Q-learning for Optimal Deployment Strategies of Frequency Controllers using the Aggregated Storage of PHEV fleets

Spyros Chatzivasileiadis, *Student Member, IEEE,* Matthias D. Galus, *Student Member, IEEE,* Yves Reckinger, Göran Andersson, *Fellow, IEEE*

Abstract—As more renewable energy sources (RES) get connected to the electric power network, the stability of the system is gradually put into increasing risk. RES lack stabilizing characteristics, such as inertia, which are inherent to conventional synchronous machines. Mimicking inertia techniques, by appropriately controlling an external power source such as a large battery storage, improve the stability of the system. Since large battery storage is costly, a distributed battery storage, based on Plug-In Hybrid Electric Vehicles (PHEVs) seems an attractive option. This paper investigates the use of the aggregated storage from large, distributed PHEV fleets for frequency control by inertia-mimicking techniques. The focus is on the saturation limits of the aggregated storage, as well as on the controller placement and speed. An algorithm based on Q-learning is developed to determine an optimal controller placement strategy.

Index Terms—Frequency Control, Plug-In Hybrid Electric Vehicles (PHEV), Renewable Energy Sources, Q-learning

I. Introduction

During the last one or two decades, power systems probably face the most fundamental changes in expansion, planning and operation. The increasing share of renewable energy sources (RES) in the generation portfolio introduces challenges in power system operation and planning. The majority of RES are connected through power electronics to the power grid and hence, they do not contribute substantially to the overall system inertia. As a result, larger frequency excursions can be observed [1]. Different countermeasures have been suggested in order to tackle this problem, including taking advantage of the pitch control of the wind generators or appropriately controlling an external power source.

The role of this external power source could assume fleets of plug-in hybrid electric vehicles (PHEVs), which are expected to be widely integrated into the power grid in the future. Besides smart charging strategies that avoid system overload and minimize charging costs [2], ongoing studies focus on so called vehicle-to-grid (V2G) services. V2G services are operation modes where vehicles draw power from or feed power back to the power system [3]. As battery costs are generally assumed to be high, V2G services become attractive mainly when taking advantage of the available power while avoiding high energy turnovers [4]. The latter occurs in e.g. wind balancing [5] while load frequency control mainly takes

advantage of the available power [6]. Facilitating inertia-like behavior of RES is another potential application area. The use of the aggregated PHEV battery storage as the desired external power source seems expedient, since the control performance for inertia mimicking relies mainly on the power infeed while the overall energy turnover remains very small due to the transient time scales of operation.

Herein, the use of PHEVs for such services is studied and limiting factors are assessed. These factors depend mostly on the temporal and spatial variability of the PHEVs and their available aggregated power capacity. The goal is to identify how different fault sizes and controller locations affect the frequency response. A *Q*-learning algorithm is developed in order to determine the optimal deployment strategy of the controllers for ensuring the stability of the system at all times. The *Q*-learning algorithm is capable of identifying the node that should offer the control power at each time of day and can decide at which time a change of the controller location should take place, so that the optimal frequency response can be achieved.

This paper is organized as follows. Section II describes the concept of inertia mimicking for RES and gives the principles of the Youla parametrization technique. Section III derives the PHEVs potential for mimicking inertia. Section IV illustrates how the different factors, e.g. available power from PHEVs, fault size and location, influence the frequency response. Section V focuses on the *Q*-learning principle, which allows to learn what are the optimal controller characteristics at each time step. Section VI presents a simple example and Section VII concludes.

II. Mimicking Inertial Behaviour through an Additional Control Path

This paper adopts the solution of an additional control path, in order inverter-connected RES, such as wind-generators, to be able to contribute to frequency. More specifically, through the control of an external power source, an inverter-connected RES can demonstrate an inertial-like behaviour similar to the one of a conventional synchronous machine. The control structure has been based on the work presented in [7] and is illustrated in Fig. 1. The controller is represented by the block C(s) and controls an external power source, which is modelled by the actuator block A(s). Frequency deviation $\Delta\omega$ is measured through the block M(s), which corresponds to a

S. Chatzivasileiadis, M. D. Galus, Y. Reckinger and G. Andersson are with the Department of Electrical and Computer Engineering, ETH Zurich, Switzerland e-mail: {spyros, galus, andersson}@eeh.ee.ethz.ch, yvesr@student.ethz.ch.
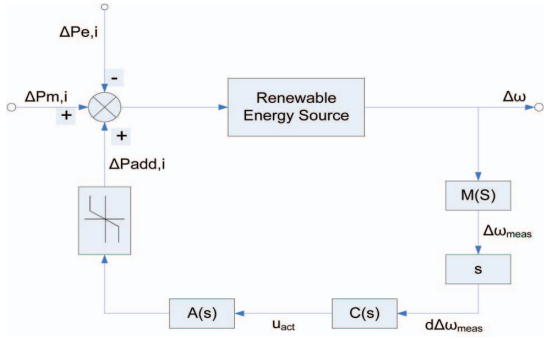
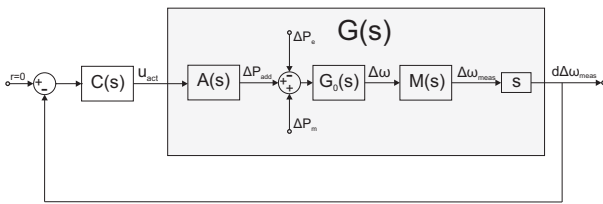Fig. 1: Overview of the implemented control structure with the additional power path concept

Fig. 2: Representation of the control path from the controller's point of view

fast first order filter. The frequency derivative $\Delta\dot{\omega}$ is given as an input to the controller. The saturation limits of the external power source are also taken into account.

An extensive study has been carried out in [7], concluding that a controller design based on Youla Parametrization [8] demonstrates improved frequency response in comparison with a PI (proportional-integral) controller. In the study, different topologies and characteristics of additional power paths were studied as well as different control structures with master-slave controllers. The present work will focus on the effect that one controller with one additional power source has on the frequency response of the system after a disturbance.

The principles of the controller design based on Youla parametrization are summarized in the following. The interested reader can find a detailed description of the approach in [7]. Fig. 2 depicts the same system as in Fig. 1, but now from the controller's point of view. The transfer function $G_0(s)$ represents the dynamic characteristics of the RES.

The objective is to find a sophisticated controller $C(s)$, so that the closed loop transfer function T_{cl} of the system has the desirable performance. First, a stable transfer function from theory is selected that fulfils the desired closed-loop behavior. This is expressed as:

$$H(s) \approx T_{cl} = \frac{G(s)C(s)}{1 + G(s)C(s)} \quad . \tag{1}$$

Based on the study carried out in [7], the selected transfer function $H(s)$ has the following form:

$$H(s) = \frac{5.14\omega_n^3 s + \omega_n^4}{s^4 + 2.41\omega_n s^3 + 4.93\omega_n^2 s^2 + 5.14\omega_n^3 s + \omega_n^4} \quad . \tag{2}$$

In the second step, the equivalent open loop transfer function is defined:
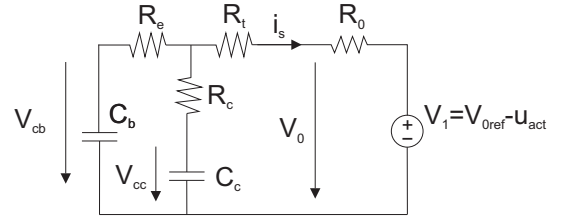
$$T_{eq}(s) = Q(s)G(s) \quad , \tag{3}$$

Fig. 3: Battery Model

where $T_{eq} = H(s)$. Hence:

$$Q(s) = G(s)^{-1}T_{eq}(s) = G(s)^{-1}H(s) \quad . \tag{4}$$

The inversion of $G(s)$ can only be performed if $G(s)$ does not contain any zeros in the right half plane. Otherwise $Q(s)$ will be unstable. In the final step, equations (1) and (3) are set equal, which results to:

$$C(s) = \frac{Q(s)}{1 - Q(s)G(s)} \quad . \tag{5}$$

As discussed in [7], the performance of the controller is highly affected from the location of the controller (i.e. on which bus it is installed), the saturation limits of the external power source (i.e. the power constraints of the storage device) as well as the speed of the controller's response, which can be influenced from the parameter ω_n in (2). The goal of this paper is to provide a technique where these characteristics can be optimally determined. In the next section, the modelling of the actuator block $A(s)$ will be described.

A. Deriving a Battery model for PHEVs

In this section, the focus is on the battery model for modelling a PHEV fleet as the actuator $A(s)$, which is depicted in Fig. 1. Li-Ion technology is a promising candidate for PHEV battery applications. Therefore, the model for the actuator is based on a High-Power Li-Ion battery model, developed in [9]. The circuit diagram of the battery model is depicted in Fig. 3 and the state-space representation, used for the dynamic simulations, can be formulated as:

$$\begin{bmatrix} \dot{V}_{cb} \\ \dot{V}_{cc} \end{bmatrix} = \begin{bmatrix} \frac{-(R+R_c)}{C_b(R_cR_e+R(R_e+R_c))} & \frac{R}{C_b(R_cR_e+R(R_e+R_c))} \\ \frac{R}{C_b(R_cR_e+R(R_e+R_c))} & \frac{-(R_e+R)}{C_b(R_cR_e+R(R_e+R_c))} \end{bmatrix} \begin{bmatrix} V_{cb} \\ V_{cc} \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{R_c}{C_b(R_cR_e+R(R_e+R_c))} \\ \frac{R_e}{C_b(R_cR_e+R(R_e+R_c))} \end{bmatrix} V_1 \quad ,$$
$$\begin{bmatrix} i_s \end{bmatrix} = \begin{bmatrix} \frac{R_c}{R_cR_e+R(R_e+R_c)} & \frac{R_e}{R_cR_e+R(R_e+R_c)} \end{bmatrix} \begin{bmatrix} V_{cb} \\ V_{cc} \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{-(R_c+R_e)}{R_cR_e+R(R_e+R_c)} \end{bmatrix} V_1 \quad . \tag{6}$$

TABLE I: Values of the battery model elements[1]

Parameter	Value	Parameter	Value
C_b	82 kF	C_c	4.074 kF
R_e	1.1 mΩ	R_c	0.4 mΩ
R_t	1.2 mΩ	R_0	10 mΩ
V_{0ref}	3.6V		

[1]The capacitance values might seem unrealistic, but these parameters help only in the derivation of a model with realistic behaviour and do not reflect actual capacitors; the actual battery cell is more complex.
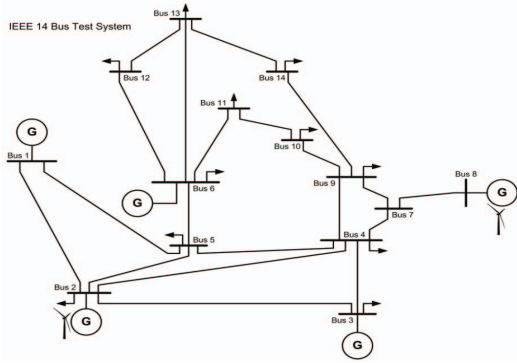
Fig. 4: IEEE 14 bus system with renewable energy generators at bus 2 and 8

Fig. 5: Control power availability profile for two differently sized PHEV fleets

The variables used in Eq. 6 correspond to the respective battery model elements in Fig. 3. The input signal which is generated by the controller is denoted by u_{act} in the figure and the variable V_{0ref} represents the reference voltage, fixed at 3.6V. As discussed in [7], a series connection of such battery cells can be built up without introducing a significant change in the fast response of the actuator system. The parameters of the model are given in Table I.

III. THE POWER SYSTEM INCLUDING WIDE SCALE PHEV PENETRATION AND RES

As the individual dynamic battery model has been derived above, it is now important to determine how the PHEVs temporally and spatially behave in the power system in order to define the saturation limit of the potential actuators.

Fig. 4 shows the IEEE 14-bus power system [10], to which RES are connected at buses 2 and 8. It is assumed that PHEVs are adopted on a wide scale and connect to the system on lower network levels when arriving at their individual destinations. Further, the PHEVs are assumed to travel only from home to work and back. In order to simulate the driving behavior, the model described in [11] is used to individually simulate energy consumption for every car. To capture differing driving behaviors, a choice of 84 possible drive cycles, which are composed of realistic drive cycles published by the Environmental Protection Agency (EPA), is made and leads to a state of charge (SOC) at the arrival at the workplace. The same drive cycle is used to simulate the way back to the home location. Once connected, either at the home or work location, it is assumed that the cars start charging at a power rate of 1.75kW. The available amount of positive and negative control power of the actuator $A(s)$ at time t, i.e. the saturation limits of the actuator, can be derived according to:

$$
\begin{aligned}
&\begin{aligned}
Positive \\
Control \\
Power
\end{aligned}
\left\{
\begin{aligned}
&P_{\text{pos}}^{\text{sat}}(\tau) = \sum_k \min \left\{ \begin{aligned} & C_k^{\text{p}} \\ & (SOC - 0.2)\frac{C_k^{\text{B}}}{\delta\tau} \end{aligned} \right\} + \mathfrak{p}_k(\tau) \\
&SOC_k(\tau + \delta\tau) = SOC_k(\tau) + \frac{(\mathfrak{p}_k(\tau) - \mathfrak{p}_k^{\text{c}})\eta_{\text{eff}}\delta\tau}{C_k^{\text{B}}}
\end{aligned}
\right. \\
&\begin{aligned}
Negative \\
Control \\
Power
\end{aligned}
\left\{
\begin{aligned}
&P_{\text{neg}}^{\text{sat}}(\tau) = \sum_k (C_k^{\text{p}} - \mathfrak{p}_k(\tau)) \\
&SOC_k(\tau + \delta\tau) = SOC_k(\tau) + \frac{(\mathfrak{p}_k(\tau) + \mathfrak{p}_k^{\text{c}})\eta_{\text{eff}}\delta\tau}{C_k^{\text{B}}}
\end{aligned}
\right. \\
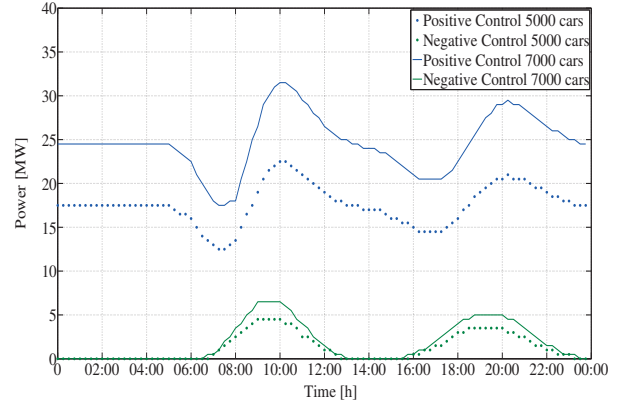&\qquad\qquad \forall\, k \in \mathcal{PHEV}(\tau) = \{1 \ldots N_{PHEV_n}\}
\end{aligned}
$$

(7)

where $P_{\text{pos}}^{\text{sat}}(\tau), P_{\text{neg}}^{\text{sat}}(\tau)$ give the positive and negative available control power in time step τ. Further, $\mathfrak{p}_k(\tau)$, $C_k^{\text{p}}, C_k^{\text{B}}, SOC_k(\tau), \eta_{\text{eff}}$ denote the charging power, the power connection capacity, the battery capacity, the actual state of charge and the charging efficiency for the k-th car from the set of connected PHEVs denoted $\mathcal{PHEV}(\tau)$, respectively. The variable $\mathfrak{p}_k^{\text{c}}$ denotes the actually demanded power by the controller. The efficiency η_{eff} is equal to 0.9 when the PHEV is being charged and $1/0.9$ when it is being discharged.

For assessing the available control power, time intervals of 15 minutes, are used. The aggregation of connected PHEVs is hence performed 96 times a day resulting in the respective number of control power values. This approach assumes that the available control power and hence the number and state of apparent PHEVs does not change during this time frame of 15 minutes. The aggregation scheme results in profiles of varying control power availability which are exemplarily given in Fig. 5 for two PHEV fleets. Observe that the available positive power from 00.00 to 05.00 for a fleet of 7'000 cars is about 25 MW. As a large number of PHEVs depart after 05.00, the positive control power decreases to ca. 18 MW. The amount of negative control power is zero in the morning as the PHEVs are assumed to have a full battery and hence cannot be further charged. The amount of available positive control power increases to 32 MW until 11.00 as PHEVs arrive at their work location and start to recharge. The available positive power is larger than before because, according to Eq. (7), more power can be drawn from the vehicles when they recharge. Obviously during this time, more negative control power is available, as the vehicles charge with a rather low charging rate which can be increased through the controller if desired.

In order to model the daily traveling behavior, which varies between regions of a power system, the availability profile depicted in Fig. 5 is shifted in time. Furthermore, the number of PHEVs per node is assumed to differ (see Table II). In total, 72'500 PHEVs are maximally connected at the same time throughout the power system.

IV. IMPACT OF THE FAULT LOCATION AND SIZE ON THE CONTROLLER LOCATION SELECTION

As already mentioned, several factors, such as the control speed ω_n, the available PHEV control power at a certain

TABLE II: Max. number of PHEVs connected at each node

Bus No.	# PHEVs	Bus No.	# PHEVs	Bus No.	# PHEVs
1	0	6	6500	11	6500
2	5000	7	0	12	6000
3	6000	8	5000	13	6500
4	7000	9	7000	14	5500
5	5500	10	6000		
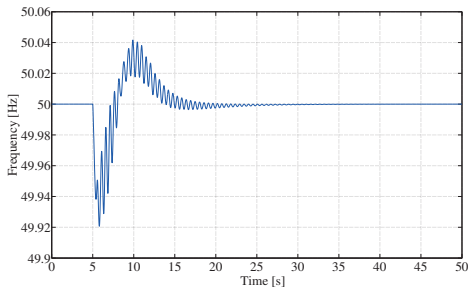
Fig. 6: Frequency response in the reference scenario [fault: - 0.2 p.u. @ bus 4; controller @ bus 6, $\omega_n = 4$.

bus and the bus type itself (i.e. generator or load bus), play a decisive role when trying to achieve an optimal system response, i.e. a fast frequency recovery. In order to investigate the potential effects of fault location and size and subsequently derive a general rule for the controller design and location, several case studies have been performed.

In the following, it is assumed that only the controller speeds $\omega_n = [2, 4, 6, 8]$ are used. It is assumed that a controller is installed on every bus where PHEVs can connect, but at each time interval only one of these controllers can be active. The controller location can change 96 times during a day, i.e in every PHEV aggregation interval. The frequency can be measured in the generator and RES infeed buses. In order to be able to perform control actions from a load bus, the frequency must also be measured. This has been achieved by including the specific bus in the differential equations assuming an infinitesimal inertia value.

Fig. 6 shows the system response for the reference case. A generation loss of 0.2 pu is simulated for 50 seconds at bus 4. The fault occurs at $t =5$ seconds. The controller is situated at bus 6, is tuned with $\omega_n = 4$ and the control power is limited by the respective PHEV fleet behavior.

In order to measure the controller performance, the following criterion is applied. The system should be stable at all times and the frequency should recover to its initial steady-state value of 50 Hz. The controller objective is then to achieve the minimal oscillation area. The criterion can be formulated with the following mathematical expression:

$$Area = \sum_{t=15s}^{50s} |f(t) - 50\text{Hz}| \quad . \tag{8}$$

It should be noted that the summation starts after 15 seconds in order to focus on the fast restoration of the frequency without taking into account the initial underfrequency and overshoot.

The most favorable controller location and ω_n value for each fault in each aggregation interval can be easily found when
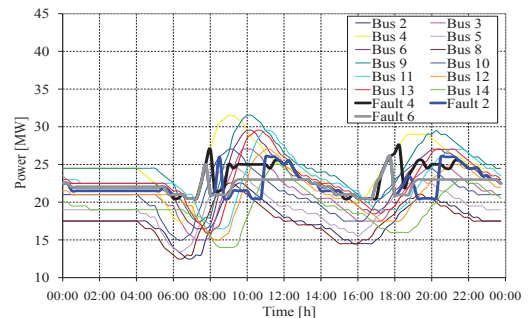

Fig. 7: Saturation limits of the controller for the three different fault locations when the most favorable controller is selected at each time interval.

comparing the frequency responses for all possible controller locations and speeds.

A. Impact of fault location on the controller location selection

For studying the effect of the fault location on the controller location, the fault, described above and referred to as the reference case, occurs at three different locations, namely at bus 2, where a RES has been installed, at bus 4, which is a load bus and at bus 6, which is a generator bus. The fault is applied on each of the 96 time intervals at the specific buses. All possible controller locations and their effectiveness in terms of minimizing the oscillation area, given by Eq. 8, are assessed.

Obviously, as the load situation in the system and the availability of control power vary during the day, the most favorable controller location and the power which is drawn from it change as well. Fig. 7 illustrates the optimal control power trajectories over time and over the number of system buses for the different fault cases. The saturation limits at each controller location for a fault at bus 2 are displayed in blue, while the trajectories for the faults at bus 4 and bus 6 are displayed in black and in grey, respectively. It can be seen that although the fault size remains the same, the temporal distribution of the necessary control power varies and is obviously dependent on the location of the controller. This is because the impact of the power infeed varies depends on the infeed location (i.e. controller location). The ability of the controller to damp the oscillations is also closely related to the available power that needs to be fed into the system. Other factors, such as the actual power flow and the thermal losses are also of importance. Observe that the power consumed is less if a fault occurs at a generator bus than at a load bus, which is due to the high inertia of the generator connected to the respective bus.

Fig. 8 shows the bus selection and Fig. 9 shows the oscillation area for the three different fault locations. The trajectory of the controller location, depicted in Fig. 8, is roughly the same when the fault occurs at bus 2 or at bus 4. This is because of the grid design. Buses 2 and 4 are peripheral buses in contrast to bus 6 which is a central bus. Therefore, for bus 2 and bus 4, there is only a limited number of locations where the controller can be installed/activated in order to mitigate the
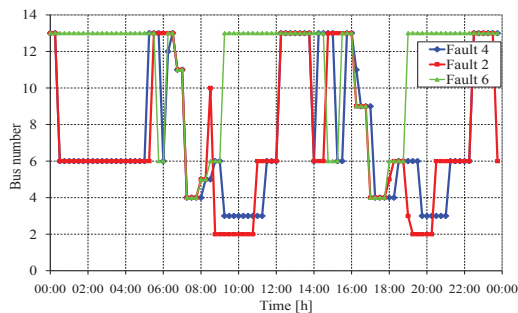
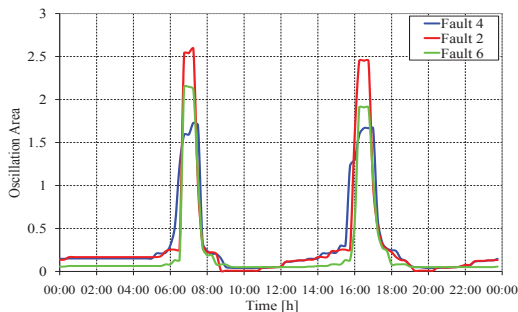Fig. 8: Controller location for the three different fault locations for optimal frequency response.

Fig. 9: Oscillation area for the three different fault locations when the most favorable controller is selected at each time interval.
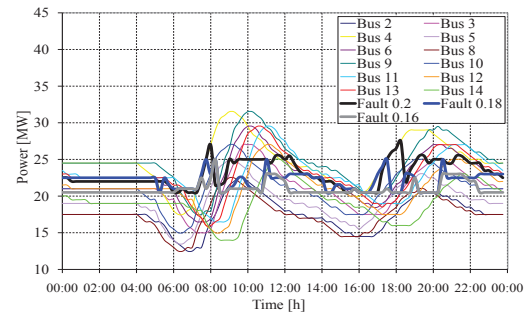

Fig. 10: Power drawn from the controller for the three different fault sizes when the most favorable controller is selected at each time interval.
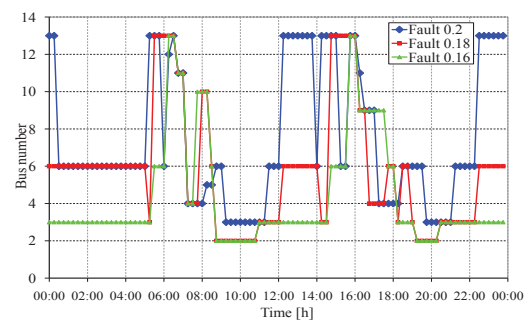

Fig. 11: Controller location for the three different fault sizes for optimal frequency response.

fault most effectively. The centrally located bus 6 incorporates many interconnections through which the necessary power in a fault case can be delivered. Hence, its stability and the ability of the system to restore the steady-state frequency is better than at the other buses for most of the time as illustrated in Fig. 9. Only during 07.00–07.15 and 16.00–16.15 the oscillation area is smallest at bus 4. In both situations, either bus 4 or bus 9 are able to optimally compensate the fault, both located in the close vicinity of bus 4.

B. Impact of fault size on the controller location

Not only the fault location is crucial when selecting the optimal controller location. It is also important how large the occurring fault actually is. Here, three different faults, with size of 0.16 pu, 0.18 pu and 0.2 pu, have been simulated. Faults larger than 0.2 p.u. led the system to instability when not enough control power is available, e.g. between 06.00 and 08.00 and between 16.00 and 17.30. Then, the controller, no matter at which location, would not be able to restore the steady-state frequency. Smaller fault sizes have not been simulated since then, no matter at which location the controller would be activated, it would always be possible to restore the steady-state frequency quickly because of the large amount of available control power.

Fig. 10 illustrates the optimal control power trajectories over time and over the number of system buses for the different fault sizes. Observe the strong correlation between fault size

and consumed control power. For large fault sizes one can generally see a higher consumption of control power. However, there are minor exceptions which are due to the controller location. It is found that the consumed power is typically higher than the simulated fault. This results from the control objective, as a higher, although spatially dependent, power infeed damps the frequency oscillations faster.

Fig. 11 depicts the controller location changes. Obviously, for every fault size frequent location variations occur. In case of the 0.16 pu and the 0.18 pu fault the number of bus changes is marginally lower than for the largest fault size. Fig. 12 depicts the oscillation areas for the different fault sizes. Not surprisingly, the oscillation area grows with increasing fault size. This suggests that the available control power should be as high as possible and larger than the maximum fault size. Evidently, the frequency can be faster stabilized with increased power infeed through the controller.

V. Q-LEARNING FOR SELECTING AN OPTIMAL CONTROLLER LOCATION

As seen in the previous section, the extraction of an explicit relationship between the optimal controller location and the different fault sizes and locations is a formidable task, dependent on several factors, such as the number of PHEVs connected to each node. The motivation for using reinforcement learning techniques lies exactly on the fact, that they allow to determine an optimal policy through trial-and-error
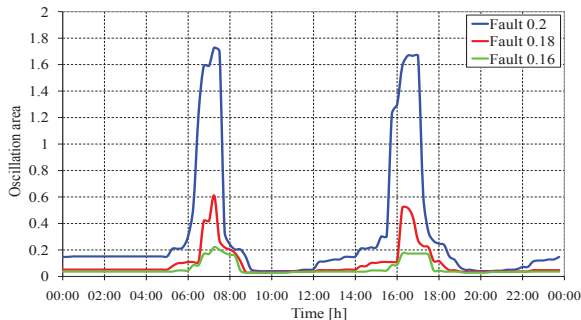
Fig. 12: Oscillation area for the three different fault sizes when the most favorable controller is selected at each time interval.

iterations without the need to have an explicit representation of the environment. Here, the optimal location and ω_n value of the controller have to be determined at every time step, given the fault location and the number of PHEVs connected to each node.

A reinforcement learning algorithm, suitable for this task, is the Q-learning algorithm [12]. An agent tries an action a at a particular state s, and evaluates its consequences in terms of the immediate reward *and* its estimate of the value of the state to which it is taken, according to the following rule:

$$Q(s,a) := (1-\alpha)Q(s,a) + \alpha(r + \gamma(max_{a'}Q(s',a'))). \quad (9)$$

Defining the notation, s is the current state and a is the action that the agent decided to take and leads him from the state s to the next state s'; a' is the action in the state s' that corresponds to the highest Q-value; α is the learning rate and γ is a discount factor. According to equation (9) the Q-value is dependent on three factors. The value $Q(s,a)$ computed in the previous iteration, the reward r for taking action a when in state s (this is given as an input) *and* the maximum Q-value that it can achieve in the future (i.e. next state), when it decides to take action a in the present.

In the case studied in the present paper, it is assumed that each state is a 15-min time interval for *each* bus. Thus, for a whole day there are in total $96 \cdot 12$ states (buses 1 and 7 are excluded). The action that needs to be decided in each case is the location and the ω_n of the controller.

If each action is executed in each state an infinite number of times on an infinite run and α is decayed appropriately, the Q-values will converge with probability 1 to Q^* [12]. In our case, we introduced a small tolerance value. When the Q-values differed less than the tolerance value during two consecutive iterations, we assumed that the algorithm has converged. When the learning phase is over (i.e. Q-values are nearly converged to their optimal values), the agent selects in each state, the action with the highest Q-value.

During learning, however, there is a difficult exploitation versus exploration trade-off to be made. The agent should, on the one hand, learn how to select the actions with the highest reward, but on the other hand it should explore as many actions as possible in every state, since actions of low rewards might lead to states where a higher future reward can

be attained. There are no good, formally justified approaches to this problem in the general case [13]. Towards the end of this section the developed method to tackle this problem will be mentioned.

The objective of the algorithm is to find the optimal location for the controller at each time step, so that, in case of a disturbance, the oscillations can be damped as fast as possible. The constraint is, however, that only 3 changes of the controller location are allowed during a day. In other words, if one could install 4 controllers in the power system, which buses would be the optimal for that? In order to enable the agent to learn through the Q-learning process, a fault is applied (e.g. to bus 4) and the frequency response of the system is measured for all the different controller locations and for different values of the ω_n. The reward r, that the Q-learning rule needs, is computed by the following relationship:

$$r = \frac{Minimal\,Area}{Area} \frac{1}{n^2} \quad (10)$$

In (10) $Area$ stands for the total oscillation area of the current state (bus and time step). The best possible total oscillation area at that precise time step among all the buses is denoted by $Minimal\,Area$. The factor n represents the number of buses that can deliver the needed power to compensate for the fault. In this way, the more buses that can provide the control service, the less will be the reward, because n will be high. If only one bus can deliver the necessary power, the reward will be equal to 1.

The learning rate α influences the impact of the new rewards to the old values. A high value of α stands for an algorithm that weighs a lot the current action. A small value of α takes more into account the old actions. In this work, a rather small value of α is selected [α=0.2], since the old actions (when a location change took place) must have an influence on the present and future actions. The discount factor γ determines the impact of the future gains to the current rewards. A high value of γ means that future actions will have a strong influence on the current state. In this work, a rather high value of γ is selected [γ=0.8] because the system behaviour until the next changing point is of importance. For example, a controller location where not enough power is available at some future point of the day is not acceptable.

As the Q-learning algorithm runs, at every time step the agent has the option either to keep the controller in the same location or to change it, according to a certain probability p. If a uniform probability distribution is assumed, then the joint probability of changing the location in e.g. the fifth step is equal to $P = (1-p)^4 \cdot p$, substantially less than in the first step. As it can be easily understood, assuming a uniform distribution for the location changes would result in an agent that tries to change the controller location quite early in time and then, as it has reached the limit of 3 changes, it must stay with its last choice until the end of the day. For a more appropriate exploration of the search space, we assumed the following probability distribution:

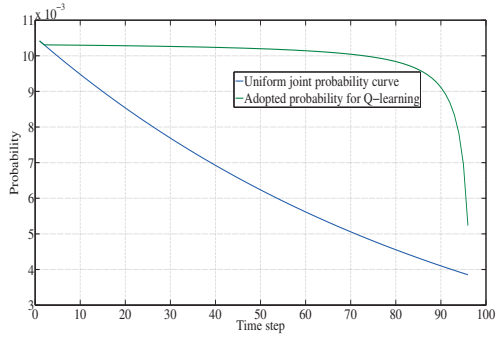$$p_\tau = \sum_{i=1}^{\tau} p_{\tau-1}^i \quad (11)$$

Fig. 13: Joint probability distributions.

According to (11), the probability of changing the controller location is low in the beginning, and it increases as time passes. Defining the notation, p_τ is the individual probability of changing in step τ and $p_{\tau-1}$ is the probability of changing the controller location from the previous time step. The initial value p_0 is selected equal to $p_0 = 1/96$, as there exist 96 time steps during the day. The joint probability resulting from this distribution, as well as from the uniform distribution, are plotted in Fig. 13.

VI. Q-LEARNING FOR THE SELECTION OF CONTROLLER LOCATION: AN EXAMPLE

As illustrated in Fig. 8 and in Fig. 11, in order to achieve optimal frequency response at every time step, the controller location must change often during the day. This exhibits certain drawbacks since it makes the power system operation much more complex. At the same time it increases the infrastructure costs, as a controller might need to be installed almost on every bus of the system. In order to limit the controller location changes, an optimal controller location policy should be found, which has to determine where the controller should be located and at which time step should the controller location change. As already mentioned, an explicit relationship between controller location and optimal frequency response is difficult to derive. But such a policy can be easily determined using the Q-learning approach, described above.

In the simple example that will be presented, the location changes are limited to maximally 3 changes. A fault at bus 4 with size of 0.2 pu is simulated. In order to limit the search space, it is assumed that in every location, the controller with the optimal ω_n is always selected. The Q-learning algorithm has to decide when to make the change and which bus to select. Of course, in a more general example, the algorithm could also decide for the ω_n. Fig. 14 illustrates the trajectories of control power over time and over the number of system buses for the Q-learning scheme. These are compared with the results for the most favorable controller location. The Q-learning method largely follows the control power availability profiles of 3 different buses and takes advantage of their maximally possible control power infeed instead of changing too often the controller location to a more efficient one. The number of location changes is illustrated in Fig 15. Q-learning locates the controller on buses 6, 11, 5 and 9 over the day. Note, that both methods select the same buses during the critical times.
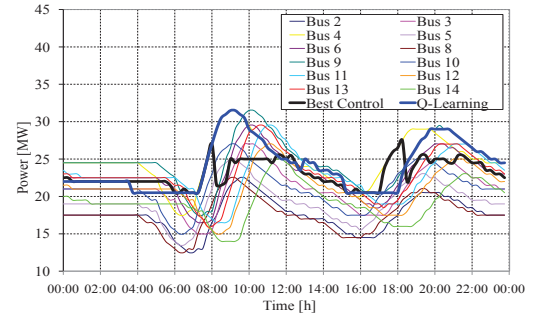

Fig. 14: Q-learning approach vs. unlimited changes in the controller location (optimal).
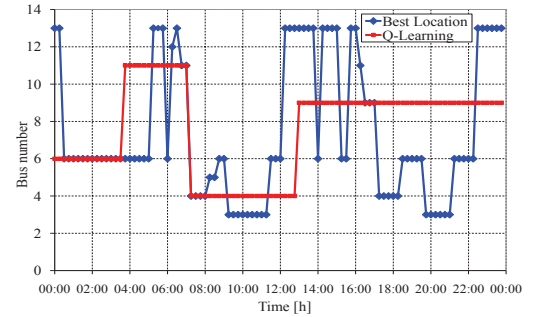

Fig. 15: Most favorable controller location (optimal) and Controller location selected by the Q-learning method.

Between 07.00 and 07.15 and between 16.15 and 16.30 there is only one bus which can compensate the fault.

The effect of not being able to always select the optimal controller location is depicted in Fig. 16. The overall oscillation area for the Q-learning method is larger than in the optimal case. It deviates from the optimum especially between 04:00 and 06:00 and between 14:30 and 18:15. This happens due to the fact that Q-learning needs to switch to bus 11 and bus 9, respectively, in order to be able to stabilize the system in the critical time steps.

The performance of the Q-learning algorithm can be increased by either increasing the limit of possible location changes or the available power at the critical buses.
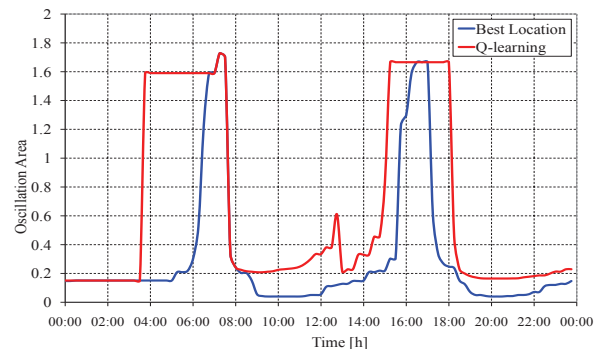

Fig. 16: Oscillation area: Q-learning vs. most favorable controller location.

VII. Conclusions and Outlook

This paper shows that the aggregated storage of distributed, mobile electric vehicles can be an effective and relatively inexpensive solution (in comparison with large battery storages) for assuming the role of the external power source, which if appropriately controlled, can emulate the missing inertial behaviour of RES. Thereby, PHEVs can contribute to overall power system stability and improve the frequency response of the system.

However, the performance of the implemented controller is largely dependent on the availability of control power, inherently temporally and spatially variable for PHEV fleets, on the controller location in the power system, on the fault size and on the fault location.

The study which was carried out within this paper showed that the location of the controller (i.e. the bus from which the power is drawn or absorbed) has to change frequently if the optimal frequency response is sought at every time step. This would imply a controller installation on almost every bus. The formulation of an explicit relationship between controller location and optimal frequency response is not straightforward. Therefore, in order to determine an optimal placement strategy for a limited number of controllers, a technique based on Q-learning has been developed. The Q-learning algorithm delivers a solution that ensures stability at all times while avoiding additional infrastructure costs and increased complexity in the power system operation that come with the controller installation on every bus of the power system. As a side-effect, the overall performance when damping the frequency deviations is, as it would be expected, somewhat decreased, compared to the most favorable controller location selection in each time step.

The proposed Q-learning approach can be easily extended, by evaluating the frequency response of the system when different faults occur on different buses. These faults could then be combined with certain probabilities of occurrence which can be integrated in the rewards vector that the Q-learning algorithm receives as input. Such an approach could then be applied to larger power systems in order to determine the most inexpensive solution to ensure system stability in case of a wide scale deployment of RES as well as PHEVs.

References

[1] G. Lalor, A. Mullane, and M. O'Malley. Frequency control and wind turbine technologies. *IEEE Transactions on Power Systems*, 20(4):1905 – 1913, 2005.

[2] M. D. Galus, R. A. Waraich, F. Noembrini, K. Steurs, K. Boulouchos, K. W. Axhausen, and G. Andersson. Integrating power systems, transport systems and vehicle technology for load, behavioral and environmental analysis of electric mobility. *submitted to IEEE Transactions on Smart Grids*, 2010.

[3] W. Kempton and J. Tomic. Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. *Journal of Power Sources*, 144(1):268–279, 2005.

[4] S. L. Andersson, A. K. Elofsson, M. D. Galus, L. Göransson, S. Karlsson, F. Johnsson, and G. Andersson. Plug-In hybrid Electric Vehicles as Regulating Power Providers: Case Studies of Sweden and Germany. *Energy Policy*, pages 2751–2762, 2010.

[5] M. D. Galus, R. La Fauci, and G. Andersson. Investigating PHEV wind balancing capabilities using heuristics and model predictive control. In *IEEE Power and Energy Society (PES) General Meeting*, Minneapolis, Minnesota, USA, 2010.

[6] M. D. Galus, S. Koch, and G. Andersson. Provision of load frequency control by PHEVs, controllable loads and a co-generation unit. *accepted to IEEE Transactions on Industrial Electronics, Special Issue on Smart Grids*, 2010.

[7] M. Gautschi and L. Friedrich. Grid stabilization control and frequency regulation for inverter-connected distributed renewable energy sources (Master thesis, available online). University of Wisconsin-Madison and ETH Zurich, 2009.

[8] J.M. Maciejowski. *Multi-variable Feedback Design*. Addisson-Wesley, 1989.

[9] V.H. Johnson, A.A. Pesaran, and T. Sack. Temperature-dependent battery models for high-power lithium-ion batteries. In *17th Annual Electric Vehicle Symposium*, Montreal, Canada, 2000.

[10] Power Systems Test Case Archive, College of Engineering, University of Washington. http://www.ee.washington.edu/research/pstca/.

[11] M. D. Galus and G. Andersson. Power system considerations of plug-in hybrid electric vehicles based on a multi energy carrier model. In *Proceedings of IEEE Power and Energy Society (PES) General Meeting*, Calgary, Canada, 2009.

[12] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

[13] L. Pack Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

Spyros Chatzivasileiadis (S'04) was born in Athens, Greece in 1985. He received the diploma in Electrical and Computer Engineering from the National Technical University of Athens in 2007, where he worked as a research assistant at the Power Systems Laboratory until August 2008. In September 2008, he joined the Power Systems Laboratory of ETH Zurich, where he is currently a PhD student. His research interests include power systems control, operation and planning, and machine learning applications for power systems.

Matthias D. Galus (S'07) was born in Swientochlowitz, Poland. He received a Dipl.-Ing. degree in electrical engineering and a Dipl.-Ing. degree in industrial engineering from the RWTH Aachen, Germany, in 2005 and in 2007, respectively. He joined the Power Systems Laboratory of ETH Zurich, Switzerland in 2007 where he is working towards a PhD. His research is dedicated to modeling, optimization and efficient integration of PHEV into power systems. He is a student member of the IEEE and VDE (German society of electrical engineers).

Yves Reckinger was born in Luxembourg, Luxembourg in 1987. He received a Bachelor Degree in Electrical Engineering and Information Technology from the ETH Zurich in 2009. He now continues on the same master programme, where he specializes in electrical power systems and mechatronics.

Göran Andersson (F'97) was born in Malmö, Sweden. He obtained his MSc and PhD degree from the University of Lund in 1975 and 1980, respectively. In 1980 he joined the HVDC division of ASEA, now ABB, in Ludvika, Sweden, and in 1986 he was appointed full professor in electric power systems at the Royal Institute of Technology (KTH), Stockholm, Sweden. Since 2000 he has been full professor in electric power systems at ETH Zurich, Switzerland, where he heads the Power Systems Laboratory. His research interests are in power system analysis and control, in particular power system dynamics and issues involving HVDC and other power electronics based equipment. Prof. Dr. Göran Andersson is a member of the Royal Swedish Academy of Engineering Sciences and Royal Swedish Academy of Sciences. He was the recipient of the IEEE PES Outstanding Power Educator Award 2007.