Enabling cooperative behavior for building demand response based on extended joint action learning

Citation for published version (APA):

Hurtado Munoz, L. A., Mocanu, E., Nguyen, H. P., Gibescu, M., & Kamphuis, I. G. (2018). Enabling cooperative behavior for building demand response based on extended joint action learning. *IEEE Transactions on Industrial Informatics*, 14(1), 127-136. [8039194]. <https://doi.org/10.1109/TII.2017.2753408>

DOI:

[10.1109/TII.2017.2753408](https://doi.org/10.1109/TII.2017.2753408)

Document status and date:

Published: 01/01/2018

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Enabling cooperative behavior for building demand response based on extended joint action learning

L.A. Hurtado*, E. Mocanu[†], P.H. Nguyen[†], M. Gibescu[†], I.G. Kamphuis[†]

*DNV GL - Energy Advisory, the Netherlands

[†]Eindhoven University of Technology, the Netherlands

luis.hurtado@dnvgl.com*, e.mocanu@tue.nl[†], p.nguyen.hong@tue.nl[†], m.gibescu@tue.nl[†], i.g.kamphuis@tue.nl[†]

Abstract—This paper explores the use of distributed intelligence to assist the integration of the demand as a flexible resource, to mitigate the emerging uncertainty in the power system, while fulfilling the customer’s local needs, i.e., comfort management. More exactly, our contribution is two-fold. Firstly, we propose a novel cooperative and decentralized reinforcement learning method, dubbed extended joint action learning (eJAL). Secondly, we perform a comparison between eJAL to non-cooperative decentralized decision making strategies, i.e., Q-learning, and a centralized game theoretic approach, i.e., Nash n-player game. This comparison has been conducted on the basis of grid support effectiveness and the loss of comfort for each customer. Various metrics were used to analyze the advantages and disadvantages of each method. We demonstrated that a range of flexibility requests can be met by providing an optimal energy portfolio of buildings without substantially violating comfort constraints. Moreover, we showed that the proposed eJAL method achieves the highest fairness index.

Index Terms—Demand flexibility, Demand Response, Decision-making, Cooperation, Distributed control, Multi-agent systems, Smart grid.

I. INTRODUCTION

The increased penetration of renewable energy sources (RES) and distributed energy resources (DER) into the power system can potentially lead to problems due to their intrinsically stochastic nature [1]–[4]. These problems can be either at the system level such as supply-demand mismatch, or network level like transport congestions. Among the different mechanisms developed to deal with the increasing operational uncertainty, *Demand flexibility* from demand response (DR) can be considered as an alternative resource to provide support services to the grid operation in a reasonable time [5]–[7]. It is generally defined as ‘*the changes in consumption/injection of electrical power from/to the power system from their current/normal patterns in response to certain signals, either voluntarily or mandatory*’ [8]. In literature, demand flexibility has been shown to: contribute to network congestion resolution; increasing asset utilization; peak demand reduction; and energy balancing; aiding in accommodating a higher penetration of RES [9]–[13].

Being responsible for about one-third of the electrical energy consumed in cities [14], non-residential commercial customers play a significant role in providing demand flexibility [15]–[18]. However, this potential is limited by the comfort requirements of the occupants [19]. Different studies have investigated the relationship between demand flexibility

and comfort, showing the flexibility potential in either comfort optimization; passive thermal energy storage; or heating ventilation and air conditioning system (HVAC) control [20], [21]. Some other studies have tried to define how comfort defines and affect flexibility [22], [23]. Nevertheless, a clear use of comfort as part of the decision variables is still lacking. Especially, to guarantee the correct operation of the building when acting as a flexible resource for DR.

Furthermore, the distributed nature and size of active customers require the demand flexibility to be aggregated (the demand flexibility potential of commercial buildings is a scarce resource, that is bounded by comfort) [15], [16]. Therefore, it is paramount to define an adequate decision making strategy to maximize the welfare of the different stakeholders in DR. In literature, reinforcement learning (RL) has become an increasingly popular technique for decision making due to its capacity to solve problems without initial knowledge of the environment [24], [25]. Inspired by behavioral psychology, RL aims to map the link between actions and environmental states in order to maximize a reward, using only the experience of the decision maker in an unsupervised learning process. This makes RL an interesting option for the coordination of DR resources in a dynamic and large scale environment, such as the emerging power system [26], [27].

This paper proposes a scalable architecture based on Multi-Agent Systems (MAS) to exploit the demand flexibility from a cluster of buildings, while minimizing the individual impacts on comfort. Accordingly, this paper proposes a novel cooperative and decentralized method for decision making based on reinforcement learning, dubbed: *extended joint action learning* (eJAL). This method is compared to a decentralized non-cooperative method, i.e., Q-learning, and to a centralized non-cooperative one, i.e., Nash n-player game, on the basis of grid support effectiveness, fairness, and comfort loss.

To summarize, the main contributions of this paper are:

- A novel cooperative and decentralized method for decision making strategies in the smart grid context based on reinforcement learning, namely eJAL.
- A comparison of centralized versus decentralized and cooperative versus non-cooperative strategies. This comparison is limited to one type of method for each strategy.
- A conceptualization of the relation between comfort and demand response, based on empirical observations of how comfort affects demand response and vice versa.

These lead to the development of an agent based platform to exploit and manage the demand flexibility potential of non-residential buildings, while taking into account the individual building dynamics.

The remainder of this paper is structured as follows. Section II explores the concept of demand flexibility and comfort, in order to formulate the problem. Section III introduces the agent based control framework, and discuss the various decision making strategies. Section IV describes the proposed method, followed by Section V where the results are shown. Lastly, the paper is closed with conclusions drawn from the work.

II. BUILDING DEMAND FLEXIBILITY

Within buildings, the main objective is to provide and maintain a comfortable and healthy environment for the occupants, regardless of the specific function of the building or building zone, e.g., retail shop, office building, conference room, or toilette [28]. This means that demand flexibility could potentially have a negative impact on the building operation [19]. Thus, to avoid any negative impacts from DR, it is paramount to establish the relationship between the comfort dynamics of buildings and the available demand flexibility.

A. Comfort and energy

Comfort is a complex and subjective human perception, defined mostly by the thermal, indoor air, visual, and acoustic characteristics of the building. Based on the previous work [29], the comfort index (C) is conceptualized as a function of the indoor temperature (T) and relative humidity (R)¹, expressed by a combination of two Gaussian functions representing thermal and air quality comfort, and varying between 0.0, i.e., no comfort, and 1.0, i.e., complete satisfaction, as shown in the following equation:

$$C = \underbrace{\xi e^{\left[\frac{-(T-\mu_T)^2}{2\sigma_T^2}\right]}}_{\text{Thermal comfort}} + \underbrace{(1-\xi)e^{\left[\frac{-(R-\mu_R)^2}{2\sigma_R^2}\right]}}_{\text{Air quality comfort}} \quad (1)$$

where, $\xi \in [0, 1]$ is a weight factor; μ_T is the mean temperature value or the optimal temperature set point; σ_T is the thermal comfort standard deviation, i.e., comfort band; μ_R is the mean humidity; and σ_R is the standard deviation for air quality comfort.

The first part in eq. (1) represents the thermal comfort, i.e., the change in time of the temperature in the building. This change can be modeled applying the energy conservation principle, as a function of the internal heat gains, the contributions of the air handling unit (AHU) and heating system, and the heat losses to the outdoor environment. The second part represents the air quality comfort measured through the relative humidity dynamics in the air, which can be represented through the mass component balances in the air volume. More details on the relation between comfort management and energy demand are given in [29].

¹ The CO_2 concentration levels is an important part of the air quality comfort. However, in this paper the CO_2 concentration level is used as a system constraint according to [29].

The total power consumption (P_{total}) in kilowatts [kW] of a building is the result of the operation of the different systems, i.e., comfort and non-comfort, present in the building.

$$P_{total} = \underbrace{P_{Aq} + P_{Tc}}_{\text{comfort}} + \underbrace{\sum_{j=1}^z P_j}_{\text{non-comfort}} \quad (2)$$

where, P_{Aq} represents the power demand of the air handling unit (AHU) to control the indoor humidity level through the removal/addition of water particles to the indoor air; P_{Tc} is the power consumed to control the indoor temperature through the removal/addition of heat to the indoor environment; and P_j represents the power consumed by the zone's devices in the Z zones, e.g. lights, computers, etc.

In this work, the AUH and the heater are the active comfort systems in charge of providing flexibility. Thus, the flexibility offer of a building (F_s) can be expressed as follows:

$$F_s = \frac{P_C^f}{P_C} \quad (3)$$

$$P_C = P_{Aq} + P_{Tc} \quad (4)$$

where P_C is the nominal power demanded by the comfort systems in building; and P_C^f is the amount of power that can be shifted, curtailed, or increased by the comfort systems in building.

B. Problem formulation

In this article, the problem is formulated as a single optimization problem that originates from the need to aggregate demand flexibility. To do so, the concept of an aggregator is introduced as the entity responsible for invoking and collecting sufficient demand flexibility from a portfolio of buildings to satisfy a given flexibility request, and at the lowest cost for the end user, e.g., comfort loss. This is equivalent to the minimization of the difference between the sum of all flexibility offers (F_s), i.e., aggregated flexibility, and a flexibility request (F_r), with a hard comfort constraint for all the buildings. This problem is solved at every time interval t for a given F_r , as follows:

$$\begin{aligned} & \text{Minimize} \quad \sum_{t=1}^n \left(F_r - \sum_{i=1}^B (F_{s,i} P_{C,i}) \right) \\ & \text{subject to} \quad C_{t,i} \geq C_{t,i}^{min}, \forall i \in \mathcal{B}, \forall t \in [1, \dots, n], \\ & \quad \quad \quad F_{s,i} \geq 0. \end{aligned} \quad (5)$$

where, $\mathcal{B} = \{1, \dots, B\}$ is the set of buildings offering flexibility; F_r is defined as the total amount of power to be shifted, curtailed, or increased at a given moment of time; $F_{s,i}$ is the flexibility offer of building i ; C_i is the comfort pertaining to the i th building as in (1); and $C_{i,min}$ is the minimum allowable comfort for the i th building.

Despite the fact that this research focuses on the problem described by eq. (5) and solved for $F_{s,i}$ to offer grid support services, it relies on the ability of each building to optimize its comfort and energy use. This is a multi-objective problem, with two conflicting objective functions, i.e., comfort maximization and energy use minimization. For the purposes of

this paper, this problem is solved using a weighted aggregation approached combined with particle swarm optimization. However, the details of the problem formulation and its solution for this problem are not presented in this article as it is beyond the our scope, but can be found in [29]. The aim of this research is to define the flexibility offer of each building at a given moment of time and given the specific comfort situation of the building. Furthermore, this paper explores the technical uses of flexibility which occurs in the operation time scale, i.e., near real-time, and within a local distribution network area. In this case, the flexibility is used to facilitate local grid operations, through appropriate decision making and the use of distributed and decentralized intelligence. More details on how this problem is solved are given in the next sections.

III. PRELIMINARIES

The demand flexibility aggregation problem (see eq.(5)) is highly complex, not only because its complexity increases with the number of buildings, but also because it is potentially conflicting to optimization task of each building. To cope with this, we harness the advantages of MAS, which provide a natural framework for solving complex and conflicting tasks relying on self-interested or cooperative behaviors.

A. Agent based control

In power systems, MAS have been applied to a wide range of applications, e.g., condition monitoring, system restoration, market simulation, network control and automation, and demand monitoring and control [30], [31]. Moreover, MAS have been also widely studied in the area of building automation, building energy management, and building control and operation [32], [33]. In general, a MAS is formed by two or more intelligent agents, that divide the large control task into smaller sub-tasks. However, the distributed nature of MAS ensures system openness to be able to cope with physical and functional changes, i.e., uncertainty, by offering flexibility from the local level, through the agents capability to tackle complex problems, based on cooperation, coordination and negotiation. Therefore, agent coordination and decision making becomes paramount in the MAS design process. In turn, an adequate agent decision strategy can be determined based on three main aspects: the number of agents (players) in the MAS, the required information, and the goal of the agents.

B. Decision making strategies

In contrast to centralized solutions with available information about the whole environment, in a MAS knowledge and operational responsibility are compartmentalized in two or more agents embedded in the environment. These agents are able to perceive their environments, and to autonomously react to changes in it. As a consequence, the performance of the MAS is defined by the set of actions taken by agents in response to changes in the agent's environment, and information exchanged with other agents.

Additionally, if the agents in a MAS are not allowed to share their observations and beliefs about the environment, i.e.,

non-cooperative process, the objective is to solve the general problem through the individual actions of the agents as a result of local information. On the other hand, if communication is granted and agents are able to share their observations and beliefs over the environment, i.e., cooperative process, the objective is to find the set of optimal joint actions that maximizes the reward function of all the agents. In the particular context described in this work, the optimization problem has to be a function of both the flexibility offer and user comfort. To do so, we explore the performance of non-cooperative agents in both centralized and decentralized settings. These results are compared to an cooperative approach in a self-organizing setting, which is therefore decentralized.

1) *Centralized non-cooperative, the n-player game:* In general, the control problem for each agent can be formalized using Markov Decision Processes (MDPs).

Definition 1. A Markov Decision Processes is defined by a 4-tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$, where $\mathcal{S}$ is a set of states, $\forall s \in \mathcal{S}$; $\mathcal{A}$ is a set of actions, $\forall a \in \mathcal{A}$; $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the transition function given by the probability that by choosing action a in state s at time t the system will arrive at state s' at time $t + 1$, such that $p_a(s, s') = p(s_{t+1} = s' | s_t = s, a_t = a)$; and $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ is the reward function, where $\mathcal{R}_a(s, s')$ is the immediate reward received by the agent after it performs the transition to state s' from state s .

Under the Markov property, MDPs assume that the state transitions are only dependent on the previous state of the system, and are conditionally independent of any previous environment states or actions of other agents on their own environment. Therein, a rational agent i acts in order to maximize its own utility, or reward function R_i , such as $V(\pi) = \sum_{t=0}^{\infty} \gamma^t \mathbb{E}(r_t | \pi, s_0 = s)$, where γ is a discount factor. This means to find the optimal policy, π^* , from the set of available policies π given by the actions to the agent in response to an observed environmental state s . Then the agent's behavior is simply a mapping function of environment states to actions. The standard solution is following the Bellman equation in the case of fully observable environment. In our case, e.g., partial observable environment, a model-free reinforcement learning solution is imposed by the information available in the system and detailed further.

IV. EXTENDED JOINT ACTION LEARNING

In this section we present a decentralized and non-cooperative strategy based on the standard Q-learning method. Moreover, we extend this method and propose a decentralized and cooperative reinforcement learning method, dubbed extended joint action learning (eJAL).

A. Q-learning

The Q-learning method introduced by Watkins [34], is one of the most popular solution for MAS, where the action rules are often stochastic. This algorithm is based on a function which calculates the quality of a state-action combination defined by $Q : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$. Before the learning phase, the Q matrix takes an initial value. Then, each time the agent selects

an action, it observes a reward, $r(s,a)$, and a new state, s' , that may depend on the previous state and the selected action, $p(s'|s,a)$. The action-value function of a fixed policy π with the value function $V^\pi : \mathcal{S} \rightarrow \mathbb{R}$ is $Q^\pi(s,a), \forall s \in \mathcal{S}, a \in \mathcal{A}$, such that:

$$Q^\pi(s,a) = r(s,a) + \gamma \sum_{s'} p(s'|s,a) V^\pi(s') \quad (6)$$

The value of state-action pairs, $Q^\pi(s,a)$, represents the expected outcome when one agent is starting from s , executing a , and then following the policy π afterwards, such that $V^\pi(x) = Q^\pi(x, \pi(x))$. Thus, the optimal value is obtained for $\forall s \in \mathcal{S}$, such that $V^*(\pi(s)) = \max_a Q^*(s,a)$ and $\pi^*(s) = \arg \max_a Q^*(s,a)$.

The value of state-action pairs is given by the same formal expectation value, $\mathbb{E}_\pi$, of the total return r_t , such that $Q(s,a) = \mathbb{E}_\pi(r_t | s_t = s, a_t = a)$. The Q-learning algorithm has the update rule defined by:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t [r_{t+1} + \gamma \max_a Q_t(s_{t+1}, a) - Q_t(s_t, a_t)] \quad (7)$$

where the discount factor $\gamma \in [0, 1]$ trades off the importance of rewards, while the learning rate $\alpha \in (0, 1]$ determines the rate at which new information overrides the old one.

In the non-cooperative setting, the problem described in II-B is tackled through the actions taken by the individual agents, without information on the actions of others. Thus, the agent's reward r_{t+1} observed after performing a_t in s_t is inversely proportional to the difference between the total flexibility request and the amount of flexibility offered by agent i , and it's given by:

$$r_{t+1}(s, a_i) = \frac{1}{|F_r - F_{s,i}|} \cdot c_{i,t} \quad (8)$$

where the $c_{i,t}$ is a binary value for agent i at time t , such that it takes a value of zero if comfort limits of the building i are violated, and a value of one otherwise. So far we considered a relaxed version of the multi-agent learning problem where the optimal policies of the other players is considered *stationary*. Furthermore, we did not consider multi-agent cooperation, i.e., the agents are choosing their actions greedily by maximizing their individual reward.

B. Extended joint action learning (eJAL)

Q-learning is typically used for equilibria in case of one agent, and it is unfeasible for cooperative multi-agent settings, where the joint action spaces will grow exponentially with the number of agents. Consequently, if the design is generalized properly, a special case of partially observable Markov decision process (POMDP) can be used. The joint action learning (JAL) method [35] extends the Q-learning method, from one agent to multi-agents, by including in the optimal policies the actions of the other (competing) agents. Successfully applied in many stochastic games with two or three agents, e.g. to design a flexible market [36], JAL requires further two slight modifications in order to make it suitable for our demand flexibility game. We call this extended version eJAL.

Algorithm 1 Extended Joint Action Learning (eJAL)

- 1: initialize model parameters ($\alpha \in (0, 1]$ and $\gamma \in [0, 1]$)
 - 2: initialize $Q(s, a_i)$ random
 - 3: $\zeta(s, a_{others}) \leftarrow 0, \forall s \in \mathcal{S}$ and $\forall a_{others} \in \mathcal{A}_{others}$
 - 4: $\zeta(s) \leftarrow 0, \forall s \in \mathcal{S}$
 - 5: **for** iteration = 1 to arbitrary number **do**
 - 6: observe and collect state s , $\zeta(s) \leftarrow \zeta(s) + 1$
 - 7: execute actions a_i by solving
 - 8: $\arg \max_{a_i} \sum_{a_{others}} \frac{\zeta(s, a_{others})}{\zeta(s)} Q(s, \langle a_i, a_{others} \rangle)$
 - 9: observe and collect the others agents action a_{others}
 - 10: observe and collect the reward $r(s, a_i)$
 - 11: observe and collect the next state s_{t+1}
 - 12: **if** episode is finished **then**
 - 13: $Q \leftarrow (1 - \alpha)Q + \alpha(r + \gamma V(s_{t+1}))$
 - 14: $\zeta(s, a_{others}) \leftarrow \zeta(s, a_{others}) + 1$
 - 15: **Until** Q converge, such that $|Q_{t+1} - Q_t| < \epsilon$
-

Firstly, due to MAS scalability reasons, the conditional action space is defined using a tuple over all the others building, such as $\langle a_i, a_{others} \rangle$. This allows to improve coordination among building agents and maintain an explicit model of the *other* agents for each state. The Q-values are updated for all possible joint actions at a given state, under the assumption that the other buildings are stationary. Thus, the probability of a joint action is calculated for the *other* agents by counting frequencies of the joint actions they executed in the past, as follows:

$$p(a_{others}) = \frac{\zeta(a_{others})}{\sum_{a_{others} \in \mathcal{A}_{others}} \zeta(a_{others})} \quad (9)$$

where $\zeta(a_{others})$ is the number of times the *others* agents has played action a_{others} . Secondly, a preferential attachment process is included for every agent in a multi-agent full cooperative game in order to use a joint reward function. The major difference between Q-learning and eJAL is the joint reward function $r_i(a_i, s_i | a_{others})$ obtained by agent i as it executes action a_t and as the other agent executes action a_{others} at each time step t . Thus, the reward observed by agent i is inversely proportional to the difference between the total flexibility request and total amount of flexibility offered by all the agents. This is further defined by:

$$r_{t+1}(a_i, s_i | a_{others}) = \frac{1}{|F_r - (F_{s,i} + F_{s,others})|} \cdot c_{i,t} \quad (10)$$

where the $c_{i,t}$ is a binary value for agent i at time t , such that it takes a value of zero if comfort limits of the building i are violated, and a value of one otherwise. Consequently, the updated equation expressed in Eq. 7 becomes for eJAL:

$$Q(s, \langle a_i, a_{others} \rangle) = (1 - \alpha)Q(s, \langle a_i, a_{others} \rangle) + \alpha[r(a_i, s_i | a_{others}) + \gamma V(s_{t+1})] \quad (11)$$

where the corresponding value of the game is given by:

$$V(s_{t+1}) = \max_{a_i} \frac{\zeta(s_{t+1}, a_{others})}{\zeta(s_{t+1})} Q(s_{t+1}, \langle a_i, a_{others} \rangle). \quad (12)$$

Fig. 1. System model, showing the schematic of the network, and the agents associated.

Algorithm 1 exemplifies on eJAL how it can be implemented using Eq.10, Eq.11, and Eq.12.

V. CASE STUDY

This section describes the simulation setup implemented to demonstrate the performance of the proposed solutions. The competence of each agent decision making strategy is assessed on basis on transformer overloading time, flexibility request and offer match, and comfort loss. The case study consists of two parts, a Matlab[®]/Simulink distribution grid serving as the physical environment, while the agents are developed in the middleware environment of Java/JADE.

A. The environment

The general schematic of the proposed system is shown in Fig. 1, while the cable data and building information can be found in Table I and II respectively. The figure represents a three-phase balanced system formed by seven non-residential commercial loads in a single MV sub-ring, and an aggregation of 27 residential loads into a single LV load. There are two voltage transforming steps, a HV-MV (150kV-10kV) step in a distribution substation, and a MV-LV (10kV-0.4kV) in a distribution transformer.

The LV aggregated load is modeled using a static hourly profile obtained from measured data of 27 typical Dutch LV loads connected to the same feeder [37]. The commercial loads are modeled as single comfort zones. The power demand of these loads consist of two parts:

- Flexible power demand:** result of the operation of the comfort system in response to variation in the outdoor conditions and the occupancy levels. The 2013 Dutch weather data from the KNMI (Koninklijk Nederlands Meteorologisch Instituut) is used for the outdoor temperature and relative humidity with hourly resolution.
- Base load consumption:** result of the operation of the non-comfort systems, which follow a predefined operation schedule. Occupancy, lights, device usage, and thermostat profiles with hourly resolution are defined for each industrial load based on the large office building

TABLE I
CABLE INFORMATION.

Sections	Length[m]	R[Ω/km]	L[mH/km]	Description
1	1140	0.07	0.246	MV bus cable
6	287 [avg]	0.532	0.311	MV-subring cable

TABLE II
BUILDING CHARACTERISTIC INFORMATION.

Building	Floor Area [m^2]	No. Floors	Floor Height [m]
1	554	3	3
2	1005	12	3
3	4535	9	3
4	5233	6	3
5	9034	5	3
6	554	3	3

schedules of the NREL (National Renewable Energy Laboratory) platform OpenStudio[®] for EnergyPlus.

Finally, construction data from the DOE 2004 standard for the Netherlands is used for all the buildings.

B. The MAS platform

In this paper, the objective of the agent platform is the coordination of the operation of the non-residential systems to act timely and appropriately to a flexibility request, based on agent reactive behaviors. To do so, the MAS platform is formed by a total of 8 agents, divided in three main types of agents [29]:

- The distribution agent:** It is in charge of monitoring the distribution network operation and responsible for establishing the flexibility request, F_r .
- The aggregator agent:** The role of the aggregator is to procure flexibility from the flexible demand to meet the flexibility demand in an economic and efficient way. It collects information from the portfolio of buildings, and the distribution agent.
- The building energy management system (BEMS) agent:** It is responsible for the building operation² and determining the flexibility potential of the building, at every moment in time, based on the comfort state. The BEMS agent is a high level representation of several agents organized in a hierarchy to represent the operational structure of a building. Nonetheless, the BESM structure is simplified to a single agent for the purpose of this paper. This agent is able to accept and prioritize requests made by the aggregator agent.

The set-up presented in the paper is a single hierarchical structure with two agents on the same level, i.e., the distribution agent and the aggregator, and the BEMS agents under the aggregator agent. The distribution agent senses the state of its environment $P[W] \rightarrow \mathcal{S}$ and decides which action to take, i.e., how much demand flexibility to request $\mathcal{A} \rightarrow F_r$. Whereas the BEMS agent i perceives the building state, defined by an inner product space between building comfort $[0,1]$ and building

² A PSO is used for the optimization of the operation of each building, and to find the balance between comfort and energy. More details can be found in [29].

Fig. 2. Learning rate and discount factor parameter selection, showing the error reduction rate for the three comfort states in the Q-learning algorithm.

flexibility $F_{s,i}$. For each building agent the state $s_i \in \mathcal{S}$ is based on the comfort value at time t , and the flexibility request F_r . More specifically, the comfort state of the building is used to define an initial four group of states:

No comfort (NC)	$0.0 \leq \mathcal{C} < 0.6$
Low comfort (LC)	$0.6 \leq \mathcal{C} < 0.75$
Medium comfort (MC)	$0.75 \leq \mathcal{C} < 0.85$
High comfort (HC)	$0.85 \leq \mathcal{C} \leq 1.0$

The ratio between the flexibility request and the power demand of the building is used to define 12 states for each comfort group. In total 48 states are defined for each building i : $\{s_1, \dots, s_{48}\} \in \mathcal{S}$. Each BEMS agent decides how much flexibility can be offered, by choosing one of the following actions $\{F_{s,i}^0, F_{s,i}^5, F_{s,i}^{10}, F_{s,i}^{15}, F_{s,i}^{20}, F_{s,i}^{30}\}$. The superscript of each action indicates a percentage of the building's power demand at time t that is offered as flexibility. In turn, every action leads to a different reward.

We implemented the n-player game, Q-learning, and eJAL algorithms in the Java/JADE environment using the mathematical details described in sections IV-A and IV. Fig 2 shows the exploration of different learning rates, α , and discount factors, γ . After a series of simulations, according to a *trial and error* methodology it was found that the best combination for this specific application was $\alpha = 0.49$ and $\gamma = 0.6$ for both cases. Both parameters have a direct influence on the convergence performance of the algorithms. The figure shows that a discount factor of 0.5 leads to a faster convergence of the learning process. However, this is at the expense of reducing the influence of the future rewards in the learning process, which lead to a Q-matrix that reflected mostly the instantaneous rewards set for each agent.

Fig. 3. Flexibility offer and request, showing the stacked flexibility offers from the six buildings and the flexibility request in time. The x-axis does not represent time, as the time at which a request is received by an BEMS agent changes according to the strategy, due to the differences in the flexibility offer for the previous time.

TABLE III
AGGREGATED FLEXIBILITY RESPONSES AND REQUEST RESULTS

	demand flexibility offer [%]						$\sum F_{s,i}$ MW	F_r MW	η [p.u.]
	B_1	B_2	B_3	B_4	B_5	B_6			
n-player	0	0.8	5.6	35.7	57.8	0	6.3	7.4	0.84
Q-learning	8.9	19.3	25.2	48.7	27.3	11.4	8.9	4.5	1.95
eJAL	11.8	24.1	35.5	39	24.8	13.0	9.4	6.29	1.48

VI. RESULTS

As previously mentioned, three decision making strategies are evaluated and compared on basis of transformer overloading and the flexibility offer. The experiments aim to analyze comprehensively the effect of cooperation in decentralized decision making strategies, and compare the proposed decentralized method to a centralized solution. Consequently, the two non-cooperative methods, i.e., Nash-NPG and Q-learning, are compared against the proposed decentralized cooperative method, i.e., eJAL. The learning process for the decentralized strategies was carried out in a period of 100 days, while the actual result correspond to a simulation for five days. Furthermore, one scenario without flexible behavior is defined as a benchmark. This scenario exhibits an overload time of 21 *hrs* over the five days. Finally, an analyze is performed between Q-learning and eJAL for 365 days.

A. Request-offer equilibrium

The relationship between flexibility offer and request can be seen more clearly in Fig. 3 for the first 20 request. Whereas, the demand flexibility offer of all the buildings, and the flexibility request aggregated over the five days are depicted in Table III.

Table III also shows the ratio, η , between the total offer ($\sum F_{s,i}$), and the flexibility request (F_r). As the table shows,

Fig. 4. Response match density function, showing the probability density function of the three strategies.

the centralized strategy offers the best match of all the strategies investigated here, as it was expected. Additionally, Fig. 4 also shows that this strategy has the smallest variance of the solutions proposed, making it the best among the three methods presented here³. However, as mentioned the better performance of the centralized strategy comes at some cost. In the decentralized setting, decision making strategies rely less on communication and do not require centralized data management. In Fig. 3 and Fig. 4 it is shown that cooperative behavior, i.e., eJAL, leads to a decrease of the flexibility over-offer, decreasing the median by 22.1% to a value of 1.69. Nonetheless, Table III shows that Q-learning reduces significantly the number of request made by the aggregator to the buildings, while eJAL exhibits 54% more request. This can be attributed to a consequence of cooperation. When the others take actions that seem to be enough, a building might choose to ignore a request, resulting in insufficient offers. This can be seen in Fig. 3.

B. Fairness of the game

In our case, all the players differ in size and energy demand, thus it is difficult to find an adequate measure of fairness. In this articles, we used the Jain's fairness index [38] defined as follows:

$$\mathcal{J}(F_{s,i}) = \frac{\left(\sum_{i=1}^B \frac{\omega_{f,i}}{\omega_{p,i}}\right)^2}{B \sum_{i=1}^B \left(\frac{\omega_{f,i}}{\omega_{p,i}}\right)^2} \quad (13)$$

where the Jain's fairness index is calculated by considering the *responsibility* and the *commitment* of allocation values. The responsibility of a building is defined as $\omega_{p,i} = \frac{P_i}{\sum_{i=1}^B P_i}$, where P_i is the total power consumed by building i . The commitment is defined as the ratio between the flexibility offer $F_{s,i}$ and the amount of flexibility requested by the aggregator, F_r , such that $\omega_{f,i} = \frac{F_{s,i}}{F_r}$.

Fig. 5 shows the distribution of the fairness index for the different decision strategies. From this figure, it is noticeable that eJAL is the most fair strategy with a mean index of 0.84, followed by Q-learning with a mean index of 0.78. In contrast,

³The reason for the flexibility mismatch for this strategy lies in the discrete nature of the flexibility action.

Fig. 5. Fairness index distribution per strategy, showing that the decentralized strategies result in a more fair distribution of the flexibility offer than the centralized strategy.

TABLE IV
COMFORT LOSSES AVERAGED OVER A DAY FOR ALL SIX BUILDINGS, USING DIFFERENT DECISION STRATEGIES.

	Comfort loss [%]						mean [%]
	B_1	B_2	B_3	B_4	B_5	B_6	
n-player	0.0	1.05	0.82	1.12	1.33	0.0	0.722
Q-learning	0.34	0.8	1.4	1.5	1.0	0.72	0.96
eJAL	0.3	0.9	1.06	1.02	0.78	0.50	0.76

Fig. 6. Building comfort profile for a single day, showing the comfort loss (shaded area) in each building under Q-learning and eJAL.

the centralized strategy proves to be the least fair with a mean index of 0.22.

C. Comfort loss

The averaged comfort losses corresponding to the demand flexibility offers are presented in Table IV. Furthermore, the comfort profile of every building for a day, and including the losses are shown in Fig. 6. As it is observable, these losses are quite small. While this is the preferred outcome, the main reason behind this can be explained by a short rebound effect in the time step that follows a request. In terms of averaged comfort loss, the Q-learning solution seems to be the cause for the largest amount of comfort losses. This can be attributed to this strategy offering more flexibility than the others.

TABLE V
SUMMARY OF THE EXPERIMENTS WITH RESPECT TO DIFFERENT STRATEGIES AS WELL DIFFERENT INDICATORS, FOR ALL SIX BUILDINGS

Method	Type of strategies	Level of agents cooperation	Time period	Flexibility offer	Comfort loss	Overload duration	Fairnesses of the game
n-player	Centralized	Non-cooperative	day	III	III	III	III
Q-learning	Decentralized	Non-cooperative	day	II	I	I	II
			year	I	II	I	II
eJAL	Decentralized	Cooperative	day	I	II	II	I
			year	II	I	II	I

Fig. 7. Transformer loading curve, showing the result for the three decision making strategies and the defined limit.

Fig. 8. Response match density function, showing the probability density function of the Q-learning and eJAL strategies over one year.

D. Overload duration

Fig. 7 shows the overload duration results for all the strategy, including the benchmark scenario. Ideally, a perfect strategy should have a fast decrease for the values above the limit, but once this value has been crossed it should return the original curve. The figure shows that this is close to what the N-player strategy does. However, it only achieves a overload reduction of 5.2%, so the total overload time of 19.9 hrs. In contrast, Q-learning achieves a reduction of 22.6%, or 4.74 hrs. Nonetheless, it is also noticeable that this strategy results in an over offer, which bring other issues, like rebound energy effects, and higher loss of comfort. Finally, the eJAL strategy reduces the overload time with 16.3%, or 3.42 hrs, while having a slightly lower over offer.

For all the methods used the computational time required is in the order of 1-3 minutes for a simulation interval. However, this is mainly due to the complexity of the network and building models used. Nonetheless, the eJAL method involves more communication between agents than Q-learning, and could bring important benefits if the network dynamics, e.g. physical constrains or price benefits, impose a multi-agent coalition.

E. Long term effects

The long-term learning characteristics of Q-learning and eJAL methods are investigated over a period of 365 days. The experiments aim to demonstrate the ability of these methods to cope with a changing environment (e.g. Winter - Summer). In general, when considering an electricity based comfort systems, i.e., electricity based cooling and heating, the total power demand of the system (set of building) will depend greatly on the temperature variations. For instance, hot and

Fig. 9. Comfort loss and flexibility action response distribution per building, showing the responses for a year for each decision strategy.

humid weathers, like the one in Texas, have a demand peak during the summer days, while cold-humid weathers, like in the Netherlands, will have a peak during the winter days. The results shows, that for the Dutch weather case, the long-term request-offer equilibrium eJAL is slightly better than Q-learning, as shown in Fig.8.

Finally, Fig.9 shows the comfort loss and flexibility response per building and according to strategy. It is shown that both strategies result in acceptable comfort losses ($\Delta C \leq 0.4$), while Q-learning result in larger flexibility offers per building, with the exception of buildings 5 and 6.

All the experiments performed in this study are summarized in Table V. Therein the level of performance acquired at the aggregated level is indicated by (I) for the first performer, (II) for the second performer, and (III) for the third performer. Overall, eJAL is the most stable strategy in long-term.

VII. CONCLUSIONS AND FUTURE WORK

This paper aims to enable a cooperative behavior for buildings to become active in the operation of the power system. Based on a MAS platform, building agents can submit flexibility offers, in order to satisfy as much as possible the grid operator's flexibility request, while taking into account the buildings' objectives and constraints, with a minimum communication burden. By introducing a joint reward function in joint action learning (JAL), we propose a novel type of cooperative and at the same time decentralized game, dubbed extended JAL (eJAL). In addition, a comparison of eJAL with a decentralized non-cooperative method (i.e. Q-learning) and with a centralized one (i.e. n-player game) is performed. It is shown that a range of flexibility requests can be met by a portfolio of buildings without significant comfort losses. This is possible for both centralized and decentralized decision-making. However, this depends greatly on the choice and definition of the reward function, which determines the flexibility offer given the local parameters. The analysis performed over 100 days showed that Q-learning is a good non-cooperative method. However, it results in the highest over-offer of flexibility, despite having the most pronounced decrease of the overload duration within the expected comfort constraints. In the end, we highlight, that eJAL achieves the highest average fairness index, which is crucial in a fully decentralized setup.

Future work: The two proposed decentralized approaches do not require the aggregating agent to have complete information about the participating buildings, which is a clear advantage from a scalability perspective. In respect with this, we expect that given the physical network constraints reflected in a limited possible number of building agents the proposed methods will show good scalability capabilities. However, in order to generalize these results, further investigation could be performed with focus on the characterization of the scalability bounds by performing simulations with a higher number of agents.

Moreover, the following topics are identified as future work:

- The implications of the extra communication burden brought by the cooperative decision making strategies in a decentralized setting.
- The role of proactive behaviors and forecasting in the provision of flexibility, in the synchronization between request and offer to reduce the overloading even further, and the number of unmet request hours.

ACKNOWLEDGMENT

This research has been performed within the project on Energy Efficient Buildings as part of the Smart Energy Regions-Brabant program subsidized by the Province of Noord Brabant, the Netherlands, and by the NL Enterprise Agency under the TKI SG-BEMS project of Dutch Top Sector.

REFERENCES

- [1] H. Quan, D. Srinivasan, A. M. Khambadkone, and A. Khosravi, "A computational framework for uncertainty integration in stochastic unit commitment with intermittent renewable energy sources," *Applied Energy*, vol. 152, pp. 71–82, 2015.
- [2] T. Zhang, R. Baldick, and T. Deetjen, "Optimized generation capacity expansion using a further improved screening curve method," *Electric Power Systems Research*, vol. 124, pp. 47–54, 2015.
- [3] H. Kondziella and T. Bruckner, "Flexibility requirements of renewable energy based electricity systems - a review of research results and methodologies," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 10–22, 2016.
- [4] M. I. Alizadeh, M. Parsa Moghaddam, N. Amjadi, P. Siano, and M. K. Sheikh-El-Eslami, "Flexibility in future power systems with high renewable penetration: A review," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 1186–1193, 2016.
- [5] N. G. Paterakis, A. Tascikaraoglu, O. Erdinc, A. G. Bakirtzis, and J. P. S. Catalao, "Assessment of Demand-Response-Driven Load Pattern Elasticity Using a Combined Approach for Smart Households," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1529–1539, 2016.
- [6] P. Faria, J. Spinola, and Z. Vale, "Aggregation and Remuneration of Electricity Consumers and Producers for the Definition of Demand-Response Programs," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 952–961, 2016.
- [7] O. Ma, N. Alkadi, P. Cappers, P. Denholm, J. Dudley, S. Goli, M. Hummon, S. Kiliccote, J. MacDonald, N. Matson, D. Olsen, C. Rose, M. D. Sohn, M. Starke, B. Kirby, and M. O'Malley, "Demand Response for Ancillary Services," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1988–1995, 2013.
- [8] Cenelec, "SG-CG / M490 / L Flexibility Management Overview of the main concepts of flexibility management," CENELEC-Smart Grid Coordination Group, Tech. Rep., 2014.
- [9] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, "A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 1–1, 2015.
- [10] F. Xu and L. L. Lai, "Novel Active Time Based Demand Response for Industrial Consumers in Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 3203, no. c, pp. 1–1, 2015.
- [11] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419–4426, dec 2008.
- [12] X. Xue, S. Wang, C. Yan, and B. Cui, "A fast chiller power demand response control strategy for buildings connected to smart grid," *Applied Energy*, vol. 137, pp. 77–87, 2015.
- [13] G. Reynders, T. Nuytten, and D. Saelens, "Potential of structural thermal mass for demand-side management in dwellings," *Building and Environment*, vol. 64, pp. 187–199, 2013.
- [14] K. Park, Y. Kim, S. Kim, K. Kim, W. Lee, and H. Park, "Building Energy Management System based on Smart Grid," in *2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC)*. IEEE, oct 2011, pp. 1–4.
- [15] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, 2010.
- [16] H. S. V. S. K. Nunna, S. Member, and S. Doolla, "Demand Response in Smart Distribution System With Multiple Microgrids," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1641–1649, 2012.
- [17] M. B. Kjrgaard, K. Arendt, A. Clausen, A. Johansen, M. Jradi, B. N. Jrgensen, P. Nellen, F. C. Sangogboye, C. Veje, and M. G. Wollens, "Demand response in commercial buildings with an assessable impact on occupant comfort," in *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2016, pp. 447–452.
- [18] S. Althaher, P. Mancarella, and J. Mutale, "Automated demand response from home energy management system under dynamic pricing and power and comfort constraints," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1874–1883, July 2015.
- [19] L. A. Hurtado, E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Comfort-constrained Demand Flexibility Management for Building Aggregations using a Decentralized Approach," in *4th International Conference on Smart Cities and Green ICT systems SMARTGREENS*, 2015, pp. 157–166.
- [20] P. Morales-Valdés, A. Flores-Tlacuahuac, and V. M. Zavala, "Analyzing the effects of comfort relaxation on energy demand flexibility of buildings: A multiobjective optimization approach," *Energy and Buildings*, vol. 85, pp. 416–426, dec 2014.
- [21] D. Christantoni, D. Flynn, and D. P. Finn, "Modelling of a Multi-purpose Commercial Building for Demand Response Analysis," *Energy Procedia*, vol. 78, pp. 2166–2171, 2015.
- [22] R. Yin, E. C. Kara, Y. Li, N. DeForest, K. Wang, T. Yong, and M. Stadler, "Quantifying flexibility of commercial and residential loads for demand response using setpoint changes," *Applied Energy*, vol. 177, pp. 149–164, 2016.

- [23] E. Nyholm, S. Puranik, É. Mata, M. Odenberger, and F. Johnsson, "Demand response potential of electrical space heating in Swedish single-family dwellings," *Building and Environment*, vol. 96, pp. 270–282, 2016.
- [24] R. Sun and G. Zhao, "An Efficient Multi-Agent Q-learning Method Based on Observing the Adversary Agent State Change," in *IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, 2006, pp. 4169–4174.
- [25] L. Busoniu, R. Babuska, B. De Schutter, and B. D. Schutter, "A comprehensive survey of multiagent reinforcement learning," *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 2, pp. 156–172, 2008.
- [26] L. Kraemer and B. Banerjee, "Multi-agent reinforcement learning as a rehearsal for decentralized planning," *Neurocomputing*, vol. 190, pp. 82–94, 2016.
- [27] S. Zarrabian, R. Belkacemi, and A. A. Babalola, "Reinforcement learning approach for congestion management and cascading failure prevention with experimental application," *Electric Power Systems Research*, vol. 141, pp. 179–190, 2016.
- [28] P. Zhao, S. Suryanarayanan, and M. G. Simões, "An Energy Management System for Building Structures Using a Multi-Agent Decision-Making Control Methodology," *IEEE Transactions on Industry Applications*, vol. 49, no. 1, pp. 322–330, 2013.
- [29] L. A. Hurtado, P. H. Nguyen, and W. L. Kling, "Smart grid and smart building inter-operation using agent-based particle swarm optimization," *Sustainable Energy, Grids and Networks*, vol. 2, pp. 32–40, 2015.
- [30] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziaargyriou, F. Ponci, and T. Funabashi, "Multi-Agent Systems for Power Engineering Applications - Part I : Concepts , Approaches , and Technical Challenges," *IEEE Transactions on Power systems*, vol. 22, no. 4, pp. 1743–1752, 2007.
- [31] P. H. Nguyen, W. L. Kling, and P. F. Ribeiro, "A Game Theory Strategy to Integrate Distributed Agent-Based Functions in Smart Grids," *IEEE Transactions on Smart Grids*, vol. 4, no. 1, pp. 568–576, 2013.
- [32] A. Dounis and C. Caraiscos, "Advanced control systems engineering for energy and comfort management in a building environment-A review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6-7, pp. 1246–1261, aug 2009.
- [33] R. Yang and L. Wang, "Development of Multi-agent System for Building Energy and Comfort Management Based on Occupant Behaviors," *Energy and Buildings*, vol. 56, pp. 1–7, oct 2013.
- [34] C. J. C. H. Watkins and P. Dayan, "Technical note: q-learning," *Mach. Learn.*, vol. 8, no. 3-4, pp. 279–292, May 1992.
- [35] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," in *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, ser. AAAI '98/IAAI '98. Menlo Park, CA, USA: American Association for Artificial Intelligence, 1998, pp. 746–752.
- [36] N.-P. Yu, C.-C. Liu, and J. Price, "Evaluation of market rules using a multi-agent system method," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 470–479, 2010.
- [37] L. A. Hurtado, P. H. Nguyen, and W. L. Kling, "Multiple Objective Particle Swarm Optimization Approach to Enable Smart Buildings-Smart Grids," in *18th Power Systems Computation Conference, PSCC*, 2014, pp. 1–8.
- [38] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *CoRR*, vol. cs.NI/9809099, 1998.

comfort. He received the MSc. degree in Sustainable Energy Technology at the same university in 2012. He received his BSc. In Electrical and Electronics Engineering in 2009 from the Universidad de los Andes, Colombia.

University of Technology, The Netherlands.

researcher. He has been a visiting researcher with the Real-Time Power and Intelligent Systems (RTPIS) Laboratory, Clemson University, US in 2012 and 2013. His research interests include distributed state estimation, control and operation of the power system, multi-agent system, and their applications in the future power delivery system.

sor with the Electrical Energy Systems Department, Eindhoven University of Technology, the Netherlands. Her research interests are in the area of smart and sustainable power systems.

mand matching of energy and for implementation in energy and electricity markets. His research interests include ICT architecture, design, and development for SmartGrids living labs.

Luis A. Hurtado (1987) PhD, from Colombia, is a consultant in renewable energy integration at DNV GL-Netherlands, and a former PhD student at the Electrical Energy Systems (EES) research group of the Electrical Engineering department at the Eindhoven University of Technology. During his PhD, his research focused on the role of demand flexibility in the emerging power system. The project aims to develop an Building Energy Management Systems under a Multi-Agent System framework to optimize energy consumption as well as user

Elena Mocanu received the B.Sc. degree in mathematics and physics from Transilvania University of Brasov, Romania, the M.Sc. degree in physics from University of Bucharest, Romania, in 2011 and the M.Sc. degree in operational research from Maastricht University, The Netherlands, in 2013. She has been a Assistant Lecturer within the Department of Information Technology, University of Bucharest, Romania from September 2008 to January 2011. She is currently a Ph.D. student in the Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands.

Phuong H. Nguyen (M06) was born in Hanoi, Vietnam, in 1980. He is assistant professor of Electrical Energy Systems group at the Eindhoven University of Technology, the Netherlands. He received the B.E. degree in electrical engineering at Hanoi University of Technology, Vietnam, in 2002 and the M.Eng. degree in the electrical engineering department, at the Asian Institute of Technology, Thailand, in 2004. In 2010, he received his Ph.D. at the Eindhoven University of Technology, the Netherlands, and was employed at the same group as a postdoctoral researcher. He has been a visiting researcher with the Real-Time Power and Intelligent Systems (RTPIS) Laboratory, Clemson University, US in 2012 and 2013. His research interests include distributed state estimation, control and operation of the power system, multi-agent system, and their applications in the future power delivery system.

Madeleine Gibescu received her Dipl.Eng. in Power Engineering from the University Politehnica, Bucharest, Romania in 1993 and her MSEE and Ph.D. degrees from the University of Washington, Seattle, WA, U.S. in 1995 and 2003, respectively. She has worked as a Research Engineer for Clear-Sight Systems, and as a Power Systems Engineer for the Alstom Grid, Washington, U.S. From 2007, she has worked as an Assistant Professor for the Electrical Sustainable Energy Department, Delft University of Technology. Currently she is an Associate Professor with the Electrical Energy Systems Department, Eindhoven University of Technology, the Netherlands. Her research interests are in the area of smart and sustainable power systems.

René (I. G.) Kamphuis (A00M09) was born in 1952. He received the Ph.D. degree in chemical physics from University of Groningen, Groningen, the Netherlands, in 1983. He is a Senior Research Technologist with Netherlands Organization for Applied Scientific Research (TNO), Groningen, the Netherlands, and a part-time Professor in "Smart operation of electricity systems through ICT with Eindhoven University of Technology, Eindhoven, the Netherlands. He is the Initiator and Developer of agent-based coordination techniques for supplyde-