Decentralised control of thermostatic loads for flexible demand response

Simon H. Tindemans, *Member, IEEE*, Vincenzo Trovato, *Student Member, IEEE* and Goran Strbac, *Member, IEEE*

***Abstract*—Thermostatically controlled loads (TCLs) such as refrigerators, air-conditioners and space heaters offer significant potential for short-term modulation of their aggregate power consumption. This ability can be used in principle to provide frequency response services, but controlling a multitude of devices to provide a measured collective response has proven to be challenging. Many controller implementations struggle to manage simultaneously the short-term response and the long-term payback, whereas others rely on a real-time command-and-control infrastructure to resolve this issue. In this work we propose a novel approach to the control of TCLs that allows for accurate modulation of the aggregate power consumption of a large collection of appliances through stochastic control. By construction, the control scheme is well suited for decentralised implementation, and allows each appliance to enforce strict temperature limits. We also present a particular implementation that results in analytically tractable solutions both for the global response and for the device-level control actions. Computer simulations demonstrate the ability of the controller to modulate the power consumption of a population of heterogeneous appliances according to a reference power profile. Finally, envelope constraints are established for the collective demand response flexibility of a heterogeneous set of TCLs.**

***Index Terms*—power system control, load management, demand response, thermostatically controlled loads, frequency response, stochastic control**

NOTATION

A. Model parameters

T_{on}	asymptotic cooling temperature
T_{off}	room temperature
$T_{\max}$	maximum temperature threshold
$T_{\min}$	minimum temperature threshold
α	temperature relaxation constant
$v_{\mathrm{on}}(T)$	heating rate of *on*-appliances
$v_{\mathrm{off}}(T)$	heating rate of *off*-appliances
P_{on}	maximum power consumption

B. Steady state descriptors

π_0	steady state duty cycle
P_0	steady state average power consumption
$\bar{T}_0$	steady state average temperature

C. Control variables

$\Pi(t)$	target relative power consumption
$v(T,t)$	average heating rate of *on* and *off* appliances
$\beta(t)$	linear control variable

D. Dependent variables

$P(t)$	actual power consumption
$T_{\mathrm{low}}(t)$	temperature of coldest appliance ($\geq T_{\min}$)
$T_{\mathrm{high}}(t)$	temperature of warmest appliance ($\leq T_{\max}$)
$\theta(t)$	temperature of particular appliance
$\bar{T}(t)$	ensemble average temperature
$f(T,t)$	probability density of all appliances
$f_s(T,t)$	density of appliances in state $s \in \{\text{on, off}\}$
$\Phi(T,t)$	net on-off density flux
$r^{\mathrm{on}}_{\mathrm{off}}(T,t)$	stochastic switch-off rate
$r^{\mathrm{off}}_{\mathrm{on}}(T,t)$	stochastic switch-on rate

I. INTRODUCTION

THERMOSTATICALLY controlled loads (TCLs) consist of an electrical heating or cooling element that is controlled by a thermostat. The thermostat modulates the power used for heating/cooling in order to maintain a system's temperature near a setpoint value. In its most common implementation, the thermostat makes use of a temperature deadband around the setpoint value. When the upper deadband threshold is exceeded the system switches to a cooling state, either by switching on the cooling mechanism or by switching off the heating mechanism, and conversely, when the lower deadband threshold is exceeded the system switches to a heating state.

For TCLs that are controlled using a temperature deadband, small fluctuations in temperature are acceptable as long as the target temperature is approximately maintained over time. This insensitivity to temperature fluctuations means it is possible to shift demand from one moment in time to another without noticeably affecting the quality of service. Because a large number of TCLs is connected to the grid at all times, tapping into the collective flexibility offered by these loads has large potential benefits.

In 1979, Schweppe proposed to use such 'energy type usage devices' for the provision of frequency services [1] to the electricity grid. By monitoring the grid frequency devices can respond to frequency deviations by decreasing (in case of low frequency) or increasing (in case of high frequency) their instantaneous power consumption. This use case was analysed in detail by Short *et al.* [2] for fridge-freezers on the Great Britain (GB) network. Recently, Aunedi *et al.* [3] have carried out an assessment of the economical and environmental impacts of frequency regulation by frequency-responsive refrigerators for the GB system. Their study has found substantial benefits, especially for future generation portfolios with an increased penetration of renewables.

This illustrates the collective ability of responsive TCLs to contribute to the efficient operation of the grid. However, it is

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK. Corresponding e-mail: s.tindemans@imperial.ac.uk.

Simon Tindemans has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 274387.

not straightforward to design a satisfactory control algorithm for these applications, because individual appliances typically have only two power states (on and off). A multitude of devices must therefore be controlled in harmony to provide a dependable service to the network. There are four main challenges that must be addressed to make optimal use of the demand response potential offered by TCLs.

1) *Accurate control across a range of time scales.* TCLs must respond to frequency events within seconds and execute a well-managed return to normal operation (the 'payback' phase) over a time span of approximately one hour. Even longer periods are required for other dynamic demand opportunities such as energy arbitrage.
2) *Freedom to design complex responses.* The system operator or demand response aggregator should have the freedom to design complex responses in accordance with the technical and commercial requirements of the network. This design should take into account the physical capabilities of the TCLs.
3) *Reliable distributed response.* Real-time communication between a central controller and a large number of appliances requires a costly communications infrastructure that is sensitive to disturbances. It is therefore desirable to implement decentralised control based on locally available control signals (frequency, time). In cases where the collective response must be subject to centralised adjustment (for example by the TSO) the controller should be insensitive to latency of the communication channel.
4) *Satisfy per-appliance quality of service constraints.* Delivering the collective response to the grid should not impact the primary function of the TCLs: dependable temperature control. Therefore any demand response controller must deliver the response without violating the quality of service requirements (e.g. temperature limits) for *each* individual appliance.

Currently available methods address one or more of these challenges, but not all at the same time.

Initial controller designs [2], [4] have been heuristic modifications of regular deadband controllers. Although such controllers provide an effective initial response to frequency deviations, careful analysis of their long term response shows a tendency for devices to synchronise their cooling cycles. This interferes with the diversity of demand on the network, and may result in self-reinforcing frequency oscillations [5]. Ad hoc solutions to synchronisation have been proposed (e.g. [6], [7]) but those generally require careful tuning of parameters for specific scenarios, and still do not provide full control over the power profile beyond the initial response.

The traditional deadband controller can be extended to allow for tracking of arbitrary power profiles by adjusting the temperature setpoint in real time [8]–[10]. This approach usually assumes the availability of a real-time communication infrastructure [11]. A drawback of this approach is that the single control parameter (temperature setpoint) limits the range of available response actions, especially when strict per-appliance temperature limits are enforced. Furthermore, current implementations rely on a number of analytical and numerical approximations [12] that complicate the analysis and design of response actions.

Sinitsyn *et al.* [13] take a very different, non-perturbative approach to constructing load profiles. The on-off sequence resulting from a regular deadband controller is modified in order to provide a qualitatively desirable response whilst guaranteeing a return to diversified steady state operations. This results in a class of control strategies ('safe protocols') that can provide short-term power pulses without long-term synchronisation. However, the small number of safe protocols constructed in this way provides only a limited set of building blocks for the design of complex responses.

The stochastic controller by Angeli and Kountouriotis [5] is suitable for fully decentralised implementation. It ensures through random switching that devices do not become synchronised, thereby avoiding the long-term instability of simple setpoint controllers. The controller adjusts the properties of the steady state distribution and thereby exerts a slow control over the temperature distribution and power consumption of a population of fridges. This design fully eliminates the payback phase, but in doing so it prolongs the time it takes for appliances to regain their steady state temperature, and limits the ability to implement rapid load changes.

In this paper we present a novel control strategy for TCLs that for the first time addresses all four challenges listed above. It results in an ability to track *arbitrary* power profiles (within limits) using *independent* actions of a *heterogeneous* collection of TCLs, which enables an unprecedented range of demand response scenarios without onerous constraints on the communications system. To illustrate the power of this approach we further introduce a simple controller for which the resulting control actions can be determined analytically. This provides an additional level of insight into the theoretical control framework, and allows for an explicit computation of the contribution that the TCLs can make collectively in terms of power and energy levels. This analysis is extended to derive envelope constraints for the collective flexibility of heterogeneous appliances.

In section II we present the overall control framework, wherein TCLs are controlled in a fully decentralised manner by letting each appliance independently target a reference power profile. It is discussed how this feature may be used to enable either fully autonomous control or semi-autonomous control, depending on requirements. The (statistically) accurate tracking of the reference power profile is enabled by a transformation of the dynamic equations of appliance temperatures that exposes the net heating rate $v(T, t)$ as a suitable control parameter. The collective power consumption of TCLs can thus be modulated by controlling the population-averaged rate of heating or cooling. Finally, we demonstrate that appropriate device-level control actions can be derived from this population-level control approach.

Section III considers the specific case of TCLs with linear first order thermal models. For this common class of models we determine the relation between temperature and power consumption. This is used to derive generic limits on the flexibility of appliances, expressed in terms of energy and

(instantaneous) power consumption.

In section IV we present an illustrative example of a controller that is designed according to the framework presented in section II. The control function $v(T,t)$ is defined to be a linear function of T, which enables an analytical derivation of the controller's properties. For an arbitrary target power profile we derive expressions for the temperature distributions, required global and local control actions, and bounds on the flexibility (power and energy) of the controller. At an appliance level the implementation of the controller requires only basic mathematical operations that can easily be performed by unsophisticated appliances.

An algorithm for simulating individual device actions is given section V, followed by results that demonstrate the ability of a heterogeneous population of appliances to follow a complex reference power curve. Finally, the aggregate demand response flexibility of such a heterogeneous population is assessed and it is shown that this can be represented by envelope constraints that are characterised by four aggregate parameters.

II. Control framework

A. General approach

We propose to control the aggregate power consumption of TCLs in a decentralised manner by letting each appliance independently target a relative power curve $\Pi(t)$. Each appliance a controls its power consumption $P^a(t)$ in such a way that

$$E[P^a(t)] = P_0^a \Pi(t), \tag{1}$$

where P_0^a is its time-averaged steady state power consumption. The notion of expectation used in this expression is with respect to the statistical ensemble of all possible initial conditions of appliance a (temperatures, states) and – for stochastic controllers – all possible control sequences. The reference curve $\Pi(t)$ is identical for all appliances, and by definition a dynamic response action starts and ends with $\Pi(t) = 1$ (steady state).

If the response of each appliance satisfies Eq. (1), then it is easy to see that $\Pi(t)$ also modulates the global power consumption $P^{\text{total}}(t)$:

$$\begin{aligned} E[P^{\text{total}}(t)] &= \sum_a E[P^a(t)] \\ &= \sum_a P_0^a \Pi(t) \\ &= P_0^{\text{total}} \Pi(t), \end{aligned} \tag{2}$$

where P_0^{total} is the steady state aggregate power consumption. Furthermore, because the appliances are statistically independent of each other, the relative deviations from the expectation will decrease approximately as $1/\sqrt{N}$, where N is the number of appliances. For large N we may therefore assume

$$P^{\text{total}}(t) \approx P_0^{\text{total}} \Pi(t). \tag{3}$$

The resulting high-level control framework is illustrated in Fig. 1. The block on top represents the demand response designer that establishes the desired demand response pattern. In the following we will assume that this role is fulfilled by the system operator, but depending on regulatory and market arrangements this task could also be performed by demand response aggregators.

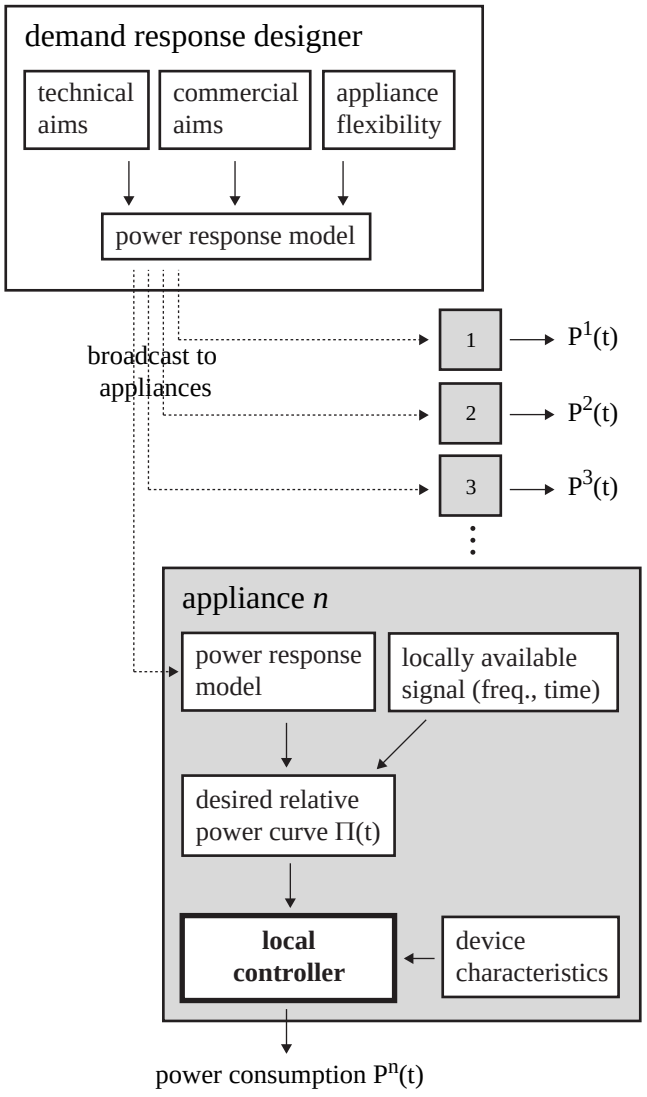

Figure 1. High level overview of the control framework. The demand response designer establishes a power response model based on technical and commercial aims, and constrained by the appliances' flexibility. The power response model is distributed to each appliance. In combination with a locally available global signal this enables each appliance to independently establish the desired relative power curve $\Pi(t)$. A local controller then modulates the power consumption of the appliance in accordance with Eq. (1).

The desired relative power curve $\Pi(t)$ may by linked dynamically to the condition of the power system, e.g. for the provision of frequency response services. Broadcasting such a curve in real time to all appliances would require a reliable low-latency communication channel, which is expensive and sensitive to disruptions. Instead, we propose that the system operator creates a *power response model* which is distributed to the appliances. The power response model specifies how the relative power curve $\Pi(t)$ can be computed as a function of a locally available signal, such as the grid frequency or the current time. This way, each appliance is able to independently construct the desired relative power curve $\Pi(t)$. For example, the appliances may be instructed to modulate their power

consumption linearly in response to changes in the grid frequency.

In a basic implementation the power response model may be embedded in the appliance during production. This approach results in fully autonomous operation and lends itself well to the provision of hardwired primary response to grid frequency deviations. However, this mode of operation ignores the potentially significant benefits that result from an ability to update the power response model periodically. Updating the power response model on an hourly or daily basis would allow the system operator to schedule primary and secondary response services that are optimally suited to characteristics of the power system, which change throughout the day, week and year. Furthermore, it would enable the provision of energy arbitrage services that reduce peak load, generation cost and/or ramp constraints.

We refer to this as 'semi-autonomous' operation, because the updates occur on a time scale (hours - days) that is larger than that of the actual response (seconds - minutes). In contrast with direct centralised control there is no need for real-time communication with the appliances, so the latency of the communication channel is not critical; it is sufficient for the channel to be dependable. The smart metering infrastructure that is currently being rolled out has such characteristics, as its primary aim is to support dependable retrospective billing. It is therefore well suited to support the proposed semi-autonomous control of TCLs.

When allocating demand response it is critical that the system operator knows that the desired curve $\Pi(t)$ can be realised by the appliances without violating their quality of service requirements. This prerequisite has two consequences. First, the system operator needs access to information regarding the aggregate capability of the participating appliances, including their total load and their ability to modulate their power levels and temperatures. If two-way communication is available, such information may be obtained directly from the appliances or their aggregators. Otherwise, it may need to be inferred from shipping numbers and targeted experiments. Second, the system operator needs efficient means to assess the compatibility of device parameters and any proposed power curves. For the controllers under consideration in this paper, a sufficient set of aggregate parameters and conditions to establish compatibility with a proposed power curve $\Pi(t)$ are provided in section V-C.

Taking Eq. (1) as a starting point, the following sections focus on the design of the local controller for a *single* appliance that modulates its power consumption according to the desired relative power curve $\Pi(t)$. Because only a single appliance is considered the use of the superscript a is suspended until section V, where the case of multiple and heterogeneous appliances is explicitly reintroduced.

B. Generic appliance model

In this work we consider thermal appliances with an internal state that is fully characterised by the temperature and can therefore be modelled by a first order ODE. For simplicity we use the example of refrigerators throughout this paper, although the same ODE model can trivially be applied to other thermostatically controlled loads such as air conditioning units, resistive space heaters, etc. Furthermore, we assume that each device exists in either an 'on' or 'off' state, but extensions to devices with fractional power states are conceivable. The evolution of the temperature $\theta(t)$ of an appliance in state s is thus determined by the differential equation

$$\frac{\mathrm{d}\theta(t)}{\mathrm{d}t} = \begin{cases} v_{\mathrm{on}}(\theta(t)), & \text{when } s = \text{on} \\ v_{\mathrm{off}}(\theta(t)), & \text{when } s = \text{off} \end{cases} \tag{4}$$

where $v_{\mathrm{on}}(\cdot) < 0$ is the (active) cooling rate and $v_{\mathrm{off}}(\cdot) > 0$ the (passive) heating rate.

A typical refrigerator controller uses a deadband $[T_{\min}, T_{\max}]$ around a temperature setpoint T_{set}. When the temperature reaches the lower bound $T_{\min}$ the appliance switches off, and when the temperature reaches the upper bound $T_{\max}$ it switches back on. A popular approach to enabling demand response from TCLs is to extend this control strategy by shifting the upper and lower temperature bounds in unison [2], [8]–[11]. In this work we enlarge the scope for control by including both deterministic and stochastic switching as follows:

1) *Variable thresholds.* The appliance always switches off when the lower temperature threshold $T_{\mathrm{low}}(t)$ is reached, and switches on when the upper temperature threshold $T_{\mathrm{high}}(t)$ is reached. This guarantees that the temperature of an individual appliance never exceeds the interval $[T_{\mathrm{low}}(t), T_{\mathrm{high}}(t)]$, provided that the rate of change of the thresholds does not exceed the maximum heating/cooling rate of the appliances.
2) *Stochastic switching.* For intermediate temperatures $\theta \in (T_{\mathrm{low}}(t), T_{\mathrm{high}}(t))$ switching is controlled by stochastic switching rates $r_{\mathrm{off}}^{\mathrm{on}}(\theta, t)$ (on $\rightarrow$ off) and $r_{\mathrm{on}}^{\mathrm{off}}(\theta, t)$ (off $\rightarrow$ on). These rates may be set to zero to recover conventional setpoint controllers.

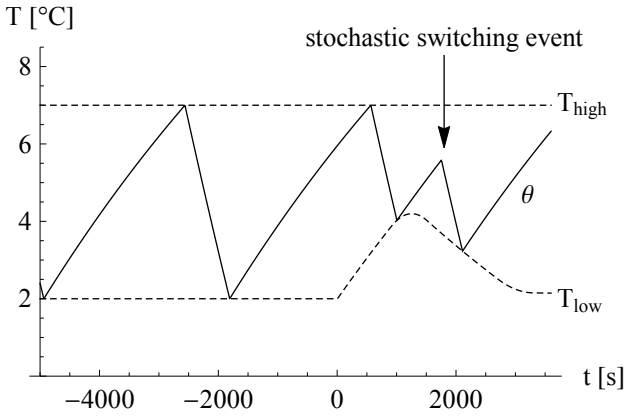

Figure 2. Illustrative temperature trace of a single refrigerator controlled by a hybrid threshold-stochastic controller. A linear thermal model as in (19) was employed, using model parameters from Table I. The device is controlled by the linear controller of section IV with the control signal shown in Fig. 5.

This control approach is illustrated in Fig. 2, which plots the temperature trace of a single refrigerator. For $t < 0$ the controller acts like a regular deadband controller and the temperature oscillates between the constant bounds $T_{\mathrm{low}}(t) = T_{\min}$

and $T_{\text{high}}(t) = T_{\max}$. After $t = 0$ the lower temperature limit $T_{\text{low}}(t)$ changes gradually and the threshold switching is augmented by occasional stochastic switching.

C. Temperature distribution formulation

Instead of individual appliance temperatures we will consider the probability density $f_s(T,t)$ of a population of fridges over temperatures T and the binary state variable $s \in \{\text{on}, \text{off}\}$. $f_{\text{on}}(T,t)$ describes the distribution of fridges that are on at time t, and $f_{\text{off}}(T,t)$ those that are off. Together, they satisfy the normalising constraint

$$\int_{-\infty}^{\infty} [f_{\text{on}}(T,t) + f_{\text{off}}(T,t)]\,\mathrm{d}T = 1. \tag{5}$$

The concept of a probability density on the temperature axis has also been used in [5], [8], [14]. Intuitively, the probability density is that of a large population of identical appliances with randomised (independent) internal states. However, the density $f_s(T,t)$ can also be interpreted as the probability distribution for the state of a *single* appliance with random initial conditions [5]. This subtle change in interpretation is significant, as it reveals the fact that we do not physically require a large population of identical appliances to make statements regarding *expected* behaviour.

Analogous to [5], we write down the evolution equations for $f_{\text{on}}(T,t)$ and $f_{\text{off}}(T,t)$, but we include unspecified temperature-dependent heating and cooling rates $v_{\text{off}}(T)$ and $v_{\text{on}}(T)$ and switching rates $r_{\text{off}}^{\text{on}}(T,t)$ and $r_{\text{on}}^{\text{off}}(T,t)$.

$$\begin{aligned}\frac{\partial}{\partial t} f_{\text{on}}(T,t) = &-\frac{\partial}{\partial T}\left[v_{\text{on}}(T) f_{\text{on}}(T,t)\right] \\ &- r_{\text{off}}^{\text{on}}(T,t) f_{\text{on}}(T,t) + r_{\text{on}}^{\text{off}}(T,t) f_{\text{off}}(T,t),\end{aligned} \tag{6a}$$

$$\begin{aligned}\frac{\partial}{\partial t} f_{\text{off}}(T,t) = &-\frac{\partial}{\partial T}\left[v_{\text{off}}(T) f_{\text{off}}(T,t)\right] \\ &+ r_{\text{off}}^{\text{on}}(T,t) f_{\text{on}}(T,t) - r_{\text{on}}^{\text{off}}(T,t) f_{\text{off}}(T,t).\end{aligned} \tag{6b}$$

These equations are valid on the interval $[T_{\text{low}}(t), T_{\text{high}}(t)]$ and are supplemented by the following flux balance boundary conditions, representing appliance switching at the lower/upper temperature thresholds.

$$\begin{aligned}f_{\text{on}}(T_{\text{low}},t)&\left[\frac{\mathrm{d}T_{\text{low}}(t)}{\mathrm{d}t} - v_{\text{on}}(T_{\text{low}})\right] \\ &= f_{\text{off}}(T_{\text{low}},t)\left[v_{\text{off}}(T_{\text{low}}) - \frac{\mathrm{d}T_{\text{low}}(t)}{\mathrm{d}t}\right]\end{aligned} \tag{7a}$$

$$\begin{aligned}f_{\text{on}}(T_{\text{high}},t)&\left[\frac{\mathrm{d}T_{\text{high}}(t)}{\mathrm{d}t} - v_{\text{on}}(T_{\text{high}})\right] \\ &= f_{\text{off}}(T_{\text{high}},t)\left[v_{\text{off}}(T_{\text{high}}) - \frac{\mathrm{d}T_{\text{high}}(t)}{\mathrm{d}t}\right]\end{aligned} \tag{7b}$$

D. The control function $v(T,t)$

The probability density $f_s(T,t)$ is defined with respect to two variables that have a fundamentally different character: temperature and appliance state. The temperature is a physical property of the appliances that is critical to the quality of service (preserving food) and that changes only gradually. The appliance state, however, is a control lever that can be changed instantaneously. In the following, we reformulate equations (6) to reflect this qualitative difference. This will expose a convenient control parameter for the expected power consumption associated with a distribution.

First we eliminate the appliance state by summing equations (6a) and (6b). We obtain the continuity equation for the temperature-only probability density $f(T,t)$:

$$\frac{\partial}{\partial t} f(T,t) = -\frac{\partial}{\partial T}\left[v(T,t) f(T,t)\right], \tag{8}$$

where

$$f(T,t) \equiv f_{\text{on}}(T,t) + f_{\text{off}}(T,t) \tag{9}$$

$$v(T,t) \equiv \frac{v_{\text{on}}(T) f_{\text{on}}(T,t) + v_{\text{off}}(T) f_{\text{off}}(T,t)}{f_{\text{on}}(T,t) + f_{\text{off}}(T,t)}. \tag{10}$$

$f(T,t)$ is the probability density of temperatures, regardless of device status; $v(T,t)$ is the average heating rate of all devices with temperature T, which dictates the 'flow' of devices along the temperature axis. Clearly, this must satisfy $v_{\text{on}}(T) \leq v(T,t) \leq v_{\text{off}}(T)$ and because there are only two device states there is a one-to-one correspondence between the value of $v(T,t)$ and the relative size of $f_{\text{on}}(T,t)$ and $f_{\text{off}}(T,t)$. Using definitions (9) and (10), the boundary conditions (7) simplify to

$$\frac{\mathrm{d}T_{\text{low}}(t)}{\mathrm{d}t} = v(T_{\text{low}}(t),t) \tag{11a}$$

$$\frac{\mathrm{d}T_{\text{high}}(t)}{\mathrm{d}t} = v(T_{\text{high}}(t),t). \tag{11b}$$

Equations (8) and (11) show that the evolution of the temperature density $f(T,t)$ is fully determined by the heating rate $v(T,t)$, which is illustrated using a schematic example in Fig. 3.

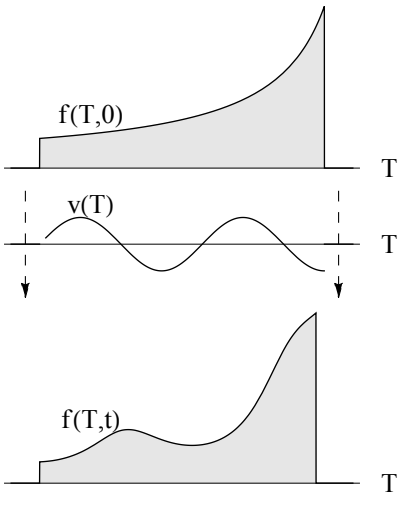

Figure 3. Schematic representation of the way the heating rate profile $v(T)$ controls the evolution of the temperature distribution $f(T,t)$. For this illustrative example a constant heating rate profile $v(T,t) = v(T)$ was used.

The main innovation of the control approach introduced in this paper is the *use of* $v(T,t)$ *itself as a control parameter*.

This approach inverts the standard formulation of the control problem in which the population properties are derived from and expressed in terms of the appliance-level switching rates $r^{\mathrm{on}}_{\mathrm{off}}(T,t)$ and $r^{\mathrm{off}}_{\mathrm{on}}(T,t)$ and the temperature bounds $T_{\mathrm{low}}(t)$ and $T_{\mathrm{high}}(t)$. In this approach, the lower/upper temperature setpoints are defined implicitly through (11). Similarly, the on/off switching rates can be determined using the asymmetric complement of equation (8). Subtracting (6b) from (6a) yields

$$\frac{\partial}{\partial t}\left[f_{\mathrm{on}}(T,t) - f_{\mathrm{off}}(T,t)\right] = - \frac{\partial}{\partial T}\left[v_{\mathrm{on}}(T) f_{\mathrm{on}}(T,t) - v_{\mathrm{off}}(T) f_{\mathrm{off}}(T,t)\right] - 2\Phi(T,t), \tag{12}$$

where we have introduced the shorthand

$$\Phi(T,t) = r^{\mathrm{on}}_{\mathrm{off}}(T,t) f_{\mathrm{on}}(T,t) - r^{\mathrm{off}}_{\mathrm{on}}(T,t) f_{\mathrm{off}}(T,t) \tag{13}$$

representing the net switching flux of appliances from the *on* to the *off* state. We solve (12) for $\Phi(T,t)$ and use the equalities

$$f_{\mathrm{on}}(T,t) = \left(\frac{v_{\mathrm{off}}(T) - v(T,t)}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t) \tag{14a}$$

$$f_{\mathrm{off}}(T,t) = \left(\frac{v(T,t) - v_{\mathrm{on}}(T)}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t) \tag{14b}$$

to express it as a function of $f(T,t)$ and $v(T,t)$:

$$\begin{aligned}\Phi(T,t) =& \frac{\partial}{\partial t}\left[\left(\frac{v(T,t) - \frac{1}{2}(v_{\mathrm{off}}(T) + v_{\mathrm{on}}(T))}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t)\right] \\ &+ \frac{\partial}{\partial T}\left[\left(\frac{\frac{1}{2} v(T,t)(v_{\mathrm{off}}(T) + v_{\mathrm{on}}(T))}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t)\right] \\ &- \frac{\partial}{\partial T}\left[\left(\frac{v_{\mathrm{off}}(T) v_{\mathrm{on}}(T)}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t)\right].\end{aligned} \tag{15}$$

Note from Eq. (13) that the value of $\Phi(T,t)$ does not fully determine the switching rates $r^{\mathrm{on}}_{\mathrm{off}}(T,t)$ and $r^{\mathrm{off}}_{\mathrm{on}}(T,t)$. This is to be expected because the flux contributions of devices switching on and off (at any given temperature T) cancel out. This indeterminacy is resolved by minimising the total rate of switching. In combination with the fact that switching rates must be positive this results in the following choice:

$$r^{\mathrm{on}}_{\mathrm{off}}(T,t) = \max\left(0, \frac{\Phi(T,t)}{f_{\mathrm{on}}(T,t)}\right) \tag{16a}$$

$$r^{\mathrm{off}}_{\mathrm{on}}(T,t) = \max\left(0, \frac{-\Phi(T,t)}{f_{\mathrm{off}}(T,t)}\right) \tag{16b}$$

We note that the choice to minimise the overall rate of switching is also beneficial from a device perspective, as it reduces the mechanical stress on the compressors of the refrigerators.

E. Two stage control of the expected power consumption

We return to the interpretation of $f(T,t)$ as the probability density for the temperature of a single appliance with random initial conditions. At each instant an appliance's compressor is either on or off (in case of a refrigerator), so its power consumption is either P_{on} (maximum power consumption) or nil. However, we can compute the *expected* fractional power consumption $\pi(t) = E[P(t)]/P_{\mathrm{on}}$ as

$$\begin{aligned}\pi(t) &= \int_{-\infty}^{\infty} f_{\mathrm{on}}(T,t)\,\mathrm{d}T \\ &= \int_{-\infty}^{\infty} \left(\frac{v_{\mathrm{off}}(T) - v(T,t)}{v_{\mathrm{off}}(T) - v_{\mathrm{on}}(T)}\right) f(T,t)\,\mathrm{d}T.\end{aligned} \tag{17}$$

Here, the expectation is an average over all possible initial conditions and - for stochastic controllers - all possible control sequences. The fractional power consumption $\pi(t) \in [0,1]$ and the relative power consumption $\Pi(t)$ are related by

$$\Pi(t) = \frac{\pi(t)}{\pi_0}, \tag{18}$$

where π_0 is the steady state fractional power consumption, equal to the duty cycle.

Note in Eq. (17) that $\pi(t)$ depends on both $f(T,t)$ and $v(T,t)$. Whereas the former is a slowly evolving density, the latter can be controlled directly by switching appliances on and off. Furthermore, we know how to compute device level switching rates (16) and temperature limits (11) from the variables $f(T,t)$ and $v(T,t)$. Together, these properties enable us to formulate a two-stage approach for the control of the expected power consumption in accordance with the desired relative power curve $\Pi(t)$. This is illustrated in Fig. 4 and further explained below.

- The first stage is the *ensemble control problem*. At this stage, each device considers a (hypothetical) ensemble of identical appliances. Based solely on its own thermal model and temperature limits it computes a suitable population-level response for achieving the desired relative power curve $\Pi(t)$. The steady state temperature distribution $f(T,0)$ is computed as a starting point, and a heating profile $v(T,t)$ is chosen in accordance with the target power consumption $\Pi(t)$ and Eq. (17). The co-evolution of the temperature distribution $f(T,t)$ and heating rate profile $v(T,t)$ follows from the continuity equation (8), as illustrated in Fig. 3.
- The second stage is the *device control problem*. This determines an appliance-level control strategy that is compatible with $f(T,t)$ and $v(T,t)$ as determined in the first stage. Specifically, for the hybrid deterministic-stochastic controller described in section II-B, this stage computes the switching rates $r^{\mathrm{on}}_{\mathrm{off}}(T,t)$ and $r^{\mathrm{off}}_{\mathrm{on}}(T,t)$ from the on-off flux $\Phi(T,t)$ using Eq. (16). The deterministic switching temperatures $T_{\mathrm{low}}(t)$, $T_{\mathrm{high}}(t)$ are computed from Eq. (11).

Finally, having determined a device-specific controller, individual appliances are switched on or off in accordance with this control strategy: switching always occurs when the temperature reaches the bounds $T_{\mathrm{low}}(t)$ and $T_{\mathrm{high}}(t)$, and stochastically according to $r^{\mathrm{on}}_{\mathrm{off}}(T,t)$ and $r^{\mathrm{off}}_{\mathrm{on}}(T,t)$ for intermediate temperatures.

III. Linear thermal model

In this section we determine properties of generic TCL controllers in combination with the common linear first order

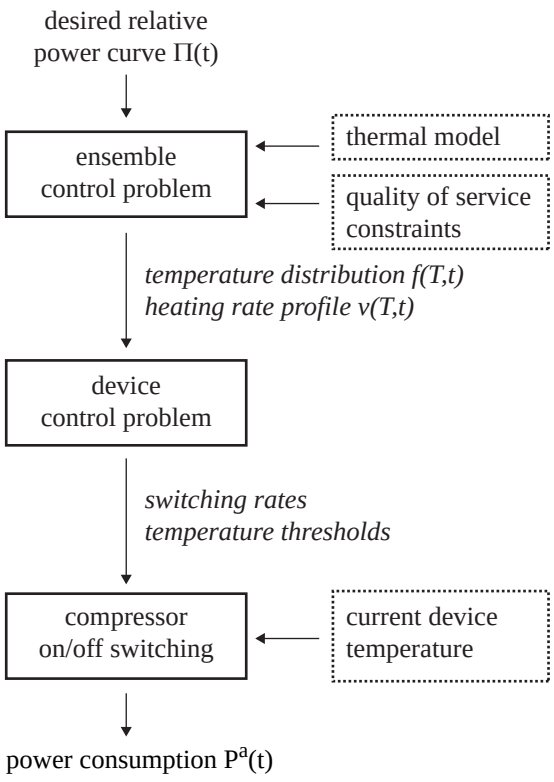

Figure 4. Components of the local controller that is embedded in each appliance. For a relative power curve $\Pi(t)$ the controller consecutively solves two sub-problems: first it determines an adequate ensemble response and for this response it computes a compatible set of switching rates and temperature thresholds. This two-stage procedure results in a control model that is used to switch the state of the compressor on and off. The controller modulates the power consumption $P^a(t)$ in such a way that $E[P^a(t)] = P_0^a \Pi(t)$ (Eq. (1)). The inputs in various stages are indicated by dotted boxes. See Fig. 1 for the embedding of the local controller into the overall control framework.

refrigerator model [4], [15], for which the heating/cooling rates are given by

$$v_{\rm on}(T) = -\alpha(T - T_{\rm on}) \tag{19a}$$
$$v_{\rm off}(T) = \alpha(T_{\rm off} - T). \tag{19b}$$

Here $T_{\rm off}$ and $T_{\rm on}$ are the asymptotic temperatures for the off and on states, respectively, and α is a temperature relaxation constant constant that quantifies the rate with which the temperature of the appliance equilibrates with its surroundings. The quantity $1/\alpha$ is also known as the thermal time constant. The appliance temperature evolution in the absence of switching is thus given by

$$T^a(t) = \begin{cases} T_{\rm on} + (T^a(t_0) - T_{\rm on})\mathrm{e}^{-\alpha(t-t_0)}, & \text{when } s^a = \text{on} \\ T_{\rm off} + (T^a(t_0) - T_{\rm off})\mathrm{e}^{-\alpha(t-t_0)}, & \text{when } s^a = \text{off} \end{cases} \tag{20}$$

We note that the population-level results in the remainder of this section do not depend on the details of the controller, and are thus applicable to *any* on-off controller for first order linear thermal models.

A. Temperature and power consumption

The fractional power consumption $\pi(t)$ as a function of $f(T,t)$ and $v(T,t)$ is determined by substituting (19) into (17).

$$\pi(t) = \int_{-\infty}^{\infty} \left(\frac{T_{\rm off} - T - v(T,t)/\alpha}{T_{\rm off} - T_{\rm on}} \right) f(T,t)\,\mathrm{d}T \tag{21}$$
$$= \frac{1}{T_{\rm off} - T_{\rm on}} \left[T_{\rm off} - \bar{T}(t) - \frac{\bar{v}(t)}{\alpha} \right], \tag{22}$$

where $\bar{T}(t)$ is the average temperature and $\bar{v}(t)$ is the average rate of heating at time t. In the steady state $\bar{v} = 0$, resulting in the following relation between the steady state average temperature $\bar{T}_0$ and the duty cycle π_0:

$$\pi_0 = \frac{T_{\rm off} - \bar{T}_0}{T_{\rm off} - T_{\rm on}}. \tag{23}$$

By definition $\bar{v}(t) = \mathrm{d}\bar{T}(t)/\mathrm{d}t$, so Eq. (22) can be interpreted as a differential equation for $\bar{T}(t)$. Expressed as a function of the relative power consumption $\Pi(t) = \pi(t)/\pi_0$ it reads

$$\frac{\mathrm{d}\bar{T}(t)}{\mathrm{d}t} = -\alpha \left[\bar{T} - T_{\rm off} + (T_{\rm off} - \bar{T}_0)\Pi(t) \right]. \tag{24}$$

It is easily verified that the ensemble averaged temperature $\bar{T}$ changes as that of a single appliance with a variable relative cooling rate $P(t) = \pi_0 \Pi(t) P_{\rm on}$. Solving the differential equation results in

$$\bar{T}(t) = T_{\rm off} + (\bar{T}(t_0) - T_{\rm off})\mathrm{e}^{-\alpha(t-t_0)} - \alpha(T_{\rm off} - \bar{T}_0) \int_{t_0}^{t} \Pi(t')\mathrm{e}^{-\alpha(t-t')}\,\mathrm{d}t'. \tag{25}$$

Taking the limit $t_0 \to -\infty$ further simplifies this to

$$\bar{T}(t) = T_{\rm off} - \alpha(T_{\rm off} - \bar{T}_0) \int_{-\infty}^{t} \Pi(t')\mathrm{e}^{-\alpha(t-t')}\,\mathrm{d}t'. \tag{26}$$

This result shows that the average temperature is determined by the exponentially smoothed reference power curve $\Pi(t)$.

B. Energy and power constraints

The primary function of thermostatically controlled loads such as refrigerators is to maintain a compartment at a desired temperature. The provision of system services should not excessively impact this task, so it is natural to require that the temperature each individual appliance must remain within an acceptable interval $[T_{\min}, T_{\max}]$ at all times. This constraint limits the permissible power profiles $\Pi(t)$, both in terms of energy and power.

1) Capacity for energy services: The energy constraint reflects the ability to sustain a reduced power level, effectively allowing the power system to 'borrow' energy from the appliances. This effective store of energy is fully depleted when *each* appliance attains the upper temperature limit $T_{\max}$, so that $\bar{T} = T_{\max}$. Conversely, the amount of stored energy is maximised when each appliance is at its lower temperature limit $T_{\min}$, so that $\bar{T} = T_{\min}$. An optimal controller that is able to fully utilise the temperature range is therefore constrained to the interval

$$T_{\min} \le \bar{T}(t) \le T_{\max}. \tag{27}$$

Using equation (26) this translates into the following constraint for the reference power profiles:

$$1-\left(\frac{T_{\max}-\bar{T}_0}{T_{\text{off}}-\bar{T}_0}\right) \leq \alpha \int_{-\infty}^{t} \Pi(t')\mathrm{e}^{-\alpha(t-t')}\,\mathrm{d}t' \leq 1+\left(\frac{\bar{T}_0-T_{\min}}{T_{\text{off}}-\bar{T}_0}\right). \quad (28)$$

This expression clearly expresses the link between the temperature margin ($T_{\max}-T_{\min}$) and the energy bounds. By setting $\Pi(t)$ to a constant value $\Pi_{\text{sustained}}$ we find the following inequality for power levels that can be sustained indefinitely without violating Eq. (27):

$$1-\left(\frac{T_{\max}-\bar{T}_0}{T_{\text{off}}-\bar{T}_0}\right) \leq \Pi_{\text{sustained}} \leq 1+\left(\frac{\bar{T}_0-T_{\min}}{T_{\text{off}}-\bar{T}_0}\right). \quad (29)$$

Note that Eq. (26) implies that operating at a constant power level eventually results in a constant average temperature. Specifically, this means the steady state temperature can always be restored asymptotically by operating at the steady state power level $\Pi = 1$. Therefore it is not a strict requirement to 'pay back' the borrowed energy after an initial demand reduction. This convenient property was used successfully in [5] to avoid fluctuating power levels during the recovery phase. However, even when a payback phase is not strictly required it may nevertheless be useful to temporarily boost power consumption in order to speed up the recovery of the average temperature.

2) Instantaneous power consumption: It is not possible to derive similar generic bounds for the instantaneous power consumption without specifying (elements of) the controller implementation. In principle, the full range of fractional power levels between $\pi(t) = 0$ and $\pi(t) = 1$ can be attained simply by randomly selecting a fraction $\pi(t)$ of appliances that will be in the 'on' state.

The exception to this simple result are appliances that cannot be switched on and off because doing so would *immediately* cause a violation of their quality of service constraints. The fraction of appliances for which this applies is usually vanishingly small, especially in the initial stages of a demand response action. Therefore we state the following approximate result in the absence of a specific controller implementation:

$$0 \lessapprox \Pi(t) \lessapprox \frac{1}{\pi_0}. \quad (30)$$

In section IV-D2 we will determine instantaneous power bounds for the specific controller developed in the next section.

IV. Linear controller

Section II introduced a two stage approach to controlling the aggregate power consumption of thermostatically controlled loads, and section III derived generic bounds for the collective demand response abilities of loads that have linear thermal models. In this section we build on these results and introduce a particular controller implementation. The proposed linear controller does not make full use of the flexibility of the appliances, but its simple functional form results in an analytically tractable control problem. This property makes it an attractive choice as an illustrative example. Furthermore, the linear controller has two desirable features:

1) It reduces to the traditional setpoint controller in the steady state.
2) It does not permit any fridges to exceed their maximum temperature $T_{\max}$.

A. Steady state: setpoint controller

Most thermostatic loads are controlled by a traditional setpoint controller, where – in the case of a refrigerator – the cooling is switched on when an upper temperature bound $T_{\max}$ is reached, and switched off when the lower temperature bound $T_{\min}$ is reached. This type of controller minimises the number of start-stop cycles for a given temperature interval $[T_{\min}, T_{\max}]$. We take this setpoint controller as the basis for our responsive demand controller, requiring identical behaviour in the steady state. Specifically, we take the steady state switching temperatures to be equal to the strict temperature limits: $[T_{\text{low}}(0), T_{\text{high}}(0)] = [T_{\min}, T_{\max}]$, with the assumption that the system is in steady state at $t = 0$.

The steady state temperature distribution $f(T, 0)$ follows from Eq. (12), using $\Phi(T, t) = 0$ (no stochastic switching) and $v(T, t) = 0$ (steady state). This implies $v_{\text{off}}(T)v_{\text{on}}(T)f(T, 0) =$ constant, and normalising $f(T, 0)$ on the interval $[T_{\min}, T_{\max}]$ results in

$$f(T,0) = \frac{C}{(T_{\text{off}}-T)(T-T_{\text{on}})} \quad (31)$$

with

$$C = \frac{T_{\text{off}}-T_{\text{on}}}{\log\left(\frac{(T_{\max}-T_{\text{on}})(T_{\min}-T_{\text{off}})}{(T_{\min}-T_{\text{on}})(T_{\max}-T_{\text{off}})}\right)}. \quad (32)$$

The steady state duty cycle (fractional power consumption) π_0 is determined by inserting (31) into (21) and setting $v(T, 0) = 0$:

$$\pi_0 = \frac{\log\left(\frac{T_{\max}-T_{\text{on}}}{T_{\min}-T_{\text{on}}}\right)}{\log\left(\frac{(T_{\max}-T_{\text{on}})(T_{\min}-T_{\text{off}})}{(T_{\min}-T_{\text{on}})(T_{\max}-T_{\text{off}})}\right)}. \quad (33)$$

The average temperature $\bar{T}_0$ then follows from Eq. (23).

B. Control parameter

In order to modulate the power consumption we parametrise the heating rate profile $v(T, t)$ (the control *function*) to obtain a control *parameter*. Eq. (22) suggests choosing $v(T, t)$ as a linear function of T, so we propose

$$v(T,t) = \alpha\beta(t)(T-T_{\max}). \quad (34)$$

The heating/cooling rate of the refrigerator population is therefore controlled by the dimensionless control parameter $\beta(t)$, with the convention that $\beta(t) < 0$ corresponds to a net heating effect (reduced power consumption) and $\beta(t) > 0$ corresponds to a net cooling effect (increased power consumption).

Because $v(T_{\max}, t) = 0$ for any choice of β, we find that $T_{\text{high}}(t) = T_{\max}$ at all times (cf. Eq. (11b)). As a result, appliance temperatures will not exceed $T_{\max}$, in line with the quality of service requirements. Unlike the upper temperature

threshold, the lower temperature threshold $T_{\text{low}}(t)$ is dynamic under this controller. Its evolution is computed by solving the differential equation (11a) with the initial condition $T_{\text{low}}(0) = T_{\min}$, resulting in

$$T_{\text{low}}(t) = T_{\max} - (T_{\max} - T_{\min})\mathrm{e}^{\int_0^t \alpha\beta(t')\,\mathrm{d}t'}. \tag{35}$$

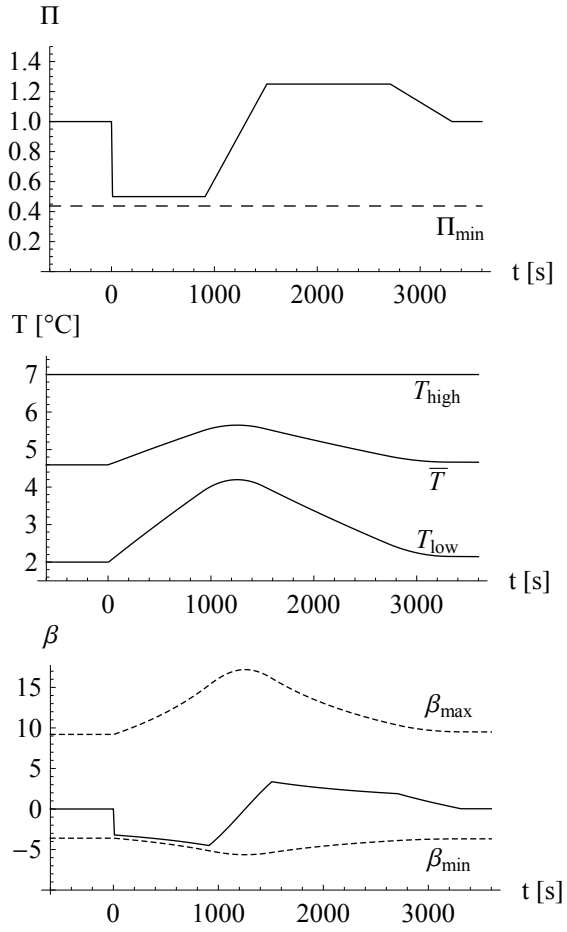

Figure 5. Ensemble control using the appliance thermal model as specified in Table I. Top: reference relative power curve and the lower limit given by Eq. (46). Middle: Minimum, mean and maximum temperatures of the ensemble. Bottom: Value of the control parameter $\beta(t)$ plotted alongside upper and lower limits representing the constraints in Eq. (43).

Inserting (34) into (22) and solving for $\beta(t)$ expresses the control parameter as a function of the ensemble-averaged relative power consumption $\Pi(t) = \pi(t)/\pi_0$:

$$\beta(t) = \frac{\Pi(t)(T_{\text{off}} - \bar{T}_0) - (T_{\text{off}} - \bar{T}(t))}{T_{\max} - \bar{T}(t)}. \tag{36}$$

Together, equations (26) and (36) provide the means to determine the mean temperature $\bar{T}(t)$ that results from a given power profile $\Pi(t)$, and the value of $\beta(t)$ that is required to achieve this power profile. Note that this response depends only on the constant system parameters α, $T_{\max}$, T_{on} and T_{off}. As a result the controller is able to solve the ensemble control problem without resorting to numerical integration of PDEs, which is especially appealing for embedded applications.

Figure 5 shows the evolution of the average temperature $\bar{T}(t)$ and the control parameter $\beta(t)$ for a particular relative power curve $\Pi(t)$ that reduces the devices' power consumption to 50% of the steady state level and maintains this for 15 minutes. A mild increase in power consumption (payback) is used to reinstate the initial condition almost exactly by 55 minutes after the initial power reduction.

C. Temperature distribution

For the particular control scheme defined by (34) the differential equation (8) can be solved on the interval $[T_{\text{low}}(t), T_{\max}]$ to obtain the temperature distribution at any time t. Using (31) as the initial condition we obtain

$$\begin{aligned} f(T,t) = -CS(t)\left[(T - T_{\max}) + S(t)(T_{\max} - T_{\text{off}})\right]^{-1} \\ \times \left[(T - T_{\max}) + S(t)(T_{\max} - T_{\text{on}})\right]^{-1} \end{aligned} \tag{37}$$

with

$$S(t) = \mathrm{e}^{\int_0^t \alpha\beta(t')\,\mathrm{d}t'} \tag{38}$$

We remark that the full solution $f(T,t)$ corresponds to the steady state solution $f(T,0)$ scaled by a factor $S(t)$ around $T = T_{\max}$ and subsequently renormalised. This important result and the related expression (35) are of course a direct consequence of the restrictive form of (34). Figure 6 shows four snapshots of the temperature distribution $f(T,t)$, illustrating this linear scaling.

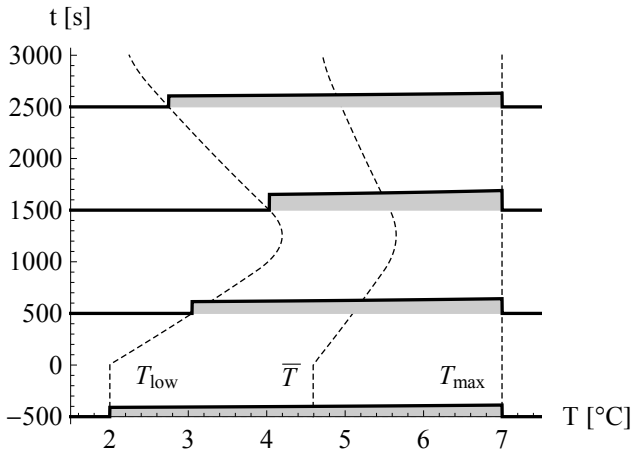

Figure 6. Snapshots of the temperature distribution $f(T,t)$ at $t = \{-500s, 500s, 1500s, 2500s\}$, for the response in Fig. 5. Dotted lines indicate the evolution of $T_{\text{low}}(t)$, $\bar{T}(t)$ and $T_{\max}$.

It is clear that the distribution returns to its steady state form at any time t^* on which the following equality holds

$$\int_0^{t^*} \beta(t')\,\mathrm{d}t' = 0. \tag{39}$$

Because the physical state of an appliance is fully determined by its temperature T, having a steady state distribution implies that the ensemble *is* in the steady state. We conclude that the time t^* marks the end of a control action.

D. Energy and power constraints

The linear controller defined by Eq. (34) was selected for its analytical tractability and designed specifically for dynamic demand *reduction* and the subsequent recovery process. For this reason it does not make full use of the physical capacity of the appliances. In this section we analyse the energy and power constraints associated with the linear controller, and compare these with the generic bounds for appliances with linear thermal models obtained in section III-B. We note that extensions to more elaborate control schemes that also permit a net absorption of energy from the grid are possible but beyond the scope of this paper.

1) Energy: The linear controller results in a scaling of the steady state temperature distribution $f(T,0)$ by a factor $S(t)$ around $T = T_{\max}$. In order to ensure that no appliance exceeds the temperature range $[T_{\min}, T_{\max}]$ we must restrict the controller to $S(t) \leq 1$ (see Eqs. (35) and (38)). This is especially relevant for refrigerators, in which inadvertent freezing must be avoided. In terms of the average temperature, this implies

$$\bar{T}_0 \leq \bar{T}(t) \leq T_{\max}. \tag{40}$$

As a result, the energy bounds in Eq. (28) are reduced to

$$1 - \left(\frac{T_{\max} - \bar{T}_0}{T_{\text{off}} - \bar{T}_0}\right) \leq \alpha \int_{-\infty}^{t} \Pi(t') \mathrm{e}^{-\alpha(t-t')}\, \mathrm{d}t' \leq 1. \tag{41}$$

A comparison with Eq.—(28) demonstrates that the controller is optimal with respect to available energy for demand *reduction*. Note that whereas increased power consumption levels $\Pi(t) > 1$ are not permitted initially, they may occur during the recovery phase that follows on an initial demand reduction (i.e. when $\bar{T}(t) > \bar{T}_0$).

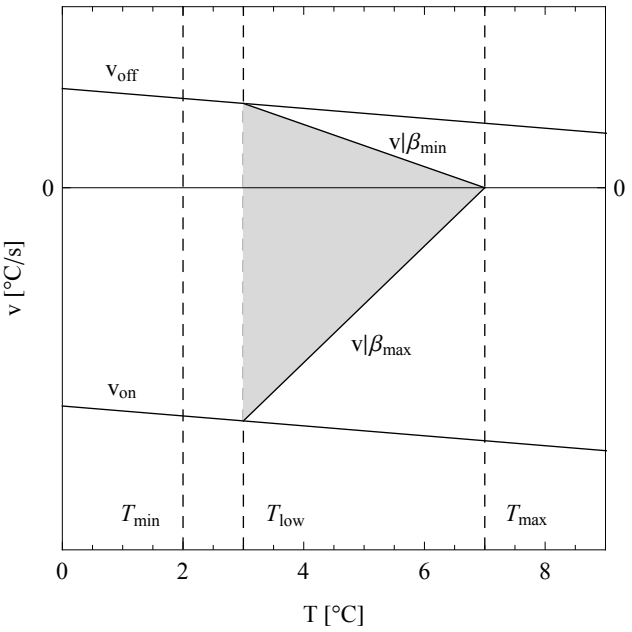

Figure 7. Graphical construction of the physical constraints on the control parameter $\beta(t)$ using parameters from Table I. Appliance temperatures are distributed between $T_{\text{low}}(t) \geq T_{\min}$ and $T_{\max}$ and their heating/cooling rates are bounded by v_{off} and v_{on}. The permissible range of $v(T,t) = \alpha\beta(t)(T - T_{\max})$ is shown in grey.

2) Power: The range of accessible instantaneous power levels is related to the permissible range of $\beta(t)$. The linear controller exerts increasing control over devices as their temperatures deviate more from $T_{\max}$. The range of β is constrained by those values at which all devices at $T_{\text{low}}(t)$ (the coldest appliances at time t) are forced to be either on or off. This range is illustrated in Figure 7. In particular, the minimum and maximum values of $\beta(t)$ are given by the solutions of

$$v_{\text{off}}(T_{\text{low}}(t)) = \alpha\beta_{\min}(t)(T_{\text{low}}(t) - T_{\max}) \tag{42a}$$
$$v_{\text{on}}(T_{\text{low}}(t)) = \alpha\beta_{\max}(t)(T_{\text{low}}(t) - T_{\max}). \tag{42b}$$

This results in the following constraint for $\beta(t)$:

$$-\frac{T_{\text{off}} - T_{\text{low}}(t)}{T_{\max} - T_{\text{low}}(t)} \leq \beta(t) \leq \frac{T_{\text{low}}(t) - T_{\text{on}}}{T_{\max} - T_{\text{low}}(t)}. \tag{43}$$

This constraint on the control parameter $\beta(t)$ can be transformed into a constraint on instantaneous power consumption by using expression (36) to substitute $\beta(t)$ and using the scaling solutions

$$\bar{T}(t) = T_{\max} - (T_{\max} - \bar{T}_0)S(t) \tag{44}$$
$$T_{\text{low}}(t) = T_{\max} - (T_{\max} - T_{\min})S(t) \tag{45}$$

to eliminate $T_{\text{low}}(t)$ and $\bar{T}(t)$. Surprisingly this results in constant lower and upper bounds for the normalised power level $\Pi(t)$:

$$\frac{(T_{\text{off}} - T_{\max})(\bar{T}_0 - T_{\min})}{(T_{\max} - T_{\min})(T_{\text{off}} - \bar{T}_0)} \leq \Pi(t) \leq \frac{(T_{\max} - \bar{T}_0)(T_{\max} - T_{\text{on}}) + (T_{\text{off}} - T_{\max})(T_{\max} - T_{\min})}{(T_{\max} - T_{\min})(T_{\text{off}} - \bar{T}_0)}. \tag{46}$$

Provided that this power constraint and the temperature (energy) constraint (41) are satisfied, it is guaranteed that a particular appliance can track the reference power profile $\Pi(t)$ in terms of its expected power consumption. The important implication is that there is no need to perform detailed device-level simulations to ascertain whether a certain response is feasible.

The parameter values from Table I result in a duty cycle $\pi_0 = 0.24$ and average temperature $\bar{T}_0 = 4.59°C$. The corresponding temperature and power constraints for the linear controller are

$$4.59°C \leq \bar{T}(t) \leq 7.00°C$$
$$0.84 \leq \Pi_{\text{sustained}} \leq 1$$
$$0.44 \leq \Pi(t) \leq 2.44.$$

E. Appliance switching rates

The solution of the ensemble control problem, represented by the control parameter $\beta(t)$, gives rise to a device-level control strategy through the temperature thresholds $T_{\max}$, $T_{\text{low}}(t)$ and the switching rates $r^{\text{on}}_{\text{off}}(T,t)$ and $r^{\text{off}}_{\text{on}}(T,t)$. The switching rates

follow from the switching flux $\Phi(T,t)$. Substituting the linear heating/cooling rates defined in Eq. (19) into Eq. (15) yields

$$\begin{aligned}\Phi(T,t) =&\frac{\partial}{\partial t}\left[\left(\frac{v(T,t)/\alpha + T - \frac{1}{2}(T_{\text{off}}+T_{\text{on}})}{T_{\text{off}}-T_{\text{on}}}\right) f(T,t)\right]\\ &+\frac{\partial}{\partial T}\left[\left(\frac{\frac{1}{2}v(T,t)(T_{\text{off}}+T_{\text{on}}-2T)}{T_{\text{off}}-T_{\text{on}}}\right) f(T,t)\right]\\ &-\alpha\frac{\partial}{\partial T}\left[\left(\frac{(T_{\text{off}}-T)(T_{\text{on}}-T)}{T_{\text{off}}-T_{\text{on}}}\right) f(T,t)\right]. \quad (47)\end{aligned}$$

Defining

$$\Xi(T,t) = \frac{(v_{\text{off}}(T)-v_{\text{on}}(T))\Phi(T,t)}{f(T,t)}, \quad (48)$$

allows Eqs. (16) to be rewritten as

$$r_{\text{off}}^{\text{on}}(T,t) = \max\left(0, \frac{\Xi(T,t)}{v_{\text{off}}(T)-v(T,t)}\right) \quad (49a)$$

$$r_{\text{on}}^{\text{off}}(T,t) = \max\left(0, \frac{\Xi(T,t)}{v_{\text{on}}(T)-v(T,t)}\right). \quad (49b)$$

In order to evaluate $\Xi(T,t)$ the following shorthand notation is introduced for relative and linearly scaled temperatures.

$$\begin{aligned}\tau_{\text{[label]}} &= T - T_{\text{[label]}}\\ \hat{\tau}_{\text{[label]}} &= T - T_{\max} + (T_{\max} - T_{\text{[label]}})\mathrm{e}^{\int_0^t \alpha\beta(t')\,\mathrm{d}t'}\end{aligned}$$

In terms of these,

$$\begin{aligned}v(T,t) =&\alpha\beta(t)\tau_{\max}\\ v_{\text{off}}(T) =&-\alpha\tau_{\text{off}}\\ v_{\text{on}}(T) =&-\alpha\tau_{\text{on}}\\ f(T,t) =&\frac{-CS(t)}{\hat{\tau}_{\text{on}}\hat{\tau}_{\text{off}}}.\end{aligned}$$

The derivatives in (47) can be evaluated using (8) to eliminate the $\partial f(T,t)/\partial t$ term and the fact that

$$\frac{\partial f(T,t)}{\partial T} = -\left(\frac{\hat{\tau}_{\text{off}}+\hat{\tau}_{\text{on}}}{\hat{\tau}_{\text{off}}\hat{\tau}_{\text{on}}}\right) f(T,t). \quad (50)$$

This results in

$$\begin{aligned}\Xi(T,t) =&\alpha\tau_{\max}\frac{\mathrm{d}\beta(t)}{\mathrm{d}t} + \alpha^2\left(\frac{\hat{\tau}_{\text{off}}+\hat{\tau}_{\text{on}}}{\hat{\tau}_{\text{off}}\hat{\tau}_{\text{on}}}\right)\\ &\times(\tau_{\text{off}}+\beta(t)\tau_{\max})(\tau_{\text{on}}+\beta(t)\tau_{\max})\\ &-\alpha^2(1+\beta(t))(\tau_{\text{off}}+\tau_{\text{on}}+\beta(t)\tau_{\max}) \quad (51)\end{aligned}$$

Rather than computing $\mathrm{d}\beta(t)/\mathrm{d}t$ using a numerical approximation, we may also use (36) and $\mathrm{d}\bar{T}(t)/\mathrm{d}t = \bar{v}(t) = \alpha\beta(t)(\bar{T}(t)-T_{\max})$ to restate it as follows

$$\begin{aligned}\frac{\mathrm{d}\beta(t)}{\mathrm{d}t} =&\frac{1}{T_{\max}-\bar{T}(t)}\left[\frac{\mathrm{d}\Pi(t)}{\mathrm{d}t}(T_{\text{off}}-\bar{T}_0)\right.\\ &\left.+\alpha\beta(t)\left(T_{\text{off}}-T_{\max}-\Pi(t)(T_{\text{off}}-\bar{T}_0)\right)\right] \quad (52)\end{aligned}$$

In summary, for a given reference power profile $\Pi(t)$ and control parameter profile $\beta(t)$, Eqs. (49)-(52) provide an analytical expression for the computation of the stochastic switching rates $r_{\text{off}}^{\text{on}}(T,t)$ and $r_{\text{on}}^{\text{off}}(T,t)$. Together with the upper temperature bound limit $T_{\max}$ and the lower temperature bound $T_{\text{low}}(t)$ as defined in (35) this fully specifies the device controller.

V. Implementation

Simulations of individual device actions are used to illustrate the results obtained in the previous sections. The behaviour of N appliances is simulated, optionally with different parameters for each device. The default parameters are listed in Table I. The devices are controlled by a common reference power curve $\Pi(t)$. Recall that $\Pi(t)=1$ represents steady state power consumption, and each response action must start and end at this value.

For simplicity, the appliances are assumed to operate in 'fire and forget' mode: they aim to deliver a predetermined power response $\Pi(t)$ that is triggered by an initial frequency event. This response is not adjusted dynamically in response to changes in the power system frequency, so the full trajectory of the control parameter $\beta(t)$ can be computed in advance. The more general case where $\beta(t)$ is adjusted on the fly to track a dynamic power profile $\Pi(t,\Delta\text{freq})$ does not lead to a fundamentally different control strategy, but would require a more elaborate algorithm for simulation.

Table I
Model parameters for simulation

T_{on}	$-44°C$
T_{off}	$20°C$
$T_{\max}$	$7°C$
$T_{\min}$	$2°C$
α	$1.37\times10^{-4}s^{-1}$

Values for the temperature thresholds and room temperature are taken from [7] (second order fridge model). The values for T_{on} and α were fitted to reproduce the duty cycle (0.24) and steady state cycle duration (52 minutes) of the second order model with an approximate first order model.

A. Algorithm

The switching actions and times for each appliance are computed using an event-driven kinetic Monte Carlo method. In the algorithm description the index a is used to indicate quantities and functions specific to appliance a. The appliances are initialised at time t_0^a with temperature T_0^a and state s_0^a according to the steady state distribution. Subsequent switching events to states s_j^a occur at times t_j^a and associated temperatures T_j^a, where $j=1,2,\ldots$. The algorithm also makes use of the function $\theta^a(t|T',t',s')$ (the solution of Eq. (4)) that reflects the temperature of appliance a at time t, starting from temperature T' and state s' at time $t'\le t$.

1) For each appliance $a\in\{1,\ldots,N\}$
   a) Compute the appliance duty cycle π_0^a using (33).
   b) Compute the steady state average temperature $\bar{T}_0^a$ using (23).
   c) Verify that the reference power profile $\Pi(t)$ satisfies the energy constraints (41) and the power constraints (46). If not, abort and readjust $\Pi(t)$ (see section V-C for more information).
   d) Compute the average temperature curve $\bar{T}^a(t)$ using (26).
   e) Compute the control parameter $\beta^a(t)$ using (36).
   f) Compute $r_{\text{off}}^{\text{on},a}(T,t)$ and $r_{\text{on}}^{\text{off},a}(T,t)$ using (49) and (51).

g) Determine initial state of the appliance
- Set $t_0^a = t_{\text{start}}$
- Randomly select a state $s_0^a \in \{\text{off}, \text{on}\}$, using the steady state duty cycle π_0^a as the probability of selecting the *on* state.
- Randomly select an appliance temperature T_0^a according to the steady state distributions $f_{\text{on}}^a(T, 0)$, $f_{\text{off}}^a(T, 0)$, using Eqs. (31) and (14) and an accept-reject sampling procedure.

h) Store $\{t_0^a, s_0^a, T_0^a\}$.

i) Set $j = 0$

j) While $t_j^a < t_{\max}$, do

i) If $s_j^a = \text{off}$

A) Compute the upper threshold switching time t_{on}^d by solving $\theta^a(t_{\text{on}}^d | T_j^a, t_j^a, s_j^a) = T_{\max}^a$ with (20).

B) Compute $R(t) = \int_{t_j^a}^{t} r_{\text{on}}^{\text{off},a}\left(\theta^a(t' | T_j^a, t_j^a, s_j^a), t'\right) \mathrm{d}t'$ for $t \in [t_j^a, t_{\text{on}}^d]$ using (49b) and (51).

C) Draw a uniform random number $u \in (0, 1)$

D) If $R(t_{\text{on}}^d) < -\log(u)$
- Set $t_{j+1}^a = t_{\text{on}}^d$.
- Set $T_{j+1}^a = T_{\max}$

else
- Solve $R(t_{\text{on}}^s) = -\log(u)$ for t_{on}^s.
- Set $t_{j+1}^a = t_{\text{on}}^s$.
- Set $T_{j+1}^a = \theta^a(t_{\text{on}}^s | T_j^a, t_j^a, s_j^a)$

E) Set $s_{j+1}^a = \text{on}$

Else (if $s_j^a = \text{on}$)

A) Compute the lower threshold switching time t_{off}^d by solving $\theta^a(t_{\text{on}}^d | T_j^a, t_j^a, s_j^a) = T_{\text{low}}(t_{\text{off}}^d)$ with (20) and (35).

B) Compute $R(t) = \int_{t_j^a}^{t} r_{\text{off}}^{\text{on},a}(\theta^a(t' | T_j^a, t_j^a, s_j^a), t') \mathrm{d}t'$ for $t \in [t_j^a, t_{\text{off}}^d]$ using (49a) and (51).

C) Draw a uniform random number $u \in (0, 1)$

D) If $R(t_{\text{off}}^d) < -\log(u)$
- Set $t_{j+1}^a = t_{\text{off}}^d$
- Set $T_{j+1}^a = T_{\text{low}}(t_{\text{off}}^d)$

else
- Solve $R(t_{\text{off}}^s) = -\log(u)$ for t_{off}^s
- Set $t_{j+1}^a = t_{\text{off}}^s$
- Set $T_{j+1}^a = \theta^a(t_{\text{off}}^s | T_j^a, t_j^a, s_j^a)$

E) Set $s_{j+1}^a = \text{off}$

ii) Store $\{t_{j+1}^a, s_{j+1}^a, T_{j+1}^a\}$.

iii) Increment $j \to j + 1$.

k) Use P_{on}^a and the stored series $\{t_i^a\}$, $\{s_i^a\}$ and $\{T_i^a\}$ to reconstruct the appliance power and temperature curves $P^a(t)$ and $T^a(t)$.

2) Compute the empirical normalised power curve $\hat{\Pi}(t) = (\sum_a P^a(t))/(\sum_a \pi_0^a P_{\text{on}}^a)$.

B. Simulation results

Figures 8-10 show results obtained using device-level simulations. Figure 8 (top) shows the on-off sequences of 20 individual appliances using the parameters from Table I. They were controlled using the piecewise linear reference power curve shown below (dashed; identical to the curve in Figure 5). It is evident that after the initial load reduction (linear reduction between $t = 0s$ and $t = 10s$) devices are less likely to be in the *on* state, reducing the aggregate power consumption. This situation is reversed during the payback phase, with devices switching back to the the *on* state. The empirical aggregate power profile generated by $N = 1000$ appliances (solid) is overlaid on the reference power profile for comparison.

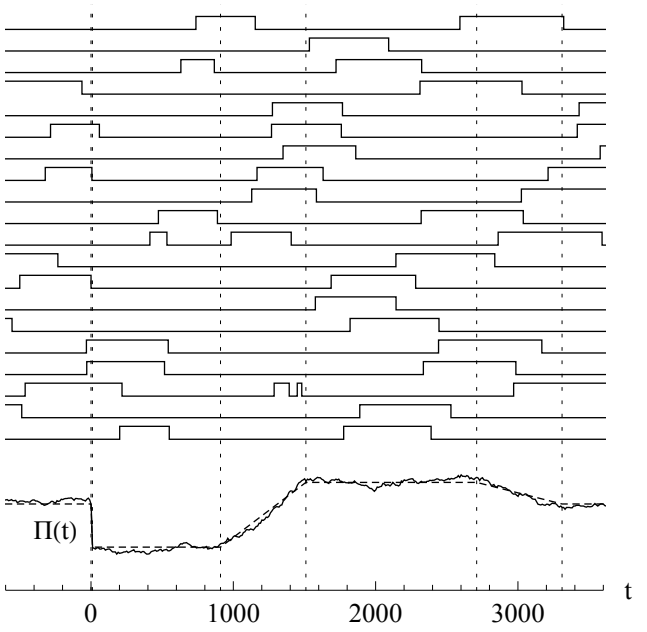

Figure 8. On/off status for 20 individual appliances (top; high=on, low=off), shown alongside the reference power profile (bottom; dashed) and the empirical normalised power profile of $N = 1000$ appliances (bottom; solid curve). For reference, kinks in the reference power profile are marked by dotted vertical lines.

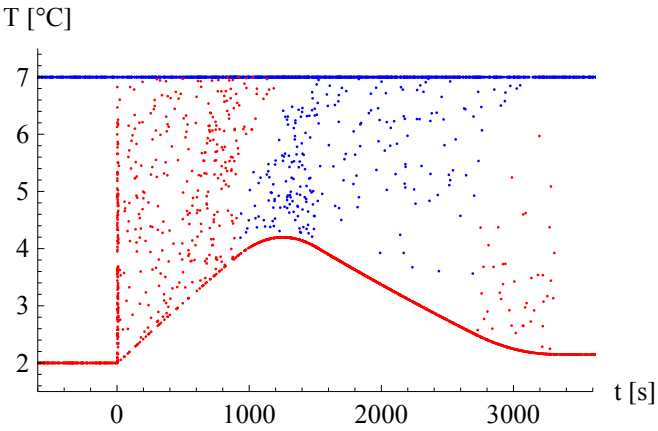

Figure 9. Graphical representation of switching behaviour of 1000 appliances for the reference power curve shown in Fig. 8. Red dots represent individual devices switching off (heating); blue dots represent devices switching on (cooling).

Figure 9 offers a global view at the switching process. In this temperature-time plot each dot represents a switching event: blue for fridges switching on (cooling) and red for fridges

switching off (heating). In the steady state ($t < 0$) devices only toggle their power state when they reach the temperature thresholds $T_{\min}$ and $T_{\max}$. After the trigger event at $t = 0$, the lower temperature setpoint $T_{\text{low}}(t)$ increases gradually before eventually returning to $T_{\min}$. During the control action a large number of appliances is seen to undergo stochastic switching at intermediate temperatures. This ensemble result may be compared to the temperature trace of a single appliance in Figure 2.

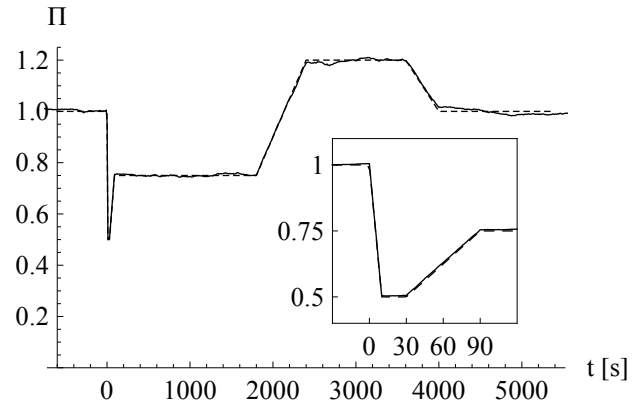

Figure 10. Aggregate power consumption of 100,000 refrigerators with randomised parameters. The reference power curve drops to 50% of nominal capacity within 10s after the trigger at $t = 0$ and remains at that level for 20 seconds (primary response), followed by another 30 minutes at 75% (secondary response). The borrowing phase is followed by a payback phase at 120% of nominal capacity.

Figure 10 presents the results from a larger study with 100,000 appliances and a different reference power profile. This study illustrates three distinct properties of the proposed controller. First, as the number of appliances is increased from 1,000 to 100,000, the statistical fluctuations around the reference power level decrease as expected. Second, the example demonstrates the potential for using more complex power curves to provide flexible services to the power grid. The durations of the first two power plateaus correspond to the classification of frequency response services by National Grid (Great Britain's Transmission System Operator): primary response (ready in 10s, maintain until 30s) and secondary response (ready in 30s, maintain for 30 mins) [16]. In this case, the controller reduces the power consumption to 50% of the nominal amount as primary response and to 75% as secondary response. Third, the results in figure 10 demonstrate that a coherent response does not require the use of identical appliances. For this simulation, the $T_{\min}$, $T_{\max}$ and α parameters for each appliance were based on the values in table I, but uniformly random multipliers between 0.8 and 1.2 were applied independently for each parameter and each device.

C. Heterogeneous appliances

The algorithm presented in section V-A starts by verifying the feasibility of the requested power profile for each appliance individually, and rejects the profile if it is not compatible with the appliance's quality of service constraints. In a well-designed implementation such a situation should never occur, as it is the responsibility of the demand response designer to ensure compatibility with all devices under its control *before* broadcasting the desired response. The designer could perform this task by comparing a proposed response with a database of appliance models, but the construction and maintenance of such a database is no trivial task. Moreover, this approach potentially results in a very large number of constraints that must be verified.

As an alternative we propose to use a simpler sufficient condition that aggregates the abilities of all appliances into a single flexibility envelope. It is clear that the power constraint (46) will be satisfied for each appliance if and only if

$$\max_a(\Pi^a_{\min}) \le \Pi(t) \le \min_a(\Pi^a_{\max}), \tag{53}$$

where $[\Pi^a_{\min}, \Pi^a_{\max}]$ represents the range of accessible power levels for appliance a. Furthermore, the energy constraint (41) for each appliance is characterised by the minimum sustained relative power level $\Pi^a_{\text{sustained,min}} = 1 - (T^a_{\max} - \bar{T}^a_0)/(T^a_{\text{off}} - \bar{T}^a_0)$ and the thermal relaxation parameter α^a. Equation (41) is satisfied for all appliances if, for all t,

$$\max_a(\Pi^a_{\text{sustained,min}}) \le \alpha_{\max} \int_{-\infty}^{t} \Pi(t')\mathrm{e}^{-\alpha_{\max}(t-t')}\,\mathrm{d}t' \le 1, \tag{54}$$

where

$$\alpha_{\max} = \max_a(\alpha^a) \tag{55}$$

reflects the appliance with the lowest thermal time constant. Note that equation (54) is a sufficient but not a necessary condition for feasibility, because the appliance with the highest value of $\Pi^a_{\text{sustained,min}}$ is not necessarily the one with the highest value of α^a.

Using these envelope constraints the response designer needs only four aggregate parameters to guarantee the feasibility of a solution: the upper and lower instantaneous power limits (53), the lower sustained power limit (54) and the highest temperature relaxation constant (55). These four parameters can be obtained either through direct communication with the appliances or through intermediate aggregators. In the example used for Fig. (10) the parameters $T_{\min}$, $T_{\max}$ and α were varied through multiplication by random factors in the range $[0.8, 1.2]$, resulting in the aggregate bounds

$$\begin{aligned} 0.90 \le \Pi_{\text{sustained}} \le 1 \\ 0.46 \le \Pi(t) \le 2.36 \\ \alpha_{\max} = 1.64 \times 10^{-4} s^{-1}. \end{aligned} \tag{56}$$

In this case, the lower bounds for instantaneous and sustained power result from the appliance with the highest $T_{\min}$ and lowest $T_{\max}$; the upper instantaneous power limit is associated with the device with the lowest values of $T_{\min}$ and $T_{\max}$.

As explained in section II-A it is not desirable to broadcast the relative power curve $\Pi(t)$ directly to the devices in real time. Instead a power response model is supplied that enables the devices to compute the desired relative power curve on the fly from local observables. The designer should therefore verify that no valid combination of input parameters can result in a violation of the energy and power constraints of the appliances under its control.

VI. Summary and discussion

A. Control framework

In this paper we have introduced a novel control framework for controlling the aggregate power consumption of thermostatically controlled loads (TCLs). It controls each appliance independently in such a way that its *expected* power consumption tracks a relative power consumption profile $\Pi(t)$. For a large number of appliances the statistical fluctuations will average out and the aggregate power consumption approaches a well-defined (diversified) value $P_0^{\mathrm{total}}\Pi(t)$. The profile $\Pi(t)$ is limited only by the physical properties of the appliances and their quality of service requirements.

The control framework permits a straightforward decentralised implementation. This can be used either for fully autonomous operation, which may be desired for primary frequency response, or semi-autonomous operation, which allows for adjustment of the devices' response to changing power system requirements. In the latter case the demand response designer (i.e. the system operator or a demand response aggregator) periodically defines a power response model that is broadcast to the devices. This power response model instructs appliances how to construct a target profile $\Pi(t)$ as a function of locally available global signals such as the grid frequency or the time. In a typical application the grid frequency may be used to trigger a frequency response action, or a clock signal may be used to trigger services at particular times. Crucially, this semi-autonomous mode of operation does not require a low-latency communication channel, because actions on short time scales are determined locally by the embedded controllers. The power response models may therefore be updated over a high-latency channel, such as a smart meter infrastructure that is built with metering as its primary purpose.

Central to the control approach in our work is the ability to accurately control the expected power consumption of an appliance. This allows for the precise shaping of the aggregate demand response, matching it to the system's requirements. The degree of control this requires is made possible by expressing the state of an appliance as a probability distribution on the temperature axis. In contrast with existing methods in the literature we have inverted the resulting control problem by taking a collective property, the net heating rate $v(T,t)$, as a control parameter. The on-off switching rates and threshold temperatures - the parameters that physically control the appliance - are then derived from $v(T,t)$ instead of the other way around. This choice enables the separation of the control problem into two stages: a first stage in which a suitable heating rate profile is determined that will result in the desired ensemble power consumption $\Pi(t)$, and a second stage in which this solution is translated to a controller for an individual appliance.

The controller in each appliance makes use of a first order thermal model for that appliance to solve both control problem stages. Such a model could be provided at the factory, for example by embedding an appropriate thermal model for the particular make and model of appliance. However it is also conceivable that smart devices tune their own internal model through measurements and thus adapt to their installation environment and usage patterns. In the case of a refrigerator, apart from the regular temperature thresholds this could include the heating/cooling rate (affected by the heat capacity of its contents) and ambient temperature.

As explained in section V-C, the system operator can guarantee the controllability of a heterogeneous group of appliances as a single entity by using a 'lowest common denominator' envelope model for their collective flexibility. In the case of the linear controller, this flexibility envelope is characterised by four parameters. Together with the average aggregate power consumption P_0 these can be used to design a feasible demand response pattern. Note that the linear controller is quite restrictive in its inability to absorb energy from the grid, resulting in an upper bound for the sustained power that is equal to the steady state value. Generalisations of this controller can relax this constraint, resulting in an additional flexibility parameter representing the upper bound. Note also that while it may be possible to control highly diverse TCLs as a single group this may severely restrict the services that can be delivered. In practice it will often be advantageous to define control clusters consisting of similar appliances so that demand response actions can be adapted to the characteristics of the appliances in each cluster.

B. Linear thermal model

In section III we have considered generic limits to flexibility resulting from the physical characteristics of TCLs and their quality of service constraints. This analysis has been applied to the case of TCLs with first order linear thermal models. For this simple but common model it is possible to determine the evolution of the population-averaged temperature from the aggregate power consumption alone - regardless of the specifics of the controller. Temperature limits for individual appliances can thus be translated into energy and power limits for the aggregate dynamic power response of thermostatic loads.

The analysis has also demonstrated that a so-called 'payback' period of increased power consumption is not strictly required after a period of reduced power consumption: operating at the steady-state power level will eventually restore the steady-state temperature. However, in practical applications it will often be desirable to include a payback period in order to speed up the return of the average temperature to its steady state value. The two stage control approach allows the system operator to optimally schedule the payback profile in accordance with the system's characteristics.

C. Linear controller

In section IV we have introduced a specific controller design where the heating profile $v(T,t)$ is a linear function of T. This functional form results in elegant scaling expressions for the temperature distributions and enables the derivation of analytical expressions for both controller stages. Moreover, the controller guarantees that no single device ever exceeds the temperature range $[T_{\min}, T_{\max}]$.

Power and energy limits have been derived specifically for this controller. A comparison with the generic bounds for linear thermal models shows that the linear controller is optimal with respect to energy provision for load reduction. However,

the controller is unable to absorb excess energy from the grid (for high-frequency services) nor can it reduce power consumption to zero. Both limitations can be overcome with more complex controller implementations based on the same framework, but the linear model sacrifices such flexibility for conceptual and notational clarity. In addition, the existence of analytical expressions for the power and energy limitations is beneficial because it allows the demand response designer to rapidly check whether TCLs can deliver a proposed response curve. This allows for a straightforward embedding of the constraints in the response design process, which typically takes the form of an optimisation problem.

e have provided an algorithm for the resulting device-level controller that does not require advanced computational methods other than numerical integration, so it can be implemented in relatively unsophisticated tools and appliances. The algorithm has been used for simulations with up to 100,000 appliances, demonstrating the ability of the controller to accurately track a reference power profile. The presented implementation makes the simplifying assumption that the desired power response $\Pi(t)$ is known in advance, but it is easily generalised to cases where $\Pi(t)$ is computed dynamically from locally available input signals.

D. Extensions and improvements

The simple linear controller that has been used to illustrate the decentralised control framework for TCLs may serve as a basis for the development of increasingly sophisticated controllers. For example, it may be extended to permit a larger range of instantaneous power levels, or to enable the absorption of excess energy from the grid (high-frequency services). Extensions - perhaps approximate - to appliances with second or higher-order thermal models may also be considered. Furthermore, although the current implementation minimises the overall switching rate of appliances, it does not strictly minimise the time between switching events for individual appliances. Because the switching in between the threshold temperatures is stochastic, an appliance may occasionally be requested to cycle more than once in a short period, and it may be unable to comply in practice if this conflicts with the compressor lockout requirements. The method described here may be expanded to allocate switching events preferentially to devices that have not recently switched and, if the lockout period is strictly enforced, slight deviations from the desired curve may need to be tolerated.

The control framework is able to accurately modulate aggregate power consumption across a range of time scales. In addition to the provision of frequency response, this property also makes it suitable for the provision of energy arbitrage services, where TCLs are used to relax constraints on the dispatch of generators [17]. TCLs can reduce their power consumption at times when the generation mix is very expensive or carbon-intensive, and shift their demand to adjacent time periods. Of course, when the high and low pricing periods are many hours apart (e.g. night and day), the thermal storage capacity of the TCLs may not be sufficient to enable a significant load transfer between these periods. If that is the case, the TCLs' flexibility may still be used to alleviate ramping constraints between such periods. The ability to specify an arbitrary relative power curve $\Pi(t)$ makes it possible to deliver both frequency and energy arbitrage services simultaneously, so long as their simultaneous provision does not exceed the physical constraints of the appliances. In such a scenario, the TCLs would commit to a certain power consumption pattern for energy arbitrage and *simultaneously* provide agreed frequency services that are activated only when a frequency drop is observed. The appliance's power response model would contain both components and should be designed in such a way that the appliance always maintains sufficient operating margins to provide the contracted frequency services if the need arises.

Finally, this paper has not attempted to determine the optimal choice of $\Pi(t)$ (and therefore the power response model). This optimal choice is the result of an interplay between technical and commercial aims, and the limits of the appliances and their controller. Determining the optimal power profile $\Pi(t)$ is a first step to determining the value of demand response services, and will be the subject of future work.

VII. Acknowledgements

The authors thank Enrique Ortega for enlightening discussions, and thank Paola Falugi and the anonymous reviewers for their helpful comments on the manuscript. Vincenzo Trovato is grateful to National Grid for supporting his PhD research.

References

[1] F. Schweppe, "Frequency adaptive, power-energy re-scheduler," U.S. Patent 4,317,049, 1979.

[2] J. A. Short, D. G. Infield, and L. L. Freris, "Stabilization of Grid Frequency Through Dynamic Demand Control," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1284–1293, Aug. 2007.

[3] M. Aunedi, P.-A. Kountouriotis, J. E. O. Calderon, D. Angeli, and G. Strbac, "Economic and Environmental Benefits of Dynamic Demand in Providing Frequency Regulation," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2013.

[4] S. Ihara and F. Schweppe, "Physically Based Modeling of Cold Load Pickup," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 9, pp. 4142–4150, Sep. 1981.

[5] D. Angeli and P.-A. Kountouriotis, "A Stochastic Approach to ?Dynamic-Demand? Refrigerator Control," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 3, pp. 581–592, May 2012.

[6] D. Hirst, "Responsive Load System," GB Patent 2 361 118 A, 2000.

[7] V. Trovato, S. H. Tindemans, and G. Strbac, "Controlling the synchronization and payback associated with the provision of frequency services by dynamic demand," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*. Stockholm: Institution of Engineering and Technology, 2013, p. 435.

[8] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, May 2009.

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

[10] C. Perfumo, E. Kofman, J. H. Braslavsky, and J. K. Ward, "Load management: Model-based control of aggregate power for populations of thermostatically controlled loads," *Energy Conversion and Management*, vol. 55, no. null, pp. 36–48, Mar. 2012.

[11] J. L. Mathieu, S. Koch, and D. S. Callaway, "State Estimation and Control of Electric Loads to Manage Real-Time Energy Imbalance," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 430–440, Feb. 2013.

[12] M. Kamgarpour, C. Ellen, S. E. Zadeh Soudjani, S. Gerwinn, J. L. Mathieu, M. Nils, A. Abate, D. S. Callaway, M. Fränzle, and J. Lygeros, "Modeling Options for Demand Side Participation of Thermostatically Controlled Loads," in *2013 IREP Symposium*, 2013, pp. 1–15.

[13] N. A. Sinitsyn, S. Kundu, and S. Backhaus, "Safe protocols for generating power pulses with heterogeneous populations of thermostatically controlled loads," *Energy Conversion and Management*, vol. 67, pp. 297–308, Mar. 2013.

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

[15] C. Chong and A. Debs, "Statistical synthesis of power system functional load models," in *1979 18th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol. 18. IEEE, 1979, pp. 264–269.

[16] National Grid, "Grid Code." [Online]. Available: http://www.nationalgrid.com/uk/Electricity/Codes/gridcode/

[17] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway, "Energy Arbitrage with Thermostatically Controlled Loads," in *Proceedings of the European Control Conference*, Zurich, Switzerland, 2013.