A Dynamic Water-filling Method for Real-Time HVAC Load Control Based on Model Predictive Control

Kan Zhou, *Student Member, IEEE* and Lin Cai, *Senior Member, IEEE*

***Abstract*—Heating Ventilation and Air-Conditioning (HVAC) system can be viewed as elastic load to provide demand response. Existing work usually used HVAC to do the load following or load shaping based on given control signals or objectives. However, optimal external control signals may not always be available. Without such control signals, how to make a tradeoff between the fluctuation of non-renewable power generation and the limited demand response potential of the elastic load, while still guaranteeing user comfort level, is still an open problem. To solve this problem, we first model the temperature evolution process of a room and propose an approach to estimate the key parameters of the model. Second, based on the model predictive control, a centralized and a distributed algorithm are proposed to minimize the fluctuation and maximize user comfort level. In addition, we propose a dynamic water level adjustment algorithm to make the demand response always available in two directions. Extensive simulations based on practical data sets show that the proposed algorithms can effectively reduce the load fluctuation.**

***Index Terms*—Model Predictive Control, Demand Response, Smart Grid**

I. Introduction

Demand response, aided by the current information and communication technologies, is anticipated to improve the grid stability and efficiency by interacting with the elastic load at users' side. By changing the elastic load w.r.t. both renewable energy generation and inelastic load variation, demand response can reduce the fluctuation of the non-renewable power generation and thus cut down the power generation cost.

To achieve this goal, existing works can be classified into two categories. In the first category, the authors assumed that how much demand response needed in each time slot is already known. Therefore, the aim of the algorithms is to use demand response to do a load following or load shaping according to a given control signal or control objective [1], [2]. However, in practice, it may be difficult to obtain the optimal control signal, in other words, to know exactly how much demand response is needed for each time slot in the future. As a result, the works in the second category usually assume the availability of some prediction information to help decide how much demand response may be needed. The key problem is that the amount of elastic load that can be adjusted at certain time (we call it "elastic load potential") may be limited. If we use too much elastic load to flatten the non-renewable power generation at the beginning, there may not be enough elastic load to use at a later time. Therefore, a tradeoff must be made between the fluctuation of non-renewable power generation and elastic load potential.

Existing work, such as [3], usually needs accurate long-term load and renewable energy generation information to obtain the optimal non-renewable energy generation, which is called the water level, for each time slot. So how much elastic load is allowed in each time slot in the future can be obtained by simply calculating the difference between the water level and the predicted non-elastic load. The traditional water filling approach is to make the elastic load in each time slot as close to this difference as possible so that the non-renewable energy generation can reach the optimal value. However, without such accurate long-term estimation, we do not know the optimal water level and thus do not know how much elastic load should be adjusted in each time slot.

In this paper, we consider the situation that the amount of elastic load to be used is limited, which is more practical at the early stage of demand response application, and only short-term (10 minutes to half an hour) renewable energy prediction is available. The HVAC systems in large buildings or houses are used as the elastic load. Due to the temperature variation constraints of each room, the HVACs cannot always provide demand response. To efficiently use the limited elastic load potential, we propose a dynamic water-filling approach based on model predictive control (MPC) to schedule the HVACs. Different from the previous works, we dynamically adjust the water level, which is the remaining load (described later) plus the reference of the HVAC load, to help reduce the fluctuation of non-renewable power generation while still keep the elastic load potential for future use. In addition, we use a different HVAC control strategy other than the existing dead band based control policy [4].

The main contributions of this paper can be summarized as follows. First, we propose a centralized algorithm to control heterogeneous HVACs in a micro-grid. The objective of this algorithm is to reduce non-renewable energy generation fluctuations while still guarantee user comfort level. An approach to estimate heterogeneous HVAC model parameters is also proposed. Second, we extend the centralized algorithm to a distributed one, which has a much lower computational complexity and is more scalable. Third, we further extend the proposed algorithms to support HVAC ON/OFF control modes other than adjusting the HVAC power level. Fourth, since the elastic load potential provided by HVACs is limited compared to the unlimited control time, a dynamic water level adjustment algorithm is proposed to reserve this elastic load potential for future demand response. Finally, extensive simulations using practical data sets obtained from Eirgrid

[5] have been conducted to evaluate the performance of the proposed algorithms. The results demonstrate the advantages of the proposed algorithms comparing to existing ones.

The rest of the paper is organized as follows. Section II discusses the existing demand response approaches and the application of MPC. A general description of the system architecture and HVAC model is given in Section III. Then we discuss the design details of the proposed centralized MPC algorithm, and the dynamic water level adjustment approach in Section IV, and propose a distributed MPC algorithm in Section V. How to support HVAC ON/OFF state control is presented in Section VI. Performance evaluation is given in Section VII, followed by concluding remarks and future research issues in Section VIII.

II. Related Work

Since load fluctuation usually adds cost to power generation and raises requirements on frequency control [6], smoothing or flatting the non-renewable power generation using demand response is one of the most important objectives.

For centralized demand response control, [3] proposed a water-filling approach to flatten the overall load assuming that perfect future load information is available. In [7], Koutsopoulos *et. al* introduced two online demand scheduling policies based on dynamic programming to minimize the long-term average power grid operation cost, without considering the variation of base load and renewable energy. He *et. al* proposed a Plug-in Hybrid Electrical Vehicle (PHEV) scheduling algorithm to minimize the total cost of electrical vehicles using a sliding window algorithm in [7]. In their algorithms, all the local controllers were coordinated by the same predicted base load model with accurate prediction in the whole time scale. [4] is similar to our work in that it aimed to smooth non-renewable electricity supply by controlling the set point and ON/OFF states of all HVACs. The algorithms proposed above try to avoid or minimize fluctuations, so they all need accurate long-term prediction information to make control decisions. They also assumed the availability of enough elastic load to perform the demand response. On the other hand, there are also some existing works which make control decisions after the power imbalance happens, so no future prediction information is needed. For example, [8]–[10] used an energy storage system to provide primary frequency control to the power grid based on the current requirement or historical data. Although the capacity of the primary frequency control system is limited, the secondary frequency control system will help reduce their burden in time. The proposed MPC based algorithms belong to the first group. Taking one step further, our work considers the situation of limited elastic load, and only short-term prediction information is needed.

Thermostatically controlled appliances, such as HVACs have been widely used for demand response in smart grid. In [11], [12], HVAC is used to minimize the economic cost by scheduling its operation time. The application of MPC to HVACs can be found in [13], with an objective to minimize the user discomfort level while keeping the economic cost within a given budget. Karmakar *et. al* introduced an online algorithm which maintained the thermal comfort-bands while keeping the total HVAC load under peak energy consumption constraint [14]. In [1], [2], HVAC is used to provide intra-hour load balancing or load following according to given control signals. Different from the existing work, our control objective is to minimize the fluctuation of the non-renewable power generation without external control signals, which is more challenging.

MPC has been used to solve various control problems in smart grid. [15] proposed an economic MPC algorithm to minimize the total cost of distributed power generation plants. The control actions are adjusting the amount of power generation from each plant. In [16], an aggregator utilizes MPC strategy to track a secondary frequency control signal by controlling heterogenous elastic loads. Different from the previous approaches, the proposed dynamic water level adjustment algorithm will make a tradeoff between the fluctuation of non-renewable power generation and elastic load potential reservation.

III. System Model

A. System Architecture

The investigated system represents a micro-grid with a high renewable energy penetration. It consists of a control center, customers with HVAC installed, and a communication network that connects them together.

The electricity power supply comes from two types of sources: conventional power generators and renewable power generators. Due to its stochastic feature, instantaneous renewable power generation is time-varying, while a good prediction over a short time period is possible [17], especially with the help of large energy buffers (batteries).

The relationship between load and power supply is shown in (1), where $S_n(t)$ is the power generation from conventional power plants at time t, S_r is the renewable energy generation, L_b is the non-HVAC load (also called base load), and L_h is the load from all HVACs.

$$S_n(t) = L_b(t) - S_r(t) + L_h(t). \tag{1}$$

In (1), $L_h(t)$ is the elastic load at time t which can be changed in each time slot; $L_b(t) - S_r(t)$ is the non-elastic load minus renewable energy generation (we call it "remaining load"). Due to the intermittent nature of renewable energy, the renewable power generation always contains a lot of fluctuations which directly affect the remaining load. As a result, the conventional power companies will need high spinning reserve or buy extra frequency regulation service to do the frequency control in the micro-grid which usually has a high cost [6]. Therefore, in this paper, we are motivated to make $S_n(t)$ change slowly and smoothly by controlling the elastic load in each time slot so the conventional power companies can save cost and improve the system efficiency.

To determine the optimal HVAC load, we also need to know the future base load. Although there is no long-term load prediction algorithm with good accuracy, short-term load prediction algorithms do exist [18]. In this work, we assume that inelastic load can be predicted in a short-term.

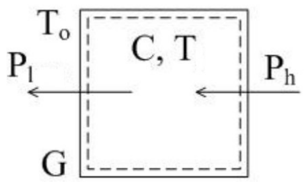


Fig. 1. HVAC model [19]

B. HVAC Model

In this paper, the HVAC model is obtained from [19] and is briefly reviewed as follows.

A simple HVAC model is illustrated in Fig. 1. In this model, P_h represents the power consumed by the HVAC in the unit of $Watt$; P_l represents the amount of heat transferred outdoors through the house boundaries in the unit of $Watt$; T and T_o represent the indoor and outdoor temperature respectively in the unit of $Kelvin$; C represents the effective heat capacity, which is the product of air heat capacity and air quantity in a house, in the unit of $Joule/Kelvin$; G represents the thermal insulation level of a house and is in the unit of $Watt/Kelvin$.

In this paper, time is divided into slots with a fixed duration of t. Define $T_i(k)$ as the indoor temperature of the house with HVAC i at the k-th time slot, and P_h^i, G_i, C_i as the corresponding parameters for this specific house. Then the indoor temperature evolves according to the following equation [19]:

$$T_i(k+1) = \left(T_i(k) - \frac{P_h^i(k)}{G_i} - T_o(k)\right) e^{-\frac{G_i}{C_i}t} + \frac{P_h^i(k)}{G_i} + T_o(k). \tag{2}$$

The parameters T_i, P_h^i, G_i, C_i, Q_i in this subsection are all related to the specific house with HVAC i, and in the following, we drop the subscription or superscription of i to simplify the notation.

Since G/C is very small (usually $< 10^{-5}$ [19]) and the indoor temperature changes during one slot is typically less than 0.5 degree, we approximate the above nonlinear model w.r.t. slot duration t by a linear one using Taylor's equations:

$$T(k+1) = T(k) - \Delta T_{off}(k) + QP_h(k), \tag{3}$$

where

$$\Delta T_{off}(k) = \frac{Gt}{C}(T(k) - T_o(k)), \tag{4}$$

and

$$Q = \frac{t}{C}. \tag{5}$$

Q is the conversion coefficient from power to temperature in a time slot for a specific room with the unit of K/J.

Since the tolerable indoor temperature variation is relatively small compared to the difference of $T(k)$ and T_o, ΔT_{off} can be approximated as a constant:

$$\Delta T_{off}(k) = \frac{Gt}{C}(T_r - T_o(k)) \tag{6}$$

where T_r is the set point temperature of an HVAC.

In this work, we consider houses and HVACs with different parameters. That is, the parameters C and G for each house may be different, and each HVAC can have a different set point T_r and a different maximum power. In addition, although our HVAC model may not be fully accurate, since the rooms' actual temperatures will be updated in each time slot, error will not be accumulated.

IV. Centralized Dynamic Water-filling Algorithm

The design objective of the centralized dynamic water-filling algorithm is to reduce the fluctuation of the power demand for conventional power plants by controlling the load of HVACs while guaranteeing HVAC user comfortable requirements.

A. Plant Model Design

(3) shows the relationship between the room temperature and the amount of power consumed by HVAC for a single house. Let the indoor temperature of each house and the total load of HVAC be the state of the plant (X), ΔT_{off} for each house be the measured disturbance (V), the input power for each HVAC be the control actions (U), and Y be the output of the plant model. The state space model of the plant is:

$$X(k+1) = AX(k) + B_u U(k) + B_v V(k), \tag{7}$$

$$Y(k) = C_x X(k) + D_u U(k) + D_v V(k), \tag{8}$$

where

$$X = \begin{bmatrix} T_1 \\ \vdots \\ T_n \\ L_h \end{bmatrix}, \quad U = \begin{bmatrix} P_h^1 \\ \vdots \\ P_h^n \end{bmatrix}, \quad V = \begin{bmatrix} \Delta T_{off}^1 \\ \vdots \\ \Delta T_{off}^n \end{bmatrix}, \tag{9}$$

A, B_u, B_v, C_x, D_u, and D_v are coefficients, k is the time slot index, n is the total number of houses with HVACs. The values of these coefficients are omitted due to space limitation. More details can be found in our technical report [20].

B. Heterogenous HVAC Parameters Estimation

To obtain the parameters for heterogeneous HVAC models, there are sensors in each house and a communication network exists between these sensors and the control center.

Let $P_h = 0$ in (3), and we can estimate ΔT_{off} when the HVAC is turned off as below:

$$\Delta \hat{T}_{off}(k) = T(k) - T(k+1), \tag{10}$$

where $\Delta \hat{T}_{off}$ is the estimated value of ΔT_{off}.

In practice, ΔT_{off} may change during different time slots, so we use the following exponentially weighted moving average (EWMA) algorithm to update ΔT_{off} for an HVAC model.

$$\Delta T_{off}(t_j) = \alpha \cdot \Delta \hat{T}_{off}(t_j) + (1-\alpha) \cdot \Delta T_{off}(t_{j-1}), \tag{11}$$

where α is the weight, and t_j is the parameter update time. Note that $\Delta T_{off}(t_j)$ is updated based on the current estimation and its last value, rather than the average of all the former values. The reason is that the status of the room may be different at different time, so the last value may be more accurate. The relationship between k, t and t_j is shown in our technical report [20]. The time duration between two parameter estimations, $t_j - t_{j-1}$, is determined by the control center. For example, the control center will choose the time slot when the HVAC is off to estimate $\Delta\hat{T}_{off}(t_j)$.

Similarly, from (3) we can estimate the value of Q as mentioned in our technical report [20].

Through this process, the control center is able to model heterogeneous HVACs by estimating different ΔT_{off} or Q for each HVAC. In the proposed work, we assume that these parameters are already available through estimation. Although this work only uses HVAC as the elastic load to provide demand response, the parameters of other elastic load can also be estimated in a similar way to provide demand response.

C. Controller Design

The objective of our control algorithm is to reduce the fluctuation of the conventional power plants' supply and guarantee user comfort level. Therefore, the objective function of the controller can be formulated as follows:

Problem I (P1)

$$\min_{P_h^i(k)} : \quad \sum_{k\in N} \left\{ (L_h(k) - r_w(k))^2 + \lambda^2 \sum_{i\in S} (T_i(k) - r_i(k))^2 \right\}, \tag{12}$$

subject to:

$$\sum_{i\in S} P_h^i(k) = L_h(k), \tag{13}$$

$$r_i(k) - \Delta T_l \leq T_i(k) \leq r_i(k) + \Delta T_u, \quad \forall i \in S, \forall k \in N, \tag{14}$$

$$0 \leq P_h^i(k) \leq u_i^{\max}, \quad \forall i \in S, \forall k \in N, \tag{15}$$

where $r_w(k)$ is the reference value of HVAC load in slot k; $P_h^i(k)$ is the consumed power by HVAC i assigned by the control center in slot k; r_i is the temperature set-point for HVAC i; ΔT_l and ΔT_u represent the maximum allowed temperature decrement and increment from the set point r_i in a house, respectively; $T_i(k)$ is the indoor temperature of the house with HVAC i in slot k; S is the set of all the HVACs; N is the prediction horizon; $u_i^{\max}$ is the maximum power consumption of the i-th HVAC; λ is the weight, and is squared to make the weight always positive.

The first part of the objective function (12) represents the deviation of the actual HVAC load from the reference HVAC load (r_w). The second part represents the sum of temperature deviation from the set-point in each house, which not only represents the influence to users' comfortableness, but also the elastic load potential because the HVAC can only provide one dimensional demand response if the room temperature reaches the upper or lower bound. Of course, we would like the room temperature be close to the set-point so the HVAC can either be turned on or off. This is quite different from the dead band based control policy which does not consider the demand response potential of HVACs [4]. λ is used to make a tradeoff between these two parts. Constraints (14) means that the controlled indoor temperature of each house should stay within a certain range of the set point during each time slot. (15) ensures that the power of each HVAC is bounded.

By solving this optimization problem, the controller can obtain a sequence of control actions corresponding to each time slot of the prediction horizon. Since the plant model is not accurate and there might be unmeasured disturbance or noise in this system, the actual indoor temperature of each house and the real load may not be the same as predicted after implementing the obtained control actions. Therefore the controller only executes the control actions in the first time slot, then it will update all the parameters and solve the optimization problem again.

D. Dynamic Water level Adjustment Algorithm

To solve the convex optimization problem **(P1)**, the reference value of HVAC load in slot k, $r_w(k)$, is needed. If $r_w(k)$ is not set appropriately, the controller may not be able to flatten the load effectively. Besides, the energy buffer capacity provided by elastic HVAC load will be consumed when all the indoor temperatures reach their upper or lower bounds. As a result, the HVACs will turn into inelastic load and lose the function of providing demand response.

In addition, since the size of energy buffer is relatively small and limited compared to the remaining load and unlimited control horizon, the value of $r_w(k)$ should not be constant. Instead, $r_w(k)$ should change according to the main trend of the remaining load so that the energy buffer will never be totally full or empty and the HVACs can always perform demand response to reduce the remaining load fluctuation.

To adjust $r_w(k)$ appropriately, we propose a dynamic water level adjustment algorithm stated as follows. In slot k, the sum of $r_w(k)$ and remaining load $L_b(k) - S_r(k)$ is the water level $W_l(k)$, and then we can calculate the reference HVAC load for each time slot according to

$$r_w(k) = W_l(k) - (L_b(k) - S_r(k)). \tag{16}$$

Actually, the water level $W_l(k)$ is the reference value for the total load of the conventional power plants. If we can keep the water level constant, then the total load is constant. However, due to the limited elastic capacity of HVAC, this is impossible. Therefore, we have to change the water level slowly and smoothly to minimize load fluctuations. Assuming a given water level for time slot k (which may not be optimal), by solving the centralized MPC problem **(P1)** we can obtain the predicted system states $X(k+N)$ for time slot $k+N$, where N is the prediction horizon. Then the water level for the next time slot $W_l(k+1)$ can be obtained using Algorithm 1. The main idea is that we adjust the water level whenever one of the room temperature may reach the upper or lower bound in the predicted future.

In Algorithm 1, $\overline{r_w}$ represents the average reference value of HVAC load for the following N slots; $P_s = \sum_{i=1}^{n} \frac{\Delta T_{off}^i}{Q_i}$ is the total amount of power needed to counteract all the houses' temperature decrease in each time slot; c is the change to the water level; $climit$ is the maximum allowed water level change in each time slot, which is determined by the power company; μ is the water level change rate and can be determined empirically.

Algorithm 1 Water Level Adjustment

Require: $X(k+N)$

$flag \leftarrow 0$
for all $i \in n$ **do**
  if $T_i(k+N) = r_i(k) + \Delta T_u$ **then**
    $flag \leftarrow 1$
    $break$
  else if $T_i(k+N) = r_i(k) - \Delta T_l$ **then**
    $flag \leftarrow -1$
    $break$
  end if
end for
if $flag \neq 0$ **then**
  $c \leftarrow \mu \cdot (\overline{r_w} - P_s)$
  if $|c| > climit$ **then**
    $c \leftarrow climit \cdot flag$
  end if
  $W_l(k+1) \leftarrow W_l(k) - c$
end if

The relationship of the MPC algorithm and the water level change can be summarized as follows: we use the MPC to predict the system status in the future. If the prediction results show that the total elastic load potential will be depleted in the future, we adjust the water level slightly to avoid the occurrence of this situation. Of course, with longer accurate prediction horizon, we can avoid unnecessary water level adjustment and therefore reduce the fluctuation.

V. Distributed Dynamic Water-Filling Algorithm

The centralized dynamic water-filling algorithm described in section IV relies on a centralized controller, which may have high computational complexity when the number of HVACs is large. To be scalable, a distributed architecture is preferable.

A. Distributed Control Architecture

Different from the centralized algorithm which has only one control center, the distributed algorithm relies on a hierarchical architecture with one central controller and several local controllers. The central controller tries to flatten the total load by adjusting the amount of power used by each local controller. The local controller will assign the amount of power designated by the central controller to each HVAC and maximize user comfort level. Note that if the population changes overtime, we can simply resize the group [1], and the algorithm still works.

B. Central Controller Design

Other than flattening the total load, the central controller should guarantee that the amount of power assigned to the local controllers in each time slot will not make any HVAC

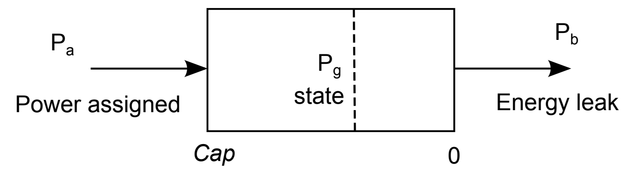


Fig. 2. Local Group Model

under that local controller violate the temperature constraints. Since there is no direct control between the central controller and the HVACs, we need a new plant model for the hierarchical MPC problem.

We consider the local controller with all the corresponding HVACs as a group. Each group has different size of the energy buffer provided by its HVACs. The model of a local group is shown in Fig. 2. The unit of all the variables in this model is J, and the group index j is omitted to simplify the notation. In each time slot, there is a total amount of energy P_b leak from all the houses, and P_a is the amount of energy assigned by the central controller to this local group. P_g represents the energy buffer state of this group. Then the evolution process of P_g can be shown as follows:

$$P_g(k+1) = P_g(k) + P_a(k) - P_b(k). \tag{17}$$

When all the indoor temperatures in the group decrease by ΔT_l from their set-points, the state of the energy buffer is 0; on the other hand, when all the indoor temperatures reach their upper bounds, the state P_g equals Cap. The energy buffer capacity can be obtained from:

$$Cap = (\Delta T_l + \Delta T_u) \sum_{i \in m} C_i, \tag{18}$$

where m is the group size, C_i is the effective heat capacity of the house with HVAC i in the unit of $Joule/Kelvin$, which can be obtained from Q_i [20].

The central controller must guarantee the states of all the local controllers between 0 and their Cap. In this state space model of the MPC problem, the control actions are the amount of power assigned to all the local controllers. The measured disturbance is the vector containing each P_b for the corresponding group, and the output is the vector including all the group states and the load of all HVACs. The exact state space model formulas are similar to (7), (8) and are omitted due to space limitation.

The objective function of the central controller is formulated as follows:

Problem II (P2)

$$\min_{P_a^j(k)} : \quad \sum_{k \in N} \left\{ (L_h(k) - r_w(k))^2 + \lambda^2 \sum_{j \in M} (P_g^j(k) - R_j(k))^2 \right\}, \tag{19}$$

subject to:

$$\sum_{j \in M} P_a^j(k) = L_h(k), \tag{20}$$

$$0 \leq P_g^j(k) \leq Cap_j, \quad \forall j \in M, \forall k \in N, \tag{21}$$

$$0 \leq P_a^j(k) \leq U_j^{\max}, \quad \forall j \in M, \forall k \in N, \tag{22}$$

[1]The groups can be resized at two levels. First, the number of HVACs in a group can be increased or decreased. Second, if the number of HVACs in each group changes too much, we can reconfigure all the groups to balance the group size.

where M is the number of groups; R_j is the reference value of the state for group j, and we set it to half of the energy buffer capacity $Cap_j/2$; $U_j^{\max} = \sum_{i \in m} u_i^{\max}$ is the maximum amount of power allowed to be assigned to local group j in each time slot.

The first part of the objective function (19) represents the deviation of the actual HVAC load from the reference HVAC load (r_w). The second part represents the sum of the energy buffer state deviation from the reference state for each local group, which *indirectly* represents the impact on users' comfortableness. λ is used to make a tradeoff between these two parts. Constraint (21) requires that the state of each group should be bounded between 0 and Cap. (22) ensures that the power assigned to each group is bounded.

Similarly, the central controller will follow the receding horizon principle and update the states of all the groups after each time slot.

C. Local Controller Design

After the local controller receives the power quota P_a assigned by the central controller for the next time slot, it assigns this amount of power to all the HVACs by maximizing user comfort level. Note that the energy buffer state of the local group is guaranteed to be bounded by the central controller, as a result the temperatures of all the houses will not violate the temperature constraint (14).

The local controller determines the amount of power for each house by solving the following optimization problem.

Problem III (P3)

$$\min_{P_h^i(k)} : \quad \sum_{i \in S_l} (T_i(k+1) - r_i(k+1))^2, \tag{23}$$

subject to:

$$\sum_{i \in S_l} P_h^i(k) = P_a, \tag{24}$$

$$r_i(k+1) - \Delta T_l \leq T_i(k+1) \leq r_i(k+1) + \Delta T_u, \quad \forall i \in S_l, \tag{25}$$

$$0 \leq P_h^i(k) \leq u_i^{\max}, \quad \forall i \in S_l, \tag{26}$$

$$T_i(k+1) = T_i(k) - \Delta T_{off}^i(k) + Q_i P_h^i(k), \quad \forall i \in S_l, \tag{27}$$

where S_l is the set of HVACs under the local controller.

Since the distributed algorithm has a hierarchical architecture, the computation complexity of the central controller can be greatly reduced. However, the central controller cannot control each HVAC directly, neither can it know the exact status of each HVAC. As a result, it may lead to some fairness problems to the HVACs because HVACs under different local controllers may be treated differently. In addition, the control variables for the central controller are reduced (from the number of HVACs to the number of local controllers), so the control precision may not be as good as the centralized algorithm. We will compare the performance of the distributed algorithm with the centralized one in Section VII.

VI. HVAC ON/OFF State Control

In problems **P1**, **P2** and **P3**, we assume that the consumed power for any HVAC in each time slot can be adjusted continuously. However, this may not be true in practice. For instance, some HVACs can only be turned on or off. Therefore, we consider how to change the proposed algorithms to support this kind of control actions.

For the proposed centralized MPC control algorithm, in order to support HVAC ON/OFF control, we can simply replace (15) with (28) in **P1**.

$$P_h^i(k) \in \{0, u_i^{\max}\}, \quad \forall i \in S, \forall k \in N. \tag{28}$$

W.r.t. the distributed MPC algorithm, we can replace (26) with (29) in **P3**, and let the local controller report the total amount of energy actually used, to the central controller which then updates the original control actions.

$$P_h^i(k) \in \{0, u_i^{\max}\}, \quad \forall i \in S_l, \tag{29}$$

This turns the original problems into multiple integer problems (MIP) which usually have much higher computational complexity. Therefore, we propose an heuristic algorithm which can obtain the control actions in polynomial time.

Algorithm 2 Determine HVAC State

Require: P_a

1: $PState[1, 2, \cdots, m] \leftarrow OFF$
2: sort $PState$ according to the difference between room temperature and the set-point from low to high
3: $sum \leftarrow 0$
4: **for all** $i \in m$ **do**
5:   **if** $T_i^-(k+1) \leq r_i(k) - \Delta T_l$ **then**
6:     $sum \leftarrow sum + P_h^i$
7:     $PState(i) \leftarrow ON$
8:   **end if**
9: **end for**
10: **for all** $i \in m$ **do**
11:   **if** $PState(i) = OFF$ and $sum < P_a$ and $T_i^+(k+1) \leq r_i(k) + \Delta T_u$ **then**
12:     **if** $P_g(k) \leq R(k)$ **then**
13:       $sum \leftarrow sum + P_h^i$
14:       $PState(i) \leftarrow ON$
15:     **else if** $sum + P_h^i < P_a$ **then**
16:       $sum \leftarrow sum + P_h^i$
17:       $PState(i) \leftarrow ON$
18:     **end if**
19:   **end if**
20: **end for**

In Algorithm 2, m is the number of HVACs in the group; $PState$ is an vector which stores the state of all the HVACs; $T_i^-(k+1)$ represents the temperature of house i in the next time slot if the state of the HVAC is OFF; $T_i^+(k+1)$ represents the temperature of house i in the next time slot if the state of the HVAC is ON; $R(k)$ is the reference value of the state in the current time slot.

Line 2 guarantees that the houses with lower temperature have a higher priority. Lines 4 to 9 of Algorithm 2 set the state of HVAC whose house temperature is very close to the lower bound to be ON. Lines 10 to 20 set the HVAC state to be ON if the total amount of power used is less than P_a. More details can be found in our technical report [20].

After obtaining the ON/OFF states of each HVAC, the local controller will report the actually used energy and the actual group state to the central controller. The central controller will update the model states and move to the next time slot.

VII. Performance Evaluation

In this section, we evaluate the performance of the proposed algorithms and compare them with the algorithm in [4] (we call it "SPDW" algorithm, which is the combination of the first letters of all the authors' names, for easy reference) and the uncontrolled one. To achieve a flatten overall non-renewable power generation, the SPDW algorithm tries to minimize the difference of the load between the current and the last time slot by adjusting the set-points of HVACs. The parameters of our algorithms are tuned using the existing approaches for MPC [21], which aims to minimize (30). After tuning, the values of the weights in our simulation are set to 3.3 and 0.28 for centralized MPC algorithm (CMPC) and distributed MPC algorithm (DMPC), respectively. The control interval is set to 2 minutes. The communication latency between the controllers and HVACs is negligible and can be ignored. For the data gathering time delay, we assume that the users will send their states to the controller at the beginning of each time slot. Without considering the communication latency and the transmission error, all the data should be obtained within seconds. With respect to the computation time of the proposed algorithms, the proposed CMPC algorithm does need a long time. However, the proposed DMPC is much faster. For example, To apply the DMPC algorithm to a community with 1000 HVACs, we can divide it into 20 groups with 50 HVACs in each group. The computation of either the central controller and the local controller for one time slot is below 0.4s. Since the room temperature will not change significantly within several seconds, the influence of the data gathering time and the computation time of the proposed algorithm are both tolerable. The wind energy generation data and users' load are obtained from Eirgrid [5]. While these data have a resolution of 15 minutes, we use shape-preserving piecewise cubic interpolation [22] to interpolate them into a resolution of 2 minutes. Besides, we scaled them down to fit a micro-grid with a population size of 1000 and let the wind energy penetration take about 50% of the total load.

For the HVAC and house model, we assume that each house only has one room and one HVAC for simplicity, and the decrement of room temperature in each time slot follows a normal distribution, with an average of 0.2 Celsius, and a standard deviation of 0.03 Celsius. The parameter Q also follows a normal distribution, with an average value of 1.65×10^{-4} and a standard deviation of 0.25×10^{-4}. The maximum power of each HVAC is uniformly distributed between $4kW$ and $6kW$. Note that the HVAC power used in our simulation may be different from that in practice, but it will not affect the effectiveness of the proposed algorithms.

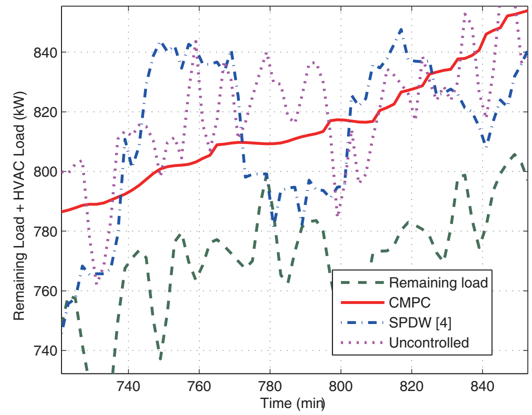


Fig. 3. Load for conventional power plants

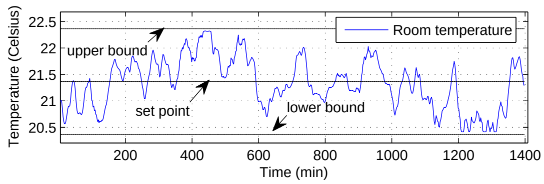
The parameters ΔT_l and ΔT_u are both set to one Celsius. The indoor set-point is uniformly distributed from 20 Celsius to 22 Celsius. Initially, all the indoor temperatures are scatted around their set-points by at most 0.5 degree.

In this simulation, the prediction horizon in the proposed CMPC is set to 30 minutes. Due to the computation complexity of the controller, the number of controlled HVAC in this simulation is only 40. We will show the simulation result with more HVACs using the DMPC later, which is much faster. The reason is that DMPC has a hierarchical architecture, so either the central controller or the local controller has much fewer variables to optimize in each time slot.

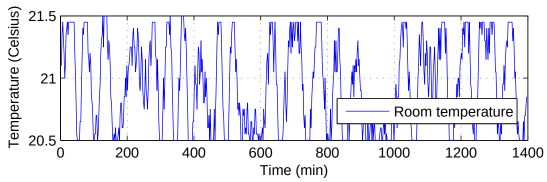
Fig. 3 shows a typical time period of the power provided by conventional power plants. As illustrated in the figure, both SPDW and CMPC can make the load smoother and flatter. Notice that the SPDW algorithm can flatten the load for a while (from minute 750 to 770), then the load suddenly decreases and is kept flat for another period of time (from minute 770 to 800) before another increase. The reason is that the SPDW algorithm will keep the load as flat as possible by adjusting the ON/OFF state of all the HVACs until the elastic capacity of all the HVACs is no longer enough to provide further demand response. Then it will make a dramatic load change to push the HVAC away from the temperature bound so they can continue to provide demand response. For example, around minute 770, the majority of HVACs have reached the temperature upper bound so they are all turned off which make the load decrease tremendously. The opposite situation happens around minute 800. The proposed CMPC algorithm predicts the status of all the rooms in the future and changes the water level beforehand so that the overall load change is much smoother.

To measure the performance of different algorithms numerically, we define a criterion "average fluctuation (AF)" to represent the amount of load fluctuation as follows (similar criterion can be found in [2]):

$$AF = \frac{1}{N-1}\sum_{i=1}^{N-1}\frac{|data(i+1) - data(i)|}{data(i)}, \tag{30}$$



(a) Room temperature of CMPC



(b) Room temperature of SPDW [4]

Fig. 4. Room temperature

where N is the total number of time slots; $data(i)$ is the data in the i-th slot; the denominator is used to normalize the data difference.

The average fluctuation for CMPC, SPDW and uncontrolled cases are 0.0028, 0.0123 and 0.0166, respectively. We find the proposed algorithm has a much smaller AF value which mainly because of the following three reasons. First, if the HVAC buffer capacity may not be enough in the future, the proposed algorithm will get prepared by changing the water level gradually, so the average fluctuation can be reduced. Second, the HVAC models in the proposed algorithm are assumed to use any amount of energy during the time slot while the SPDW algorithm only controls the ON/OFF states of the HVACs. We will discuss the performance of the proposed distributed MPC algorithm with ON/OFF support (DMPCOF) later. Third, the HVACs in the proposed algorithm are controlled directly by the central controller while the SPDW algorithm only control the set-point of all HVACs which indirectly affect the ON/OFF states of HVACs.

Fig. 4 (a) shows the temperature variation of a typical room using the proposed CMPC algorithm. The set-point is not an integer because we set random set-point for each house. As can be seen, CMPC can effectively guarantee user comfort level by restricting room temperature's variation within a range of the set-point. Fig. 4 (b) represents the set-point of all the controlled HVACs by SPDW. Note that the individual house temperature is allowed to deviate from the set-point for 0.5 Celsius, so the temperature variation range of a house is the same as that in CMPC. Comparing these two figures, we will find the temperature variation of the proposed algorithm is much smoother which leads to a higher user comfort level.

For the DMPC and DMPCOF algorithms, we set the number of local controllers to be 4, and each local controller manages 8 to 12 HVACs, with a total number of 40. The prediction horizon is also 30 minutes.

Fig. 5 illustrates the zoomed-in load of SPDW, CMPC, DMPC and DMPCOF algorithms. The load of CMPC is slightly below the others due to different water levels. Since the control actions and system model are a bit different, the water level of CPMC and DMPC may not be the same all the time. The curve of DMPCOF contains more fluctuations

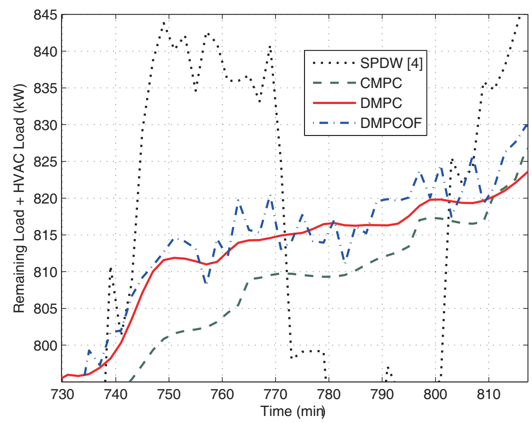


Fig. 5. Load comparison of proposed three algorithms

TABLE I
AVERAGE FLUCTUATION

Prediction Horizon	DMPC with 40 HVACS	DMPCOF with 40 HVACS	DMPCOF with 80 HVACS
10min	0.0032	0.0058	0.0052
20min	0.0030	0.0052	0.0048
30min	0.0029	0.0050	0.0046

because the load change is discrete rather than continuous. However, the fluctuation is still much smaller than that of SPDW.

The AF of DMPC and DMPCOF are 0.0029 and 0.0050 respectively, and slightly larger than that of CMPC but still much smaller than that of SPDW and the uncontrolled cases. The reason why the AF of CMPC is smaller may be that DMPC has fewer control variables for the central controller.

In addition, we define the user comfortable influence factor (UIF) as the root mean square of all the room temperature deviation from the set-point.

We find the UIF of DMPC and DMPCOF are about 8.23% and 72.4% larger than that of CMPC. The reason is that CMPC minimizes UIF directly while DCMP and DCMPOF minimize it indirectly by controlling the energy buffer state for each local controller. Besides, the HVACs of DCMPOF can only be turned ON or OFF which makes its UIF even larger.

With a different prediction horizon, the total load for the conventional power plants is anticipated to be different. Fig. 6 shows the situation when the prediction horizon is 10 minutes and 30 minutes, respectively. The total number of HVACs are both 40.

From Fig. 6, the curve DMPC-30 (corresponding to 30-minute prediction horizon) is smoother than that of DMPC-10 (corresponding to 10-minute prediction horizon) because it contains fewer ups and downs. The reason is that with a longer prediction horizon, the controller has more information about the future load change and thus can get prepared earlier.

Table I shows the average fluctuation for DMPC, DMPCOF with 40 HVACs and DMPCOF with 80 HVACs (4 groups, each group has 16 to 24 HVACs). Obviously, the average fluctuation of DMPCOF is larger than that of DMPC with the

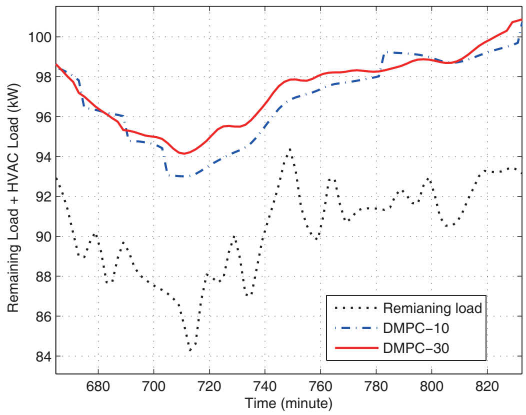


Fig. 6. Load with different prediction horizon and HVAC number

same prediction horizon, and with a longer prediction horizon the average fluctuation is smaller. We also notice that with more HVACs, the average fluctuation of DMPCOF is smaller under the same prediction horizon too. The reason is that with more HVACs to be controlled in a group, the amount of actual HVAC load can have a better chance to be closer to the reference HVAC load value.

VIII. Conclusion

In this paper, we have proposed two algorithms to control HVACs for demand response based on MPC. The centralized approach directly controls all the HVACs while the distributed approach uses a hierarchical architecture. Both of them can effectively reduce load fluctuations while keeping all the room temperature within a range of the set-point. Moreover, the proposed distributed algorithm has been extended under a more practical assumption that each HAVC can only support ON and OFF.

There are several open issues left behind. First, how to extend the current MPC algorithms to control other types of elastic load requires further investigation. The key difference is that the initial states of HVACs are already known, while for other types of elastic load, such as PHEV, the arrival time and departure time in the future may not be available. Second, the water level change rate μ in Algorithm 1 is determined empirically and is a constant in this paper. If we can adjust μ w.r.t. history statistic information, such as peak time etc., we may achieve an even better performance. Third, the computational complexity will increase with more control variables. Therefore, how to make a tradeoff between the number of groups, the size of each group, load fluctuation, and the influence to user comfort level is also an important problem left to future research.

References

[1] N. Lu, "An evaluation of the hvac load potential for providing load balancing service," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1263–1270, 2012.

[2] N. Lu and Y. Zhang, "Design considerations of a centralized load controller using thermostatically controlled appliances for continuous regulation reserves," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 914–921, 2013.

[3] M. Shinwari, A. Youssef, and W. Hamouda, "A water-filling based scheduling algorithm for the smart grid," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 710–719, 2012.

[4] S. Parkinson, D. Wang, C. Crawford, and N. Djilali, "Comfort-constrained distributed heat pump management," *Energy Procedia*, vol. 12, pp. 849 – 855, 2011.

[5] EirGrid, "Wind energy generation & System demand data sets." [Online]. Available: http://www.eirgrid.com/operations/systemperformancedata/

[6] O. SOLBERG, "A new wind turbine control method to smooth power generation," Master's thesis, Department of Energy and Environment, Chalmers University of Technology, 2012.

[7] I. Koutsopoulos and L. Tassiulas, "Optimal control policies for power demand scheduling in the smart grid," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1049–1060, 2012.

[8] A. Oudalov, D. Chartouni, and C. Ohler, "Optimizing a battery energy storage system for primary frequency control," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1259–1266, 2007.

[9] T. Borsche, A. Ulbig, M. Koller, and G. Andersson, "Power and energy capacity requirements of storages providing frequency control reserves," in *IEEE Power and Energy Society General Meeting*, 2013, pp. 1–5.

[10] C. Jin, N. Lu, S. Lu, Y. Makarov, and R. Dougal, "Coordinated control algorithm for hybrid energy storage systems," in *Power and Energy Society General Meeting, 2011 IEEE*, 2011, pp. 1–7.

[11] H. T. Nguyen, D. Nguyen, and L. B. Le, "Home energy management with generic thermal dynamics and user temperature preference," in *S-mart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, 2013, pp. 552–557.

[12] M. Tasdighi, H. Ghasemi, and A. Rahimi-Kian, "Residential microgrid scheduling based on smart meters data and temperature dependent thermal load modeling," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 349–357, 2014.

[13] Z. Yu, L. McLaughlin, L. Jia, M. Murphy-Hoye, A. Pratt, and L. Tong, "Modeling and stochastic control for home energy management," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1–9.

[14] G. Karmakar, A. Kabra, and K. Ramamritham, "Coordinated scheduling of thermostatically controlled real-time systems under peak power constraint," in *Real-Time and Embedded Technology and Applications Symposium (RTAS), 2013 IEEE 19th*, 2013, pp. 33–42.

[15] T. Hovgaard, K. Edlund, and J. Bagterp Jorgensen, "The potential of economic mpc for power management," in *49th IEEE Conference on Decision and Control (CDC),*, 2010, pp. 7533–7538.

[16] M. D. Galus, S. Koch, and G. Andersson, "Provision of load frequency control by phevs, controllable loads, and a cogeneration unit," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4568–4582, 2011.

[17] G. Giebel, "The state-of-the-art in short-term prediction of wind power: A literature overview." [Online]. Available: http://www.risoe.dk/rispubl/vea/veapdf/ANEMOS_giebel.pdf

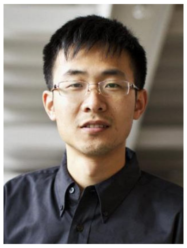
[18] K. Liu, S. Subbarayan, R. Shoults, M. Manry, C. Kwan, F. Lewis, and J. Naccarino, "Comparison of very short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 877–882, 1996.

[19] A. Rautiainen, S. Repo, and P. Jarventausta, "Using frequency dependent electric space heating loads to manage frequency disturbances in power systems," in *IEEE Bucharest PowerTech*, 2009, pp. 1–6.

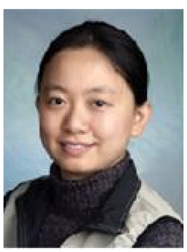
[20] K. Zhou and L. Cai, "A Dynamic Water-filling Method for Real-Time HVAC Load Control Based on MPC," E&CE at University of Victoria, Tech. Rep., 2014. [Online]. Available: http://web.uvic.ca/~kzhou/doc/waterfill-TechReport.pdf

[21] T. J. van den Boom and T. C. Backx, "Lecture Notes for the MPC DISC Course." [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.216.3393

[22] MathWorks, "Interpolation." [Online]. Available: http://www.mathworks.com/moler/interp.pdf



Kan Zhou '14'14S received the B.S. and M.A.Sc. degree in the Department of Electrical and Computer Engineering from Southeast University, Nanjing, China, in 2008 and 2011 respectively. He is currently working toward the Ph.D. degree in electrical and computer engineering in the University of Victoria, British Columbia, Canada. His fields of interest include PHEV scheduling, smart grid and artificial intelligence.



Lin Cai (S'00-M'06-SM'10) received her M.A.Sc. and PhD degrees (awarded Outstanding Achievement in Graduate Studies) in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2002 and 2005, respectively.

Since 2005, she has been an Assistant Professor and then an Associate Professor with the Department of Electrical & Computer Engineering at the University of Victoria. Her research interests span several areas in wireless communications and networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic over wireless, mobile, ad hoc, and sensor networks. She has been a recipient of the NSERC Discovery Accelerator Supplement Grant in 2010, the best paper award of IEEE ICC 2008, and the best academic paper award of IEEE WCNC 2011. She has served as a TPC symposium co-chair for IEEE Globecom'10 and IEEE Globecom'13, and the Associate Editor for IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, EURASIP Journal on Wireless Communications and Networking, International Journal of Sensor Networks, and Journal of Communications and Networks (JCN).