

Analysis of Heat Pump Water Heater Electric Load as a Dispatchable Demand Response Resource

Joseph D'Ambrosia, Dana Paresa, Robert B. Bass

Department of Electrical and Computer Engineering

Portland State University, Portland, Oregon, USA

Abstract—Heat pump water heaters (HPWHs) are becoming increasingly present in residential homes across the United States due to advancements in performance, purchase incentives, and energy efficiency regulations. As virtual power plants and demand response programs mature, electric utilities can leverage customer-owned flexible loads, like HPWHs, to balance the grid and maximize renewable energy integration. However, the potential impact of aggregating and deploying customer-owned, behind-the-meter, flexible loads is difficult to quantify. The objective of this paper is to analyze the marginal and aggregate load shapes of an arbitrary number of HPWHs. By building off the concept in distribution planning of After Diversity Maximum Demand, this paper introduces the concept of *marginal load* and calculates the *aggregate load* from the *marginal load*. The results of this study show that as the number of aggregated HPWHs increases, the uncertainty around their baseline aggregate load decreases. The results also highlight how when there are only a small number of aggregated HPWHs, the aggregate load is difficult to accurately estimate. To improve grid service planning, utilities need to improve load shape estimates for distributed flex loads in their system. By addressing these concerns, utilities can better forecast flexible load-based grid service programs.

Index Terms—distributed energy resources, heat pump water heater, flexible loads, after diversity maximum demand, virtual power plant, demand response.

I. INTRODUCTION

As demand response (DR) programs develop and mature, residential HPWHs have the potential for enacting various grid services. Grid operators can use distributed energy resources (DERs) to relieve stress on transmission and distribution grids, lowering the cost of transmission upgrades, and minimizing the curtailment of renewable energy, for example. For optimal flexible load grid service scheduling, load managing entities (LMEs) need to first quantify the baseline time-dependent load shapes of the selected flexible loads.

DERs are customer-owned, grid-interactive assets, such as flexible loads, generation, or storage, that are capable of performing various grid services [1]–[3]. LMEs can leverage participating DER to provide grid reliability services that may otherwise be performed by traditional generating resources.

In this manuscript, HPWHs are aggregated to estimate their potential as leveraged flexible loads for various grid services. HPWHs are capable of acting as unidirectional batteries by storing electrical energy as thermal energy within their storage

tanks. HPWH adoption is increasing, due to their energy efficiency and reduced emissions compared to electric resistance and natural gas water heaters. Recently, the Department of Energy passed energy conversion standards regarding water heater efficiency, which will essentially require the use of HPWH in residential homes starting May 6, 2029 [4].

As more renewable generation is connected to the grid, the variability in generating output increases. Flexible loads can be part of the solution because LMEs will be able to request grid services that will aid in balancing generation and load. One means for managing messaging to flex loads is the ANSI/CTA-2045-B (CTA-2045) protocol [5]. HPWHs can provide grid services by responding to CTA-2045 requests to shift their energy use, acting as unidirectional energy storage devices [6]. Flexible loads are, however, similar to renewable generation: variable in their energy use and therefore variable in their ability to load shift. Therefore, in this paper we consider the uncertainty of flexible loads, specifically HPWHs.

In previous work, the time dependent aggregate load shapes of HPWHs have been analyzed. However, these studies do not consider how the number of aggregated units affect load shapes [7]–[9]. After Diversity Maximum Demand (ADMD) is a well known concept in the field of distribution planning that considers how aggregate maximum demand of a number of customers on a feeder changes as the number of customers increases. It is primarily used for sizing distribution equipment by calculating the maximum demand of the next, or marginal, unit of customer load being studied. Prior research regarding ADMD has considered traditional residential customers [10], customers with electric vehicles [11], and customers with solar and electric vehicles [12]. Zakaria *et al.* considers how ADMD changes when temporal variation of electrical vehicle loads is considered, and employs Monte Carlo simulations (MCSs) to estimate uncertainty [11].

This work proposes a method for understanding the behavior of aggregated flexible loads and their potential to contribute to grid services. Relying on publicly available simulated end-use load data, we estimate marginal and aggregate load, their uncertainty, and how they change with the number of loads, N . Clear trends emerge wherein additional aggregated resources asymptotically approach certainty, highlighting the benefit of DR programs with large numbers of participants. Once the baseline load is understood, LMEs can take the next step to leverage flexible loads: provide grid services to the grid.

To enable flexible loads and unlock their potential, a better

Corresponding author: Joseph D'Ambrosia (jdambro@pdx.edu). This work is supported by Portland General Electric and the Northwest Energy Efficiency Alliance.

understanding of the baseline energy use of those loads is required. Once the baseline is understood, planning for grid services becomes more straightforward.

II. METHODOLOGY

This work builds upon the methodology developed by Zakaria *et al.* to analyze marginal and aggregate time-dependent demand of HPWHs and to quantify certainty at an increasing number of loads, N [11].

A. OCHRE

Accurate, sub-hourly, localized, metered load data for individual heat-pump-only HPWHs of a sufficiently large sample size is not yet widely available. Object-oriented controllable high-resolution residential energy model (OCHRE) is a Python-based tool that can be used to simulate thousands of DERs embedded within the distribution grid [13]. We use OCHRE to obtain location-specific load profiles for HPWHs. This work analyzes HPWHs in detached single-family residential homes in Portland, Oregon. OCHRE uses the NREL Res Stock tool to simulate location specific, statistically significant, single family homes [9]. Each simulated home includes a HPWH. OCHRE stochastically generates unique water draw profiles for each home based on the number of occupants, time of day, day of the week, and season with 15 minute resolution [14]. The water draw profiles and localized weather data are then fed into OCHRE's thermal models for both the HPWHs and homes to calculate the electric energy used by each individual HPWH. OCHRE assumes all the HPWHs models are functioning normally and uses the same temperature setpoints for each HPWH. This could cause the electric load profiles to be less diverse than they may actually be.

Each OCHRE simulation results in a set of individual, time-dependent, load profiles, $P_{t \rightarrow T}$, where T is the period of one day and t is each 15 minute timestep. These load profiles are used to calculate marginal and aggregated load mean, percentiles, and confidence intervals. This method simulates 409 HPWHs for a single winter weekday at 15-minute intervals.

B. Marginal Load

Planning a DR program requires knowledge of the time-dependent load profiles to generate grid service schedules. This work considers how load profiles change shape as the number of units changes. Though distribution planners have considered unit-load variability extensively, specific results regarding the *aggregate load* of HPWHs are not available. We use a modified ADMD approach to estimate the marginal maximum load that the next unit will have on the existing system. Analysis using ADMD reveals that marginal maximum demand of many like-loads is much smaller than the individual maximum demand of the same loads, due to diverse use patterns. Previous work describes ADMD as having a logarithmic relationship with the number of individual units [12]. As more units are considered, the additional demand from the next unit converges to a steady value that can be predicted and modeled accurately for any large number of units.

This work modifies the process of ADMD to consider per-HPWH time-dependent load shapes, rather than maximum load, while preserving the number of units as a variable. This will be referred to as the *marginal load*. The *marginal load* for a HPWH is defined as the most likely electric demand that the next, or marginal, HPWH will have on the system. In this work, the system is the aggregation of all HPWH that are participating in a given DR program.

Calculations for *marginal load*, derived from the ADMD method, consider a number of individual, time-dependent, load profiles, $P_{t \rightarrow T}$. In this work, 409 individual load profiles for a single winter weekday are simulated in OCHRE. This work uses Monte Carlo simulations, as described in previous work [11]. The mean *marginal load* at each timestep is evaluated individually. The time-dependent mean *marginal load* for a specific number of HPWHs N is shown in Equation 1.

$$ML_{t \rightarrow T, N, m \rightarrow M} = \text{mean} \left[\frac{1}{NP_{base}} \sum_{n=1}^N P_{t \rightarrow T, n, m} \right] \quad (1)$$

Marginal load, or $ML_{t, N, m \rightarrow M}$, is the mean of the set intermediate *marginal load* values. Each intermediate *marginal load* value is calculated by randomly sampling N load profiles from the set of 409 load profiles generated by OCHRE. P_{base} is the base power, which is the HPWH rated power, chosen to be 0.5 kW, the same value OCHRE uses in its modeling. Resampling and calculations are repeated for each run of the Monte Carlo simulation $m \rightarrow M$, where $M = 1000$ and stored for analysis. The final *marginal load* value is then calculated by taking the mean of the set of intermediate *marginal load* profiles generated by the MCS at each timestep $t \rightarrow T$, as shown in Table I. The process is repeated for each N in the range of units from 1 to 500.

Of interest, aside from mean *marginal load*, is uncertainty of the *marginal load*. To quantify uncertainty, the 95% confidence interval (95% CI) is used, which describes the region where 95% of *marginal load* profiles for N number of HPWHs will fall. 95% CI is defined as the region between the 97.5th and 2.5th percentiles. Percentiles are calculated using modified versions of Equation 1, which, instead of calculating the mean from the set of intermediate *marginal load* values, calculate the 97.5th and 2.5th percentiles.

A summary of the process is as follows:

- 1) Use MCS to randomly sample N available HPWH load profiles for run m .
- 2) Sum the N sampled profiles at each timestep and divide by the base power and number of units N to obtain the intermediate *marginal load* profile for the N HPWHs.
- 3) Repeat these steps for the desired number of MCS runs M . In our case $M = 1000$.
- 4) Calculate the mean, 97.5th, and 2.5th percentiles from the 1000 runs at each timestep independently.
- 5) Repeat for another number of HPWHs N for all units in the range 1 to 500.

Table I shows an example of the intermediate *marginal load* values at each timestep and how the mean and 95% CI are

TABLE I

DEMONSTRATION OF HOW MEAN AND PERCENTILE VALUES ARE CALCULATED. TABLE I(A) SHOWS THE INTERMEDIATE *marginal load* VALUES FOR EACH RUN OF THE MCS WHERE $N=500$. TABLE I(B) SHOWS HOW THE MEAN AND PERCENTILE VALUES ARE CALCULATED AT EACH TIMESTEP FROM THE SET OF INTERMEDIATE *marginal load* VALUES AS SHOWN BY THE ARROWS FROM TABLE I(A) TO TABLE I(B). THE VALUES ARE IN PER-UNIT POWER, WHERE $P_{base}=0.5$ kW.

$N = 500$ MCS	00:00	00:15	...	23:45	
1	0.177	0.139	...	0.166	
2	0.152	0.125	...	0.169	(A)
...	...	...	...	...	
1000	0.164	0.128	...	0.173	
	↓	↓	...	↓	
$N = 500$ Stats	00:00	00:15	...	23:45	
97.5 th %	0.183	0.148	...	0.185	(B)
Mean	0.163	0.131	...	0.168	
2.5 th %	0.145	0.116	...	0.151	

calculated for each timestep $t \rightarrow T = 24$ hours for a single value of N . The intermediate values are calculated for each run of the MCS from $m \rightarrow M = 1000$ and each occupies one row of Table I(A). The resulting mean, 97.5th, and 2.5th percentiles are shown in Table I(B) with arrows going from the sets they were calculated from.

The *marginal load* highlights the individual unit's contribution. However, *aggregate load* is of significant interest to electric utilities and aggregators when planning grid services.

C. Aggregate Load

Aggregate load of HPWHs is estimated at each timestep of day for N units. The *aggregate load* is defined as the sum of all individual time-dependent HPWH loads. The calculation is similar to that of *marginal load*, however *aggregate load* continues to scale with the number of loads. The 95% CI of the *aggregate load* is calculated using the 97.5th and 2.5th percentiles and in this case modified versions of Equation 2.

$$AL_{t \rightarrow T, N, m \rightarrow M} = \text{mean} \left[\sum_{n=1}^N P_{t \rightarrow T, n, m} \right] \quad (2)$$

The uncertainty of the *aggregate load* differs from the uncertainty of the *marginal load*. This is because the *aggregate load* is increasing as more units are included in the aggregation. Therefore, we defined the aggregate uncertainty as the percentage of the 95% CI to the mean. This scales the width of the uncertainty by the total *aggregate load*, painting a more representative picture of the variability in the *aggregate load*.

III. RESULTS

A. Marginal Load Results

The *marginal load* for a generic winter weekday is shown in Figure 1. To focus on the impact of the load shape as the number of units changes, simulation duration is limited to a single day. As the number of units increases, *marginal load* decreases, although not uniformly across all hours of the day.

Marginal load uncertainty is quantified by the 95% CI, bounded by the gray 97.5th and 2.5th percentile surfaces, as

shown in Figure 1, which narrows with increasing N . The mean surface is presented in blue. As the number of units increases, *marginal load* uncertainty decreases, which speaks to the value of very large DR programs.

The highlighted edge of the surfaces in Figure 1 where $N = 500$, show the *marginal load* along with the 95% CI of the 500th HPWH. The width of the 95% CI fluctuates throughout the day, which is to be expected, due to variations in water draw usage periods. Generally, as energy use decreases, uncertainty also decreases, and as energy use increases, uncertainty also increases.

Consider the far-right edge of the curves in Figure 1, where $t = 23:45$. Here, *marginal load* trend lines are shown using thick black curves. The width of the 95% CI is decreasing logarithmically as the number of units increases, where the 97.5th and 2.5th percentile surfaces are converging towards the mean surface. This logarithmic decrease shows that at a small number of units, the 95% CI is large, meaning the uncertainty of where the *marginal load* will fall is quite high. However, when the number of units is large, the uncertainty is significantly smaller. This relationship between the width of the 95% CI and the number of units is consistent for each timestep throughout the selected winter weekday. The maximum width of the 95% CI of the *marginal load* at $N = 100$ is relatively small compared to the where $N = 1$. The maximum width of the 95% CI of the *marginal load* at $N = 1$ is 0.93 per-unit and at $N = 100$ is 0.12 per-unit. At $N = 500$ the maximum 95% CI has decreased to 0.05 per-unit. These values are summarized in Table II.

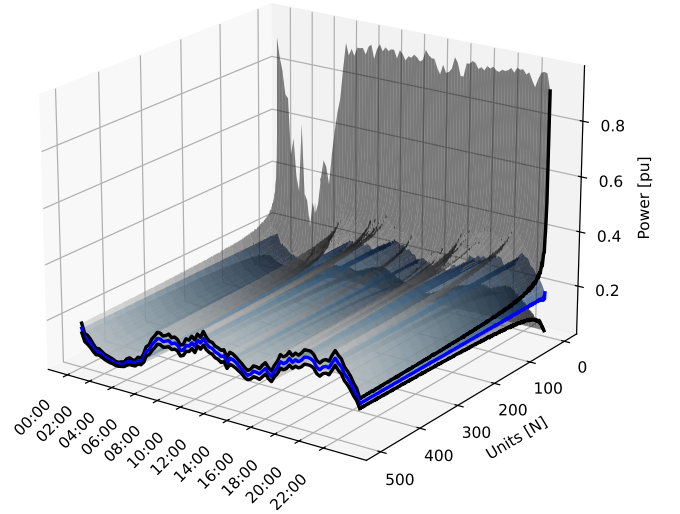


Fig. 1. *Marginal load* of HPWHs during a selected day in winter. The mean of the *marginal load* is shown as the blue surface. The 95% confidence interval of the *marginal load* is shown bounded by the gray surfaces on top of and underneath the mean surface. The edge where $N=500$ and $t \rightarrow T$ in steps of 15 minutes is highlighted in blue and black and shows the estimated *marginal load* for each timestep in the selected day where there are 500 HPWHs being considered. The edge where $t = 23:45$, the last timestep in T , is similarly highlighted. N ranges from 1 to 500 units. The vertical axis is show in power per-unit, where the base power is 0.5 kW.

A non-normal distribution is clearly present when the num-

ber of units is small, noting the distance from the mean of the 97.5th and 2.5th percentiles. Further analysis conducted but not presented here shows the non-normal distribution is still present to a lesser degree as the number of units grows larger. For example, when $t = 21:45$, the skew at $N = 1$ is 1.02 and the skew at $N = 500$ is 0.17.

B. Aggregate Load Results

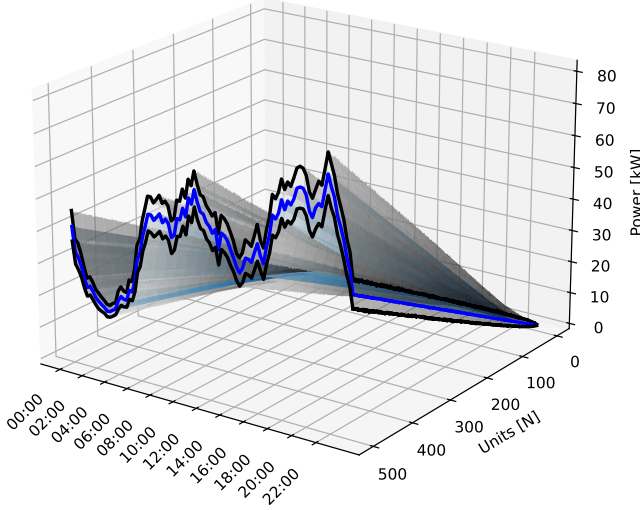


Fig. 2. Aggregate load of HPWHs during a selected day in winter. The mean of the aggregate load is shown as the blue surface. The 95% CI of the aggregate load is shown bounded by the gray surfaces on top of and underneath the mean surface. The edge where $N=500$ and $t \rightarrow T$ is highlighted in blue and black and shows the estimated aggregate load for each timestep in the selected day where there are 500 HPWHs being considered. The edge where $t = 23:45$, the last timestep in T , is similarly highlighted but N ranges from 1 to 500 units. The vertical axis is shown in kW.

Figure 2 shows the estimated time coincident aggregate load in kW of $N = 1$ to $N = 500$ HPWHs by plotting the mean of the aggregate load in blue as well as the 97.5th and 2.5th percentiles in gray above and below the mean, respectively. The region between the 97.5th and 2.5th percentiles is the 95% CI of the aggregate load for the selected winter weekday.

The edges of Figure 2 are highlighted similarly to Figure 1. Aggregated and marginal loads behave similarly at $N = 500$, as shown in Figure 1 and 2. The other highlighted edge of Figure 2, where $t = 23:45$ is different from the marginal load plot. The aggregated load mean is increasing linearly as the number of units increases. The 95% CI is also increasing as the number of units grows. However, when considering the aggregate load, we present the proportion of the 95% CI to the mean to quantify uncertainty of the aggregate load. We call this the aggregate uncertainty. Figure 3 presents a slice at $t = 21:45$, showing aggregated load and uncertainty as a ratio of the width of the 95% CI to mean load magnitude. The aggregate uncertainty is quantified at $N = 1, 100$, and 500 when $t = 21:45$ as seen in Table II.

Figure 3 shows the mean of the aggregate load increasing linearly with the number of units, as expected from the results in Figure 2. A linear best fit line and corresponding R^2

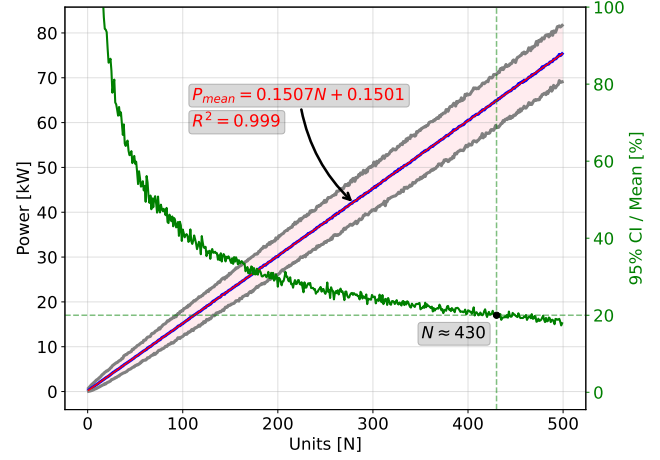


Fig. 3. Aggregate load at $t = 21:45$. The straight red line is the best fit line of the mean which is shown underneath in blue. The equation for the linear best fit along with the R^2 value are both shown also in red. Above and below the mean are the 97.5th and 2.5th percentiles respectively plotted in gray. The region in pink is the 95% CI. Shown in green is the percentage of the width of the 95% CI to the mean.

value is shown in red in Figure 3. The secondary axis plots aggregate uncertainty of the aggregate load at 21:45 in green. At $N \approx 430$ the percentage of the 95% CI to the mean is less than about 20% of the mean value.

TABLE II
MARGINAL AND AGGREGATE UNCERTAINTY AT $N = 1, 100$, AND 500.

Units	Marginal Uncertainty	Aggregate Uncertainty
1	0.93 [pu]	260%
100	0.12 [pu]	42%
500	0.05 [pu]	18%

IV. DISCUSSION

The marginal load as it is presented here will not be directly used in the assessment of flexible load potential by LMEs for DR programs. It is used as a key step in constructing the aggregate load, which is the estimation of the baseline load of an aggregation of an arbitrary number of flexible loads. The marginal load is, by definition, the slope of the aggregate load.

The mean marginal load, shown in blue in Figure 1, varies throughout the time of day but remains constant as the number of units increases, except for a small exponential region where $N < 5$. This relationship between the marginal load and the number of units points to a linear relationship between the number of units and the aggregate load.

Marginal load uncertainty, quantified by the 95% CI, decreases logarithmically as the number of units increases, showing that marginal load uncertainty is large when N is small, but quickly converges as N increases. The 95% CI width decreases by 87% from $N = 1$ to $N = 100$ and again decreases by 55% from $N = 100$ to $N = 500$. Thus, as more units are added to the aggregation, we see diminishing returns on the decreasing uncertainty.

This result highlights that when there are only a small number of units participating in a given DR program, the LMEs will have less ability to predict the aggregate baseline load of the flexible loads participating in the demand response program. It also shows that the ability to predict the load gets more certain as the number of participants increases.

Figure 1 highlights the mean of the *aggregate load*, showing a linear increase. The linear relationship between the number of units and the *aggregate load* is confirmed at each timestep by fitting a linear trend line with a high coefficient of determination, R^2 . This relationship, together with the decreasing proportional uncertainty, is of particular importance to LMEs when estimating baseline demand of HPWHs, enabling better estimations for an arbitrary number of HPWHs. The *aggregate load* can be estimated for potential scenarios where $N = 1, 100, 10,000$, or even $100,000$, aiding in DR program decision making investments. Recall that the *aggregate load* of a given flexible load is the maximum load that can be shifted by a DR program. This upper bound is useful for LMEs looking to estimate the shifting potential of an aggregation of flexible loads.

The aggregate uncertainty, as expressed as a percentage of the 95% CI to the mean *aggregate load* value, decreases as the number of units increases. Therefore, there exists a large enough number of units so that the proportional uncertainty of the *aggregate load* is below a tolerable threshold. For demonstrative purposes in this paper, we chose a proportional uncertainty value of 20% as a sample threshold. Figure 3 shows the number of units required to meet the desired threshold, at 430 HPWHs.

This result highlights that when there are only a small number of units participating in a given DR program, there is less certainty in the size of resources that can be dispatched during a DR event. It also shows that the ability to predict the dispatchable aggregate load gets better as the number of participants increases.

These results clearly demonstrate how the number of individual flexible loads impacts of the uncertainty of the load profile of the aggregate demand response resource. LMEs will need to consider the uncertainty of flexible loads in their decision-making surrounding the sizing of demand response programs. The results above show that as the number of units enrolled in a given DR program, the uncertainty of the aggregation of those flexible loads decreases logarithmically. LMEs can use the proportional uncertainty of the *aggregate load* to optimize the sizing of a DR program in order to balance flexible capacity, defined initially by the existing load, with the uncertainty of that load and the investment required to enroll each participant.

V. CONCLUSION

Understanding how HPWH load shapes converge as more are aggregated informs LMEs and DR program planners. The number of units significantly impacts marginal and aggregate load uncertainty. The results from this paper show that when there are only a small number of units, *marginal load* and

aggregate load are largely unpredictable in their behavior, and that as the number of units increases, the 95% CI or proportion of the 95% CI to the mean improve, respectively. Considering the *marginal load*, when $N = 1$, the width of the 95% CI is 0.93 power per-unit, when $N = 100$, the width of the 95% CI has decreased to 0.12 power per-unit, which is an 87% decrease with the addition of 99 HPWHs. When $N = 500$ the width of the 95% CI is 0.054, a 55% decrease from when $N = 100$. The width of the *marginal load* 95% CI converges to zero as the number of aggregated units approaches infinity. In a similar trend, the aggregate uncertainty falls from 247% when $N = 1$ to 20% when $N \approx 430$. This analysis highlights the impact of the number of participating units for load forecasting in virtual power plants or DR programs.

Future analysis that includes more days to compare seasonality, peak, and low load days will shed light on how other contributing variables impact HPWH *aggregate load*. Additionally, work can be expanded to estimate various end-use aggregate load shapes, including hybrid-HPWHs, electric resistance water heaters, or HVAC units. Furthermore, an analysis of how the uncertainty changes when the flexible loads are being controlled will offer more insights into the capacity of flexible loads to provide grid services.

REFERENCES

- [1] S. Widergren, R. Melton, A. Khandekar, B. Nordman, and M. Knight, "The plug-and-play electricity era: Interoperability to integrate anything, anywhere, anytime," *IEEE P.&En. Mag.*, vol. 17, no. 5, pp. 47–58, 2019.
- [2] "IEEE guide for distributed energy resources management systems (DERMS) functional specification," *IEEE Std 2030.11-2021*, 2021.
- [3] *FERC Order No. 2222-A*. U.S. Department of Energy, Federal Energy Regulator Commission, Docket No. RM18-9-002, March 18, 2021.
- [4] U.S. Department of Energy, "Energy conservation standards for consumer water heaters; final rule," 89 Fed. Reg. 37778 (May 6, 2024), May 2024.
- [5] anon, "ANSI/CTA-2045-B: Modular communications interface for energy management," Consumer Technology Association, Tech. Rep., November 2020.
- [6] D. Paresa, M. Adham, and R. B. Bass, "Evaluating the role of heat pump water heaters in demand response," in *2025 IEEE PES GM*, 2025.
- [7] anon, "Field testing of pre-production prototype residential heat pump water heaters," U.S. Department of Energy, Office of Energy Efficiency Renewable Energy, Tech. Rep. DOE/EE-0317, May 2007.
- [8] H. Gong, T. Rooney, O. M. Akeyo, B. T. Branecky, and D. M. Ionel, "Equivalent electric and heat-pump water heater models for aggregated community-level demand response virtual power plant controls," *IEEE Access*, vol. 9, pp. 141 233–141 244, 2021.
- [9] anon, "ResStock general reference documentation," National Renewable Energy Laboratory, Tech. Rep., jan 2024.
- [10] S. Bielecki, "Estimation of maximum loads of residential electricity users," in *E3S Web of Conferences, Research & Development in Power Engineering*, vol. 137, 2019.
- [11] A. Zakaria, C. Duan, and S. Z. Djokic, "After diversity maximum demand and daily load profiles of maximum demand for uncontrolled residential EV charging," in *2023 IEEE PES Innovative Smart Grid Technologies Europe*, 2023, pp. 1–5.
- [12] C. Barteczko-Hibbert, "After diversity maximum demand (ADMD) report," Customer-Led Network Revolution, Tech. Rep., Feb. 2015.
- [13] M. Blonsky, J. Maguire, K. McKenna, D. Cutler, S. P. Balamurugan, and X. Jin, "OCHRE: The object-oriented, controllable, high-resolution residential energy model for dynamic integration studies," *Applied Energy*, vol. 290, p. 116732, 2021.
- [14] R. Hendron, J. Burch, and G. Barker, "Tool for generating realistic residential hot water event schedules: Preprint," in *Proceedings of SimBuild 2010*, no. NREL/CP-550-47685. New York, NY, USA: National Renewable Energy Laboratory, Aug. 2010.