Zachary E. Lee
Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853
e-mail: zel3@cornell.edu

Qingxuan Sun
School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: qs86@cornell.edu

Zhao Ma
School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: zm64@cornell.edu

Jiangfeng Wang
School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China
e-mail: jfwang@xjtu.edu.cn

Jason S. MacDonald
Ernest Orlando Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720
e-mail: JSMacDonald@lbl.gov

K. Max Zhang[1]
Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853
e-mail: kz33@cornell.edu

Providing Grid Services With Heat Pumps: A Review

The integration of variable and intermittent renewable energy generation into the power system is a grand challenge to our efforts to achieve a sustainable future. Flexible demand is one solution to this challenge, where the demand can be controlled to follow energy supply, rather than the conventional way of controlling energy supply to follow demand. Recent research has shown that electric building climate control systems like heat pumps can provide this demand flexibility by effectively storing energy as heat in the thermal mass of the building. While some forms of heat pump demand flexibility have been implemented in the form of peak pricing and utility demand response programs, controlling heat pumps to provide ancillary services like frequency regulation, load following, and reserve have yet to be widely implemented. In this paper, we review the recent advances and remaining challenges in controlling heat pumps to provide these grid services. This analysis includes heat pump and building modeling, control methods both for isolated heat pumps and heat pumps in aggregate, and the potential implications that this concept has on the power system. [DOI: 10.1115/1.4045819]

Keywords: air conditioning, building, control systems, energy, grid, optimal controls, optimization

1 Introduction

The US electrical grid has experienced a rise in renewable energy generation capacity in recent years, increasing by more than 50% in the past ten years.[2] In addition, some states are beginning to adopt aggressive clean energy goals with high percentages of wind and solar energy. This large and rapid shift in electricity generation sources poses difficult new problems for the electrical grid. Conventional grid operation relies on the practice that generators can be reliably controlled to match electrical supply and demand, while ensuring grid stability. However, with the diminishing percentage of electrical capacity provided by thermal generators and the increasing percentage of variable generation sources like wind and solar, the grid becomes much more difficult to predict and control. Therefore, to maintain a reliable electrical grid in high renewable energy scenarios, the grid requires a significant addition of supporting technology such as energy storage and demand management [1].

A potential source of demand management is through controlling heat pumps. Heat pumps are an efficient, electric source of building heating and cooling. Instead of converting electrical energy directly to heat, e.g., an electric resistance heater, heat pumps use a compressor-driven vapor-compression cycle to move heat from a low-temperature source to a high-temperature sink, which can provide both heating and cooling through the use of a reversing valve. A heat pump's main efficiency metric, the coefficient of performance (COP), is defined as the ratio of the amount of heat moved to the amount of electrical input. The COP is inversely related to the difference between the indoor and outdoor temperatures, and therefore, heat pumps perform poorly in extreme environments, particularly in cold climates. Despite this, recent advancements in heat pump technology have significantly increased the COPs at both extreme high and low temperatures [2,3], expanding heat pump technical feasibility to new geographical regions. However, in many regions of the United States, it is still not economically feasible to replace a natural gas heating system with a heat pump, and given the current electrical generation mix, displacing natural gas heating with a heat pump could actually increase greenhouse gas (GHG) emissions [4]. Nevertheless, heat pump adoption is the cornerstone of many aggressive GHG emission reduction policies, such as New York City's 80×50 plan [5]. Rapid and widespread adoption of heat pumps in areas like this is likely to create significant new operational challenges for the electrical grid, and therefore, these heat pumps must be correctly managed and integrated into an increasingly renewable grid.

As long as indoor thermal comfort is maintained, heat pumps have inherent operational flexibility. This flexibility has already been

[1]Corresponding author.
[2]https://www.eia.gov/totalenergy/data/monthly/index.php

Manuscript received August 1, 2019; final manuscript received December 14, 2019; published online December 20, 2019. Assoc. Editor: Jorge Gonzalez.

harnessed by utilities in the form of thermostat-based demand response programs. These programs generally consist of utilities turning off heat pumps during extreme peak load hours, either through setpoint modification or direct load control. In addition, the use of thermal energy storage (TES) has grown in popularity particularly in Europe and allows for load shifting to accommodate high levels of renewable energy [6]. However, new research shows the potential for heat pumps to provide more complex grid services by operating in ancillary service markets. Ancillary services, which are often provided by controllable thermal generators, are essential for power system stability and maintain the instantaneous balance of electricity supply and demand on the grid. Providing these services often involves following a specific power trajectory sent by the system operator requiring response on the order of seconds to minutes. However, controlling heat pumps to provide ancillary services can require installation of a significant amount of additional hardware. For example, building temperature, heat pump power consumption, external disturbances, and grid signals must all be collected and processed in real time. Much of these data can now be collected and transmitted using Internet of Things devices like smart thermostats and electricity meters. Smart thermostats have seen a rapid rise in adoption [7], and advanced metering infrastructure smart meters have now been installed for 47% of US customers as of 2016 [8]. Harnessing the potential of these devices is a key component in widescale implementation of heat pump ancillary services.

These three driving factors—renewable energy variability, heat pump integration, and smart grid implementation—have sparked many recent studies into the capability of providing ancillary services from heat pumps. Heat pumps have been shown experimentally to have the physical capability of providing ancillary services without significant occupant discomfort [9–12]. Simulations have shown that aggregating together hundreds or thousands of variable-speed or single-stage heat pumps significantly increases their ability to provide ancillary services. However, despite many studies showing the capability and potential of heat pumps to provide ancillary services, to the best of our knowledge, there have been no experimental results for large-scale heat pump aggregation.

While a detailed review of the role of heat pumps in a smart grid was given in Ref. [13], this study reviews the various methods for modeling and controlling heat pumps specifically for ancillary services. Section 2 outlines the various ancillary services that heat pumps can provide. Section 3 describes various methods for modeling and controlling heat pumps both locally and in aggregate. Section 4 shows how heat pumps participate in ancillary service markets. Section 5 analyzes potential performance, capacity, and economics. Section 6 concludes the paper and presents opportunities for the future work.

2 Ancillary Services

The reliability of the electrical grid hinges on the ability of grid operators to match electricity generation and consumption on a variety of timescales and under many contingencies. Grid operators control this balance through several types of ancillary services, which are broadly defined based on their time scale, presented in Table 1. In deregulated markets, grid operators procure these services through ancillary service markets. In contrast to energy markets, where generators are only paid for the energy they produce, ancillary service markets are primarily capacity markets, where a grid operator also pays for the capacity of a generator to alter its production. A technical review of ancillary services is given in Ref. [14], while Ref. [15] gives a review of various US ancillary service market structures. While ancillary services are often provided by generators, they can also be provided through demand response. *Demand response* is the process of controlling the demand to respond to grid signals. Reference [16] describes the role of demand response in ancillary service markets and the effects of market policies on demand response participation. The following text introduces the particular ancillary services that can be provided by heat pumps.

Table 1 Summary of ancillary services that can be potentially provided using heat pumps

Service	Time scale	Details
Frequency regulation	Seconds	Power must track a regulation signal sent every 2–4 s.
Load following	Minutes to hours	Used to balance load on longer time scale than frequency regulation. Can be in response to a grid signal or real-time energy prices.
Reserve	Minutes to hours	Load must curtail within 10 min in response to dispatch signal. Used for contingencies.

Heat pumps can provide ancillary services in a similar way to other energy storage devices like electrochemical batteries or pumped hydroelectric storage. Heat pumps can store energy by injecting or removing heat from the building's thermal mass. For example, in summer, a heat pump can increase its power consumption and charge its storage by removing heat and cooling the building to its lower thermal comfort limit. By doing so, the heat pump now has the flexibility to reduce its future power consumption and allow the indoor temperature to drift up to its upper thermal comfort limit. This increase or reduction in heat pump power consumption results in a net removal or injection of power onto the grid, achieving a similar result to a generator lowering or increasing its power output, respectively. The building then acts as a virtual battery, where the indoor temperature relative to the upper and lower thermal comfort limits acts as a state of charge, and the building's thermal mass acts as a measure of the energy storage capacity [17]. These unique attributes introduce several key control considerations that differentiate heat pumps from generators in providing ancillary services:

(1) *Controlling strategy:* When a generator is required to reduce generation, the heat pump should increase load and vice versa.
(2) *Controlling limits:* Heat pumps must not violate indoor temperature constraints and therefore cannot operate above or below their setpoint for an extended period of time.
(3) *Capacity:* Heat pumps are much smaller than generators and therefore must be aggregated together to satisfy the 100 kW to 1 MW minimum requirement to participate in ancillary service markets[3] [16].

Depending on the service, these differences can have both beneficial and detrimental effects on heat pumps' ability to provide ancillary services. The reminder of Sec. 2 describes the potential services that heat pumps can provide and how their operation differs from a conventional generator.

2.1 Frequency Regulation. A stable grid frequency is ensured by an instantaneous balance between electrical supply and demand. The frequency will drop if demand exceeds supply and will rise if supply exceeds demand. If system frequency drifts more than 1–2 Hz from normal levels (60 Hz in the United States), the equipment can be severely damaged and generators can trip, causing cascading failures [18]. Because of this, frequency regulation requires response on the order of seconds. Generators providing frequency regulation must be equipped with telemetry and control technology to follow an automatic generator control (AGC) signal from the grid operator, which is usually sent every 2–4 s. Demand-side frequency regulation providers can also follow this AGC signal by reducing the load when it calls for an increase in generation.

Frequency regulation is the highest priced ancillary service and is operated primarily as a capacity market. A service provider must bid a certain capacity for regulation often in the day-ahead market and,

[3]PJM is currently the only US operator that allows aggregation for frequency regulation participants.

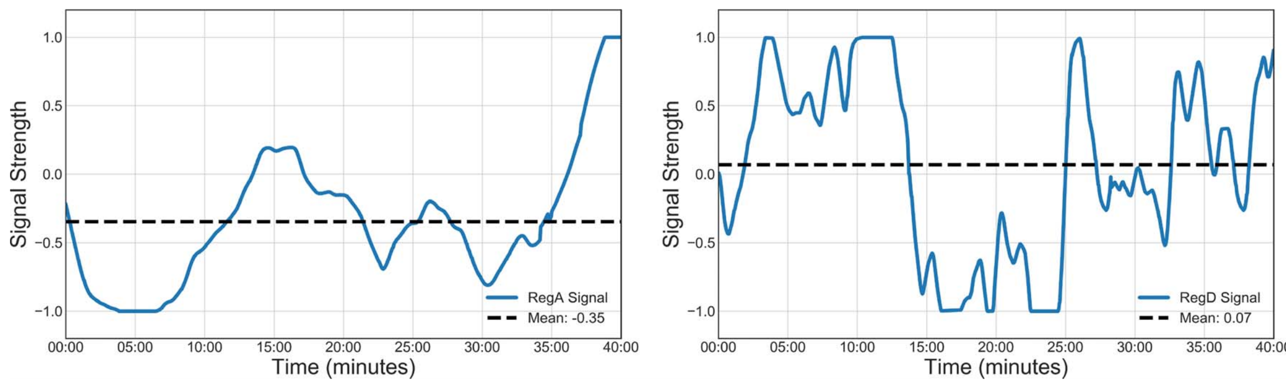

Fig. 1 PJM self-test signals for RegA (left) and RegD (right). RegA has low-frequency fluctuations and a nonzero-mean, making it more suitable for steam generators. RegD contains higher frequency fluctuations and is close to zero-mean, making it more suitable for energy storage and demand response systems.[4]

if accepted, must follow the power signal sent by the system operator. Depending on the system operator, the regulation market either has separate markets for regulation-up and regulation-down or requires symmetric regulation capacity (equal up and down regulation capacity). Currently, California Independent System Operator (CAISO) and Electric Reliability Council of Texas (ERCOT) are the only US system operators that operate separate up and down regulation markets. For generators, these two methods are equivalent [14]. However, for energy storage and load control, significant differences in revenue can occur based on the market type.

Another challenge for demand response and energy storage systems is that the frequency regulation signal is not necessarily zero-mean. For heat pumps providing frequency regulation, a signal bias can cause the heat pump to run consistently below or above its baseline consumption, potentially violating comfort constraints. To resolve this, some system operators have introduced fast regulation signals that are designed to be zero-mean [19]. For example, Pennsylvania, Jersey, Maryland Regional Transmission Organization (PJM) has filtered its signal into two, called RegA and RegD, which are shown in Fig. 1. The RegA signal has a slower time constant and was designed to accommodate steam generators with relatively low ramping capability. RegD consists of higher frequency fluctuations and often converges to zero-mean within 15 min [20]. For this reason, many studies on the technical capability of providing frequency regulation with heat pumps follow the RegD signal [11,12,21,22].

In addition, many US system operators have implemented a pay-for-performance pricing structure in response to FERC Order 755 [23]. In addition to paying for capacity, this pricing structure also pays for mileage and performance. *Mileage*, or movement, is calculated as the sum of the absolute values of the regulation control signal movements and given in ΔMW/MW. Given capacity $P_{\max}$ and power outputs $\{P_1 \ldots P_n\}$, the mileage for n time steps is calculated as [19] follows:

$$M = \sum_{i=1}^{n} |P_i - P_{i-1}|/P_{\max} \tag{1}$$

Performance is given as a score between 0 and 1 and represents how well the participant follows a regulation signal. A frequency regulator must achieve a minimum performance score to qualify, and depending on the market structure, higher performance scores can lead to higher payments. PJM's performance score is often used as a benchmark for frequency regulation control algorithms and is calculated using a combination of three subscores involving delay, correlation, and precision [24]. More information on frequency regulation policies for specific Independent System Operators can be found in Ref. [15].

2.2 Load Following. Load following consists of generators following the slower, more predictable fluctuations in electricity demand on a time scale of several minutes to several hours. This is often procured through economic dispatch, where generators are dispatched according to their generation cost [25]. However, as wind and solar supply an increasing percentage of electricity on the grid, this service could become much more important, particularly for ramping in the mornings and evenings [26]. This could be a potential service provided by heat pumps, either as a reserve capacity similar to CAISO's flexible ramping product [27] or through responsiveness to a real-time price disseminated by the system operator. For example, when solar and wind energy are readily available, electricity prices can drop significantly due to a surplus in supply, encouraging loads to operate during these times. In grids with high solar penetration, such as in the CAISO, there is a growing frequency of negative wholesale electricity prices [28], where generators must pay to produce electricity. This poses a unique opportunity for heat pumps to potentially be paid for operation.

Since heat pumps operate in the retail electricity market, they are often charged a static electricity price, giving no incentive to shift operation toward times of high energy supply. Time-of-use rates, which have predefined price tiers for peak and off-peak hours, have had some success in providing consumers' indirect access for providing a load following service by shifting load away from peak hours [29]. Connected thermostat demand response programs such as Austin Energy's Power Partner^{sm} program[5] have also been widely deployed. These programs allow the utility to turn off heat pumps for short periods of time to reduce the peak load. However, these methods are simplified and therefore do not capture the full potential of heat pumps to provide a load following service.

A second challenge to provide this service is the relatively low frequency of a load following signal. If the frequency of the load following signal is on the same order of the building's thermal response, comfort constraints can be violated [30]. This severely limits the capacity that heat pumps can offer for load following compared with a higher frequency signal like frequency regulation. However, this time constant has the added benefit of reducing the need for fast response controllers.

2.3 Reserve. Power systems are required to maintain a certain amount of reserve margin to ensure reliability in case of contingencies. For example, if a large generator unexpectedly trips, the system might need to dispatch reserves. To provide this service, a heat pump or heat pump aggregation bids a reserve capacity into the reserve capacity market. This contract requires the system to curtail its full capacity offering for a certain amount of time

[4] http://bit.ly/2yieHWA

[5] http://bit.ly/2YvJvlu

determined by the reserve dispatch signal. After the signal ends, the heat pump system can recover back to its baseline energy consumption.

The reserve market can be split into two main categories: spinning and nonspinning reserve. While different system operators can sometimes have different definitions [15], spinning (or synchronous) reserves primarily consist of online generators synchronized to the grid and capable of dispatching to full capacity within 10 min. Nonspinning reserves must respond within 10–30 min but are not necessarily connected to the grid. Providing spinning reserve is preferred over nonspinning reserves for two main reasons. First, heat pumps are already connected to the grid and have high ramping capabilities relative to thermal generators. Second, spinning reserve is priced an order of magnitude higher than nonspinning reserve. However, since reserve dispatch signals result from contingencies, the frequency and the duration can be quite unpredictable. From 2013 to 2018, PJM dispatched spinning reserve anywhere from 0 to 8 times each month with a duration anywhere from 3 to 50 min.[6] For this reason, accounting for uncertainty is a vital component of providing reserve.

3 Modeling and Control

Studies on the modeling and control of heat pumps for ancillary services cover a wide range of scale and complexity. Frequency regulation requires a fast and accurate controller that can track a signal on the order of seconds. Load following controllers can be slower and simpler, while reserve controllers can be as simple as an on/off controller. However, it is important to note that for all ancillary services, the underlying goal is to track a given ancillary service signal. For this reason, many control schemes and methods of determining ancillary service capacity can work for several types of services. Sections 3.1 and 3.2 discuss how heat pumps are modeled and controlled on both local and aggregate levels.

3.1 Local Modeling and Control. On a local level, heat pumps and their buildings can be described by high-fidelity models and directly controlled to follow an ancillary service signal. This often involves directly controlling the fan speed or compressor speed to change the power consumption. Therefore, depending on the type of the system, different models and control methods must be used.

3.1.1 Modeling. There are several types of heat pumps and many different ways to model heat pump systems [31]. For residential applications, local control for ancillary services focuses on variable-speed heat pumps (VSHP). VSHPs modulate the compressor speed to heat or cool the indoor coil. A constant speed fan then blows air over the coil to distribute conditioned air throughout the home. VSHP dynamics are governed by nonlinear mass, momentum, and energy balances of the refrigerant flowing throughout the system [32]. However, these equations are unsuitable for control, and simpler models are required. By using the experimental data from Ref. [32] for many types of VSHPs, Ref. [33] developed simplified steady and dynamic VSHP models. For steady operation, the heat pump power P can be described by,

$$P = k_\omega \omega + k_c T_c + k_e T_e + k_{\text{offset}} \tag{2}$$

where T_C is the ambient air temperature at the condenser, T_e is the ambient air temperature at the evaporator, ω is the compressor shaft speed, and k_i are coefficients that can be fit to the performance data for the specific heat pump using multiple linear regression. The dynamic VSHP model is expressed as the transfer function:

$$\Delta P(s) = \frac{n_{\omega 1} s + n_{\omega 0}}{s^2 + d_{\omega 1} s + d_{\omega 0}} \Delta\omega(s) \tag{3}$$

The coefficients $n_{\omega 1}$, $n_{\omega 0}$, $d_{\omega 1}$, and $d_{\omega 0}$ can similarly be fit from the experimental data. Another simplified model for the fast dynamics of a water-based VSHP is given in Ref. [34]. While this control model assumes a steady-state response, the nonlinear transient dynamics is accounted for with an estimated "lost thermal energy." These simplified models allow for manipulation of compressor speed in control algorithms.

Variable air volume (VAV) heating, ventilating, and air conditioning (HVAC) systems are most often used in large commercial buildings. A heat pump sometimes called a chiller provides a central cooling or heating coil used to condition air, which is then distributed through ducts via a variable-speed fan. Since the coil temperature remains relatively constant, the fan alters its speed to maintain the setpoint. Therefore, this type of the HVAC system uses the fan to provide ancillary services. Fan power $P(t)$ increases with the cube of fan speed $u(t)$ [35]:

$$P(t) = c_1 (u(t))^3 \tag{4}$$

where c_1 is a constant. While the rate of change of the fan speed has inherent limitations from the variable frequency drive to prevent equipment damage, only a time constant of 0.1 s was observed between the controller input and the power output in Ref. [36]. Because of this fast response time, VAV HVAC systems are most often evaluated for frequency regulation. Other similar models for VAV HVAC systems for ancillary services are explained in Refs. [37,38]. Water pumps in water-based heat pump systems can operate in a similar way [39], but are sometimes neglected due to their low energy consumption relative to other components [40].

An accurate building thermal model is also important to determine the amount of thermal energy that can be stored and to prevent violation of thermal comfort constraints. Modeling complexity varies widely based on the building type and size. Detailed reviews on various building modeling techniques are given in Refs. [41,42]. For large commercial buildings, building information modeling (BIM) is often available to provide detailed white box models based on known material properties and building dimensions. However, both accurate identification of each of these parameters and using detailed models for control can be difficult and expensive to obtain. Reference [43] gives a simple method for converting a more complex EnergyPlus [44] model to a reduced-order model usable in model predictive control. Meta-model–based optimization is used in Ref. [45] to identify optimal reduced-order model parameters for a building that are suitable for control.

For smaller buildings or buildings without BIM, gray-box models are often used. The most popular gray-box building modeling technique is through a thermal circuit, sometimes called equivalent thermal parameters. These thermal circuits contain resistors, which represent resistance to heat transfer, and capacitors, which represent heat storage capability. The values of each of these components can be identified from either experimental or physical data [46]. Common circuits for small buildings include either 1R1C (1 resistor and 1 capacitor) or 3R2C. In a 1R1C model, the entire building is lumped into one thermal mass represented by the single capacitance. For a 3R2C, however, the thermal masses of the indoor air and the building material are separate, giving a more accurate prediction over longer time scales. Figure 2 shows an example of a 3R2C model. For larger buildings with many different zones, higher order models containing more capacitors and resistors can also be used [47].

By adding TES to a building, additional thermal capacitance is introduced, significantly increasing the potential for providing ancillary services. The most common type of TES takes the form of water tanks and has been shown to increase the power flexibility for frequency regulation [48], as well as allow flexibility over longer

[6]http://bit.ly/2GwDlaz

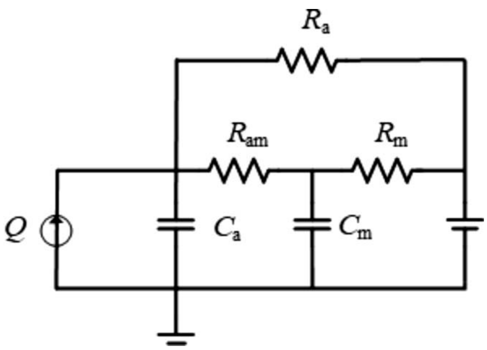

Fig. 2 Example of a 3R2C thermal circuit building model. The subscript *a* represents indoor air temperature, while *m* represents the building mass. *Q* represents the combined heat input from the heat pump and indoor loads.

time scales [49]. Other forms of TES technology involve phase change material, either in a tank coupled with the heat pump or directly embedded in the building construction walls [50]. Since phase change material stays at a relatively constant temperature during operation, additional modeling considerations must be taken into account [51].

3.1.2 Control. Based on the heat pump system, various components can be controlled to alter the power consumption. Feedback controllers are typically used for local control, but common difficulties in implementation are determining optimal controller gains and accounting for time delays.

In Ref. [30], a commercial VAV HVAC system was experimentally shown to be capable of following a frequency regulation signal through control of the fan. The signal was first filtered to exclude low frequencies and high amplitude oscillations. Low frequencies that are of similar order to the building's thermal response can cause temperature constraint violations, while high amplitude oscillations can have harmful effects on the fan's reliability, decreasing its useful life. By perturbing the existing controller's fan speed and airflow setpoints, this controller was able to achieve PJM performance score of 0.83, exceeding PJM's test performance requirement of 0.75. The fan speed for a commercial VAV HVAC system was also controlled to provide frequency regulation in Refs. [9,10]. In this study, the authors use a novel switched controller to maximize speed while ensuring stability. If the desired power output is within some error tolerance from the existing output, a standard proportional-integral (PI) controller is used. Otherwise, a model-based feed-forward controller is used. This controller resulted in much higher test performance scores between 0.94 and 0.98.

For a VSHP, the compressor consumes a majority of power and can be controlled to provide ancillary services. However, due to manufacturer limitations, it is usually difficult to control the compressor directly. In Ref. [11], the supply water temperature for an air-to-water VSHP was used to control the power consumption using a PI controller with nonlinear signal processing to ensure stability. While controlling supply water temperature setpoints was not as effective as simulations involving the direct compressor speed control, the controller was still able to achieve performance scores around 0.8. In Ref. [21], the VSHP compressor was directly controlled using feedback controllers and operated in a small-scale experimental microgrid, showing the feasibility of participation with other distributed energy resources.

3.2 Aggregate Modeling and Control. By aggregating together many heat pumps, the combined capacity of ancillary services can be greatly increased. However, in aggregate heat pump control, the detailed parameters of each individual building and heat pump are difficult to obtain. Therefore, aggregate control studies often contain high-level control schemes using simplified heat pump and building models. The main objective in aggregate control is to determine which heat pumps to modulate to accurately track an ancillary service signal while maintaining thermal comfort and reliability constraints. Note that while these aggregation control studies assume that each heat pump serves a single building, district heating and cooling systems can also provide ancillary services while serving an aggregation of buildings. These systems are much larger and more complex, and a review of controlling district heating and cooling systems for grid services is given in Ref. [52].

3.2.1 Modeling. Early work on controlling heat pump aggregations modeled single-stage heat pumps as thermostatically controlled loads (TCLs), which cycle on and off to maintain temperature within a deadband. TCLs, which also include water heaters, space heaters, and refrigerators, have inherent operational flexibility allowing the power to be modulated to track an ancillary service signal. The general TCL model for cooling is given as follows [53]:

$$m_{t_{n+1}} = \begin{cases} 0, & T_{t_n} < T_- \\ 1, & T_{t_n} > T_+ \\ m_{t_n}, & \text{otherwise} \end{cases} \tag{5}$$

where m_t is a binary variable representing the state of the TCL, T_- and T_+ are the lower and upper temperature limits, and T_{t_n} is the thermostat temperature. The thermostat temperature response can then be modeled according to the individual building and heat pump model.

Due to the simplicity of this model, heat pumps are often modeled using constant COPs, providing a constant amount of heat regardless of external conditions. Buildings containing these TCLs were most often modeled using a 1R1C thermal circuit model. Reference [54] presented an example of the 1R1C model, which describes the internal temperature as follows:

$$T_i(t) = \frac{1}{C_i R_i}(T_{\infty,i} - T_i(t) - s_i(t) R_i P_i), \quad i = 1, 2, \ldots, N_L \tag{6}$$

where $s_i(t) \in 0, 1$ is the on/off signal of the ith TCL. T_i, C_i, and R_i show the temperature, thermal capacitance, and resistance, respectively.

TCL aggregations are often modeled as a virtual battery, with both power and energy capacities. The power capacity is the instantaneous flexibility that TCLs can provide, while the energy capacity is related to the cumulative time that TCLs can operate above or below its baseline. Virtual battery models for a TCL aggregation are given in Refs. [55–57]. In Ref. [57], a method of characterizing the aggregate flexibility of a large collection of TCLs is given through a generalized battery model. The models were grouped into two types: (1) individual models of TCLs to model temperature and power consumption and (2) a generalized battery model that characterizes flexibility. The set of acceptable perturbations of each TCL $\mathbb{E}^k$ is given by

$$\mathbb{E}^k = \left\{ e^k(t) \middle| \begin{array}{l} 0 \le P_0^k + e^k(t) \le P_m^k \\ P_0^k + e^k(t) \text{ maintains } |\theta^k(t) - \theta_r^k| \le \Delta^k \end{array} \right\} \tag{7}$$

where $e^k(t)$ is an acceptable perturbation such that the perturbation will not cause the power P_0^k to exceed its maximum P_m^k and that the temperature $\theta^k(t)$ maintains a distance Δ^k from the setpoint θ_r^k. The total flexibility $\mathbb{U}$ is then defined as the Minkowski sum:

$$\mathbb{U} = \sum_k \mathbb{E}^k \tag{8}$$

3.2.2 Control. The control of TCLs for ancillary services has been widely studied [53,57–67]. In Refs. [53] and [54], a feedback controller was used to control a global thermostat setpoint that turns on or off a certain number of TCLs based on statistical state predictions. This method is difficult in practice, though, as it can rely on setpoint changes down to 0.0025 °C, which is far below the measurement resolution for thermostats. In Ref. [57], a priority stack control method was used to directly control TCL status. This method prioritized turning on or off the TCLs that were closest to automatically turning on or off, respectively. Finally, Ref. [66] explored the stability of TCLs as a result of significant perturbations during control for demand response.

However, the majority of these TCL controllers use simplified, simulated models that neglect many important differences between heat pumps and other TCLs like electric heaters. For example, to avoid damaging the compressor and reducing efficiency, heat pumps have minimum on and off times, which can be the most financially and physically limiting factor for ancillary service provision [68]. Moreover, heat pump COP can vary drastically even among the same heat pump model [69]. Finally, there are many different types of heat pumps, including VSHPs, which do not follow the standard TCL model. Because of these additional complexities, the use of heat pump aggregations for grid services has not been commercially implemented in the same way that other TCLs like water heaters have been implemented [70].

A solution to the minimum off time is given in Ref. [71], which adds an additional "lock-out" state between the on and off states. Variable-speed heat pumps are used in Refs. [72] and [34] by dividing a frequency regulation signal equally among each heat pump. A rule-based controller is used in Ref. [68] to provide frequency regulation from an aggregate of ground-source heat pumps in conjunction with thermal energy storage. Finally, Ref. [64] shows the effect that changes in ambient temperature can have on a population of air conditioners functioning as TCLs.

For ancillary services that require fast response like frequency regulation, control and communication delays can become a serious issue. For aggregations, a reference signal must be received from the system operator and processed to determine the corresponding control decision, and then the control decision is distributed to each heat pump. Moreover, for control systems that communicate with the thermostat rather than the heat pump directly, uncertain time delays can accumulate based on internal thermostat and heat pump control systems. Without delay compensation, tracking accuracy was found to be reduced by as much as 40% for a 20-s delay in Ref. [73]. However, a Kalman filter-based state estimation technique was used in Ref. [74] to mitigate this effect and produce no performance deterioration for delays up to 20 s.

While these heat pump aggregation studies are beginning to include more realistic constraints, they still require some significant assumptions, and there is little experimental validation. For example, the transient power profile of heat pumps and heat pump reliability considerations are relatively unexplored and are an avenue for further research.

4 Market Participation

While Sec. 3 describes control methods for providing ancillary services, the heat pump must establish both a baseline and flexibility capacity to bid into either the day-ahead or real-time electricity markets [16]. A *baseline* is the future power trajectory that the heat pump plans to follow for the length of the ancillary service contract. A *capacity*, sometimes called flexibility, is the amount of power that the heat pump can go above or below its baseline without violating constraints. This is an important difference between generators and heat pumps providing ancillary services. A generator can operate indefinitely within its declared power capacity limits and thus can ignore the energy impact of the ancillary service signal, i.e., the generator can run at 10% above its baseline for an indefinite amount of time if required. A heat pump cannot do this without eventually violating temperature constraints. Therefore, the amount of capacity that a heat pump can offer for ancillary services is heavily dependent on the energy content of the ancillary service signal.

4.1 Baseline. In the context of ancillary services, a baseline is analogous to a generator setpoint and must be determined ahead of time such that the contracted ancillary service capacity can be maintained. This baseline definition is slightly different from a traditional demand response counterfactual baseline, which uses the historical data to estimate what the unmodified energy consumption would have been to measure the amount of demand response provided. In contrast, an ancillary service baseline can be decided by the ancillary service provider based on market and weather conditions to optimize a user-defined objective. Model-predictive control (MPC) is among the most widely used methods to determine an ancillary service baseline. MPC is an iterative control scheme that optimizes a model-based objective function over a given time horizon. The optimal control for the first time step is then implemented, and the MPC reoptimizes with updated inputs. Possible optimization objective functions could be to maximize total profit, maximize thermal comfort, or a combination of the two.

There is a large amount of research on determining optimal power trajectories for heat pump systems [75]. However, it is important to note that the energy optimal power trajectory does not always provide an adequate flexibility for providing ancillary services. In Ref. [76], a contract for declaring a baseline and flexibility capacity for ancillary services in real time is given. A robust MPC determines a baseline and flexibility determination that minimizes the energy cost minus the ancillary service revenue. One key feature in this contract is that the building owner pays only for its baseline energy consumption and not for the altered consumption based on an ancillary service signal, hedging the utility and building owner from any nonzero-mean ancillary service signal.

However, the uncertainty of disturbance predictions and the fidelity of the model can significantly degrade the performance and must be carefully considered. Common prediction methods for disturbances for heat pump control include numerical weather predictions, occupancy schedules [77,78], autoregressive regression, and neural networks [79,80]. The effect of model fidelity on MPC performance was explored in Ref. [46].

4.2 Capacity Determination Without Uncertainty. As previously stated, the flexibility available at a given time step is heavily dependent on the content of the ancillary service signal in previous time steps. One way to simplify this analysis is to assume that the ancillary service signal is zero-mean over the time step, which allows for independent time-wise optimization of flexibility capacity, i.e., each time step does not depend on the ancillary service signal from the previous time step. Since this method does not consider any uncertainty of the mean of the ancillary service signal, it is the most aggressive capacity determination method and can potentially overestimate the actual capacity available. For fast frequency regulation signals such as PJM's RegD, this assumption can be valid since it is designed to be zero-mean over a 15-min period [20]. However, for slower frequency signals such as RegA, load following, and reserve, this method can be unfeasible.

The limitation of this assumption is often addressed by calculating the general flexibility characteristics of a heat pump or building. Reference [81] develops a flexibility index suitable for control on both an individual and aggregate level. Thermal energy storage is added in Ref. [48] to increase the flexibility of a heat pump. Finally, Ref. [82] determines the load reduction flexibility using behind the meter electricity data. By developing battery-like models for flexibility, these types of studies provide the basis for modeling heat pump flexibility for control.

4.3 Capacity Determination With Uncertainty. There are two primary methods of accounting for uncertainty during capacity determination: robust and scenario based. Robust determination is the most conservative approach. This approach ensures that the flexibility offered by the heat pump can be met under the worst case ancillary service signal or disturbances. This method is of particular importance in providing reserve, since the heat pump must be able to reduce its full capacity offering for an unknown amount of time.

Robust distributed optimization is used in Ref. [83] for day-ahead and intra-day scheduling of flexibility capacity for an aggregation of flexible loads. Reference [72] uses robust MPC to determine flexibility capacity for frequency regulation while considering uncertainty in both external disturbances and the frequency regulation signal. Reference [84] provides a robust control strategy for managing uncertain communication time delays for an aggregation.

Another way of dealing with uncertainty is scenario-based optimization. In this method, the capacity determination must not violate temperature constraints under a set of disturbance scenarios that are developed based on historical conditions. By satisfying a certain number of these scenarios, the controller can provide the flexibility it offers with a certain confidence level [85]. While this can be computationally intensive, scenario-based optimization can provide a less conservative flexibility capacity than robust optimization while still considering uncertainty. Reference [86] gives a scenario-based MPC for determining optimal energy consumption of a building, while Ref. [87] gives a scenario-based method for determining the flexibility of a population of controllable loads. Research on accounting for uncertainty for heat pumps in both local and aggregate control are relatively limited, and this is an area for the future work.

4.4 Hierarchical Control. Since MPC requires optimization of a sometimes complex objective function, it alone is not fast enough to ensure response to fast ancillary service signals. Many studies use a hierarchical control scheme to solve this problem [9,34,47,83,88–91]. This hierarchical control scheme combines the strategies for local and aggregate control with prediction methods used for baseline and capacity determination. For example, a three-tier hierarchical controller was used in Ref. [89] to control an aggregation of single-stage heat pumps consisting of (1) a load aggregator that interacts with the power system and ancillary service markets, (2) a central controller that prioritizes which heat pumps to turn on or off, and (3) a local controller that considers local constraints. Figure 3 shows a common layout for hierarchical controllers.

Level 1 is sometimes referred to as a virtual power plant (VPP) and acts as the interface to the grid. From a power system operator's perspective, a VPP acts and is controlled similar to a conventional power plant: It bids into day-ahead ancillary service markets and its aggregate power responds to grid control signals. The VPP passes grid signals to the central controller, level 2, for real-time aggregate control. The central controller can take the form of various aggregate control schemes outlined in Sec. 3.2.1. The control signal sent from the central controller to the local controller, level 3, can take the form of setpoint change or direct load control. The local controller then responds to this control in accordance with local constraints and disturbances. Together, these controls allow an aggregation of small, distributed heat pumps to provide ancillary services to the grid as if it were a large-scale energy storage resource.

5 Performance, Capacity, and Economics

While heat pumps have the physical capability to provide ancillary services to the grid, whether there is an adequate economic incentive to do so is still an open question. With the vast amount of heat pumps already in operation, there is an enormous potential capacity available for ancillary services. However, the revenues from providing services do not always justify the accompanying capital costs and potential efficiency losses. Therefore, a holistic view of costs and performance comparison with other energy storage technologies must be considered to determine whether providing ancillary services is attractive to both heat pump owners and grid operators.

5.1 Performance and Capacity. TCLs have been both experimentally and numerically shown to have potential capacity to provide ancillary services [9,10,57,92,93]. Reference [94] calculates that the ancillary service capacity provided by residential, such as refrigerators, heat pumps, and electric water heaters, can reach 10–40 GW and 8–12 GWh in California, which can easily satisfy the energy storage mandate of 1325 MW to support their renewable portfolio. This estimated capacity was heavily dependent on the climate zone: Some of the zones could only provide flexibility during either winter or summer, while those in more balanced climates could provide a higher average capacity throughout the year. While a large amount of capacity is estimated to be available, Ref. [95] concludes that given current technology and regulatory frameworks, widespread utilization of this flexibility is insufficient for high renewable energy portfolios.

However, using heat pumps as a form of energy storage is not necessarily 100% efficient. Perturbing the power consumption to follow an ancillary service signal can consume extra energy due to excessive cycling or modulation. One key efficiency metric used to rate a variety of grid-scale energy storage devices is the round-trip efficiency (RTE). For conventional energy storage devices like batteries, RTE is defined as the ratio of energy released to energy stored during a charge/discharge cycle. RTEs for common

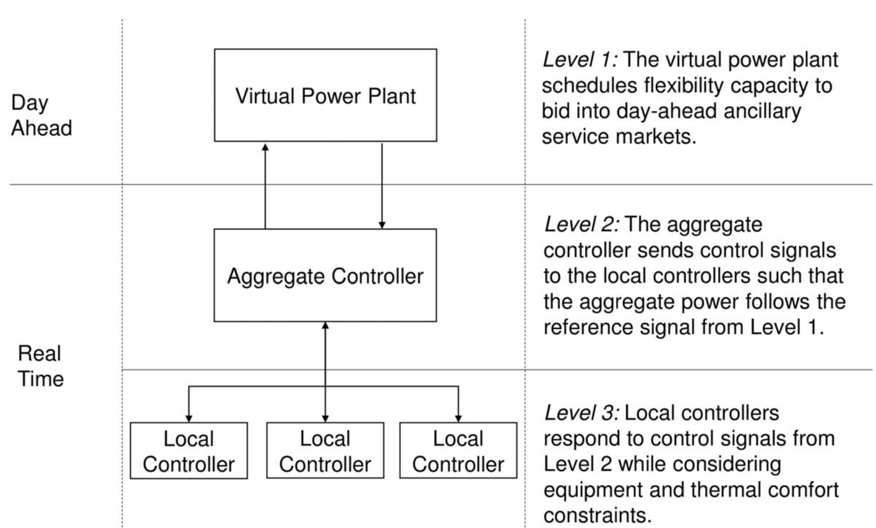

Fig. 3 Common control hierarchies to provide ancillary services from a system of aggregated heat pumps

energy storage devices include redox flow batteries (65–85%), lithium-ion batteries (85–95%), flywheels (93–95%), and pumped hydro storage (70–82%) [96]. For a heat pump providing a symmetrical ancillary service request, the RTE can be defined similarly [92],

$$\mathrm{RTE} = \frac{E_{\mathrm{out}}}{E_{\mathrm{in}}}$$

where E_{out} is the energy reduction with respect to the baseline due to the ancillary service and E_{in} is the increase with respect to the baseline. In calculating RTE, the baseline is set to be the counterfactual baseline or the amount of power that the heat pump would have consumed without providing the service. Therefore, for RTEs less than 1, there is additional energy consumption associated with providing the service.

Several studies have experimentally tested the RTE performance of single heat pumps following regulation service signals with very different results. In Ref. [92], an experimental study controlling a VAV HVAC system to provide a fast, symmetrical service, similar to a charge/discharge cycle in a battery, found that the extra energy consumption was significant. The RTE was only 46% for fan power and 42% for the combined power of the chiller and fans. While this RTE seems low, analysis of space conditioning data from Ref. [97] gave almost identical RTEs at around 46% [92]. Relatively low RTEs were also found in Ref. [98], where experimentally controlled VAV HVAC systems showed RTEs ranging from 34% to 81%. Both experimental studies relied on open-loop global temperature setpoint control mechanisms, in contrast to the MPC approaches previously discussed. However, Refs. [9,10] found that the energy loss associated with following the much faster PJM frequency regulation, RegD, signal was negligible.

The causes of inefficiency were explored through physics-based modeling in Ref. [99], which gave RTEs of less than 100% when the power is increased and then decreased, but greater than 100% when the power is decreased and then increased. This effect can be explained by differences in efficiency due to indoor air temperature variation. Furthermore, Ref. [100] found that when the HVAC system is repeatedly used, the RTE converges to 100%. They attributed the low RTE values reported from experiments [92] to the fact that the experiment ran only one cycle. Therefore, more experimental results are required to accurately define the RTE for a heat pump.

In addition to RTE, there are efficiency losses associated with providing flexibility capacity. To provide flexibility, the heat pump might need to deviate from the energy optimal control schedule. The amount of energy increase compared with an energy optimal controller in Refs. [9,10] was 68 % for the fan and 11% for the chiller. However, by including payments for providing ancillary services, this controller provided the cost optimal solution despite increases in energy. Moreover, Ref. [34] found that the ratio of reserve payment to electricity cost must be above a threshold to incentivize deviating from the energy optimal control to provide flexibility for ancillary services.

This wide variety of results show that there is still no consensus on the total efficiency of a heat pump providing ancillary services. They reveal that the 100% efficient assumption may not be justified in control simulations, and flexibility capacity could be significantly overestimated. In addition, the ancillary service efficiency of an aggregation of heat pumps, as well as variable-speed and single-stage heat pumps, is relatively unstudied. Therefore, more experimental work is needed to determine how potential efficiency losses affect the actual performance of heat pumps providing ancillary services.

5.2 Economical Potential. By receiving payments for providing ancillary services, heat pump owners can have additional revenue streams, reducing the net present cost of heat pump installations. These revenue streams are modest but not negligible. Table 2 presents a summary of potential revenues for a variety of heat pump types, locations, and markets. Revenue varies significantly depending on type of load, climate zone, and regional ancillary service prices. In Ref. [101], residential heat pumps providing frequency regulation in a TCL model were estimated to earn \$1–52/unit/year for cooling and \$11–46/unit/year for heating under the pay-for-performance pricing structure. The wide range of variation is primarily due to the difference in the climate zone. For example,

Table 2 Revenue summary of ancillary service provision by heat pumps and other TCLs

Reference	Market	Benefit	Details
Reference [12]	PJM—RegA and RegD	Offsets 46% of the electricity cost for RegA	2–4.75 kW VSHP power
		Offsets 56% of the electricity cost for RegD	
Reference [105]	PJM—RegA and RegD	Offsets 20–48% of the electricity cost	44.0-kWth variable-speed rooftop unit
			35.2-kWth split heat pump
Reference [101]	CAISO—regulation market	AC: \$0.31–9.36/kW/year	AC with electrical capacity of 4–7.2 kW
		HP: \$2.04–8.31/kW/year	HP with electrical capacity of 4–7.2 kW
		Water heater: \$33.72/kW/year	Water heater with electrical capacity of 4–5 kW
		Refrigerator: \$36/kW/year	Refrigerator with electrical capacity of 0.1–0.5 kW
Reference [94]	CAISO—regulation market	AC: \$0–5.71/kW/year	AC with electrical capacity of 4–7.2 kW
		HP: \$3.93–10/kW/year	HP with electrical capacity of 4–7.2 kW
		Electrical heater: \$5.33/kW/year	Electrical water heater with electrical capacity of 4.5 kW
		Refrigerator: \$31.43/kW/year	Refrigerator with electrical capacity of 0.2–0.5 kW
Reference [68]	Germany—residential frequency reserve	Not financially viable	Electrical storage system of 5 kWh
			3.7 kW heat pump
			Water heat storage of 400 L
Reference [106]	Netherlands—frequency containment reserve	\$26.56/kW/year in "always available" scenario	Heat pump with electrical capacity of 0.5 kW
		\$115.44/kW/year in "always reliable" scenario	

Note: AC, a heat pump providing air conditioning; HP, a heat pump in heating mode.

heat pumps in more extreme climates like Bakersfield and Sacramento, CA, could earn significantly more than those in mild climates like San Francisco, CA.

Spinning reserve revenues are significantly lower due to the much lower spinning reserve capacity prices. Spinning reserve revenues were estimated to be less than $5/unit/year in Ref. [94] and therefore is not attractive under current market policies. There are relatively few revenue studies specifically for load following, but significant energy costs savings are possible by indirect participation though dynamic energy pricing and thermostat-based utility demand response programs. For example, electricity costs were reduced by up to 30% using a price-based controller in a real-time retail electricity market [102]. Utility demand response programs primarily used for reducing peak load also give monetary incentives. The SmartAC™ program of PG&E (Pacific Gas and Electric Company) provided one-time signup bonus of $50 to each participating unit.[7] The OnCall™ program of Florida Power and Light Company provides a monthly credit on bill, totaling up to $83 annually for each participating unit.[8]

However, these revenues must be compared with both instrumentation costs and opportunity costs for providing services. Basic telemetry devices are needed to connect the heat pump to the grid or aggregator, including a real-time electricity meter and controllable thermostat. Reference [101] estimated that this instrumentation could cost between $100 and $250. In addition, heat pumps could be incentivized to consume more energy during times of high ancillary service prices to provide more service despite the possibility of high energy prices or less-efficient operating conditions. A opportunity cost model was given in Ref. [103] that provides a rational goal for optimizing energy consumption, benefit, and ancillary service provision.

Given these revenue and cost results, providing ancillary services may not be attractive for many heat pump owners. Policy changes or price increases could have a positive impact on adoption. For example, CAISO doubled their regulation requirements in Feb. 2016 in response to increasing levels of intermittent renewable energy [104]. This roughly tripled the regulation price, and it has continued to increase each year. Since previous studies referenced in this paper use now outdated price data, future price trends should be taken into account when assessing economic feasibility. Other policy changes that provide energy storage or demand response specific ancillary services such as PJM's RegD and the pay-for-performance market structure could also play a part in increasing heat pump participation.

6 Conclusion

Heat pumps can be controlled to provide stability to the electrical grid in the form of ancillary services. These services range from response on the order of seconds to hours, and heat pumps can be paid for this provision. Local control of VSHPs and VAV HVAC systems has been experimentally shown to track the fastest ancillary service signal, frequency regulation. Aggregations of heat pumps have been numerically shown to be able to provide a variety of ancillary services. Heat pumps also have some key advantages compared with other energy storage systems and generators providing ancillary services, such as reduced costs, increased cycle life, and higher ramp rates.

While a large amount of research has proven the capability for heat pumps to provide ancillary services, there are still significant challenges to large-scale implementation. Recommendations for future research are as follows:

(1) Experimental results are primarily on a local scale, controlling only a single heat pump rather than an aggregation. To our knowledge, there are no experimental heat pump aggregation studies. As a result, single-stage heat pumps, which represent a majority of residential heat pumps, have not been experimentally shown to be capable of providing ancillary services.
(2) Dealing with uncertainty is vital for accurate forecasting of flexibility capacity and is relatively unstudied. Stochastic optimization techniques like robust and scenario-based optimization should also be considered when determining flexibility.
(3) Aggregate control models, specifically for single-stage heat pumps, are relatively simple and do not capture the full dynamics of individual heat pumps and their buildings. Better parameter identification methods and higher order models that are scalable to heat pump aggregations could significantly improve flexibility estimation and ancillary service tracking.
(4) Efficiency losses due to both ancillary service tracking and capacity scheduling are not completely understood. Gaps still remain between experimental and simulation results, and therefore RTE is not well defined. A high RTE is an underlying assumption in many control simulations and therefore has broad implications.
(5) Communication latency issues are a significant barrier to frequency regulation since the system must respond on the order of seconds. Predictive methods or hardware retrofits could be a potential solution.
(6) Revenue estimates are still quite low and represent a barrier to implementation. Trends in ancillary service prices should be considered, as well as new policy and incentive structures.

[7]http://bit.ly/2LPd6k5
[8]http://bit.ly/332Ttug

Acknowledgment

The authors acknowledge the support from the National Science Foundation (NSF) (Grant No. 1711546).

References

[1] Shaner, M. R., Davis, S. J., Lewis, N. S., and Caldeira, K., 2018, "Geophysical Constraints on the Reliability of Solar and Wind Power in the United States," Energy Environ. Sci., **11**(4), pp. 914–925.
[2] Aldrich, R., Grab, J., and Lis, D., 2017, *Northeast/Mid-Atlantic Air-Source Heat Pump Market Strategies Report 2016 Update*, Northeast Energy Efficiency Partnerships.
[3] Chua, K. J., Chou, S. K., and Yang, W. M., 2010, "Advances in Heat Pump Systems: A Review," Appl. Energy, **87**(12), pp. 3611–3624.
[4] Billimoria, S., Henchen, M., Guccione, L., and Louis-Prescott, L., 2018, *The Economics of Electrifying Buildings*, Rocky Mountain Institute, Basalt, CO.
[5] Chaudet, E., 2017, *New York City's Roadmap to 80 × 50*, New York City Mayor's Office of Sustainability.
[6] Arteconi, A., Hewitt, N. J., and Polonara, F., 2013, "Domestic Demand-Side Management (DSM): Role of Heat Pumps and Thermal Energy Storage (TES) Systems," Appl. Therm. Eng., **51**(1–2), pp. 155–165.
[7] Rotondo, J., Johnson, R., Gonzalez, N., Waranowski, A., Badger, C., Lange, N., Goldman, E., and Foster, R., 2016, *Overview of Existing and Future Residential Use Cases for Connected Thermostats*, Building Technologies Office, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy.
[8] Mey, A., and Hoff, S., 2017, *Nearly Half of All Us Electricity Customers Have Smart Meters*, US Energy Information Administration, Washington, D.C.
[9] Vrettos, E., Kara, E. C., MacDonald, J., Andersson, G., and Callaway, D. S., 2018, "Experimental Demonstration of Frequency Regulation by Commercial Buildings Part I: Modeling and Hierarchical Control Design," IEEE Trans. Smart Grid, **9**(4), pp. 3213–3223.
[10] Vrettos, E., Kara, E. C., MacDonald, J., Andersson, G., and Callaway, D. S., 2018, "Experimental Demonstration of Frequency Regulation by Commercial Buildingsp Part II: Results and Performance Evaluation," IEEE Trans. Smart Grid, **9**(4), pp. 3224–3234.
[11] Kim, Y. J., Fuentes, E., and Norford, L. K., 2016, "Experimental Study of Grid Frequency Regulation Ancillary Service of a Variable Speed Heat Pump," IEEE Trans. Power Syst., **31**(4), pp. 3090–3099.
[12] Cai, J., Braun, J. E., and Lin, Y., 2018, "Laboratory Demonstration of Variable-Speed Packaged Air-Conditioning Units for Provision of Frequency Regulation Services," 2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, Feb. 19–22, IEEE, pp. 1–5.
[13] Fischer, D., and Madani, H., 2017, "On Heat Pumps in Smart Grids: A Review," Renewable Sustainable Energy Rev., **70**, pp. 342–357.

[14] Kirby, B., 2016, *Ancillary Services: Technical and Commercial Insights*, Wärtsilä.
[15] Zhou, Z., Levin, T., and Conzelmann, G., 2016, "*Survey of U.S. Ancillary Services Markets*," Argone National Laboratory, Lemont, IL.
[16] MacDonald, J., Cappers, P., Callaway, D., and Kiliccote, S., 2012, "Demand Response Providing Ancillary Services," Proceedings of the Grid-Interop Forum., Irving, TX, Dec. 3–6.
[17] Le Dréau, J., and Heiselberg, P., 2016, "Energy Flexibility of Residential Buildings Using Short Term Heat Storage in the Thermal Mass," Energy, **111**, pp. 991–1002.
[18] Makarov, Y. V., Reshetov, V. I., and Stroev, V. A., 2005, "Blackout Prevention in the United States, Europe, and Russia," Proc. IEEE, **93**(11), pp. 1942–1955.
[19] Xu, B., Dvorkin, Y., Kirschen, D. S., Silva-Monroy, C. A., and Watson, J., 2016, "A Comparison of Policies on the Participation of Storage in U.S. Frequency Regulation Markets," 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, July 17–21, pp. 1–5.
[20] Benner, S., 2015, "A Brief History of Regulation Signals at PJM," http://goo.gl/9UUQts
[21] Kim, Y. J., and Wang, J., 2018, "Power Hardware-in-the-Loop Simulation Study on Frequency Regulation Through Direct Load Control of Thermal and Electrical Energy Storage Resources," IEEE Trans. Smart Grid, **9**(4), pp. 2786–2796.
[22] Su, L., and Norford, L. K., 2015, "Demonstration of HVAC Chiller Control for Power Grid Frequency Regulation Part 1: Controller Development and Experimental Results," Sci. Technol. Built Environ., **21**(8), pp. 1134–1142.
[23] FERC, 2011, "Order No. 755: Frequency Regulation Compensation in the Organized Wholesale Power Markets," Federal Energy Regulatory Commission (FERC), WA.
[24] PJM, 2016, *PJM Manual 12: Balancing Operations*, Pennsylvania, Jersey, Maryland Power Pool.
[25] PJM, and DOE, 2012, *Economic Dispatch and Technological Change*, U.S. Department of Energy, Washington, DC.
[26] CAISO, 2019, *What the duck curve tells us about managing a green grid*, California Independent System Operator.
[27] CAISO, 2016, *CAISO Flexible Ramping Product*, California Independent System Operator.
[28] Bajwa, M., and Cavicchi, J., 2017, "*Growing Evidence of Increased Frequency of Negative Electricity Prices in U.S. Wholesale Electricity Markets*," IAEE Energy Forum.
[29] Eid, C., Koliou, E., Valles, M., Reneses, J., and Hakvoort, R., 2016, "Time-Based Pricing and Electricity Demand Response: Existing Barriers and Next Steps," Utilities Policy, **40**, pp. 15–25.
[30] Lin, Y., Barooah, P., Meyn, S., and Middelkoop, T., 2015, "Experimental Evaluation of Frequency Regulation From Commercial Building HVAC Systems," IEEE Trans. Smart Grid, **6**(2), pp. 776–783.
[31] Afram, A., and Janabi-Sharifi, F., 2014, "Review of Modeling Methods for HVAC Systems," Appl. Therm. Eng., **67**(1–2), pp. 507–519.
[32] He, X.-D., 1996, "Dynamic Modeling and Multivariable Control of Vapor Compression Cycles in Air Conditioning Systems," Ph.D. thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.
[33] Kim, Y.-J., Norford, L. K., and Kirtley, J. L., 2015, "Modeling and Analysis of a Variable Speed Heat Pump for Frequency Regulation Through Direct Load Control," IEEE Trans. Power Syst., **30**(1), pp. 397–408.
[34] Vrettos, E., and Andersson, G., 2016, "Scheduling and Provision of Secondary Frequency Reserves by Aggregations of Commercial Buildings," IEEE Trans. Sustainable Energy, **7**(2), pp. 850–864.
[35] ASHRAE, 2016, *The ASHRAE Handbook: HVAC Systems and Equipment*, American Society of Heating, Refrigerating and Air-Conditioning Engineers.
[36] Hao, H., Lin, Y., Kowli, A. S., Barooah, P., and Meyn, S., 2014, "Ancillary Service to the Grid Through Control of Fans in Commercial Building Hvac Systems," IEEE Trans. Smart Grid, **5**(4), pp. 2066–2074.
[37] Blum, D. H., and Norford, L. K., 2014, "Dynamic Simulation and Analysis of Ancillary Service Demand Response Strategies for Variable Air Volume HVAC Systems," HVAC&R Res., **20**(8), pp. 908–921.
[38] Goddard, G., Klose, J., and Backhaus, S., 2014, "Model Development and Identification for Fast Demand Response in Commercial HVAC Systems," IEEE Trans. Smart Grid, **5**(4), pp. 2084–2092.
[39] Wang, H., Wang, S., and Tang, R., 2019, "Investigation on the Use of Pumps in HVAC Systems for Providing Ancillary Services in Smart Grids," Energy Procedia, **159**, pp. 219–224.
[40] Lin, Y., Barooah, P., and Mathieu, J. L., 2017, "Ancillary Services Through Demand Scheduling and Control of Commercial Buildings," IEEE Trans. Power Syst., **32**(1), pp. 186–197.
[41] Li, X., and Wen, J., 2014, "Review of Building Energy Modeling for Control and Operation," Renewable Sustainable Energy Rev., **37**, pp. 517–537.
[42] Kramer, R., van Schijndel, J., and Schellen, H., 2012, "Simplified Thermal and Hygric Building Models: A Literature Review," Front. Arch. Res., **1**(4), pp. 318–325.
[43] Cole, W. J., Hale, E. T., and Edgar, T. F., 2013, "Building Energy Model Reduction for Model Predictive Control Using Openstudio," Proceeding of the 2013 American Control Conference, Washington, DC, June 17–19, pp. 449–454.
[44] Crawley, D. B., Pedersen, C. O., Lawrie, L. K., and Winkelmann, F. C., 2000, "Energyplus: Energy Simulation Program," ASHRAE J., **42**(4), pp. 49–56.
[45] Eisenhower, B., O'Neill, Z., Narayanan, S., Fonoberov, V. A., and Mezić, I., 2012, "A Methodology for Meta-Model Based Optimization in Building Energy Models," Energy Buildings, **47**, pp. 292–301.
[46] Blum, D. H., Arendt, K., Rivalin, L., Piette, M. A., Wetter, M., and Veje, C. T., 2019, "Practical Factors of Envelope Model Setup and Their Effects on the Performance of Model Predictive Control for Building Heating, Ventilating, and Air Conditioning Systems," Appl. Energy, **236**, pp. 410–425.
[47] Wang, D., Parkinson, S., Miao, W., Jia, H., Crawford, C., and Djilali, N., 2013, "Hierarchical Market Integration of Responsive Loads as Spinning Reserve," Appl. Energy, **104**, pp. 229–238.
[48] Finck, C., Li, R., Kramer, R., and Zeiler, W., 2018, "Quantifying Demand Flexibility of Power-to-Heat and Thermal Energy Storage in the Control of Building Heating Systems," Appl. Energy, **209**, pp. 409–425.
[49] Alimohammadisagvand, B., Jokisalo, J., Kilpeläinen, S., Ali, M., and Sirén, K., 2016, "Cost-Optimal Thermal Energy Storage System for a Residential Building With Heat Pump Heating and Demand Response Control," Appl. Energy, **174**, pp. 275–287.
[50] De Gracia, A., and Cabeza, L. F., 2015, "Phase Change Materials and Thermal Energy Storage for Buildings," Energy Buildings, **103**, pp. 414–419.
[51] Al-Saadi, S. N., and Zhai, Z., 2013, "Modeling Phase Change Materials Embedded in Building Enclosure: A Review," Renewable Sustainable Energy Rev., **21**, pp. 659–673.
[52] Vandermeulen, A., van der Heijde, B., and Helsen, L., 2018, "Controlling District Heating and Cooling Networks to Unlock Flexibility: A Review," Energy, **151**, pp. 103–115.
[53] Callaway, D. S., 2009, "Tapping the Energy Storage Potential in Electric Loads to Deliver Load Following and Regulation, With Application to Wind Energy," Energy Convers. Manage., **50**(5), pp. 1389–1400.
[54] Bashash, S., and Fathy, H. K., 2011, "Modeling and Control Insights Into Demand-Side Energy Management Through Setpoint Control of Thermostatic Loads," Proceedings of the 2011 American Control Conference, San Francisco, CA, June 29–July 1, pp. 4546–4553.
[55] Hughes, J. T., Domínguez-García, A. D., and Poolla, K., 2016, "Identification of Virtual Battery Models for Flexible Loads," IEEE Trans. Power Syst., **31**(6), pp. 4660–4669.
[56] Zhao, L., Zhang, W., Hao, H., and Kalsi, K., 2017, "A Geometric Approach to Aggregate Flexibility Modeling of Thermostatically Controlled Loads," IEEE Trans. Power Syst., **32**(6), pp. 4721–4731.
[57] Hao, H., Sanandaji, B. M., Poolla, K., and Vincent, T. L., 2015, "Aggregate Flexibility of Thermostatically Controlled Loads," IEEE Trans. Power Syst., **30**(1), pp. 189–198.
[58] Oldewurtel, F., Borsche, T., Bucher, M., Fortenbacher, P., Haring, M. G. V. T., Mathieu, J. L., Megel, O., Vrettos, E., and Andersson, G., 2013, "A Framework for and Assessment of Demand Response and Energy Storage in Power Systems," IREP Symposium: Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid, IREP, Rethymnon, Greece, Aug. 25–30, pp. 1–24.
[59] Bashash, S., and Fathy, H. K., 2013, "Modeling and Control of Aggregate Air Conditioning Loads for Robust Renewable Power Management," IEEE Trans. Control Syst. Technol., **21**(4), pp. 1318–1327.
[60] Tindemans, S. H., Trovato, V., and Strbac, G., 2015, "Decentralized Control of Thermostatic Loads for Flexible Demand Response," IEEE Trans. Control Syst. Technol., **23**(5), pp. 1685–1700.
[61] Perfumo, C., Kofman, E., Braslavsky, J. H., and Ward, J. K., 2012, "Load Management: Model-Based Control of Aggregate Power for Populations of Thermostatically Controlled Loads," Energy Convers. Manage., **55**, pp. 36–48.
[62] Papadakis, G., Ruiz, N., and Cobelo, I., 2009, "A Direct Load Control Model for Virtual Power Plant Management," IEEE Trans. Power Syst., **24**(2), pp. 959–966.
[63] Mathieu, J. L., Koch, S., and Callaway, D. S., 2013, "State Estimation and Control of Electric Loads to Manage Real-Time Energy Imbalance," IEEE Trans. Power Syst., **28**(1), pp. 430–440.
[64] Mahdavi, N., Braslavsky, J. H., and Perfumo, C., 2018, "Mapping the Effect of Ambient Temperature on the Power Demand of Populations of Air Conditioners," IEEE Trans. Smart Grid, **9**(3), pp. 1540–1550.
[65] Koch, S., Mathieu, J. L., and Callaway, D. S., 2011, "Modeling and Control of Aggregated Heterogeneous Thermostatically Controlled Loads for Ancillary Services," 17th Power Systems Computation Conference., Stockholm, Aug. 22–26, pp. 1–7.
[66] Chertkov, M., and Chernyak, V., 2017, "Ensemble of Thermostatically Controlled Loads: Statistical Physics Approach," Sci. Rep., **7**(1), pp. 1–9.
[67] Chang, C. Y., Zhang, W., Lian, J., and Kalsi, K., 2013, "Modeling and Control of Aggregated Air Conditioning Loads Under Realistic Conditions," 2013 IEEE PES Innovative Smart Grid Technologies Conference, Washington, DC, Feb. 24–28, pp. 1–10.
[68] Rodríguez, L. R., Brennenstuhl, M., Yadack, M., Boch, P., and Eicker, U., 2019, "Heuristic Optimization of Clusters of Heat Pumps: A Simulation and Case Study of Residential Frequency Reserve," Appl. Energy, **233–234**, pp. 943–958.
[69] Korn, D., Walczyk, J., Jackson, A., Machado, A., Kongoletos, J., and Pfann, E., 2016, *Ductless Mini-Split Heat Pump Impact Evaluation*, Cadmus, Boston, MA.
[70] Chen, H., Baker, S., Benner, S., Berner, A., and Liu, J., 2017, "PJM Integrates Energy Storage: Their Technologies and Wholesale Products," IEEE Power Energy Mag., **15**(5), pp. 59–67.
[71] Zhang, W., Lian, J., Chang, C.-Y., and Kalsi, K., 2013, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," IEEE Trans. Power Syst., **28**(4), pp. 4655–4664.

[72] Vrettos, E., Oldewurtel, F., and Andersson, G., 2016, "Robust Energy-Constrained Frequency Reserves From Aggregations of Commercial Buildings," IEEE Trans. Power Syst., **31**(6), pp. 4272–4285.

[73] Hao, H., Sanandaji, B. M., Poolla, K., and Vincent, T. L., 2014, "Frequency Regulation From Flexible Loads: Potential, Economics, and Implementation," Proceeding of the 2014 American Control Conference, Portland, OR, June 4–6, pp. 65–72.

[74] Ledva, G. S., Vrettos, E., Mastellone, S., Andersson, G., and Mathieu, J. L., 2018, "Managing Communication Delays and Model Error in Demand Response for Frequency Regulation," IEEE Trans. Power Syst., **33**(2), pp. 1299–1308.

[75] Afram, A., and Janabi-Sharifi, F., 2014, "Theory and Applications of HVAC Control Systems—A Review of Model Predictive Control (MPC)," Building Environ., **72**, pp. 343–355.

[76] Maasoumy, M., Rosenberg, C., Sangiovanni-Vincentelli, A., and Callaway, D. S., 2014, "Model Predictive Control Approach to Online Computation of Demand-Side Flexibility of Commercial Buildings HVAC Systems for Supply Following," Proceedings of the 2014 American Control Conference, Portland, OR, June 4–6, pp. 1082–1089.

[77] Yuan, S., and Perez, R., 2006, "Multiple-Zone Ventilation and Temperature Control of a Single-Duct VAV System Using Model Predictive Strategy," Energy Buildings, **38**(10), pp. 1248–1261.

[78] Henze, G. P., Kalz, D. E., Liu, S., and Felsmann, C., 2005, "Experimental Analysis of Model-Based Predictive Optimal Control for Active and Passive Building Thermal Storage Inventory," HVAC&R Res., **11**(2), pp. 189–213.

[79] Florita, A. R., and Henze, G. P., 2009, "Comparison of Short-Term Weather Forecasting Models for Model Predictive Control," HVAC&R Res., **15**(5), pp. 835–853.

[80] Lee, Z. E., Gupta, K., Kircher, K. J., and Zhang, K. M., 2019, "Mixed-Integer Model Predictive Control of Variable-Speed Heat Pumps," Energy Buildings, **198**, pp. 75–83.

[81] Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., and Madsen, H., 2018, "Characterizing the Energy Flexibility of Buildings and Districts," Appl. Energy, **225**, pp. 175–182.

[82] Ziras, C., Heinrich, C., Pertl, M., and Bindner, H. W., 2019, "Experimental Flexibility Identification of Aggregated Residential Thermal Loads Using Behind-the-Meter Data," Appl. Energy, **242**, pp. 1407–1421.

[83] Diekerhof, M., Peterssen, F., and Monti, A., 2018, "Hierarchical Distributed Robust Optimization for Demand Response Services," IEEE Trans. Smart Grid, **9**(6), pp. 6018–6029.

[84] Zhu, Q., Jiang, L., Yao, W., Zhang, C.-K., and Luo, C., 2017, "Robust Load Frequency Control With Dynamic Demand Response for Deregulated Power Systems Considering Communication Delays," Electric Power Compon. Syst., **45**(1), pp. 75–87.

[85] Campi, M. C., Garatti, S., and Ramponi, F. A., 2018, "A General Scenario Theory for Nonconvex Optimization and Decision Making," IEEE Trans. Automat. Control, **63**(12), pp. 4067–4078.

[86] Zhang, X., Schildbach, G., Sturzenegger, D., and Morari, M., 2018, "Scenario-Based MPC for Energy-Efficient Building Climate Control Under Weather and Occupancy Uncertainty," 2013 European Control Conference (ECC), Zurich, July 17–19, pp. 1029–1034.

[87] Ottesen, S. O., and Tomasgard, A., 2015, "A Stochastic Model for Scheduling Energy Flexibility in Buildings," Energy, **88**, pp. 364–376.

[88] Gao, Y., Li, S., and Lin, B., 2019, "Economic and Hierarchical Control Multi-thermal Load for Bidding Ancillary Service," Clemson University Power Systems Conference, PSC, Charleston, Sept. 4–7, pp. 1–5.

[89] Wu, X., Member, S., He, J., Xu, Y., Lu, J., Member, S., Lu, N., Member, S., and Wang, X., 2018, "Hierarchical Control of Residential HVAC Units for Primary Frequency Regulation," IEEE Trans. Smart Grid, **9**(4), pp. 3844–3856.

[90] Schirrer, A., Konig, O., Ghaemi, S., Kupzog, F., and Kozek, M., 2013, "Hierarchical Application of Model-Predictive Control for Efficient Integration of Active Buildings Into Low Voltage Grids," 2013 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, Berkeley, May 20, pp. 2–7.

[91] Wei, W., Wang, D., Jia, H., Wang, C., Zhang, Y., and Fan, M., 2017, "Hierarchical and Distributed Demand Response Control Strategy for Thermostatically Controlled Appliances in Smart Grid," J. Modern Power Syst. Clean Energy, **5**(1), pp. 30–42.

[92] Beil, I., Hiskens, I., and Backhaus, S., 2015, "Round-Trip Efficiency of Fast Demand Response in a Large Commercial Air Conditioner," Energy Buildings, **97**, pp. 47–55.

[93] Beil, I., Hiskens, I., and Backhaus, S., 2016, "Frequency Regulation From Commercial Building HVAC Demand Response," Proc. IEEE, **104**(4), pp. 745–757.

[94] Mathieu, J. L., Dyson, M. E., and Callaway, D. S., 2015, "Resource and Revenue Potential of California Residential Load Participation in Ancillary Services," Energy Policy, **80**, pp. 76–87.

[95] Kohlhepp, P., Harb, H., Wolisz, H., Waczowicz, S., Müller, D., and Hagenmeyer, V., 2019, "Large-Scale Grid Integration of Residential Thermal Energy Storages as Demand-Side Flexibility Resource: A Review of International Field Studies," Renewable Sustainable Energy Rev., **101**, pp. 527–547.

[96] Kintner-Meyer, M., Balducci, P., Colella, W., Elizondo, M., Jin, C., Nguyen, T., Viswanathan, V., and Zhang, Y., 2013, "*National Assessment of EnergyStorage for Grid Balancing and Arbitrage*," Pacific Northwest National Laboratory, Richland, WA.

[97] Hammerstrom, D. J., Ambrosio, R., Carlon, T. A., DeSteese, J. G., Horst, G. R., Kajfasz, R., Kiesling, L. L., Michie, P., Pratt, R. G., Yao, M., Brous, J., Chassin, D. P., Guttromson, R. T., Jarvegren, O. M., Katipamula, S., Le, N. T., Oliver, T. V., and Thompson, S. E., 2008, "*Pacific Northwest GridWise™ Testbed Demonstration Projects; Part I. Olympic Peninsula Project*," Pacific Northwest National Laboratory, Richland, WA.

[98] Keskar, A., Anderson, D., Johnson, J. X., Hiskens, I. A., and Mathieu, J. L., 2019, "Do Commercial Buildings Become Less Efficient When They Provide Grid Ancillary Services?" Energy Efficiency, **12**(4), pp. 1–15.

[99] Lin, Y., Mathieu, J. L., Johnson, J. X., Hiskens, I. A., and Backhaus, S., 2017, "Explaining Inefficiencies in Commercial Buildings Providing Power System Ancillary Services," Energy Buildings, **152**, pp. 216–226.

[100] Raman, N. S., and Barooah, P., 2019, "On the Round-Trip Efficiency of an HVAC-Based Virtual Battery," IEEE Trans. Smart Grid, **11**(1), pp. 1–8.

[101] Hao, H., Sanandaji, B. M., Poolla, K., and Vincent, T. L., 2015, "Potentials and Economics of Residential Thermal Loads Providing Regulation Reserve," Energy Policy, **79**, pp. 115–126.

[102] Yoon, J. H., Baldick, R., and Novoselac, A., 2014, "Dynamic Demand Response Controller Based on Real-Time Retail Price for Residential Buildings," IEEE Trans. Smart Grid, **5**(1), pp. 121–129.

[103] Blum, D. H., Zakula, T., and Norford, L. K., 2016, "Opportunity Cost Quantification for Ancillary Services Provided by Heating, Ventilating, and Air-Conditioning Systems," IEEE Trans. Smart Grid, **8**(3), pp. 1264–1273.

[104] CAISO, 2016, *Q1 2016 Report on Market Issues and Performance*, California Independent System Operator.

[105] Cai, J., and Braun, J. E., 2019, "Laboratory-Based Assessment of HVAC Equipment for Power Grid Frequency Regulation: Methods, Regulation Performance, Economics, Indoor Comfort and Energy Efficiency," Energy Buildings, **185**, pp. 148–161.

[106] Posma, J., Lampropoulos, I., Schram, W., and van Sark, W., 2019, "Provision of Ancillary Services From an Aggregated Portfolio of Residential Heat Pumps on the Dutch Frequency Containment Reserve Market," Appl. Sci., **9**(4), pp. 590–606.