Q-Learning-Based Model Predictive Control for Energy Management in Residential Aggregator

Kianoosh Ojand and Hanane Dagdougui, *Member, IEEE*

Abstract—**This article presents a demand response scheduling model in a residential community using an energy management system aggregator. The aggregator manages a set of resources, including photovoltaic system, energy storage system, thermostatically controllable loads, and electrical vehicles. The solution aims to dynamically control the power demand and distributed energy resources to improve the matching performance between the renewable power generation and the consumption at the community level while trading electricity in both day-ahead and real-time markets to reduce the operational costs in the aggregator. The problem can be formulated as a mixed-integer linear programming problem in which the objective is to minimize the operation and the degradation costs related to the energy storage system and the electric vehicles batteries. To mitigate the uncertainties associated with system operation, a two-level model predictive control (MPC) integrating Q-learning reinforcement learning model is designed to address different time-scale controllers. MPC algorithm allows making decisions for the day-ahead, based on predictions of uncertain parameters, whereas Q-learning algorithm addresses real-time decisions based on real-time data. The problem is solved for various sets of houses. Results demonstrated that houses can gain more benefits when they are operating in the aggregate mode.**

Note to Practitioners: **Residential buildings besides commercial and public sectors are among the building sectors responsible for high-energy consumption. Numerous measures have been considered to construct more energy-efficient buildings, such as implementing new effective insulation materials and increasing the utilization ratio of sunlight. However, there is also a need for practical solutions to reduce the greenhouse gas emissions and avoid power peak from the residential sector. Under this situation, energy management system aggregator (EMSA) offers the opportunity to exploit the flexibility potential of various houses and other available distributed energy resources, promoting their participation in ancillary services and benefiting from rewards and lower energy bills. In this article, an innovative and comprehensive model predictive control-based scheduling optimization that considers uncertainties of renewable resources and weather conditions is formulated. It can be considered as a practical solution in order to optimally control the operation of a residential community. We proposed a curtailable demand response (DR), where customers agree to participate in DR programs defined by the EMSA in response to price changes.**

Index Terms—**Demand response (DR), distributed energy resources (DERs), electric vehicle (EVs), mixed-integer linear programming (MILP), model predictive control (MPC), reinforcement learning, residential community, thermostatically controlled loads (TCLs).**

NOMENCLATURE

DERs	Distributed energy resources.
DG	Distributed generation.
DR	Demand response.
DTC	Direct thermostat control.
EMSA	Energy management system aggregator.
ESS	Energy storage system.
EVs	Electric vehicles.
EWH	Electric water heater.
GHG	Greenhouse gas.
HVAC	Heating, ventilation, and air conditioning.
MILP	Mixed-integer linear programming.
MILP-MPC	Mixed-integer linear programming-model predictive control.
MPC	Model predictive control.
MPP	Maximum power point.
MPPT	Maximum power point tracking.
PCA	Principle component analysis.
SVM	Support vector machine.
TCLs	Thermostatically controlled loads.
V2H	Vehicle-to-home.

Indices I	Set of houses.
J	Set of electric vehicles.
i	ith house $\forall\ i \in I$.
j	jth electric vehicle $\forall\ j \in J$.
n	nth noncontrollable load in house ith.
m	mth controllable load in house ith.
t	Time.

Manuscript received February 25, 2021; accepted March 25, 2021. Date of publication July 9, 2021; date of current version January 6, 2022. This article was recommended for publication by Editor M. Robba upon evaluation of the reviewers' comments. *(Corresponding author: Hanane Dagdougui.)*

Kianoosh Ojand is with the Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada (e-mail: kianooshojand@polymtl.ca).

Hanane Dagdougui is with the Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada, and also with the GERAD Research Center, Montreal, QC H3T 2A7, Canada (e-mail: hanane.dagdougui@polymtl.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2021.3091334.

Digital Object Identifier 10.1109/TASE.2021.3091334

Parameters

α_k	System inertia of each HVAC.
β	Degradation cost of ESS charge/discharge.
η_{MPPT}	Efficiency of the PV's dc/ac converter and the MPPT system.
$\eta_{b,\mathrm{ch}}$	Charging efficiency of ESS.
$\eta_{b,\mathrm{dch}}$	Discharging efficiency of ESS.
η_{cooling}	Working efficiency of HVACs in cooling mode.
η_{heating}	Working efficiency of HVACs in heating mode.
$\eta_{\mathrm{ev,ch},j}$	Charge efficiency of the jth EV.
$\eta_{\mathrm{ev,dch},j}$	Discharge efficiency of the jth EV.
λ	Constant value related to the degradation cost ofcharging and discharging of EVs.
$\theta_{i,\max}^{\mathrm{ewh}}$	Max power of EWH (kW) in the ith house.
$\theta_{i,\min}^{\mathrm{ewh}}$	Min power of EWH (kW) in the ith house.
ρ_{water}	Water density.
Ψ_t	Percentage of state of charge (%).
Δt	Time interval (30 min).
A_i	Set of appliances of ith house.
a_1	Threshold value of the SOC to switch charging behavior.
c_1	Constant value representing charging of ESS.
C_k	Thermal capacity of HVAC system (kWh/°C).
C_p	Specific heat water [4.2157 kJ/(kg.°C)].
$C_g^{p,s}$	Cost of electricity buy/sell to/from grid.
d_1	Constant value related to the charging behavior of ESS.
Db_i^{ewh}	Dead band of set point temperature of ith EWH (°C).
Db_i^{hvac}	Dead band of the set point temperature of the HVAC in the ith house (°C).
$D_{\mathrm{wt},i}$	Water demand (m^3/h) in ith house.
G_T	Irradiance level (W/m^2).
$I_{\mathrm{mode},t}^a$	$I_{\mathrm{made},t}^a = 1$ means that HVAC is working in heating mode and $I_{\mathrm{made},t}^a = 0$ means that it is working in the cooling mode.
N	Total number of houses.
N_{ev}	Total number of EVs.
$P_{b,\max}$	ESS Maximum charge/discharge power (kW).
$P_{g,\max}$	Maximum power transmissible from/to grid.
$P_{\mathrm{ev},j}^{\max}$	Max charge/discharge power of jth EV (kW).
$P_{k,i}^{\min}$, $P_{k,i}^{\mathrm{Max}}$	Min and max power of ith house HVAC (kW).
$P_{\mathrm{PV,Rated}}$	PV system rated power at $G_t = 1000\mathrm{W/m}^2$.
P_{pv}	Photovoltaic system output power (kW).
R	Thermal insulation.
R_k	Thermal resistance of HVAC system (°C/kW).
$\mathrm{SOC}_{b,\min}$	Min level of SOC of ESS (kWh).
$\mathrm{SOC}_{b,\max}$	Max level of SOC of ESS (kWh).
$\mathrm{SOC}_{\mathrm{ev},j}^{\min}$	Minimum SOC of battery of jth EV (kW).
$\mathrm{SOC}_{\mathrm{ev},j}^{\max}$	Maximum SOC of battery of jth EV (kW).
S_{res}	Reservoir surface area (m^2).
$T_{k,t}^a$	Ambient temperature (°C).
t_j^A, t_j^D	Arrival and departure time of the jth EV.
$T_{\mathrm{set},i}^{\mathrm{ewh}}$	Thermostat set point of EWH (°C).
$T_{\mathrm{set},i}^{\mathrm{hvac}}$	ith house HVAC set point temperature (°C).
m_j^A, m_j^D	jth EV SOC at arrival and departure (kWh).
V	Water reservoir capacity (m^3).
$x_i(t)$	Total load of the ith house (kW).
$x_{i,n}^{\mathrm{nc}}$, $x_{i,m}^c$	ith house noncontrollable and controllable load power consumption (kW).
$x(t)$	Total power consumption of all houses in time t.

Binary Variables

$I_{\mathrm{ch},b}$	Charging and discharging status of ESS.
$I_{\mathrm{ev,ch},j}$	Charging and discharging status of EVs.

Continuous Variables

θ_i^{ewh}	Power consumption of each EWH (kW).
$P_{k,t}^a$	Power consumption of each HVAC (kW).
$P_{b,\mathrm{ch}}(t)$	ESS charge power (kW).
$P_{b,\mathrm{dch}}(t)$	ESS discharge power (kW).
$P_{\mathrm{ev,ch},j}$	Charging power of the jth EV (kW).
$P_{\mathrm{ev,dch},j}$	Discharging power of the jth EV (kW).
$P_g(t)$	Power purchased/sold from/to the grid (kW).
$\mathrm{SOC}_b(t)$	State of charge stored in the ESS (kWh).
$\mathrm{SOC}_{\mathrm{ev},j}$	State of charge of the jth EV (kWh).
T_i^{hvac}	Indoor temperature of each house (°C).
T_i^{ewh}	Hot water temperature inside the tank of EWH in each house at time t (°C).
$x_{i,m}^c$	Controllable load in the ith house (kW).

I. Introduction

ENERGY consumption in buildings worldwide accounts for approximately 40% of global energy consumption [1]. Besides commercial and public institutions, residential buildings are among the buildings sector responsible for high-energy consumption. Numerous measures have been considered to construct more energy-efficient buildings, such as implementing new effective insulation materials and increasing the utilization ratio of sunlight [2]. However, there is also a need for practical solutions to reduce the GHG emissions and avoid power peak from the residential sector. Smart building integrated with DERs is a prevalent green approach that allows buildings to use the local available renewable energy resources, distribute, and regulate the flow of electricity to local consumers using advanced information and communication technologies and building energy management system. The basic control objectives for this latter are to maintain the high comfort level while reducing total energy consumption and peak of electricity demand. DR is an alternative solution that will allow the building energy operator to reduce, manage, and regulate energy consumption and peak loads in response to supply conditions [3]–[5].

The large integration of renewable generation, and specially that of solar energy, has led to much concern over the power system stability and reliability. Therefore, the flexibility of power system to respond to the variation in load and generation becomes challenging for those systems with high shares in renewables. One possible option to support power system flexibility is through residential communities. For uncoordinated residential community with many houses, energy management system of every house would engage in a selfish manner to manage and control its resources, having sometimes similar actions to maximize their benefits from demand-side management. These similarities in DR actions can lead sometimes to a rebound peak load and other grid instability issues [6]. In this case, DR strategies need to be managed by entities that have better grid conditions visibility and can aggregate load and generation flexibility in one DR scheme. EMSA offers the opportunity to exploit the flexibility potential of various houses, prosumers, buildings, and DERs, promoting their participation in wholesale, local, and ancillary markets. We propose an energy management system of an aggregator of smart homes and DERs that is connected to the main grid. The interaction with the grid will take place through the aggregator, along the lines of the proposed architecture [7]. In [8], a newly developed aggregator, responsible for V2H connection, is described. This new strategy allows every individual owning an EV, to participate in the regulation of their EV's batteries. Their results show that the strategy is practical and could successfully deliver the predicted power to the grid and maximize the profit of each particular player in the market.

In recent years, there have been a series of literature related to DR in buildings. Different DR controlling strategies have been proposed, which aims to control homes [9], [10], residences [11], [12] or commercial buildings [13]. Among those studies, there are papers on modeling appliances that have thermal storage capabilities, such as HVAC and EWH [14], [15]. For instance, Sun *et al.* [14] investigated the importance of various HVAC physical parameters and their distributions, affecting the aggregated response to a set of units. Lu [16] demonstrated that aggregated TCLs can provide load balancing capacity, which may enable small residential or commercial customers to participate in ancillary service markets. In [17], the capability of aggregated EWHs as a DR tool to respond to time-of-use control signals generated by the utility for load-shifting and balancing reserve is investigated. A novel algorithm, integrating DR aggregators with distribution network operators, is developed in [18] to have a secure and effective scheduling and real-time operation in residential radial feeders. Their results show that fulfilling operational constraints can lead to a significant limitation to the allocation of DR. Ruelens *et al.* [19] investigated different deep reinforcement learning techniques using convolutional neural network and recurrent neural network in order to optimally control the TCLs, and they have concluded that long short-term memory recurrent neural network gives better results. Zhang *et al.* [20] proposed a model-based reinforcement learning approach using a neural network. They adopted MPC by using the learned system dynamics to perform control with random sampling shooting method.

Moreover, many studies have focused on the control of nonthermostatically appliances, such as EVs [21], [22]. Existing studies on EVs have been mainly focused on coordinated charging/discharging allowing EVs' customers and/or utilities to schedule their charging profiles [23]. Wi *et al.* [24] proposed a hierarchical optimal control framework for charging coordination between multiple plug-in EVs parking decks. In [25], three different coordinated energy dispatching methods in the microgrid context with wind power generators and plug-in EV are proposed.

Furthermore, various studies have tackled the market bidding and power exchange with the power grid. For example, in [26], an optimal strategy for trading electricity considering uncertainties related to DG power production, load, day-ahead, and real-time market prices is introduced to minimize the operating cost of microgrid. In [27], a bidding strategy for a load aggregator considering DTC is proposed. In this study, aggregator tries to handle the uncertainties related to the weather conditions, nonthermostatic loads, and electricity prices. In [28], an energy management system algorithm from an aggregator's point of view is presented. This aggregator manages PV panels, batteries, and thermal energy storage in order to participate in the electricity trading market to bid for energy. Adjustable robust optimization is exerted to discover a robust counterpart considering uncertainties. Bruninx *et al.* [29] studied the interaction between an aggregator, its consumers, and electricity market using day-ahead prices using a bilevel optimization framework. In this model, the interaction between aggregator and consumers is modeled using the Nash bargaining game and chance-constrained programming. A forecasting model intending to help load aggregators anticipate the DR of smart houses in the electricity market in order to enable them to bid in the market is presented in [30]. Lately, a layered stochastic approach for residential DR considering real-time pricing and incentive mechanism is proposed in [31].

Although there are many literature works addressing the design, modeling, and optimization in residential aggregators, only a few works investigated the uncertainties related to weather data, demand, and PV generation. As far as the authors' knowledge, there are no works on the optimization of a unified scheme for residential aggregator that includes simultaneously various types of resources, such as TCLs, PV system, ESS, and EVs. Based on these considerations, this article presents a novel energy management approach for residential aggregator that allows a two-layer control strategy. The key contributions to the state of the art are the following.

1) An integrated architecture for a residential aggregator with various DERs and controllable loads is proposed.
2) A unified model for EMSA to optimize EVs operation, TCLs, and DER in a residential environment is presented. The effectiveness of the proposed method is verified by comparing it with the case where the house operates in solely mode and independently of the aggregator.
3) A two-level MPC integrating Q-learning model is designed to address different time-scale controllers. MPC algorithm allows making decisions for the day-ahead based on predictions of uncertain parameters, whereas Q-learning algorithm addresses real-time decisions based on real-time data. At the level of aggregator, MPC control signals are based on the predictions of uncertain parameters. However, due to communication errors and delays along with the lack of knowledge about

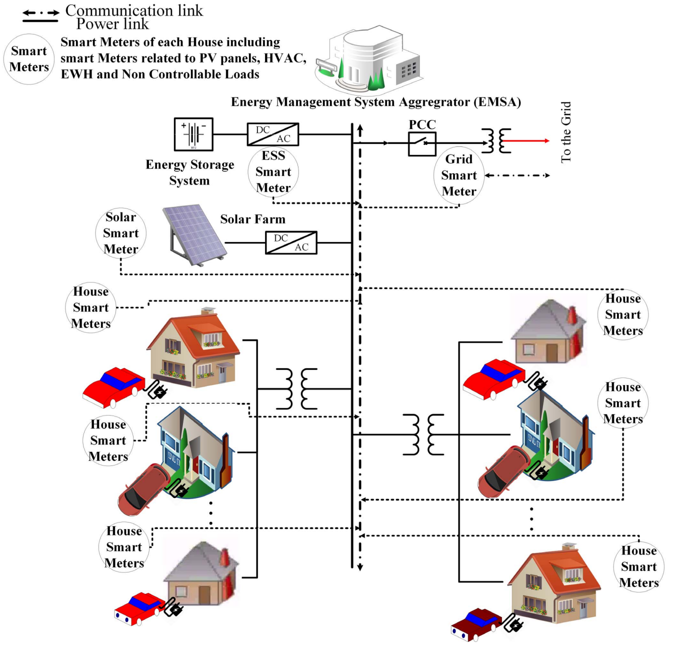

Fig. 1. Architecture of the residential microgrid.

specifications of each house, the aggregator may not have complete observability of the system, which in turn decreases the performance of the approach. Due to the model-free feature of reinforcement learning, Q-learning controllers can make real-time decisions in an uncertain situation without prior knowledge of the system model.

The rest of this article is organized as follows. Section II describes the proposed framework. Section III presents demand-side flexibility modeling, including PV system, ESS, EVs, HVACs, EWHs modeling, and residential DR mechanism. In Section IV, the EMSA, including optimization problem, and forecast information are described. Description of the case study and numerical results are presented, respectively, in Sections V and VI. We conclude this article by a conclusion in Section VII.

II. Description of the Proposed framework

In this article, we propose an energy management strategy for an aggregator of smart houses integrating DERs and EVs. The houses share DER that comprise PV arrays and ESS used as a backup for intermittent generation. The ESS has also the flexibility to be charged from the main grid. Some of the houses are equipped with EVs that can actively participate in DR through the V2H concept. Fig. 1 shows the configuration of the proposed system.

The energy management system aggregator, as shown in Fig. 1, is designed to gather and analyze data from various smart meters related to DER and loads. In addition, it communicates the optimal control signals to various appliances in the residential community and EVs. In this study, the houses have different architectures and user's behaviors, and their consumption patterns related to lighting, HVACs, EWHs, and EVs are different. We propose a curtailable DR, where customers agree to participate in DR programs defined by the EMSA in response to price changes.

III. Demand-Side Flexibility Modeling

A. Photovoltaic System Modeling

The output power of the PV array at the MPP can be modeled by a linear power source according to the irradiance level

$$P_{\mathrm{PV}}(t) = \left[P_{\mathrm{PV,Rated}} \times \frac{G_T(t)}{1000} \times \eta_{\mathrm{MPPT}}\right] \tag{1}$$

where P_{PV} and $P_{\mathrm{PV,Rated}}$ are, respectively, the output power (kW) at the MPP and the rated power at $G_T = 1000$ W/m^2 and G_T and η_{MPPT} are, respectively, the irradiance level (W/m^2) and the efficiency of the PV's dc/ac converter and the MPPT system.

B. ESS Modeling

The residential community is equipped with an ESS that provides additional flexibility for the EMSA. The discharged power is used to satisfy various appliances and available EVs in the community. The battery bank can be charged by the excess of the PV power generation or/and by the main grid depending on the energy pricing. The over-time evolution of energy stored in the microgrid is described by

$$\mathrm{SOC}_b(t+\Delta t) = \mathrm{SOC}_b(t) + \left(\eta_{b,\mathrm{ch}} P_{b,\mathrm{ch}}(t) - \frac{P_{b,\mathrm{dch}}(t)}{\eta_{b,\mathrm{dch}}}\right)\Delta t \tag{2}$$

where $\mathrm{SOC}_b(t)$ (kWh) is the state of charge in the ESS, $P_{b,\mathrm{ch}}(t)$ is the charged power, $P_{b,\mathrm{dch}}(t)$ is the discharged power, $\eta_{b,\mathrm{ch}}$ and $\eta_{b,\mathrm{dch}}$ are, respectively, the charging and discharging efficiencies, and Δt is the time interval (30 min).

It is essential to keep the ESS SOC_b between the highest allowable level, $\mathrm{SOC}_{b,\max}$, and the minimum allowable level, $\mathrm{SOC}_{b,\min}$

$$\mathrm{SOC}_{b,\min} \leq \mathrm{SOC}_b(t) \leq \mathrm{SOC}_{b,\max}. \tag{3}$$

In this study, it is assumed that the ESS is composed of Na–Ni batteries, and therefore, based on [32], the charging model of the ESS system can be described as follows:

$$P_{b,\mathrm{ch}}(t) = \begin{cases} P_{b,\max} \cdot 0.98, & \text{if } \Psi_t \leq a_1 \\ c_1 \cdot \Psi_t - d_1, & \text{if } \Psi_t \geq a_1. \end{cases} \tag{4}$$

As it is indicated in [32], c_1 and d_1 are two parameters used to simulate the characteristics of the Na–Ni batteries in the ESS. The power charged/discharged to/from the ESS is bounded by a certain maximum charging/discharging power limit of $P_{b,\max}$ (kW). Ψ_t is the percentage of state of charge and a_1 is a threshold. As it is not possible to charge and discharge ESS simultaneously, it is required to define a dummy binary variable I_{ch}, which is equal to 1 when the ESS is charging, and otherwise, it is equal to 0

$$\begin{cases} 0 \leq P_{b,\mathrm{ch}}(t) \leq P_{b,\max} \times I_{\mathrm{ch},b} \\ 0 \leq P_{b,\mathrm{dch}}(t) \leq P_{b,\max} \times \left(1 - I_{\mathrm{ch},b}\right). \end{cases} \tag{5}$$

C. EVs Modeling

Every EV $j \in N_{\text{ev}}$ (total number of EVs) is equipped with a local controller, which enables the vehicle to communicate with the aggregator through a smart metering infrastructure. Once EV is connected to the charging infrastructure, data related to the battery capacity, the desired energy at the departure time and the departure time can be transmitted to the aggregator. The state of charge $[\text{SOC}_{\text{ev},j}(t)]$ of the jth EV (kWh) can be expressed as follows:

$$\begin{aligned}\text{SOC}_{\text{ev},j}(t+\Delta t) &= \text{SOC}_{\text{ev},j}(t)\\ &\quad+\left(\eta_{\text{ev,ch},j}\times P_{\text{ev,ch},j}-\frac{P_{\text{ev,dch},j}}{\eta_{\text{ev,dch},j}}\right)\Delta t\\ &\quad\forall t\in\left[t_j^A\ t_j^D\right] \end{aligned}\tag{6}$$

where $t \in [t_j^A\ t_j^D]$ is the jth EV scheduling time horizon, t_j^A and t_j^D are, respectively, the arrival and departure times of the jth EV, $P_{\text{ev,ch},j}(t)$ and $P_{\text{ev,dch},j}(t)$ are, respectively, the charged or discharged powers of the jth EV (kW) at time step t, and $\eta_{\text{ev,ch},j}$ and $\eta_{\text{ev,dch},j}$ are the charging or discharging efficiencies of the jth EV, respectively.

The energy stored in each battery at every time slot should satisfy the battery capacity limits. The charging and discharging power also have upper bounds to meet physical constraints. Let m_j^A denote the energy left in the battery when the jth user's EV arrives home and let m_j^D denote the desired energy for the next trip for the jth user. These constraints can be stated as follows:

$$\text{SOC}_{\text{ev},j}\left(t=t_j^A\right)=m_j^A \tag{7}$$

$$\text{SOC}_{\text{ev},j}\left(t=t_j^D\right)=\overline{\text{SOC}_{\text{ev},j}}\left(t=t_j^D\right)=m_j^D \tag{8}$$

$$\text{SOC}_{\text{ev},j}^{\min}\le\text{SOC}_{\text{ev},j}(t)\le\text{SOC}_{\text{ev},j}^{\max}\quad\forall\ t\in\left[t_j^A\ t_j^D\right] \tag{9}$$

$$\begin{aligned}0\le P_{\text{ev,ch},j}(t)&\le P_{\text{ev,ch},j}^{\max}\\ &\cdot I_{\text{ev,ch},j}\quad\forall\ t\in\left[t_j^A\ t_j^D\right]\end{aligned}\tag{10}$$

$$\begin{aligned}0\le P_{\text{ev,dch},j}(t)&\le P_{\text{ev},j}^{\max}\cdot\left(1-I_{\text{ev,ch},j}\right)\\ &\forall\ t\in\left[t_j^A\ t_j^D\right]\end{aligned}\tag{11}$$

where $\text{SOC}_{\text{ev},j}^{\max}$ and $\text{SOC}_{\text{ev},j}^{\min}$ are the maximum and minimum allowable levels of SOC of battery of the jth EV (kWh), respectively, and $P_{\text{ev},j}^{\max}$ is the maximum charging and discharging power of the jth EV (kW).

D. Residential DR

The load demand in each house consists of controllable loads mainly related to TCLs (EWHs and HVAC systems) and uncontrollable loads (lighting and other appliances). Let us consider N houses in a residential community managed by the aggregator, with a set of different appliances, denoted as $\tilde{A}=\{A_1,\ldots,A_i,\ldots,A_N\}$. Each house i has a multiset of appliances (A_i), where each one can contribute to the controllable loads or noncontrollable loads of the house. The total load of the ith house at time t, $x_i(t)$ is given by

$$x_i(t)=\sum_{n\in A_i^{\text{ns}}}x_{i,n}^{\text{nc}}(t)+\sum_{m\in A_i^c}x_{i,m}^c(t) \tag{12}$$

$$x(t)=\sum_{i\in N}x_i(t) \tag{13}$$

where $x_{i,n}^{\text{nc}}(t)$ (kW) and $x_{i,m}^c(t)$ (kW) are, respectively, the power consumption of the nth noncontrollable and mth controllable loads of the ith house, A_i^{nc} and A_i^c correspond, respectively, to the set of noncontrollable and controllable load appliances available at the ith house, and $x_i(t)$ and $x(t)$ are, respectively, the total power consumption of the ith house (kW) and total power consumption of all houses in time t.

E. HVACs Load

It is assumed that there is a unit of the same HVACs in all N houses. Each house informs the aggregator about its preferred temperature range variation. Then, the aggregator tries to maintain thermal comfort by keeping the indoor temperature within specified ranges based on prediction over ambient temperature. In this article, discrete-time difference equation is used for local controllers [33]

$$\begin{aligned}T_i^{\text{hvac}}(t+\Delta t)&=\alpha_k T_i^{\text{hvac}}(t)+(1-\alpha_k)\big(T_k^{\alpha}(t)\\ &\quad-\big(\big(R_k\eta_{\text{cooling}}P_{k,i}^a(t)\big)\big(1-I_{\text{mode},t}^a\big)\\ &\quad+\big(-R_k\eta_{\text{heating}}P_{k,i}^a(t)\big)\big(I_{\text{mode},t}^a\big)\big)\big)\end{aligned}\tag{14}$$

$$\begin{cases}I_{\text{mode},t}^a=0 \text{ Cooling mode}\\ I_{\text{mode},t}^a=1 \text{ Heating mode}\end{cases}\tag{15}$$

$$T_{\text{set},i}^{\text{hvac}}(t)-Db_i^{\text{hvac}}\le T_i^{\text{hvac}}(t)\le T_{\text{set},i}^{\text{hvac}}(t)+Db_i^{\text{hvac}}\tag{16}$$

$$\alpha_k=e^{-\Delta t/C_kR_k}\tag{17}$$

$$P_{k,i}^{\min}\le P_{k,i}^a(t)\le P_{k,i}^{\max}.\tag{18}$$

In this model, when $T_k^{\alpha}(t)$ [ambient temperature (°C)] is less than a specific temperature that is a certain parameter and known as $T_{k,t}^{\text{mode}}$ [cooling and heating mode recognition (°C)], then $I_{\text{mode},t}^a=1$ (binary parameter) and it means that it is working in heating mode, and when $I_{\text{mode},t}^a=0$, it means that it is working in cooling mode. Therefore, $I_{\text{mode},t}^a$ is not a decision variable and it is used to discriminate between cooling and heating modes in the mathematical modeling. In this model, T_i^{hvac} is the indoor temperature of every house (°C), α_k is the system inertia of every HVAC, C_k is the thermal capacitance of every HVAC system (kW h/°C), R_k is the thermal resistance of every HVAC (°C/kW), and η_{cooling} and η_{heating} are working efficiency factors of each HVAC in either cooling or heating modes, respectively. HVACs local controllers try to keep temperature in a specific range that is defined by inhabitants using set point temperature $(T_{\text{set},i}^{\text{hvac}})$ and a dead band (Db_i^{hvac}), which helps to reduce power consumption. Power consumption of each HVAC $(P_{k,i}^a(t))$ should be between $P_{k,i}^{\min}$ that is minimum power consumption of HVAC (kW) and $P_{k,i}^{\max}$ that is the maximum power consumption of HVAC (kW) in ith house.

For simplicity, we make the following assumptions.

1) Each house has one HVAC unit of the same type and the house can be modeled as a large room exchanging thermal energy with the ambient environment. The indoor temperature is uniformly distributed within a house.
2) The impact of disturbances, such as humidity, solar radiation, and wind speed, on the residence thermal dynamics are assumed to be negligible compared to the influence of the ambient temperature, internal heat losses and gains, and HVAC output power.

F. EWHs Load

An EWH of the same type is available in every house, where each one has a different operational behavior. The thermostat control is set such that the output water temperature from the reservoir will fluctuate around the thermostat set point ($T_{\text{ewh}}^{\text{set},i}$) within the dead band of (Db_i^{ewh})

$$T_{\text{set},i}^{\text{ewh}}(t) - Db_i^{\text{ewh}} \leq T_i^{\text{ewh}}(t) \leq T_{\text{set},i}^{\text{ewh}}(t) + Db_i^{\text{ewh}} \tag{19}$$

where $T_i^{\text{ewh}}(t)$ is the temperature of hot water inside the tank of EWH in the ith house at time t (°C).

The temperature of hot water inside the EWH reservoir can be modeled by the following function [21], [34]:

$$\begin{aligned} T_i^{\text{ewh}}(t) = T_i^{\text{ewh}}(t-\Delta t)e^{-\left(\dfrac{\Delta t}{F_i(t).C}\right)} + \left[1 - e^{-\left(\dfrac{\Delta t}{F_i(t).C}\right)}\right] \\ \cdot\{E.F_i(t).T_{\text{env},i}(t) + G_i(t).F_i(t).T_{\text{in}}(t) \\ + Q_i(t).F(t)\} \end{aligned} \tag{20}$$

where C is the equivalent thermal mass (kJ/°C), E is the ratio of the surface area to thermal resistance of the reservoir, $T_{\text{env},i}$ is the ambient environment temperature at the ith house (°C), T_{in} is the incoming cold water temperature (°C), and $F_i(t)$, $G_i(t)$, and $Q_i(t)$ are parameters given by the following equations:

$$F_i(t) = \frac{1}{G_i(t) + E} \tag{21}$$

$$E = \frac{S_{\text{res}}}{R} \tag{22}$$

$$C = \rho_{\text{water}}.V.C_p \tag{23}$$

$$Q_i(t) = 3.4121 \cdot 10^3 \cdot \theta_i^{\text{ewh}}(t) \tag{24}$$

$$G_i(t) = \rho_{\text{water}} \cdot C_p \cdot D_{\text{wt},i}(t) \tag{25}$$

$$\theta_{i,\min}^{\text{ewh}} \leq \theta_i^{\text{ewh}}(t) \leq \theta_{i,\max}^{\text{ewh}} \tag{26}$$

where R is the thermal insulation (m^{2}°C.hr/kJ), S_{res} is the reservoir surface area (m^2) , ρ_{water} is the water density (kg/m^3), C_p is the specific heat water [4.2157 kJ/(kg. °C)], V is the water reservoir capacity (m^3), θ_i^{ewh} is power consumption of element (kW), $\theta_{i,\max}^{\text{ewh}}$ and $\theta_{i,\min}^{\text{ewh}}$ are the max and min power consumption of EWH (kW) in the ith house, respectively, and $D_{\text{wt},i}(t)$ is the water demand (m^3/h).

IV. EMSA

A. Optimization Problem

1) Objective Function: Since P_g is a decision variable that takes positive values when aggregator decides to buy electricity from the grid and takes negative values when it wants to sell electricity to the grid, it means that EMSA is trying to minimize the cost of purchase while selling more power to the grid. Also, EMSA aims to minimize the degradation costs associated with ESS and EVs

$$\begin{aligned} J = \min \ \sum_{t=1}^{T} C_g^{p,s}(t)P_g(t) + \beta \cdot \sum_{t=1}^{T}\sum_{j=1}^{N_{\text{ev}}}\big(P_{\text{ev,ch},j}(t) \\ + P_{\text{ev,dch},j}(t)\big) + \sum_{t=1}^{T} \lambda \cdot \big(P_{b,\text{ch}}(t) + P_{b,\text{dch}}(t)\big) \end{aligned} \tag{27}$$

where $C_g^{p,s}$ is the cost of electricity purchased/sold from/to the electric grid at instant t and $P_g(t)$ is the amount of power purchased/sold from/to the electric grid at instant t. It is worth mentioning that there is algebraic multiplication between $C_g^{p,s}$ and $P_g(t)$.

β and λ are degradation costs associated with the ESS and batteries of EVs, respectively. The degradation cost of batteries is evaluated as a linear function of the battery charging/discharging power [35].

2) Demand-Supply Balance Constraint: The power balance in the microgrid is given by the following equation:

$$\begin{aligned} \Delta P_{\text{bal}}(t) = P_{\text{PV}}(t) - P_{b,\text{ch}}(t) + P_{b,\text{dch}}(t) - \sum_{i=1}^{I} P_{k,i}(t) \\ - \sum_{i=1}^{I}\sum_{n=1}^{A_i^{\text{ns}}} x_{i,n}^{\text{nc}}(t) - \sum_{i=1}^{I} \Theta_i^{\text{ewh}}(t) \\ = \sum_{j=1}^{N_{\text{ev}}} \big(P_{\text{ev},c,j}(t) - P_{\text{ev,dch},j}\big) - P_g(t) \quad \forall t \in T. \end{aligned} \tag{28}$$

It is assumed that the EMSA purchases power for negative power balance and sells power for positive power balance

$$\begin{cases} P_g(t) > 0, & \text{if } \Delta P_{\text{bal}}(t) \leq 0 \\ P_g(t) \leq 0, & \text{if } \Delta P_{\text{bal}}(t) > 0. \end{cases} \tag{29}$$

3) Transaction With the Main Grid: Considering the influences on the main grid, power exchange with the main grid is constrained as follows:

$$-P_{g,\max} \leq P_g(t) \leq P_{g,\max} \tag{30}$$

where $P_{g,\max}$ represents the maximum power limit related to power exchange with the main grid based on the capacity of electricity transmission line.

B. Optimization Strategy

The conventional MPC algorithm follows the receding horizon paradigm. The MPC aims to use the current existing knowledge related to the system along with prediction of uncertain parameters such as weather conditions and solar

Algorithm 1 Proposed Algorithm

```
Function Controller_EMSA(status que states of the system):
    Solving Mixed Integer Linear Programming problem based on the equations (2),…, (30);
    return Control Signals for the next time step including: $P_g(t)$, $P_{b,\mathrm{ch}}(t)$, $P_{b,\mathrm{dch}}(t)$,…;
;
Function Plan(Parameters, control signals and real-time current situation of the system):
    Control signals related to ESS, EVs are applied as were predicted by Controller_EMSA based on equations (2) and (6);
    $T_i^{\mathrm{hvac}}$(HVACs states) and $T_i^{\mathrm{ewh}}$(EWHs states) are estimated by using control signals from Controller_EMSA
    if HVACs and EWHs temperatures are not within boundaries of equations (16) & (19) then
        Q-learning agents will apply corrective signals based on Q-TABLE learnied in Algorithm 2;
    else
        Control signals from Controller_EMSA  is applied to the HVACs and EWHs;
    end
    Deciding about $P_g(t)$ based on known control signals and corrective power signals;
    return Current states of HVACs, EWHs, EVs, ESS;
initialization of the parameters;
while time $\leq$ total simulation time do
    Function Controller_EMSA
    Function Plan
end
```

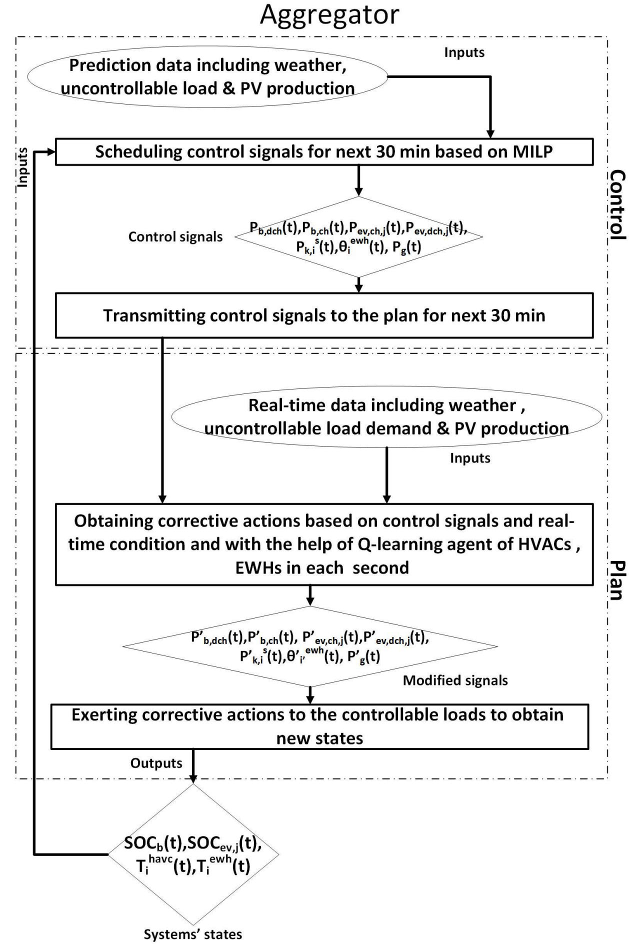
Aggregator
Inputs
Prediction data including weather, uncontrollable load & PV production
Inputs
Scheduling control signals for next 30 min based on MILP
Control
Control signals
$P_{b,dch}(t),P_{b,ch}(t),P_{ev,ch,j}(t),P_{ev,dch,j}(t), P_{k,i}{}^{s}(t),\theta_i{}^{ewh}(t), P_g(t)$
Transmitting control signals to the plan for next 30 min
Real-time data including weather , uncontrollable load demand & PV production
Inputs
Obtaining corrective actions based on control signals and real-time condition and with the help of Q-learning agent of HVACs , EWHs in each second
Plan
$P'_{b,dch}(t),P'_{b,ch}(t), P'_{ev,ch,j}(t),P'_{ev,dch,j}(t), P'_{k,i}{}^{s}(t),\theta'_{i'}{}^{ewh}(t), P'_g(t)$
Modified signals
Exerting corrective actions to the controllable loads to obtain new states
Outputs
$SOC_b(t),SOC_{ev,j}(t), T_i{}^{havc}(t),T_i{}^{ewh}(t)$
Systems' states

Fig. 2. Aggregator behavior and algorithm.

radiation to make decisions. In this regard, in the first step of MPC, the actual current state of the system should be initiated. Then, based on the current information about the states of the system and predictions, an optimal control sequence is computed for the selected prediction horizon. In the next step, the first element in the control sequence is applied to the system. After that, the information about the states of the system will be updated. Finally, the algorithm moves to the next sampling instant and the same algorithm is repeated [36], [37].

In this article, a two-level MPC algorithm integrating Q-learning, a reinforcement learning algorithm, is designed. First, in day-ahead operation, MILP-MPC tries to find the control signals for system aggregator. Second, in real-time operation, a Q-learning algorithm is implemented to address the real-time decisions based on real-time data.

In each epoch of the algorithm, based on the predicted input data, the first-level controller (MILP-MPC) finds an optimal solution in the predefined control horizon (24 h), and then, it sends control signals of the first-level controller to the local plans. In this latter, control signals related to ESS and EVs received from the controller are implemented without any changes and the next states of charges of ESS and EVs are updated.

However, the next states of EWHs and HVACs are highly dependent on the prediction of ambient temperature. Therefore, current state of HVACs and EWHs temperatures are updated based on control signals received from the first-level controller. If the temperatures of HVAC or EWH systems are within the boundaries of comfort zone of residents, then there is no need to do any corrective actions; otherwise, if the temperature delivered by the EMSA is below or beyond the comfort zone due to prediction errors, then corrective actions need to be integrated using Q-learning as a second-level fast controller. It should be mentioned that in the plan, there is a local controller for each TCL located in each house. Then, new states of the system as shown in Algorithm 1 are sent to the controller to repeat the procedure for the next time step and later based on the new collected data and states, and next input data are predicted for the next cycle. In each cycle, control horizon shifts forward one step and predictions and states of the system are updated. The proposed aggregator algorithm is shown in Fig. 2. In the proposed optimization strategy, control horizon has been chosen to be equal to 4 h with a time interval of 30 min.

C. Forecast Information

Neural networks are implemented to predict the ambient temperature. Training and test data were collected from the

Montréal-Pierre Elliott Trudeau International Airport. To train the predictor, we have used 17 465 data points for the year 2017 where 70% was used for training, 15% for validation, and 15% for testing the model. The mean square error of the proposed neural network on the validation data was nearly 1.32 °C. Due to the lack of data, white noise with mean of zero and variance of 0.05 has been added to noncontrollable loads and solar radiations to reflect the prediction errors.

D. Q-Leaning as Real-Time Power Modifier

In this study, we have implemented an agent for every TCL in each house. The agents in the plan are trained based on the comfort zones of the occupants. Q-learning is a model-free reinforcement learning approach with three important components, including possible states (S), actions (A), and rewards (R) [38]. Q-learning is based on learning a Q-table, which is a tensor having all measurable states of the system and possible actions as dimension. Moving through episodes in the training phase, the agent tries to learn a Q-table, which consists of all possible rewards for each possible action in a specific state. The agent will choose the action with the highest reward in that specific state. Algorithm 2 explains the Q-learning training algorithm.

Algorithm 2 Q-Learning Algorithm

```
initialization of $Q(s, a)$ $\forall s \in S, a \in A$, arbitrarily;
Defining reward(s) $\forall s \in S$;
while episode $\leq$ Max episode do
    Randomly choosing initial state s;
    while move $\leq$ Max number of moves in episode do
        Choose action derived from e-greedy;
        Taking action a and observing reward and next state s';
        $Q^{\text{new}}(s, a) = Q^{\text{previous}}(s, a) + \alpha[R + \gamma \max_a Q(s', a') - Q(s, a)]$;
        s $\leftarrow$ s'
    end
end
```

States in HVAC agent are the ambient temperature and the temperature inside the house. Also, states in EWH agent are the HVAC temperature, the temperature inside EWH, and the water flow. To make better decisions and minimize power consumption, the reward function is defined as follows:

$$\text{reward} = \text{reward}(s) - \zeta \times \text{selected action}. \qquad (31)$$

In order to train the EWH agent, Q-table that is state–action tuple should be formed. The hot water temperature inside the tank of EWH as an element of the Q-table tuple is discretized by step of 0.1 °C from 62 °C to 71 °C. The ambient temperature at the ith house is discretized by the steps of 0.1 °C from 18 °C to 25 °C. Water demand as the last element of the tuple is discretized by a step of 1 (USgal/h) from 0 (USgal/h) to 9 (USgal/h). It is worth mentioning that each 1 (USgal/h) is equal to 0.00378541 (m^3/h).

Whenever an agent reaches to keep the temperature inside the EWH within the comfort boundaries defined by tenants, he will receive a reward of 20, and otherwise, the reward will be -1. It is worth mentioning that the reward is obtained when the agent succeeds to choose an amount of power to keep the temperatures within boundaries defined by dwellers. The punishment of the agent occurs if temperatures fall below or exceed the upper boundaries of the comfort zone. ζ is defined to be equal to 4 in order to teach the agent that it is not always a good idea to use more power and he should keep the temperature of the HVACs and the EWHs within boundaries using minimum power consumption. Based on the importance of future actions, the value of γ is equal to 0.96. The ambient temperature (outside temperature) is discretized using a step of 0.1 °C from -20 °C to 0 °C and the temperature inside the house is chosen to be between 18 °C and 25 °C. Whenever actions selected by the HVAC agent could keep the inside temperature within the comfort zone defined by the tenants, the agent will be given the reward of 20; otherwise, it will be given the reward of -1. Other parameters are such as the ones used in the EWH agent.

V. Case Study

The proposed two-level optimization model is demonstrated for a set of 60 houses. All houses are equipped with a Level 2 charging infrastructure with the output power of 10 kW, which can fully charge the EV batteries in 4 h. An EV can be connected to each house. Six types of EVs can be found in the residential community: 1) ford focus electric car with a battery capacity of 33.5 kWh; 2) Nissan Leaf (ZE1) with a battery capacity of 40 kWh; 3) Hyundai Loniq with a battery capacity of 28 kWh; and 4) Kia Soul with a battery capacity of 30.5 kWh. Each house is equipped with HVAC and EWH systems with a maximum power consumption of 5 kW. To run the simulation, we have used MATLAB version 2018 and its built-in modules for optimization of MILP. The input data related to solar irradiation and ambient temperature are gathered for Montreal city. For the noncontrollable loads, data are generated based on usage patterns of various appliances. For the electricity pricing, one day-ahead and real-time prices of Chicago have been used, respectively, in the controller and in the plan module [39].

Based on the abovementioned parameters, we have examined two scenarios with different optimizers in the controller and the plan as follows.

1) *Scenario 1 (Aggregated Mode):* MPC as optimizer in the controller of EMSA and Q-learning reinforcement learning model in the Plan to calculate and apply corrective powers.
2) *Scenario 2 (Solely Mode):* MPC as an optimizer in the controller of home energy management system (HEMS) and Q-learning reinforcement learning in plan to calculate and apply corrective powers. In this case, the HEMS's objective is to minimize home costs related to electricity consumption and degradation cost of EV's battery.

VI. Results

In this section, the results obtained from the proposed algorithm using different optimizers will be discussed. It should

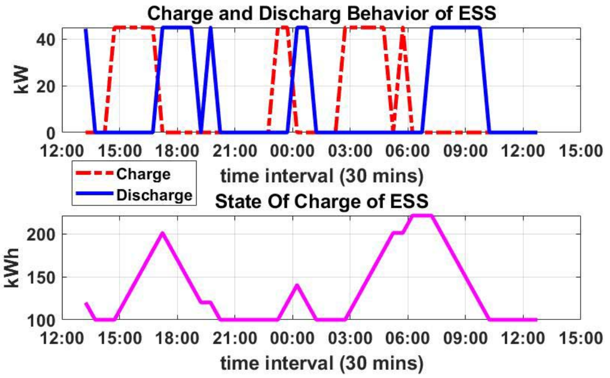

Fig. 3. ESS in aggregated mode.

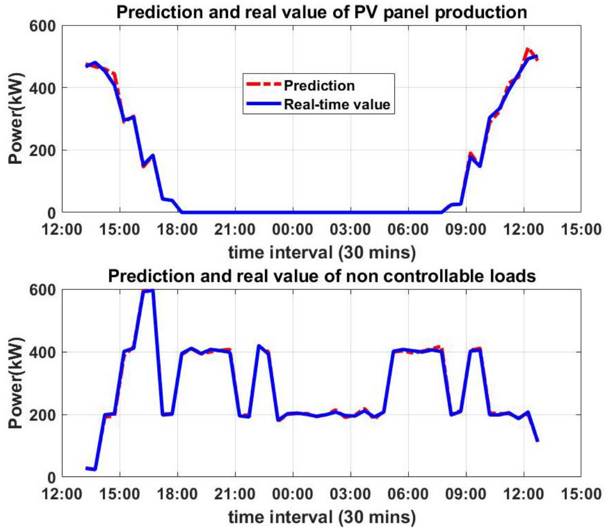

Fig. 4. PV production and noncontrollable loads.

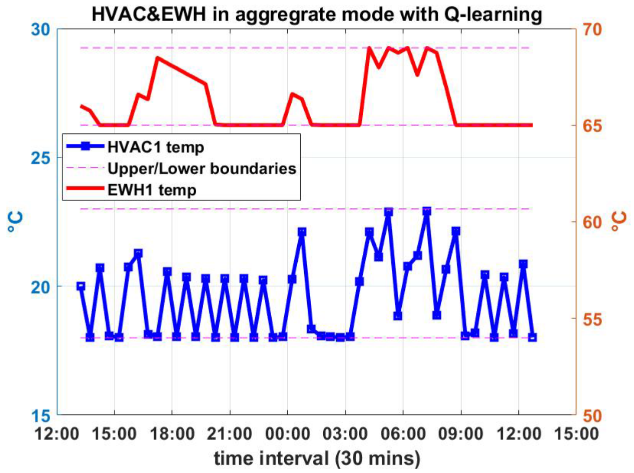

Fig. 5. Temperature changes of HVAC1 and EWH1 in aggregate mode with Q-learning.

be mentioned that all figures show the results and data for a period of 24 h from March 30, 2018, 13:14’ to March 31, 2018, 12:44’.

Fig. 3 shows the state of charge, charging and discharging behavior of the ESS available at the level of the residential community (aggregator). It can be depicted from Fig. 3 that the charging and discharging of ESS occur during the whole time period. This means that ESS can be charged from both the PV system and the main grid.

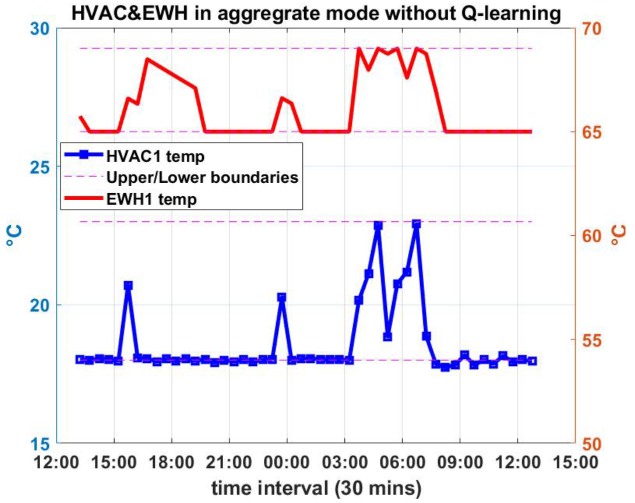

Fig. 6. Temperature changes of HVAC1 and EWH1 in aggregated mode without Q-learning.

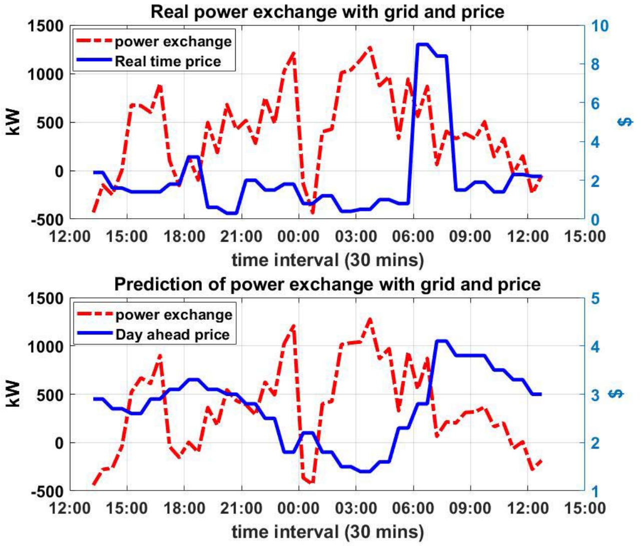

Fig. 7. Power exchange with the grid.

Fig. 4 shows the PV power generation and noncontrollable loads. As it can be seen, the first plot shows both real PV generation values and its prediction. Slight errors are observed comparing both plots. Also, as shown in Fig. 5, temperature controls of TCLs (EWH and HVAC) in house 1 using Q-learning-based reinforcement learning tend to track the minimum boundary of set points with some variations. However, Fig. 6 shows the behavior of the same TCL units in the case where no Q-learning (reinforcement learning algorithm) is used to handle the uncertainties in real time, and as it is illustrated, HVAC cannot keep the temperature within the desired boundaries. This issue is especially highlighted after 9 o’clock in Fig. 6. This remark means that tenants’ comfort can be violated where there is no Q-learning.

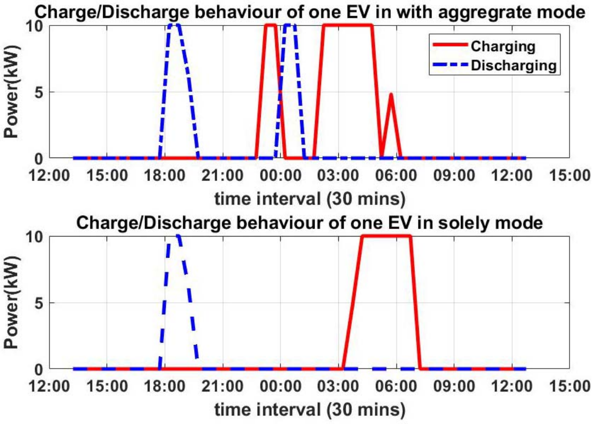

Fig. 8. Charge/discharge of an EV in with aggregate/sole.

Based on Fig. 7, the difference between power exchange with the grid for controller and plan is related to the amount of PV power generation, noncontrollable loads, and ambient temperature predictions. A good forecasting mechanism can properly predict the behavior of the abovementioned values, which leads to an appropriate microgrid management. In this article, white noise with tolerance of 10% and 8% was added to the PV production and noncontrollable loads, respectively, to investigate the effects of prediction errors. In addition, we have developed an artificial neural network model as described earlier to predict the ambient temperature. Since the decision has been taken based on one-day-ahead pricing, the power exchange with the grid tends to be influenced by one-day-ahead pricing and the only aspect that can minimize the differences between the controller output and plan is a good prediction of prices.

According to the results obtained in Fig. 5, it can be concluded that from the residents' comfort viewpoint, comfort level of the tenants is preserved. Despite having upper and lower desired bounds for HVAC and EWH, temperature variations are minimized.

As shown in Fig. 8, for house 1 in aggregated mode, the EMSA tends to frequently charge and discharge the battery compared to solely mode in order to compensate for lack of power in the grid. Also, it can be seen that the batteries of EVs tend to be charged more in the aggregated mode than in the solely mode.

Fig. 9 shows HVAC1 control signals considering Q-learning and without Q-learning as the reinforcement learning approach in the plan using MPC. According to Figs. 5, 6, and 9, we can see that besides power consumption pattern that is less when Q-learning is implemented in MPC, HVAC temperature control is different in both modes. For example, in the aggregate mode with Q-learning, we remark that temperature is always within boundaries defined by tenants, but this never happens in aggregated mode without Q-learning where comfort of dwellers is highly violated.

According to Fig. 10 and Table I, the sum of the power sold/bought to/from grid in aggregated mode is 286.25 kW and for solely mode is 297.46 kW. This means that EMSA in aggregated mode tends to buy less power and/or sell more power compared to solely mode. Moreover, it is true that we still need to buy power from grid. However, overall less power is purchased in aggregated mode. Also, Fig. 11 shows the importance of using Q-learning as a reinforcement learning approach in the plan in addition to Figs. 6 and 9. It is worth mentioning that Q-learning could deal with uncertainties and

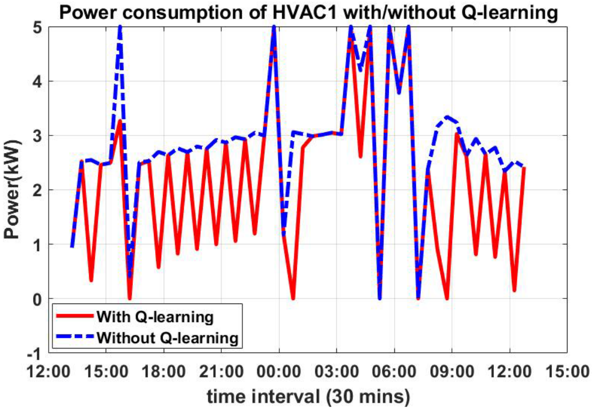

Fig. 9. One HVAC control signals in aggregate mode with/without Q-learning.

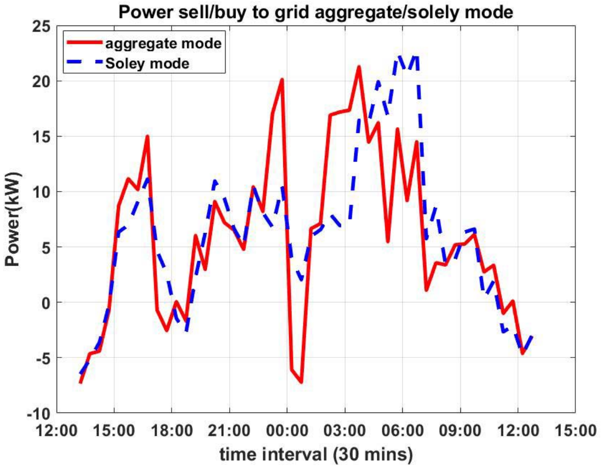

Fig. 10. Power sell/buy to/from grid in aggregate/solely mode.

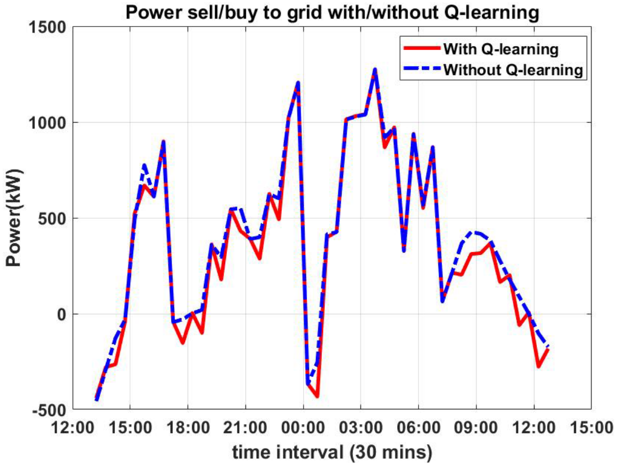

Fig. 11. Power sell/buy to/from grid aggregated mode with/without Q-learning.

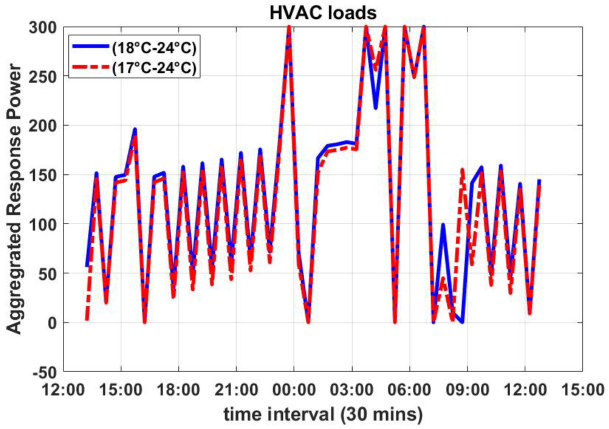

Fig. 12. Aggregated response of HVACs in different comfort zones.

TABLE I

OBJECTIVE FUNCTIONS VALUES OF ONE HOUSE

Type	Objective function value
MILP-MPC aggregated mode with Q-learning	286.25
MILP-MPC aggregated mode without Q-learning	318.19
solely mode with Q-learning	297.46

modify the control signals in real time in order to maximize the benefits of dwellers.

If tenants change their comfort zones, the system will change its power consumption behavior. As it can be seen in Fig. 12, if we assume that all houses have the same comfort zone, we can see that by increasing lower boundary of comfort zone, winter power consumption will increase because the controller tends to keep the temperature within the comfort zone of tenants. In winter, controller tries to keep the temperature near the lower boundary to reduce power consumption, and when the lower boundary of comfort zone increases, it means that compared to the previous condition, more energy is needed to maintain the temperature within the lower bound. The same behavior exists for summer period, and if tenants reduce the upper bound of comfort zone, then more power will be consumed.

Table I shows the objective function values of one house in different scenarios. Based on these results, we conclude that houses could gain more benefits when they are operating in the aggregate mode and if they have a Q-learning algorithm in plan to modify the control signals and deal with uncertainties in real time. Therefore, aggregator makes them more able to buy less power from the grid due to its increased flexibility. An objective function value for one house in aggregated mode with Q-learning is better than one house in solely mode with Q-learning, which shows the role of the aggregator. However, aggregated mode without Q-learning has a higher objective function compared to solely mode with Q-learning. This remark demonstrates the importance of Q-learning for online modification of control signals.

MPC could find the optimal solution for an aggregator supporting 60 houses within approximately 10 s for each time step. This approach could be implemented for real-time applications. The online Q-learning controller implemented in the plan is very fast and can take decisions in milliseconds.

VII. CONCLUSION

In this article, we proposed a novel energy management system using an aggregator, which can use MILP-MPC in the controller and Q-learning in the plan to take decisions in a residential community of houses. Also, in this article, we compared the results of MILP-MPC in the aggregated and solely modes. Results show that in the aggregated mode, houses gain more benefits compared to solely mode through selling more power and buying less power to/from the grid. In addition, the implementation of Q-learning algorithm in the plan allows modifying the control signals accordingly while dealing with uncertainties in real time.

For future works, we plan to develop proper prediction mechanisms using newly developed deep learning methods to better forecast all uncertain parameters, including arrival time, departure time of EVs, PV generation, and noncontrollable loads. In addition, the hierarchical control system can be proposed with more levels of decisions while addressing various energy markets and grid ancillary services.

REFERENCES

[1] *Transforming the Market: Energy Efficiency in the Buildings*, World Bus. Council Sustain. Develop., Geneva, Switzerland, 2009.

[2] X. Guan, Z. Xu, and Q.-S. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, Dec. 2010.

[3] M. C. Vlot, J. D. Knigge, and J. G. Slootweg, "Economical regulation power through load shifting with smart energy appliances," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1705–1712, Sep. 2013.

[4] Z. Zhou, F. Zhao, and J. Wang, "Agent-based electricity market simulation with demand response from commercial buildings," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 580–588, Dec. 2011.

[5] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[6] M. Muratori and G. Rizzoni, "Residential demand response: Dynamic energy management and time-varying electricity pricing," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1108–1117, Mar. 2016.

[7] G. T. Costanzo, G. Zhu, M. F. Anjos, and G. Savard, "A system architecture for autonomous demand side load management in smart buildings," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2157–2165, Dec. 2012.

[8] H. Nakano *et al.*, "Aggregation of V2H systems to participate in regulation market," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 668–680, Apr. 2021.

[9] F. Pallonetto, M. De Rosa, F. Milano, and D. P. Finn, "Demand response algorithms for smart-grid ready residential buildings using machine learning models," *Appl. Energy*, vol. 239, pp. 1265–1282, Apr. 2019.

[10] R. Lu and S. H. Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network," *Appl. Energy*, vol. 236, pp. 937–949, Feb. 2019.

[11] M. Beaudin, H. Zareipour, A. K. Bejestani, and A. Schellenberg, "Residential energy management using a two-horizon algorithm," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1712–1723, Jul. 2014.

[12] S. M. S. Basnet, H. Aburub, and W. Jewell, "Residential demand response program: Predictive analytics, virtual storage model and its optimization," *J. Energy Storage*, vol. 23, pp. 183–194, Jun. 2019.

[13] Y. Cao, J. Du, and E. Soleymanzadeh, "Model predictive control of commercial buildings in demand response programs in the presence of thermal storage," *J. Cleaner Prod.*, vol. 218, pp. 315–327, May 2019.

[14] Y. Sun, M. Elizondo, S. Lu, and J. C. Fuller, "The impact of uncertain physical parameters on HVAC demand response," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 916–923, Mar. 2014.

[15] N. Kampelis, A. Ferrante, D. Kolokotsa, K. Gobakis, L. Standardi, and C. Cristalli, "Thermal comfort evaluation in HVAC demand response control," *Energy Procedia*, vol. 134, pp. 675–682, Oct. 2017.

[16] N. Lu, "An evaluation of the HVAC load potential for providing load balancing service," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1263–1270, Sep. 2012.

[17] S. A. Pourmousavi, S. N. Patrick, and M. H. Nehrir, "Real-time demand response through aggregate electric water heaters for load shifting and balancing wind generation," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 769–778, Mar. 2014.

[18] V. Rigoni, D. Flynn, and A. Keane, "Coordinating demand response aggregation with LV network operational constraints," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 979–990, Mar. 2021.

[19] F. Ruelens, B. J. Claessens, P. Vrancx, F. Spiessens, and G. Deconinck, "Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning," *CSEE J. Power Energy Syst.*, vol. 5, no. 4, pp. 423–432, Dec. 2019.

[20] C. Zhang, S. R. Kuppannagari, R. Kannan, and V. K. Prasanna, "Building HVAC scheduling using reinforcement learning via neural network based model approximation," in *Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, Nov. 2019, pp. 287–296.

[21] C. Jin, J. Tang, and P. Ghosh, "Optimizing electric vehicle charging with energy storage in the electricity market," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 311–320, Mar. 2013.

[22] M. Raoofat, M. Saad, S. Lefebvre, D. Asber, H. Mehrjedri, and L. Lenoir, "Wind power smoothing using demand response of electric vehicles," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 164–174, Jul. 2018.

[23] L. Jian, Y. Zheng, and Z. Shao, "High efficient valley-filling strategy for centralized coordinated charging of large-scale electric vehicles," *Appl. Energy*, vol. 186, pp. 46–55, Jan. 2017.

[24] Y.-M. Wi, J.-U. Lee, and S.-K. Joo, "Electric vehicle charging method for smart homes/buildings with a photovoltaic system," *IEEE Trans. Consum. Electron.*, vol. 59, no. 2, pp. 323–328, May 2013.

[25] T. Wu, Q. Yang, Z. Bao, and W. Yan, "Coordinated energy dispatching in microgrid with wind power generation and plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1453–1463, Sep. 2013.

[26] G. Liu, Y. Xu, and K. Tomsovic, "Bidding strategy for microgrid in day-ahead market based on hybrid stochastic/robust optimization," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 227–237, Jan. 2016.

[27] S. Chen, Q. Chen, and Y. Xu, "Strategic bidding and compensation mechanism for a load aggregator with direct thermostat control capabilities," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2327–2336, May 2018.

[28] C. A. Correa-Florez, A. Michiorri, and G. Kariniotakis, "Optimal participation of residential aggregators in energy and local flexibility markets," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1644–1656, Mar. 2020.

[29] K. Bruninx, H. Pandzic, H. Le Cadre, and E. Delarue, "On the interaction between aggregators, electricity markets and residential demand response providers," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 840–853, Mar. 2020.

[30] F. Wang *et al.*, "Smart households' aggregated capacity forecasting for load aggregators under incentive-based demand response programs," *IEEE Trans. Ind. Appl.*, vol. 56, no. 2, pp. 1086–1097, Jan. 2020.

[31] Z. Wang, R. Paranjape, Z. Chen, and K. Zeng, "Layered stochastic approach for residential demand response based on real-time pricing and incentive mechanism," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 3, pp. 423–431, Feb. 2020.

[32] F. Delfino, G. Ferro, R. Minciardi, M. Robba, M. Rossi, and M. Rossi, "Identification and optimal control of an electrical storage system for microgrids with renewables," *Sustain. Energy, Grids Netw.*, vol. 17, Mar. 2019, Art. no. 100183.

[33] S. Nan, M. Zhou, and G. Li, "Optimal residential community demand response scheduling in smart grid," *Appl. Energy*, vol. 210, pp. 1280–1289, Jan. 2018.

[34] G. Goddard, J. Klose, and S. Backhaus, "Model development and identification for fast demand response in commercial HVAC systems," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 2084–2092, Jul. 2014.

[35] W. Su, J. Wang, and J. Roh, "Stochastic energy scheduling in microgrids with intermittent renewable energy resources," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1876–1883, Jul. 2014.

[36] A. Ouammi, "Optimal power scheduling for a cooperative network of smart residential buildings," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1317–1326, Jul. 2016.

[37] F. Liberati, A. D. Giorgio, A. Giuseppi, A. Pietrabissa, E. Habib, and L. Martirano, "Joint model predictive control of electric and heating resources in a smart building," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 7015–7027, Nov. 2019.

[38] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[39] Commonwealth Edison (ComEd), Chicago, IL, USA. *Live Prices: ComEd's Hourly Pricing Program*. Accessed: Jan. 2019. [Online]. Available: https://hourlypricing.comed.com/live-prices/