A Group-based Approach for Heterogeneity in Packetized Energy Management*

Adil Khurram[1], Luis Duffaut Espinosa[2] and Mads Almassalkhi[2].

Abstract— **In practice, fleets of DERs are inherently heterogeneous due to manufacturers' specifications and the effects of wear and tear. Accounting for heterogeneity is critical in the design of control policies for DER aggregations, otherwise, the system response may be inaccurate and performance can be degraded. This paper presents a group-based approach to characterize parametric heterogeneity in a fleet of aggregated DERs by grouping into homogeneous fleets. The proposed group-based approach borrows concepts from quantization in the area of signal processing and the paper particularly highlights the effect of rated power heterogeneity due to its relevance to the control policy design in packetized energy management (PEM). The reason for this is that the effective control mechanism in PEM is a function of the aggregate DER demand and the rated power of devices that are switched on. As a result, the PEM system is susceptible to tracking errors that may degrade performance in a heterogeneous fleet. Therefore, the proposed quantization-based approach provides a systematic approach to group DERs in the fleet so that the desired performance is achieved.**

I. INTRODUCTION

The increasing penetration of intermittent, renewable generation has highlighted the importance of an expanded role for real-time demand-side management [1]. Demand-side management via direct load control provides fast time-scale and predictable control opportunities as opposed to price-responsive approaches [2]. Distributed energy resources (DERs), such as thermostatically controlled loads (TCLs), energy storage systems (ESSs), and plug-in electric vehicles are suitable for demand dispatch [3] since their energy states can be manipulated [4]–[6]. However, randomization is helpful to limit synchronization in aggregate and intelligence at the local, load control layer ensures that the end-user remains unaffected by demand dispatch [6].

Fundamentally, a fleet of DERs represents a collection of heterogeneous agents. This means that the fleet's distribution of dynamic states does not change uniformly as is the case for a homogeneous population. This heterogeneity arises because DERs belong to different classes, such as TCLs and ESSs, have different manufacturer specifications, rates of wear-and-tear, and diverse end-user behaviors. Nonetheless, heterogeneity can be beneficial in the sense that it limits synchronization of the population by ensuring population mixing. In schemes where the effective control mechanism is a function of the DERs' rated powers, heterogeneity can be considered differently. As such, this work focuses on heterogeneity (e.g., model uncertainty) in the form of the rated power of a DER, which extends earlier works to heterogeneous fleets that were derived under homogeneous assumptions, from the viewpoint of the coordinator. Specifically, the results presented in this paper enhance the device-driven, demand dispatch scheme called packetized energy management (PEM). PEM essentially metes out a DER's contiguous period of constant demand into multiple, shorter epochs of demand, analogous to data transfer in packet-switched networks [7]. State bin transition models of PEM have been developed to capture the aggregated dynamics of DERs under PEM [8], [9] for homogeneous DERs and their nominal behavior of PEM is defined in [10].

Heterogeneity has been studied in the literature [11], [12]. The transition probabilities in state-bin transition models of TCLs with heterogeneous time constants are analytically derived in [11] and are shown to be identical to those identified from simulations of TCL populations. The myopic DER demand dispatch scheme of [6] is extended for heterogeneous load in [13] and the corresponding state estimation techniques are presented in [12]. The authors of [13] further show that the heterogeneity that helps with limiting synchronization comes at the cost of reduced capacity in the aggregate response. A control strategy for the aggregate population of air conditioning loads that assumes predefined heterogeneity in the load was presented in [14], [15]. The authors in [16] identified the sources of error in an aggregated model and studied the effect of noise and parametric heterogeneity. An empirical scheme to cluster heterogeneous groups of similar dynamics into homogeneous groups is presented in [17]. These approaches consider heterogeneity by grouping together the TCLs with similar parameters, however, a specific procedure to bound the modeling errors due to grouping has not been provided.

Consider a PEM coordinator managing $5,000$ TCLs with heterogeneity in rated powers only. The population is grouped based on their rated powers similar to parametric binning in [16]. The comparison between the fully heterogeneous population and the grouped population is shown in Figure 1. As the number of groups increases, the error between the actual output and approximated output decreases. Groups can therefore be used to capture the dynamics of heterogeneous fleets. However, the number of groups also affects the nature of the TCL population models and leads to aggregation errors at the PEM coordinator, which amounts to the approach becoming (1) computationally expensive as the number of groups increases, and (2) inaccurate due to

*This work was supported by the U.S. Department of Energy's Advanced Research Projects Agency—Energy (ARPA-E) Award DE-AR0000694. M. Almassalkhi is co-founder of startup Packetized Energy, which seeks to bring to market a commercially viable version of Packetized Energy Management.

[1]Adil Khurram is with Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093 USA akhurram@ucsd.edu.

[2]Luis Duffaut Espinosa and Mads Almassalkhi are with the Department of Electrical and Biomedical Engineering, University of Vermont, Burlington, VT 05405 USA {lduffaut, malmassa}@uvm.edu

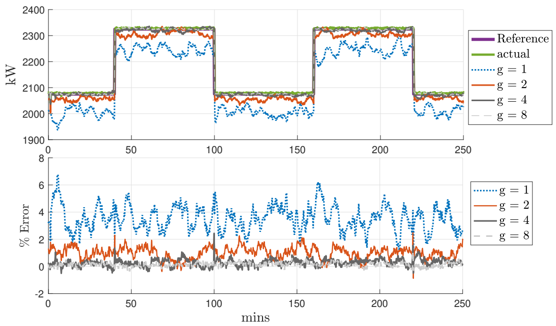

Fig. 1. TCL fleet consisting of $5,000$ devices with rated power in the range $[4, 5]$ kW divided into $g = \{1, 2, 4, 8\}$ groups at the coordinator. As the number of groups increases, the tracking error decreases. (Top) The total power demand (P_{dem}) kW (Botton) %Tracking Error $= \frac{P_{\text{ref}} - P_{\text{dem}}}{P_{\text{ref}}}$.

modeling error as the number of devices in each group decreases. Modeling the error in the state-bin transition model has been explored in [9].

Heterogeneous populations are generally characterized by the time constant of the linear system as in [11]. However, in PEM only the rated power associated with each request is visible to the coordinator by design and lends itself to analytical simplification at the coordinator's control layer. Therefore, this work develops a procedure to incorporate heterogeneity in the rated power from the viewpoint of the coordinator. The unique contributions of this paper are:

1) Heterogeneity is characterized in terms of the modeling error introduced due to grouping together DERs that are different in parameters including rated power for a nominal PEM system.
2) Conditions are provided under which the error is normally distributed and its mean and standard deviation is quantified.
3) A formula for the appropriate number of groups is given as a function of an upper bound on modeling error and the concept of nominal PEM response for a homogeneous fleet is extended to heterogeneous fleets.

This paper is organized as follows. Section II provides a summary of packetized energy management (PEM), macro-models and steady-state behavior. In section III, quantization-based heterogeneous model is presented. Section IV develops conditions for selecting the appropriate number of groups to accurately capture heterogeneity. Each section is illustrated with corresponding examples and simulations. Conclusions are given in the last section.

II. Preliminaries

In this section, background regarding PEM and the modeling of its aggregated behavior are briefly described. The focus of this research work is on TCLs and more precisely electric water heaters. Detailed descriptions are in [7]–[9].

A. DER coordination under PEM

The dynamic model for the i-th DER of a fleet of TCLs with energy state z_i at time k is

$$z_i[k+1] = \left(1 - \frac{\Delta t}{\tau}\right) z_i[k] + \frac{\Delta t z_a}{\tau} + \frac{\Delta t P_i^{\text{rate}} u_i[k]}{cdL\eta} - \frac{\Delta t w_i[k]}{cdL}, \quad (1)$$

where Δt is the discretization time step, $c = 4.186 \ \frac{\text{kJ}}{\text{kg}^\circ\text{C}}$ is the specific heat constant, z_{a} is the ambient temperature, $\tau = 150 \times 3600$ seconds is the standing loss time constant to ambient temperature and $d = 0.990 \ \frac{\text{kg}}{\text{L}}$ is the density of water when close to $50\ ^\circ C$. Furthermore, P_i^{rate} is the rated power, L is the tank size, η is the heat transfer efficiency, $u_i \in \{0, 1\}$ is the state of the thermodynamic switch, and w_i is the uncontrolled end-user power consumption. In the case of electric water heaters (EWHs) end-use is the hot-water extraction process and has been modeled using Poisson random pulse processes as described in [8].

TCLs are designed to operate within a deadband $[\underline{z}, \overline{z}]$ around a setpoint z_{set}, where $\underline{z}$ and $\overline{z}$ are the lower and upper limits of the deadband. Under PEM, TCLs can be in one of three modes; $i)$ charge, $ii)$ off or $iii)$ opt-out. In off mode, TCLs request the coordinator to consume energy for a time equal to the packet length that constitutes a *packet*. The requests are made according to the probability of request curve

$$\rho(z_i[k]) := 1 - e^{-h(z_i[k])\Delta t}, \quad (2)$$

where the rate parameter $h(z_i[k]) > 0$ is dependent on the local dynamic state [7]. The coordinator then either accepts or rejects the request depending upon grid conditions such as power reference tracking error. If the request is accepted, the TCL consumes power until the packet expires. Whereas, if the request is denied, the TCL requests again with probability given by (2). PEM also allows TCLs to opt-out of PEM if the temperature is too low in which case, the TCL consumes power until the temperature is sufficiently recovered.

B. Macro-model of PEM

A macro-model in this section refers to a state-bin transition model for the aggregate behavior of a fleet of TCLs. A PEM macro-model for TCLs is described next and more details can be found in [8]. Consider a population of homogeneous TCLs described by (1). To obtain such model, the deadband $[\underline{z}, \overline{z}]$ is divided into a set of discrete states or bins. Three identical copies of these binned states are created. That is, one for each mode of PEM (charge, off and opt-out). TCLs then evolve according to a controlled Markov chain with discrete-time dynamics given by

$$q[k+1] = f\left(\beta[k], \beta^-[k]\right) q[k], \ y[k] = Cq[k], \quad (3)$$

where $q[k]$ is the corresponding probability mass function, f represents a bilinear map of representing the PEM dynamics, $\beta[k] \in [0, 1]$ is the control signal that represents the proportion of DERs allowed to consume power (charge) from the grid and $\beta^-[k] \in [0, 1]$ is the proportion of TCLs that have consumed a packet and are transitioned from charging to off mode [8]. Finally, the output $y[k]$ with output matrix C is the total power demanded by the DER fleet based on P^{rate}.

The nominal demand of a homogeneous fleet of TCLs under PEM is defined in [8] as the minimum constant power demand (P_{nom}) for which quality of service (QoS) is satisfied where β_{nom} is the control input that achieves P_{nom}. QoS is considered to be satisfied if the average temperature of the fleet is greater than or equal to the desired setpoint (z_{set}). Furthermore, it is shown in [8] that in steady-state β^- is a constant related to the number of time steps that a packet lasts (n_p). That is, $\beta^-_{\text{nom}} = 1/n_p$. For further details on the macro-model, the reader is referred to [8], [9]. The next section deals with quantifying the statistics of the modeling error, ϵ, due to the proposed grouping approach.

III. Incorporating heterogeneity in PEM

As pointed out in the introduction, the majority of aggregation models assume homogeneity of the component DERs. The macro-model presented in Section II-B is not the exception [8]. Since decision-making under PEM relies heavily on P^{rate}, therefore, the focus is on heterogeneity in this power parameter. This assumption is supported by the fact that within a class of DERs the speed of energy delivery (heat, cooling, charge, etc) is expected to be normalized due to industry standards, which implies that the combination of DER model parameters amounts to the referred constant delivery speed within the population. For instance, two TCLs with different parameters should be able to increase the temperature of the same amount of water at the same rate but their rated power could vary due to the size of the water tank. Therefore, this section leverages the idea of PEM on diverse DER populations in [8] and proposes a path for describing a heterogeneous population of DERs on rated power through groups of macro-models of the same DER type. Figure 2 illustrates how the grouping procedure enters the PEM closed-loop setup. This model will be called *g-grouped* macro-model. The error of the g-grouped macro-model with respect to the TCL fleet is quantified via a simple quantization method when the macro-models are performing around their nominal response.

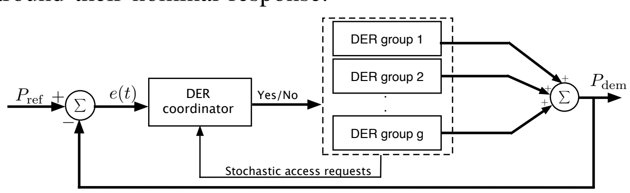

Fig. 2. Conceptual diagram of a heterogeneous fleet of TCLs under PEM

A. Modeling heterogeneity with the macro-model

When additively putting together a group of macro-models that represent a heterogeneous fleet, the coordinator assigns to each request the rated power corresponding to a group, which leads to modeling error. The question then is how to reduce this error by picking an appropriate number of groups according to a design criterion. Here the heterogeneity on P^{rate} is distributed according to some arbitrary distribution as shown in Figure 3. Let P^{rate} for all TCLs in the fleet be distributed arbitrarily within the interval $\mathbb{I} := [\underline{P}, \overline{P}]$ where $\underline{P}$ and $\overline{P}$ are the lower and upper limits of P^{rate}. By picking g points $\mathbb{I}_Q := \{\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_g\}$ evenly separated in $\mathbb{I}$, the i-th TCL with rated power P_i in $\mathbb{I}$ can be considered to have instead the rated power

$$P_i^* = \arg\min_{\bar{P}_j} \{ \left| \bar{P}_j - P_i \right| \text{ s.t. } \bar{P}_j \in \mathbb{I}_Q \}.$$

This constitutes the *quantizer* $Q_P : \mathbb{I} \to \mathbb{I}_Q$ illustrated in Figure 3. Therefore, Q_P is easily parametrized by the extremes in $\mathbb{I}$ and the width parameter ΔP which is designed to be the same for all groups. That is $\Delta P = \frac{\overline{P} - \underline{P}}{2g}$. The objective is to model a fleet of N TCLs having rated power distributed in $\mathbb{I}$ with the grouping of g macro-models. The j-th macro-model contains the N_j TCLs in the j-th group and has rated power $\bar{P}_j \in \mathbb{I}_Q$. Furthermore, TCLs in j-th group have indices in the set Υ_j, such that $\cup_{j=1}^{g} \Upsilon_j = \{1, \ldots, N\}$ and $\Upsilon_j \cap \Upsilon_i = \emptyset$, if $i \neq j$. Finally, it is assumed that in each group of TCLs, with indices in Υ_j, their rated power is approximately uniformly distributed, which constitutes a piece-wise approximation of the distribution in Figure 3.

Treating heterogeneity as a result of applying a quantizer provides a systematic approach for analyzing how well groups approach heterogeneity and finding the minimum number of groups necessary to achieve that goal. Clearly, characterizing heterogeneity by groups makes (i) the approach computationally expensive as g increases and (ii) inaccurate as $N_j, \ \forall \ j \in 1, \ldots, g$ becomes smaller since the number of TCLs in each group (N_j) is reduced [9]. Therefore, the number of groups needs to be related to the quantization error produced by the grouping approach. The quantization error is first illustrated next.

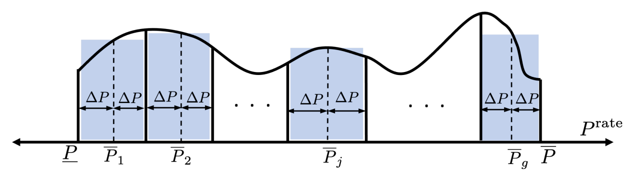

Fig. 3. Quantization of rated power interval $\mathbb{I} = [\underline{P}, \overline{P}]$ into g groups having midpoints $\bar{P}_j = \bar{P}_{j-1} + 2\Delta P$ for $j = 2, \ldots, g$ with $\bar{P}_1 = \underline{P} + \Delta P$.

B. Quantization error illustration

Quantization in signal processing is the process of mapping a set of continuous values from a large set to a set of discrete and countable values. An approach of this kind gives a venue to probabilistically characterize the *quantization error* coming from applying Q_p to the power interval $\mathbb{I}$. Consider a heterogeneous population of $1,000$ TCLs with P^{rate} uniformly distributed in the range $\mathbb{I} = [\underline{P}, \overline{P}] = [4, 5]$. Let $g = 2$ with range for the rated power quantizer $\mathbb{I}_Q = \{\bar{P}_1, \bar{P}_2\} = \{4.25, 4.75\}$ so that $\Delta P = \frac{\overline{P} - \underline{P}}{2g} = 0.25$. For a fixed $\beta = 1$, the steady-state response of the TCL fleet with 2-grouped macro-models, the histogram of relative error (in time) due to the quantizer is given in Figure 4. As expected, the quantization scheme assumes that TCLs within a group have the same rated power, which leads to error in the estimation of the total power consumed by the population. In the case where the right number of groups is chosen that minimizes quantization error, the statistics of the quantization error should be normally distributed with zero mean due to the central limit theorem. However, improperly designed quantizer translates the distribution mean towards the left ($\mu = -51.7$ kW and $\sigma = 40.7$ kW in Figure 4) since there is a difference between the request probability rates for different rated powers.

The scenario above motivates the approach to be followed in the next section in which the goal is to find the *smallest* value of g that makes the error distribution mean sufficiently close to zero. For this purpose two methods are proposed that are based on equalizing the probability of request within each group and reducing the standard deviation to a level that satisfies a predefined error tolerance.

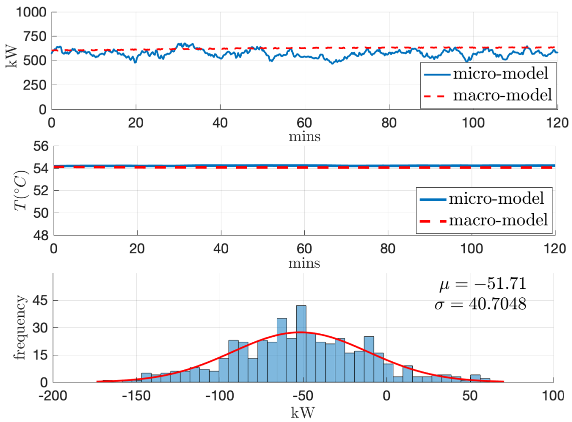

Fig. 4. Simulation of 1000 TCLs in steady-state showing the effect of grouping in a heterogeneous population (Top) Total power P_{dem} (Center) Mean temperature (^oC) (Bottom) Modeling error distribution.

IV. Quantization error modeling and analysis

In the previous section, the quantized approach to modeling heterogeneity was presented. This section describes the error introduced due to grouping heterogeneous TCLs and the minimum number of groups needed so that model error is within a pre-defined error tolerance (ϵ_ρ or ϵ_σ). The following assumptions are made in this section:

Assumption A1: The fleet is in steady-state, as defined in Section II-B, i.e., β and β^- are constant.

Assumption A2: The quantizer partitions the fleet into g uniformly distributed groups.

Assumption A3: $u_i \;\; \forall \;\; i = 1,\dots,N$ are independent, identically distributed (i.i.d).

Assumption A1 is reasonable since stationarity in the macro-model represents a meaningful starting point for analyzing aggregations of controllable load around a nominal operating point [13], [18], [19]. This baseline is defined by the quality of service considerations of DERs as described for PEM in II-B and the available flexibility is, therefore, based on the nominal operating point. Thus, around the nominal operating point, steady-state population behaviors are expected, which makes steady-state a salient operating regime for analysis. A2 is reasonable and allows the proposed procedure to be applicable to a heterogeneous fleet is which P^{rate} is distributed arbitrarily. This can happen due to the lack of information of individual power rates and the fact that these change over time due to wear and tear. A3 is a consequence of A1 since the fleet is in stationary steady-state and u_i are i.i.d.

The power drawn by TCL i within a fleet is $P_i^{\text{rate}}u_i$, $u_i \in \{0,1\}$ is the logic state in (1) that is 1 when consuming power and 0 otherwise. Let r_j denote the proportion of TCLs that are charging for the j-th group having N_j TCLs after quantization (where $N = \sum_{j=1}^{g} N_j$). If the total power demand given by TCL fleet is $\sum_{i=1}^{N} P_i^{\text{rate}}u_i$ and the total power demand from the g-group macro-model is $\sum_{j=1}^{g} \bar{P}_j N_j r_j$, then the quantization error ϵ is defined as,

$$\epsilon := \sum_{i=1}^{N} P_i^{\text{rate}}u_i - \sum_{j=1}^{g} \bar{P}_j r_j N_j, \tag{4}$$

where one can re-arrange (4) as $\epsilon = \sum_{j=1}^{g} \epsilon_j$, in which $\epsilon_j = \sum_{i\in\Upsilon_j} P_i^{\text{rate}}u_i - \bar{P}_j r_j N_j$ and, for each j and $i \in \Upsilon_j$, $\bar{P}_j - \Delta P \le P_i^{\text{rate}} < \bar{P}_j + \Delta P$. Moreover, each group can behave as a homogeneous population for a well chosen g obtained depending upon quantization error as discussed in the next section. Therefore, the macro-model accurately estimates the average behavior of the TCL population restricted to that group and $\sum_{i\in\Upsilon_j} u_i = r_j N_j$ that results in

$$\epsilon_j = \sum_{i\in\Upsilon_j} P_i^{\text{rate}}u_i - \bar{P}_j u_i = \sum_{i\in\Upsilon_j} (P_i^{\text{rate}} - \bar{P}_j)u_i. \tag{5}$$

Equation (5) constitutes the basis for the subsequent analysis of quantization error and is key for developing conditions under which the number of groups g is selected. Notice that by design $(P_i - \bar{P}_j)$ is the outcome of a uniformly distributed random variable. Therefore, under Assumption A3, the sum of products in (5) represents a sum of i.i.d. random variables.

A. Effect of probability of request on modeling error

As mentioned previously, arbitrary partitioning of an heterogeneous fleet without regarding the effect of quantization errors can result in an inaccurate estimation of the distribution of the fleet. An important observation of PEM, in steady-state, is that a TCL from the heterogeneous population described herein will request energy packets with a probability that is inversely proportional to its rated power. Thus, in this section, the relationship between request probability and rated power for devices in the population is characterized and a simple linear dependence of the form $\rho(P^{\text{rate}}) = -\alpha_j P^{\text{rate}} + \gamma_j$ is obtained, where $\alpha_j > 0$. This function then allows one to characterize the asymptotic behavior of the modeling error with respect to the number of groups. The following Theorem then relates the variation in probability of requests within a group to the width of the quantization interval ΔP.

Theorem 1: Consider a heterogeneous TCL population over the rated power interval $\mathbb{I} = [\underline{P}, \overline{P}]$ satisfying assumptions A1 and A2 under PEM, a quantizer $Q_P : \mathbb{I} \to \{\bar{P}_1,\dots,\bar{P}_{g_\mu}\}$ with g_μ partitions and parameter ΔP, the number of groups required to achieve a given probability of request tolerance $\epsilon_\rho > 0$ is given by

$$g_\mu \ge \frac{\alpha_1(\overline{P} - \underline{P})}{\epsilon_\rho}, \tag{6}$$

where $\alpha_j := \tan\left(\frac{\Delta\rho_j(\Delta P)}{2\Delta P}\right)$, $\Delta\rho_j(\Delta P) := \rho(\bar{P}_j - \Delta P) - \rho(\bar{P}_j + \Delta P)$ and ϵ_j as in (5) for all j.

Proof: From assumptions A1 and A2, the TCL population is actuated by β and β^- constants and the quantizer described in Section III-A applies to the TCL population at hand. When Q_P is applied to the TCL population, then the quantization error for all TCLs whose $P^{\text{rate}} \in \mathbb{I}_j := [\bar{P}_j - \Delta P, \bar{P}_j + \Delta P]$ is characterized by (5). Taking the expectation of ϵ_j in (5) results in

$$\mu_j = E[\epsilon_j] = \sum_{i\in\Upsilon_j} E[(P_i^{\text{rate}} - \bar{P}_j)]E[u_i],$$

where $E[.]$ is the expectation operator. Note that since u_i is a Bernoulli random variable with mean $E[u_i] = \Pr(u_i = 1)$ and P_i^{rate} for all $i \in \Upsilon_j$ are uniformly distributed random variables with mean equal to $\overline{P}_j$, therefore,

$$\mu_j = \sum_{i \in \Upsilon_j} E[(P_i^{\text{rate}} - \bar{P}_j)]\Pr(u_j = 1).$$

Furthermore, if $\Pr(u_i = 1)$ is constant for all $i \in \Upsilon_j$ then $\sum_{i \in \Upsilon_j} E[(P_i^{\text{rate}} - \bar{P}_j)]$ approaches 0 when N_j (number of elements in Υ_j) goes to infinity, which implies that μ_j approaches zero for N_j large enough. This indicates that one needs to find a ΔP so that $\Pr(u_i = 1)$ is constant for all $i \in \Upsilon_j$ TCLs with rated power in $\mathbb{I}_j$. In this paper, this is called *equalizing* the probability of request within a group. It happens that $\Pr(u_i = 1)$ is not the same for TCLs with different rated power. That is, TCLs with low power rates tend to request more often and TCLs with high-rated power request less often. The idea is then to compute a number of groups such that the probability of request within the group does not exceed a small pre-defined value. It then follows that $\Pr(u_i = 1)$ will be the same for all i in the group.

From (1) and (2), one can find the different rates at which TCLs with different rated-power make requests given that, initially, the TCLs had the same temperature. Specifically, the dependence of ρ on P^{rate} for fixed time variation Δt, constant w, initial condition $z[0]$ and $u = 1$ is $\rho(z) = \rho\left(\bar{a}P^{\text{rate}} + \bar{b}\right)$, where $\bar{a} = \Delta t/(cdL\eta)$ and $\bar{b} = \left(1 - \frac{\Delta t}{\tau}\right)z[0] - \frac{\Delta t z_a}{\tau} - \frac{\Delta t w}{cdL}$ are constant parameters obtained from (1) and ρ is given by (2). Since this expression is a function of $z[0]$, then it is pragmatic to assume that $z[0] = z^{\text{set}}$. Abusing notation, denote the probability of request as a function of P^{rate} by $\rho\left(P^{\text{rate}}\right)$ after time Δt. Although ρ is mathematically nonlinear, it is "close" to linear (see Example 1). A linear approximation of $\rho(P^{\text{rate}})$ for $P^{\text{rate}} \in \mathbb{I}_j$ gives

$$\rho(P^{\text{rate}}) = -\alpha_j P^{\text{rate}} + \gamma_j, \tag{7}$$

where $\alpha_j = \tan\left(\frac{\Delta\rho_j(\Delta P)}{2\Delta P}\right) > 0$, $\Delta\rho_j(\Delta P) = \rho(\bar{P}_j - \Delta P) - \rho(\bar{P}_j + \Delta P)$ and $\gamma_j > 0$. For a predefined tolerance ϵ_ρ in the probability of request curve within the rated power sub-interval $\mathbb{I}_j$, the value of ΔP that satisfies

$$\Delta\rho_j(\Delta P) \leq \epsilon_\rho \tag{8}$$

has the property that its probability of request cannot vary more than ϵ_ρ. Thus, using (7) in (8) gives $\Delta\rho_j(\Delta P) = 2\alpha_j\Delta P \leq \epsilon_\rho$. Recall that ρ is monotonically decreasing as a function of the dynamic state and also as a function of rated power. This fact implies that the largest $\Delta\rho_j(\Delta P)$ is produced by the leftmost partition of $\mathbb{I}$. In other words, $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{g_\mu}$ and one can focus only in the statistics of ϵ_1. Finally, to achieve this behavior, one needs the number of groups g_μ to satisfy (6) since $\Delta P = \frac{(\overline{P} - \underline{P})}{2g_\mu}$. This completes the proof. ■

Corollary 1: For a TCL population with heterogeneity according to assumption A2, as the probability of request tolerance within a group decreases, the g-grouped macro-model approaches the PEM fleet.

Proof: The proof is straightforward and follows from taking the limit of g_μ in Theorem 1 as ϵ_ρ goes to zero, which gives $\lim_{\epsilon_\rho \to 0} g_\mu = \infty$. ■

Observe, in reality, that as g_μ approaches the total number of TCLs (N) in the population, then the maximum number of groups can only be the total number of individual TCLs in the population, which becomes computationally expensive.

Example 1: Consider a fleet of $2,000$ power rate heterogeneous TCLs that are uniformly distributed in $\mathbb{I} = [3, 12]$ and a quantizer Q_P with its usual parametrization in terms of ΔP. Selecting $\epsilon_\rho \leq 0.00032$ as the probability of request tolerance produces at most a difference of 5 requests per sample time as shown in Figure 5. In this case, the linearity assumption holds for the entire interval with $\alpha_1 = 0.00072$ and $\gamma_1 = 0.049$ as shown in Figure 6. From Theorem 1, $\Delta P \leq 0.25$, which leads to $g_\mu \geq 18$. Figure 7 illustrates g_μ groups further resulting in $E[\epsilon]$ sufficiently close to 0. □

Remark: Quantifying the modeling error due to grouping on the basis of the probability of request reveals a unique feature of PEM i.e. heterogeneity is associated with the request rates. Theorem 1 showed that using the rated power associated with each request allows obtaining accurate bounds on modeling error. Although the results presented herein focus only on P^{rate}, however, it should be noted here that heterogeneity with respect to the τ in EWHs can also be incorporated within this framework. In (8), within Δt (e.g. 2 seconds), the parameter $\tau >> 1$ so that $\frac{1}{\tau} << 1$ compared to P^{rate} and L, therefore, its effect on $\rho(z[k])$ is negligible over Δt. The case of tank size (L) is more complex and requires analyzing the relationship between $\rho(z)$ and L as done in Theorem 1. However, simulation studies show that a relationship between L and $\rho(z[k])$ exists which is similar to the one in Figure 6. Also, Figure 8 shows that this is indeed the case for a fleet of $2,000$ TCL heterogeneous uniformly distributed in $\mathbb{I} = [3, 12]$ for power and in-tank capacity L over $[150, 250]$ liters. This is out of the scope of the current work and it is deferred to future research.

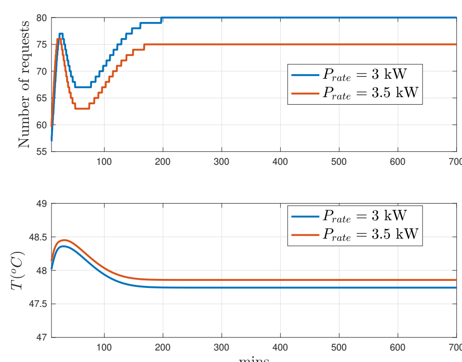

Fig. 5. Request rates (top) and dynamic state (bottom) showing the difference in steady-state conditions when $\Delta P = 0.5$ kW

B. Reducing error via standard deviation minimization

Previous section relates the modeling error due to grouping to the variation in probability of request within a group. Here the analysis for bounding the standard deviation of the quantization error distribution is provided. Consider once again a fleet of N TCLs that are heterogeneous in parameters as in assumptions A1-A3 over $\mathbb{I}$. The objective now is to provide grouping conditions under which the standard deviation, σ, satisfies a predefined error tolerance, ϵ_σ, i.e., how many groups g_σ should one pick so that $\sigma^2 \leq \epsilon_\sigma^2$.

Since the analysis is performed at steady-state (assumption A1) or nominal conditions, the following lemma is relevant

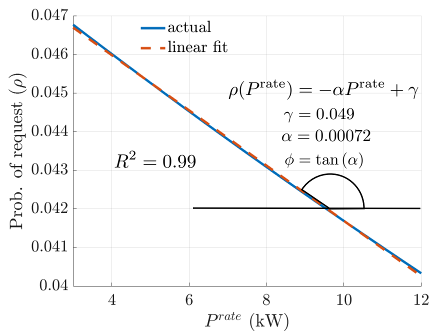

Fig. 6. ΔP vs ρ showing that P^{rate} and ρ can be approximated with a linear function, where $R^2 = 0.99$ is the correlation coefficient.

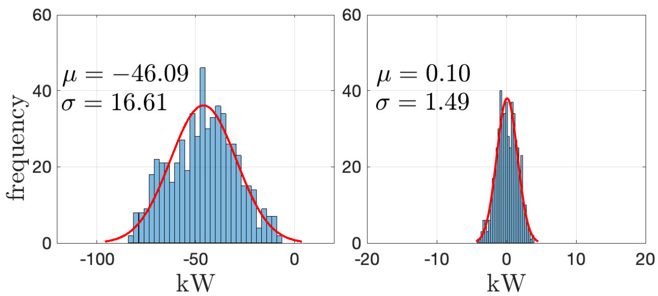

Fig. 7. Quantization error distribution with P^{rate} uniformly distributed in the interval $[3, 12]$ kW only: (Left) $g_\mu = 2$ and (right) $g_\mu = 18$.

in that it gives the probability of being in charging state ($u = 1$) for a homogeneous fleet of TCLs in terms of the invariant distribution of an underlying Markov chain.

Lemma 1: The random variable describing the logic state u taking values in $\{0, 1\}$ of an arbitrary TCL in a homogeneous fleet is Bernoulli with probability,

$$\kappa = \frac{\rho_{\text{nom}}\beta_{\text{nom}}}{\rho_{\text{nom}}\beta_{\text{nom}} + \beta^-_{\text{nom}}}. \tag{9}$$

Proof: Let u be the logic state of an arbitrary TCL in a homogeneous population and ρ be the probability that a TCL makes a request. u has only two possible outcomes $\{0, 1\}$ and can be modeled as a two-state Markov chain. Recall that β^- is the probability that a TCL has finished its packet and transitions from charging to standby mode. In nominal steady-state, $\beta^- = \beta^-_{\text{nom}}$ [10], $\rho_{\text{nom}} = \rho(z^{\text{set}})$ and $\beta = \beta_{\text{nom}}$ [8]. Therefore, the transition probabilities for the two-state Markov chain are $\mathrm{P}(u[k+1] = 0 | u[k] = 1) = \beta^-_{\text{nom}}$ and $\mathrm{P}(u[k+1] = 1 | u[k] = 0) = \rho_{\text{nom}}\beta_{\text{nom}}$. The probability of a TCL being in the charging state, $\kappa = P(u = 1)$, follows from calculating the stationary distribution of the two-state Markov chain, which gives (9). ■

Example 2: Consider the aggregate model of a fleet of homogeneous TCLs such that $\beta_{\text{nom}} = 0.1$ and $\beta^-_{\text{nom}} = 0.05$, the existence of an invariant distribution is guaranteed since the Markov chain is trivially aperiodic and irreducible. The chosen β_{nom} drives the population to steady-state at $z_{\text{set}} = 51.8°C$ Then it follows from (2) that $\rho(z_{\text{set}}) = 0.0593$. From (9), the probability that an arbitrary TCL in the fleet is consuming power is $\kappa = 0.1060$. On the other hand, a realization of the TCL fleet under the same conditions produces $\kappa = 0.1017$, which is close enough to the value produced by (9). □

The main result of the section is provided next.

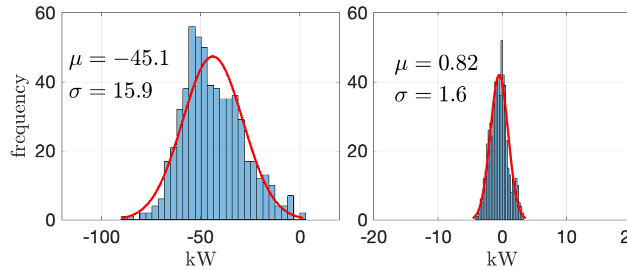

Fig. 8. Quantization error distribution with P^{rate} uniformly distributed in the interval $[3, 12]$ kW and tank size also uniformly distributed but in the interval $[150, 250]$ liters: (Left) $g_\mu = 2$ and (right) $g_\mu = 18$.

Theorem 2: Under the assumptions A1-A3, the quantization error ϵ in (4) for a heterogeneous TCL population and a quantizer $Q_P : \mathbb{I} \to \{\bar{P}_1, \ldots, \bar{P}_{g_\sigma}\}$ with g_σ partitions and parameter ΔP is distributed normally with zero mean and

$$\sigma^2 = N\frac{\kappa(\overline{P} - \underline{P})^2}{12g_\sigma^2}.$$

In addition, for a standard deviation tolerance $\epsilon_\sigma > 0$, if

$$g_\sigma \geq \frac{\sqrt{\kappa N}(\overline{P} - \underline{P})}{2\sqrt{3}\epsilon_\sigma} \tag{10}$$

then $\sigma^2 \leq \epsilon_\sigma^2$.

Proof: From assumptions A2-A3, the quantization error ϵ_j in (5) is generated by the random variable $Z_i = (P_i - \bar{P}_j)u_i$, where $(P_i - \bar{P}_j)$ is a uniformly distributed over the interval $[-\Delta P, \Delta P]$ and u_i is Bernoulli distributed with parameter κ_i. The probability distribution of Z_i can easily be computed as

$$f_{Z_i}(z) = \begin{cases} \frac{\kappa_i}{2\Delta P}, & \text{if } z \neq 0 \\ 1 - \kappa_i, & \text{otherwise.} \end{cases} \tag{11}$$

It is also straightforward from (11) that the expected value of Z_i is zero and the variance of Z_i is $\sigma^2_{Z_i} = \frac{\kappa_i \Delta P^2}{3}$. From assumption A3 ϵ_j is the sum of N_j i.i.d. random variables distributed as in (11) all having $\sigma_Z^2 = \frac{\kappa \Delta P^2}{3}$, therefore, the central limit theorem gives

$$\frac{\epsilon_j - \mathrm{E}[\epsilon_j]}{\sqrt{N_j}\sigma_Z} \xrightarrow[N_j \to \infty]{} N(0, 1).$$

Therefore, $\sigma^2_{\epsilon_j} = N_j \frac{\kappa \Delta P^2}{3}$. Given that $\Delta P = \frac{(\overline{P} - \underline{P})}{2g_\sigma}$, it follows that the variance of the total quantization error in (4) is

$$\sigma^2 = \sum_{j=1}^{g_\sigma} N_j \frac{\kappa(\overline{P} - \underline{P})^2}{12g_\sigma^2} = N\frac{\kappa(\overline{P} - \underline{P})^2}{12g_\sigma^2}. \tag{12}$$

Finally, the proof is completed by using (12) in $\sigma^2 \leq \epsilon_\sigma^2$, which gives (10). ■

Example 3: Consider a population of $10,000$ heterogeneous TCLs whose rated power is uniformly distributed in the interval $[4, 8]$. Suppose the population is in nominal steady-state. The population parameters were chosen so that a nominal steady-state response is obtained for $\beta_{\text{nom}} = 0.2$ and $\beta^-_{\text{nom}} = 0.05$. Furthermore, the set point is set at $z_{\text{set}} = 52°C$ and β drives the population to this temperature. The request rate of the distribution in steady-state is $\rho_{\text{nom}} \approx 0.0281$, which gives $\kappa = 0.101$. Theorem 2 is illustrated for

$\epsilon_{\sigma,1} = 10$ kW and $\epsilon_{\sigma,2} = 3$ kW in Figure 9. The minimum number of groups to attain a standard deviation less than $\epsilon_{\sigma,1}$ and $\epsilon_{\sigma,2}$ are 4 and 12 respectively. Figure 9 shows that the standard deviation is within the chosen tolerance. □

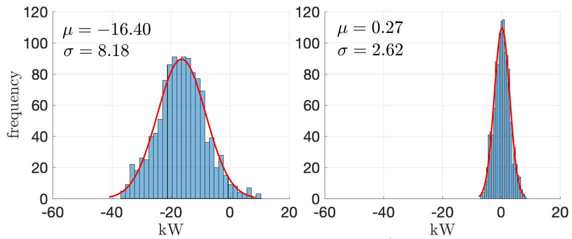

Fig. 9. Quantization error distribution with P^{rate} uniformly distributed in the interval $[4, 8]$ kW only: (Left) $g_\sigma = 4$ and (right) $g_\mu = 12$.

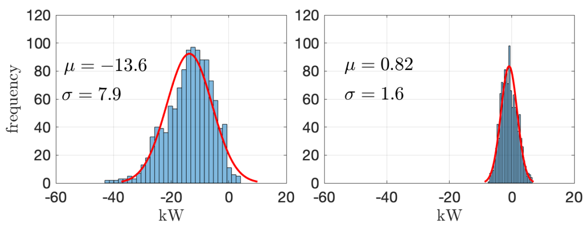

Fig. 10. Quantization error distribution with P^{rate} uniformly distributed in the interval $[4, 8]$ kW and tank size also uniformly distributed but in the interval $[150, 250]$ liters: (Left) $g_\sigma = 4$ and (right) $g_\mu = 12$.

The above example shows that with heterogeneity in P^{rate} only, Theorem 2 relates the standard deviation of quantization error to the number of groups. In this case, κ is only dependent on P^{rate}, which accounts for the design of g. However, κ is expected to depend on L and τ if these parameters are also heterogeneous. Simulation under the same scenario in Example 3 is conducted but with L now uniformly distributed in the interval $[150, 250]$ liters. Figure 10 shows the quantization error for $g = 4$ and $g = 12$ and the standard deviation corresponds to the one obtained in Example 3. This indicates that there exists a relationship between σ in (10) and g_σ. Furthermore, it is possible that for an arbitrary distribution when the number of groups increases, quantization results in a group that is empty and does not contain any TCLs. This can be accommodated in a post-processing step where such groups are not considered. Finally, the analysis presented in this paper has been carried out under the macro-model's steady-state assumption. As a result, the quantization errors can be different during transients than those predicted by Theorems 1 and 8. In order to achieve a desired quantization error behavior during transients, a different grouping scheme can be developed. By switching between grouping schemes specifically the desired aggregate quantization error response can be achieved. However, the quantization error analysis during transients requires further analysis of the macro-model. Work is ongoing to add heterogeneity in multiple parameters, quantization error during transient behavior, and practical implementation of the grouping scheme developed herein.

V. Conclusions

This paper presented a quantization-based approach to quantify the modeling error in a population of DERs with heterogeneity in rated powers. It was also shown that heterogeneity in this sense is relevant in the context of PEM since rated power associated with each request is readily available to the coordinator. A systematic approach was then provided by Theorem 2 to obtain the minimum number of groups required to achieve a specified error tolerance. The analysis provided in the paper revealed that heterogeneity in the form of request rates is a feature unique to packet-based coordination schemes, such as PEM. Furthermore, empirical results show that Theorems 1 and 2 can be extended to account for the heterogeneity in tank size (L) and standing loss time constant (τ) and will be considered in future publications.

References

[1] S. Bashash and H. K. Fathy, "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads," in *Amer. Control Conf. (ACC)*, 2011, pp. 4546–4553.

[2] D. S. Callaway and I. A. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184–199, 2011.

[3] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl, "Demand dispatch," *IEEE Power and Energy Magazine*, vol. 8, no. 3, pp. 20–29, May 2010.

[4] Q. Lénet, M. S. Nazir, and R. P. Malhamé, "An inverse nash mean field game-based strategy for the decentralized control of thermostatic loads," in *IEEE 60th Annu. Conf, Decis. Control*, 2021, pp. 4929–4935.

[5] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway, "Energy arbitrage with thermostatically controlled loads," in *European Conference on Circuit Theory and Design*, July 2013, pp. 2519–2526.

[6] J. Mathias, R. Moye, S. Meyn, and J. Warrington, "State space collapse in resource allocation for demand dispatch," in *IEEE 58st Annu. Conf, Decis. Control*, 2019, pp. 6181–6188.

[7] M. Almassalkhi, L. A. Duffaut Espinosa, P. D. Hines, J. Frolik, S. Paudyal, and M. Amini, *Asynchronous Coordination of Distributed Energy Resources with Packetized Energy Management*. New York, NY: Springer New York, 2018, pp. 333–361.

[8] L. A. Duffaut Espinosa and M. Almassalkhi, "A packetized energy management macromodel with quality of service guarantees for demand-side resources." *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3660–3670, 2020.

[9] L. A. Duffaut Espinosa, A. Khurram, and M. Almassalkhi, "Reference-tracking control policies for packetized coordination of heterogeneous DER populations." *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 6, pp. 1–17, 2020.

[10] L. A. Duffaut Espinosa, M. Almassalkhi, P. Hines, and J. Frolik, "System properties of packetized energy management for aggregated diverse resources," in *Power Syst. Comput. Conf.*, 2018.

[11] S. Koch, J. L. Mathieu, and D. S. Callaway, "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services," in *Power Syst. Comput. Conference*, 2011, pp. 1–7.

[12] Y. Chen, A. Bušić, and S. P. Meyn, "State estimation for the individual and the population in mean field control with application to demand dispatch," *IEEE Trans. Autom. Control*, vol. 62, no. 3, pp. 1138–1149, March 2017.

[13] J. Mathias, A. Bušić, and S. P. Meyn, "Demand dispatch with heterogeneous intelligent loads," in *HICSS*, 2017.

[14] W. Zhang, J. Lian, C. Chang, K. Kalsi, and Y. Sun, "Reduced-order modeling of aggregated thermostatic loads with demand response," in *IEEE 51st Annu. Conf, Decis. Control*, Dec 2012, pp. 5592–5597.

[15] W. Zhang, J. Lian, C. Chang, and K. Kalsi, "Aggregated modeling and control of air conditioning loads for demand response," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4655–4664, 2013.

[16] M. S. Nazir and I. A. Hiskens, "Noise and parameter heterogeneity in aggregate models of thermostatically controlled loads," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8888–8894, 2017.

[17] S. E. Z. Soudjani and A. Abate, "Aggregation and control of populations of thermostatically controlled loads by formal abstractions," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 975–990, 2015.

[18] S. El-Ferik and R. P. Malhamé, "Identification of alternating renewal electric load models from energy measurements," *IEEE Trans. Autom. Control*, vol. 39, no. 6, pp. 1184–1196, June 1994.

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