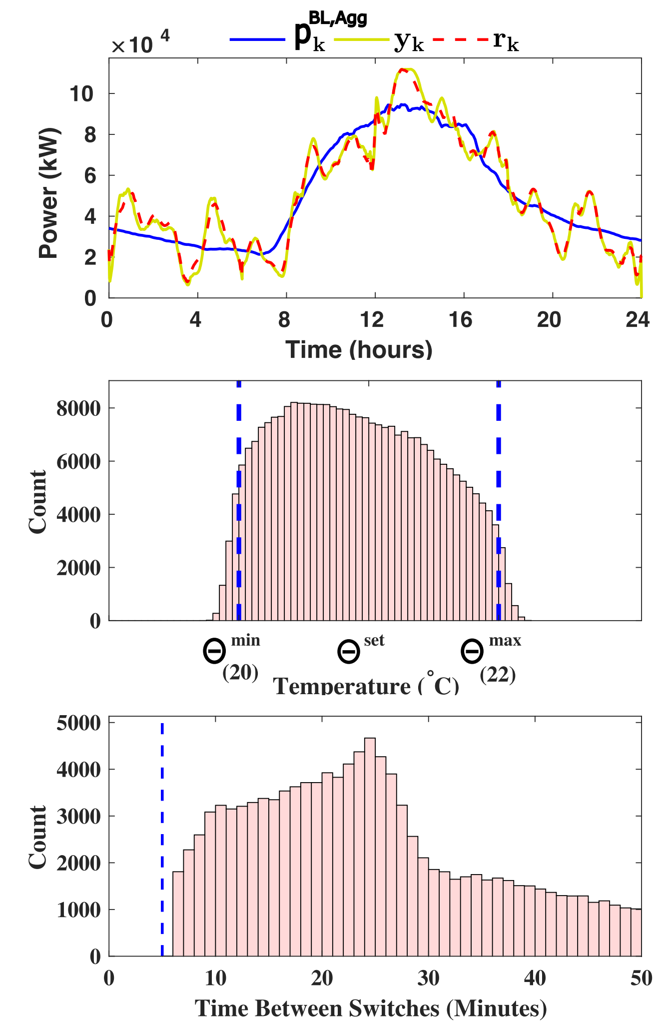


**Fig. 6.** (Top): Reference tracking results for the TCLs under the influence of the grid support control policies obtained by solving (46). (Middle): Histogram of the 200 TCL's temperature trajectories over the entire simulation horizon. (Bottom): Histogram of the time between switches over 3000 TCLs with the vertical line representing the minimum allowable time between switches.

the constructed policy satisfies (44). It remains to show that the policy constructed by the algorithm is a valid randomized policy.

The policy will automatically satisfy constraints (34)–(35) since it represents a structural constraint that specifies how to build the matrix representation of the policy from its vectorized form, which is object constructed in Algorithm 1. When $\nu_v[\Theta^j, l, k] = 0$ the algorithm choices then ensure constraints (29) and (30) and $\mathbb{1} = \Phi\mathbb{1}$ by construction. When $\nu_v[\Theta^j, l, k] > 0$ we have that

$$\sum_u \phi_v^{\mathrm{GQ}}(u \mid j, l) \tag{A.1}$$

$$= \sum_u \frac{\mathsf{P}\left(m_{k+1} = u,\ I_k = j,\ \mathsf{L}_k = l,\ m_k = v\right)}{\nu_v[\Theta^j, l, k]} \tag{A.2}$$

$$= \frac{1}{\nu_v[\Theta^j, l, k]} \sum_u \mathsf{P}\left(m_{k+1} = u,\ I_k = j,\ \mathsf{L}_k = l,\ m_k = v\right) \tag{A.3}$$

$$= \frac{\nu_v[\Theta^j, l, k]}{\nu_v[\Theta^j, l, k]} = 1 \tag{A.4}$$

by definition of the joint distribution. We also see from the above that the entries of the vector $\phi_v^{\mathrm{GQ}}$ sum to 1, and each entry is nonnegative by construction. Thus, each entry must be in [0, 1], meaning $\phi_v^{\mathrm{GQ}}$ is a valid probability vector. The equality constraints as specified in (29) and (30) will be satisfied since those appear as constraint of the form (45) in the optimization problem (46). The vector $\phi_v^{\mathrm{GQ}}$ is thus a valid randomized policy.

### A.2. Proof of Lemma 5

The proof structure is similar to the one in Benenati et al. (2019). The idea is to exploit the fact that: (i) $\nu_k^{\mathrm{GQ}}$ is a decision variable for both optimization problems (46) and (40) and (ii) the objective function is the same for both problems and solely a function of the marginal $\nu_k^{\mathrm{GQ}}$. We rewrite these problem compactly below,

$$\eta^*_{\mathrm{CVX}} = \min_{(\nu^{\mathrm{GQ}}, J) \in X} \eta(\nu^{\mathrm{GQ}}), \tag{A.5}$$

$$\eta^*_{\mathrm{NCVX}} = \min_{(\nu^{\mathrm{GQ}}, \Phi^{\mathrm{GQ}}) \in Y} \eta(\nu^{\mathrm{GQ}}), \tag{A.6}$$

where the sets $X$ and $Y$ collect all of the relevant constraints for the problems. The variables $\nu^{\mathrm{GQ}}$, $\Phi^{\mathrm{GQ}}$, and $J$ are concatenated over the considered finite time horizon and hence are not sub-scripted by $k$.

To prove that $\eta^*_{\mathrm{CVX}} \le \eta^*_{\mathrm{NCVX}}$, pick any argument minimizer that achieves value $\eta^*_{\mathrm{NCVX}}$ and denote the pair as $(\nu^{\mathrm{GQ}}_{\mathrm{NCVX}}, \Phi^{\mathrm{GQ}}_{\mathrm{NCVX}})$. Trivially construct $J$ through the relation (44) so that this constructed $J$ and $\nu^{\mathrm{GQ}}_{\mathrm{NCVX}}$ (that is optimal for (40)) are also feasible for (46), i.e., $(\nu^{\mathrm{GQ}}_{\mathrm{NCVX}}, J) \in X$. This is since $\mathbb{1}^T \mathrm{diag}(\nu_k^{\mathrm{GQ}}) = \nu_k^{\mathrm{GQ}}$ and $\mathbb{1} = \Phi_k^{\mathrm{GQ}}\mathbb{1}$. Hence we have that

$$\eta^*_{\mathrm{CVX}} = \min_{(\nu^{\mathrm{GQ}}, J^{\mathrm{GQ}}) \in X} \eta(\nu^{\mathrm{GQ}}) \le \eta(\nu^{\mathrm{GQ}}_{\mathrm{NCVX}}) = \eta^*_{\mathrm{NCVX}} \tag{A.7}$$

where the inequality follows from the fact that $\nu^{\mathrm{GQ}}_{\mathrm{NCVX}}, J^{\mathrm{GQ}}_{\mathrm{NCVX}}$ is only a feasible point and need not be a minimizer in $X$.

To prove the opposite, $\eta^*_{\mathrm{NCVX}} \le \eta^*_{\mathrm{CVX}}$, pick any argument minimizer that achieves value $\eta^*_{\mathrm{CVX}}$ and denote the pair as $(\nu^{\mathrm{GQ}}_{\mathrm{CVX}}, J^{\mathrm{GQ}}_{\mathrm{CVX}})$. Then apply Algorithm 1 to construct the corresponding conditional pmf, and call it (the matrix version) $\Phi^{\mathrm{GQ}}_{\mathrm{CVX}}$. Due to Lemma 4, this is a valid randomized policy and satisfies (44), which means that the pair $(\nu^{\mathrm{GQ}}_{\mathrm{CVX}}, \Phi^{\mathrm{GQ}}_{\mathrm{CVX}})$ is feasible for the nonconvex problem. Hence,

$$\eta^*_{\mathrm{NCVX}} = \min_{(\nu^{\mathrm{GQ}}, \Phi^{\mathrm{GQ}}) \in Y} \eta(\nu^{\mathrm{GQ}}) \le \eta(\nu^{\mathrm{GQ}}_{\mathrm{CVX}}) = \eta^*_{\mathrm{CVX}} \tag{A.8}$$

where the inequality follows from the fact that $\nu^{\mathrm{GQ}}_{\mathrm{CVX}}, \Phi^{\mathrm{GQ}}_{\mathrm{CVX}}$ is only a feasible point and need not be a minimizer over $Y$. Combining with (A.7), we have $\eta^*_{\mathrm{NCVX}} = \eta^*_{\mathrm{CVX}}$, which proves the Lemma.

## References

Bashash, Saeid, & Fathy, Hosam K. (2013). Modeling and control of aggregate air conditioning loads for robust renewable power management. *IEEE Transactions on Control Systems Technology*, *21*(4), 1318–1327.

Benenati, Emilio, Colombino, Marcello, & Dall'Anese, Emiliano (2019). A tractable formulation for multi-period linearized optimal power flow in presence of thermostatically controlled loads. In *2019 IEEE 58th conference on decision and control* (pp. 4189–4194). IEEE.

Burger, Eric M., & Moura, Scott J. (2017). Generation following with thermostatically controlled loads via alternating direction method of multipliers sharing algorithm. *Electric Power Systems Research*, *146*, 141–160.

Bušić, Ana, & Meyn, Sean (2016). Distributed randomized control for demand dispatch. In *IEEE conference on decision and control* (pp. 6964–6971).

Callaway, D. S., & Hiskens, I. A. (2011). Achieving controllability of electric loads. *Proceedings of the IEEE*, *99*(1), 184–199.

Chen, Yue, Bušić, Ana, & Meyn, Sean P. (2017). State estimation for the individual and the population in mean field control with application to demand dispatch. *IEEE Transactions on Automatic Control*, *62*(3), 1138–1149.

Chen, Yue, Hashmi, Md Umar, Mathias, Joel, Bušić, Ana, & Meyn, Sean (2017). Distributed control design for balancing the grid using flexible loads. In *IMA volume on the control of energy markets and grids* (pp. 1–26).

Coffman, Austin (2021). Distributed control of TCLs through convex optimization. https://gitlab.com/austinrc925/distributed-control-of-tcls-through-convex-optimization/-/tree/main.

Coffman, Austin, Bušić, Ana, & Barooah, Prabir (2018). Virtual energy storage from TCLs using QoS preserving local randomized control. In *5th ACM international conference on systems for built environments (BuildSys)* (p. 10).

Coffman, Austin R., Bušić, Ana, & Barooah, Prabir (2019). Aggregate capacity for TCLs providing virtual energy storage with cycling constraints. In *IEEE conference on decision and control*.

Coffman, Austin R., Bušić, Ana, & Barooah, Prabir (2021a). Control-oriented modeling of TCLs. In *American control conference (ACC)* (pp. 4148–4154). extended version in arXiv:2009.12960.

Coffman, Austin R., Bušić, Ana, & Barooah, Prabir (2021b). A unified framework for coordination of thermostatically controlled loads. arXiv preprint 2108.05840.

Coffman, Austin, Cammardella, Neil, Barooah, Prabir, & Meyn, Sean P. (2022). Aggregate flexibility capacity of TCLs with cycling constraints. *IEEE Transactions on Power Systems*, *38*(1), 52–62.

Coffman, Austin R., Guo, Zhong, & Barooah, Prabir (2021). Characterizing capacity of flexible loads for providing grid support. *IEEE Transactions on Power Systems*, *36*, 2428–2437.

Franceschelli, Mauro, Pilloni, Alessandro, & Gasparri, Andrea (2021). Multi-agent coordination of thermostatically controlled loads by smart power sockets for electric demand side management. *IEEE Transactions on Control Systems Technology*, *29*(2), 731–743.

Grant, M., & Boyd, S. (2011). CVX: Matlab software for disciplined convex programming. version 1.21. http://cvxr.com/cvx.

Hao, H., Sanandaji, B. M., Poolla, K., & Vincent, T. L. (2015). Aggregate flexibility of thermostatically controlled loads. *IEEE Transactions on Power Systems*, *30*(1), 189–198.

Kim, S., & Giannakis, G. (2013). Scalable and robust demand response with mixed-integer constraints. *IEEE Transactions on Smart Grid*, *4*(4), 2089–2099.

Lee, Zachary E., & Max Zhang, K. (2021). Scalable identification and control of residential heat pumps: A minimal hardware approach. *Applied Energy*, *286*, Article 116544.

Lee, Zachary E., Sun, Qingxuan, Ma, Zhao, Wang, Jiangfeng, MacDonald, Jason S., & Max Zhang, K. (2020). Providing grid services with heat pumps: A review. *ASME Journal of Engineering for Sustainable Buildings and Cities*, *1*(1), Article 011007.

Liu, M., & Shi, Y. (2016). Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services. *IEEE Transactions on Power Systems*, *31*(3), 1963–1971.

Liu, M., Shi, Y., & Liu, X. (2016). Distributed MPC of aggregated heterogeneous thermostatically controlled loads in smart grid. *IEEE Transactions on Industrial Electronics*, *63*(2), 1120–1129.

Mahdavi, Nariman, Braslavsky, Julio H., Seron, Maria M., & West, Samuel R. (2017). Model predictive control of distributed air-conditioning loads to compensate fluctuations in solar power. *IEEE Transactions on Smart Grid*, *8*(6), 3055–3065.

Malhame, Roland, & Chong, Chee-Yee (1985). Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system. *IEEE Transactions on Automatic Control*, *30*(9), 854–860.

Manne, Alan S. (1960). Linear programming and sequential decisions. *Management Science*, *6*(3), 259–267.

Mathieu, J. L., Koch, S., & Callaway, D. S. (2013). State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Transactions on Power Systems*, *28*, 430–440.

Meyn, Sean, Barooah, Prabir, Bušić, Ana, Chen, Yue, & Ehren, Jordan (2015). Ancillary service to the grid from intelligent deferrable loads. *IEEE Transactions on Automatic Control*, *60*, 2847–2862.

Paccagnan, Dario, Kamgarpour, Maryam, & Lygeros, John (2015). On the range of feasible power trajectories for a population of thermostatically controlled loads. In *2015 54th IEEE conference on decision and control* (pp. 5883–5888).

Soudjani, Sadegh Esmaeil Zadeh, & Abate, Alessandro (2015). Aggregation and control of populations of thermostatically controlled loads by formal abstractions. *IEEE Transactions on Control Systems Technology*, *23*(3), 975–990.

Totu, L. C., Wisniewski, R., & Leth, J. (2017). Demand response of a TCL population using switching-rate actuation. *IEEE Transactions on Control Systems Technology*, *25*(5), 1537–1551.

Versteeg, Henk Kaarle, & Malalasekera, Weeratunge (2007). *An introduction to computational fluid dynamics: the finite volume method*. Pearson education.

Dr. **Austin Coffman** received both his M.S. (Mechanical Engineering) and Ph.D. from the University of Florida in 2017 and 2021, respectively. After graduation he accepted an offer to work at Sysco as a Data Scientist. At the University of Florida, he was the recipient of the Iva and Norman Tuckett FINS fellowship in 2017, and a Chateaubriand fellowship for the 2019–2020 academic year. His research interests range from optimization and reinforcement learning to Markov chains.

**Ana Bušić** is a Research Scientist at Inria Paris Research Centre and the Computer Science Department at Ecole Normale Supérieure, PSL University, France. She received the M.S. degree in Applied Mathematics and the Ph.D. degree in Computer Science from the University of Versailles, France, in 2003 and 2007, respectively. She was a Post-Doctoral Fellow at Inria Grenoble–Rhône-Alpes and at University Paris-Diderot-Paris 7. Her research interests include stochastic modeling, simulation, optimization, and reinforcement learning, with applications to communication networks and energy systems. She is a member of the Laboratory of Information, Networking and Communication Sciences, a joint lab between Inria, Institut Mines-Télécom, UPMC Sorbonne Universities, Nokia Bell Labs, and SystemX. She received a Google Faculty Research Award in 2015.

**Prabir Barooah** is currently a Professor of Electronics and Electrical Engineering at the Indian Institute of Technology, Guwahati, India, on leave from the University of Florida, where he has been since 2007. He received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology, Kanpur, in 1996, the M.S. degree in mechanical engineering from the University of Delaware in 1999, and the Ph.D. degree from the University of California, Santa Barbara, in 2007. From 1999 to 2002, he was a Research Engineer with United Technologies Research Center, East Hartford, CT. He is the recipient of the Distinguished Professor Fellowship by the Fondation Sciences Mathe'matiques de Paris, France (2017), ASEE Southeastern Section's Outstanding Researcher Award (2013), NSF CAREER Award (2010), the Best Paper Award at the 2nd International Conference on Intelligent Sensing and Information Processing (2005), and a NASA Group Achievement Award (2003).