Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 231 – 239

Procedia
Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

# Distributed Optimization of Solar Micro-grid using Multi Agent Reinforcement Learning

Leo Raju[a*],Sibi Sankar[a], Milton R S[a]

[a]SSN College of Engineering ,OMR,Chennai,603110,India

**Abstract**

In the distributed optimization of micro-grid, we consider grid connected solar micro-grid system which contains a local consumer, a solar photovoltaic system and a battery. The consumer as an agent continuously interacts with the environment and learns to take optimal actions. Each agent uses a model-free reinforcement learning algorithm, namely Q Learning, to optimize the battery scheduling in dynamic environment of load and available solar power. Multiple agents sense the states of the environment components and make collective decisions about how to respond to randomness in load, intermittent solar power using a Multi-Agent Reinforcement Learning algorithm, called Coordinated Q Learning (CQL). The goals of each agent are to increase the utility of the battery and solar power in order to achieve the long term objective of reducing the power consumption from grid.



Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

*Keywords*: Solar micro-grid; Multi-agent Reinforcement Learning; CQ-learning; battery scheduling; optimization

## 1. Introduction.

The world economy is largely dependent on quality power. The present power grid is expected to undergo a period of rapid change in the near future. Renewable energy plays a significant role in building a green and sustainable environment. Solar and wind are the only solution to the growing energy crisis in the world [1]. The key element driving change is the emergence of smart grids in the developed countries.

*Corresponding author. Tel: 91-44-27469700; fax: 91-44-27469772

Email address: leor@ssn.edu.in



Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

doi:10.1016/j.procs.2015.02.016

Smart Grid paradigm represents a transition towards an intelligent, digitally enhanced, two way power delivery grids. The aim of smart grid is to promote and enhance the efficient management of the power generation and delivery facilities by incorporating advanced communications, Information technology and control methodologies into existing power grid. Micro-grids are the building block of smart grid. Integrating renewable energy into the micro-grid is the way forward for economic and environmental optimization, generating clean and green energy and thereby providing a solution to global warming [2].

The importance of having more reliable, efficient, smart systems is getting more public attention. In coming years consumer wants only smart machines and expect machine to think and operate autonomously and optimally. Fuzzy logic is used in energy management of micro-grids[2]. Genetic algorithm is used in smart energy management of micro-grids[3, 4]. Constrained optimization is used in energy management of hybrid renewable energy generation[5]. Expert systems and other classical and heuristic algorithms are also used in energy management of micro-grids[6, 7, 9]. Agent-based modeling approach is used to model micro-grids and energy management is done by simulation of the interactions between individual agents[2, 3, 8]. Reinforcement learning is used for energy management of micro-grid using wind energy system[10, 14]. Coordination Q learning method in Multi-Agent Reinforcement Learning (MARL) is used for coordination among agents for energy management[15]. The aim of this paper is to introduce a smart decision making system using MARL method, called Coordination Q-Learning (CQL), for optimization of the distributed energy management in the micro-grid. The system behaves in strategic manner when dealing with operational scenarios, by aiming to achieve the lowest possible cost of power generation.

The rest of the paper is organized as follows. Section 2, 3 and 4 presents the modeling framework of the solar micro-grid and the details of solar photovoltaic (PV) system. Section 5 to 9 provides a comprehensive framework of solar micro-grid energy management with reinforcement learning. Simulation results are analyzed and the performance of the distributed solar micro-grid systems in the long run is discussed in section 10. Conclusions and suggestions for possible improvements are given in the last section.

## 2. Solar Micro-grid

A micro-grid is a localized grouping of electricity sources in the distribution side which can supply power to communities, universities and other local requirements. It can operate stand alone or connected to main grid[2]. A micro-grid is low power local grid made up of renewable power generators, electrical loads, and a distribution system that allows power to flow from the sources to the loads. In case of excess demand, power is allocated from grid. The urban solar micro-grid involves a consumer with a dynamically varying load $D_t$ , a transformer providing electricity power from the external grid, a solar generator (solar photovoltaic (PV) system) with available power output $P_{sp}$ and a storage facility with a level of battery charge $R_t$. The architecture of the considered solar micro-grid is shown in Fig. 1. The consumer can cover his demand partly by using the electricity produced by the local renewable (solar) generator, store electricity in the battery when the solar source is available and can discharge the storage when needed. The consumer has the possibility to control the storage and the solar power generator. The challenge in solar micro-grid is that the power supply is intermittent in nature and with the ultra-capacitors and battery the randomness in the supply should be managed so that the load is given constant supply all the time irrespective of the nature of the solar power. The battery can be charged by the solar power or from the grid. A solar PV module is a clean energy source used in power systems which absorb solar irradiation and converts it into electric energy. A solar PV module is the basic element of each photovoltaic system. A PV system uses sun as power source and converts solar power into electricity[13]. A solar PV module consists of a number of solar cells connected in series or parallel based on the requirement. There are various factors which affect the solar power like solar irradiance (G), temperature (T), partial shading, and cloud, arrangement of the cells and the angle of tilt of the panel. Feed forward neural network is trained with G and T to get equivalent circuit parameters. Once the neural network is trained with sufficient number of examples then the current and the voltage of the solar module for unknown values of G and T is determined by generalisation

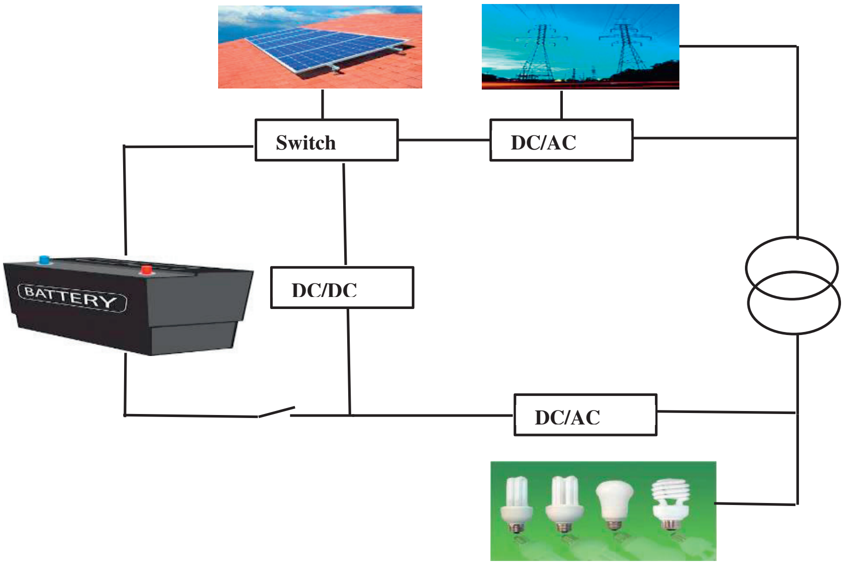


Fig. 1. Solar Microgrid.

The maximum power $P_{sp}$ is found by Maximum Power Point Tracking (MPPT) algorithm. The MPPT refers to the point with a maximum power output in the curve under specific external temperature and solar irradiation [13].

## 3. Modeling of the battery storage

A simplified model of battery storage dynamics is adopted by implementing a discrete time system for the power flow dynamics over the time step interval $D_t$.

$$R_t = R_{t-1} + R_t^{store.charge} + R_t^{store.discharge} \tag{1}$$

$R_t$ and $R_{t-1}$ are the levels of energy stored in the battery at time $t$ and $t-1$

$R_t^{store.charge}$ and $R_t^{store.discharge}$ are power flows over time step interval delta $t$ between battery and consumer and solar generator and the battery respectively [10].

## 4. Modelling of the consumer agent

The dynamic variations of load, solar power and the battery are considered as external environment. Consumer is modelled as an individual agent who makes use of reinforcement learning for its decision-making, action-taking and moving towards its goal. Reinforcement learning is about learning with limited feedback and falls between supervised and unsupervised learning. Markov Decision Process MDP has become standard formalism for learning sequential decision making. In MDP, the environment is modelled as set of states and actions can be performed to control the system state. In MDP environment the effect of an action taken in a state depends only on that state and not on the prior history. The goal is to maximize the performance criteria of the system. The consumer agents interact, adapt and take decisions towards its goals defined in the form of reward functions in a MDP environment, characterized by the available solar power output $P_{sp}$, the load $D_t$ and the battery charge $R_t$.

## 5. Markov Decision Process

Markov Decision Process (MDP) is a way to model a sequential decision making under uncertainty. The MDP is formalized with discrete states and actions. The initial state is $s_o$ and each state will have a reward r associated with it. The transition function T (s|a, s') indicates the probability of transitioning from state s to s' when action ‘a’ is taken. A discount factor $\gamma$ in the range 0 to 1 is applied to future rewards. This represents the notion that a current reward is more valuable than one in the future. If it is near zero, future rewards are almost ignored; a near one places great value on future reward. The reward from a policy is the sum of the discounted expected utility of each state visited by that policy. The optimal policy is the policy that maximizes the total expected discounted reward.

## 6. Reinforcement Learning

Reinforcement learning algorithm is used to model the consumer's adaptation to a dynamically changing environment by performing actions of battery scheduling in an MDP environment[10]. The agents observe the environment and take an action. It gets a reward or punishment from the environment. The agent takes the next action to optimize the reward in the long run. After a number of interactions, the agent finds the optimal policy to achieve long term objective. The goal of an agent is to find the optimal policy based on interactive learning with the environment. Fig. 3 shows a simple reinforcement learning scheme. The environment is taken as no of discrete states, denoted by S (t) at time t. Every state has a value dependent upon an immediate reward, denoted by R (t) at time t.

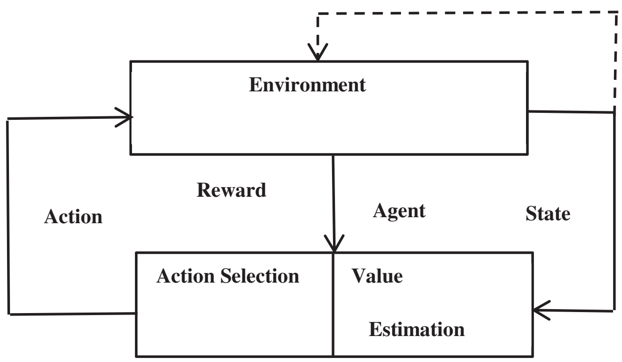


Fig.2. Reinforcement Learning

At each discrete time t, the agent takes an action out of number of possible actions, which affects the next state of the system, and therefore the next reward. After number of actions, the agent learns through experience and takes action which maximizes the long-time reward. The agent learns through mistakes as human being learns and takes better action based on accumulated experience. The agent learns the impact of each action over the period of time and choses the action, which leads to achieving goal of the agent in the long run[11].

## 7. Q learning

Q Learning (QL) is a model-free reinforcement learning where the agent explores the environment and finds the next reward plus the best the agent can do from the next state. In Q learning, the agent does not need to have any model of the environment. It only needs to know what states exist and what actions are possible in each state. We assign each state an estimated value, called a Q value[11]. When we visit a state and take an action we receive a reward. We use this reward to update our estimate of the value of that action in the long run. We visit the states infinitely often and the action values (Q values) are continuously updated till it becomes convergent. The Q learning algorithm is outlined below. In the algorithm, $\gamma$ is the discount factor and $\alpha$, learning rate. The reward functions are the response we get from the environment for the actions taken. If the battery is charging ( $a_1$ ) then the reward function is minimum

Of $P_{sp}$ and $B_{difference}$ . If it is discharging ( $a_0$ ) then the reward function is minimum of $D_t$ and $B_{level}$ . Here $B_{difference}$ is the difference between maximum possible charge and the current battery level ( $B_{level}$ ).

The Q-Learning algorithm[11] goes as follows:

1. Set the gamma parameter, and environment rewards in matrix R.
2. Initialize matrix Q to zero.
3. For each episode:
   Select a random initial state.
   Do while the goal state (convergent) hasn't been reached.

- Select one among all possible actions for the current state.
- Using this possible action, consider going to the next state.
- Get maximum Q value for this next state based on all possible actions.
- Compute: $Q(s,a) = Q(s,a) + \alpha[r + \gamma \max Q(s',a') - Q(s,a)]$ (2)
- $s \leftarrow s'$
- Set the next state as the current state.

End Do
End For

## 8. Multi Agent Reinforcement Learning

In the decentralized control, Multi Agent System (MAS) concepts are used for multiple agents acting in an environment, communicating and coordinating each other for global optimization. The agent can reason, control, learn, optimize and coordinate with other agents. Also the agent can be reactive, proactive and social. When Reinforcement Learning is inducted in Multi Agent System, MAS acquire the property of autonomous optimization and hence the concept of Multi Agent Reinforcement Learning (MARL) is developed. Distribution optimization is done by coordination and communication between the agents[15]. The agents make collective decisions about how to respond to user requests, dynamic changes in the environment and unplanned contingencies. In this paper, a method of MARL, called, Coordinated Q Learning (CQL) is used for coalition of agents. Two solar micro-grids with capacity of 150 KW and 200 KW are considered. One is at electrical department building and the other is at hostel respectively. Each agent optimizes the scheduling of the battery to increase the utility of the solar power and the battery to reach long term objective of reducing the power from the grid. The two agents communicate and co-operate to manage the various scenarios due to intermittency of the solar power supply and randomness in load.

## 9. Implementation of distributed optimization of solar micro-grids with Coordinated Q Learning

The forecast value of solar irradiance G is found from the National Renewable Energy Laboratory (NREL) for a 150 KW and 200 KW solar PV systems in our campus electrical engineering building and hostel respectively for the year 2014. The temperature T is taken from meteorological department for the same year. These two data are fed into already trained neural network to predict the solar power $P_{sp}$ for the whole year. Then the solar power distribution for the whole year in hourly basis is drawn as shown in the Fig. 4. The load pattern of electrical engineering department and hostel in our campus are considered. The hourly basis solar power ( $P_{sp}$ ) and load ( $D_t$ ) feeds the Q Learning (QL) algorithm. The agent learns optimal single agent policy when acting alone in the environment using single agent reinforcement learning method. The agents have their own model of expected rewards for every state action pair. The Coordinated Q learning algorithm checks to detect if there are changes in the observed rewards for the selected state-action pair[15]. If the change is detected in the immediate reward, then this state is taken as join state space in which collision occurs and is marked as dangerous state, where the action depends on other agents. If no changes occur in immediate reward then it marked as a safe state. So initial attempts are made to detect changes in the immediate reward. Any interference from other agents will reflect in the reward. Agent observes the changes in the reward and confirms that it is in the dangerous state and updates the Q value from the following equation (3).

$$Q_k^j(js,a) = Q_k^j(js,a_k) + \alpha t[r(js,ak) + \gamma \max_a Q(s_k^{'},a_k^{'}) - Q(js,a\}] \quad (3)$$

Where $Q_k$ stands for the Q-table containing Q values of the safe states, where no updates are required, and $Q_k^j$ contains Q values of the joint states $(js)$ or dangerous states, where updates are required. The safe and dangerous states are maintained in separate lists to decide which Q value table to use. The battery levels of each indicate the present state of the system. Thus the number of possible states is a chosen such as to increase the productivity of the algorithm while still having a reduced number of states. After acquiring the Q value from the table and maintaining the mean and variance of various states through single agent Q-Learning method, the agent undergoes reinforced Coordinating Q-Learning. Thus the agent has both local and global responsibilities for coordinating and cooperating in a distributed environment. Local optimization is done through single agent Q Learning and Global optimization is done through Coordinated Q Learning[15].

For the electrical department and the hostel the available solar power is drawn as shown in Fig. 3 and Fig. 4. Load variations in department and hostel are calculated and the load patterns for 24 hours are drawn as shown in Fig. 5 and Fig. 6. Python programming is used to program CQ learning algorithm and various scenarios the agents encounters are analysed and the optimized solution is found to meet the long term objective of reducing the cost of energy consumption. Initially single agent Q Learning is used to optimize the solar PV systems individually in the department and the hostel. Then CQ Learning is used for distributed optimization of two solar power systems in dynamic environment. Comparison is made between CQ Learning with conventional distributed optimization method and individual agents optimizing independently with Q learning. CQ Learning proved to be better way of optimizing the solar micro-grid in a distributed environment. The explorer problem is solved using Monte Carlo method where the actions are chosen stochastically and averaged over 50 times. The yearly average values of utility of solar power, battery and the power consumption from the grid of the department are found as given below and simulated for ten years as shown in Fig. 7, Fig. 8 and Fig. 9. Three indicators $B_0$, $S_0$ and $P_g$ show the improvement in the performance of the micro-grid by using the Q Learning algorithms.

$$B_0 = \sum Battery\,to\,load \,/\, \sum Load\,(D_t) \quad (4)$$

$$S_0 = \sum Solar\,to\,Battery \,/\, \sum Solar\,Power\,(P_{sp}) \quad (5)$$

$$P_g = \sum Grid - \sum Battery\,to\,load \quad (6)$$

Fig. 10 shows the reduction in power from grid in the hostel. Fig. 11 and Fig. 12 show the utility of the battery and the solar power. The reduction in power consumption from the grid when both the units (department and the hostel) are operating independently, using Q learning algorithm is observed in Fig. 13. This is compared with power consumption from grid when the two units co-operate using Coordinated Q Learning (CQL). It is found from the Fig. 14 that in Coordinated Q Learning, the power consumed from grid is relatively reduced due to distributed optimization.

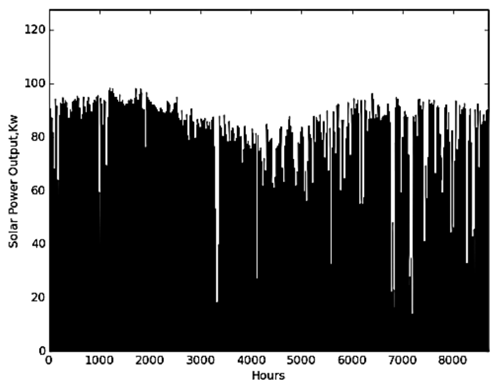


Fig. 3. Solar power for 200 KW unit

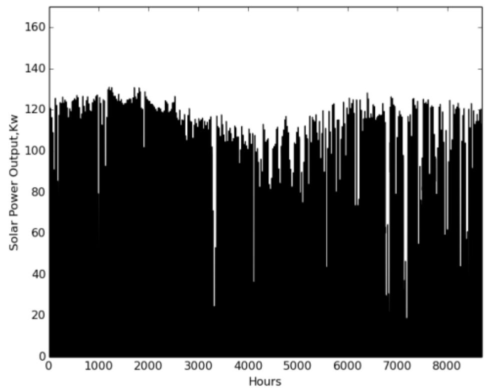


Fig. 4. Solar power for 150KW unit

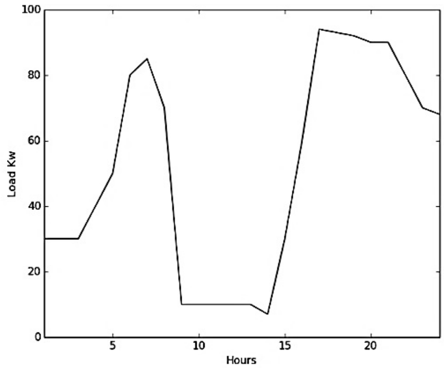


Fig. 5. Load pattern of Hostel

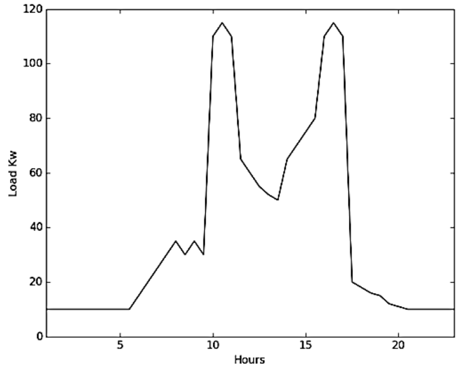


Fig. 6. Load pattern of Department

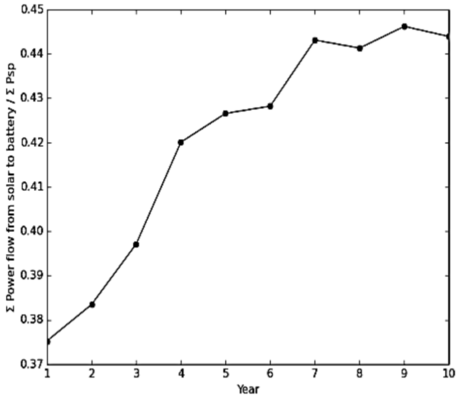


Fig. 7. Utility of the Solar power (200 KW unit)

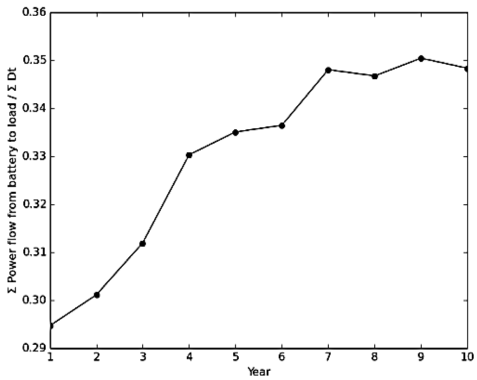


Fig. 8. Utility of the battery (200 KW unit)

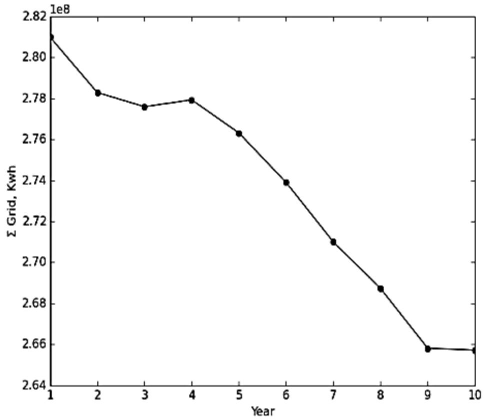


Fig.9. Power from grid with QL (Department)

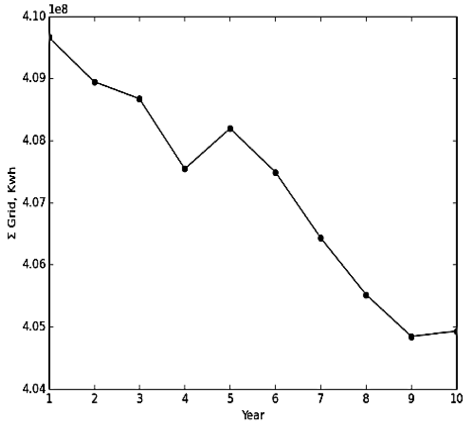


Fig. 10. Power from grid with QL (Hostel)

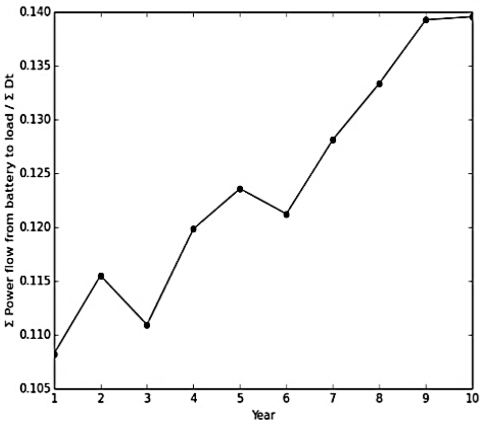


Fig.11. Utility of the battery with CQL

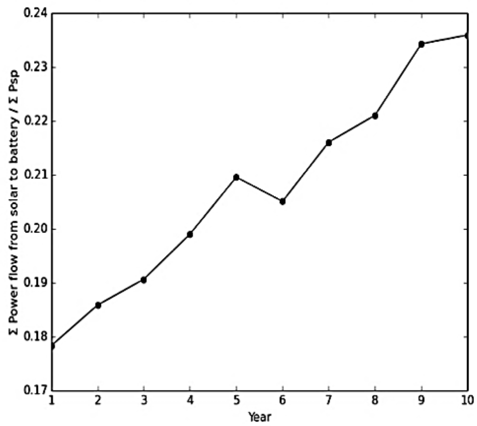


Fig.12. Utility of solar power unit with CQL

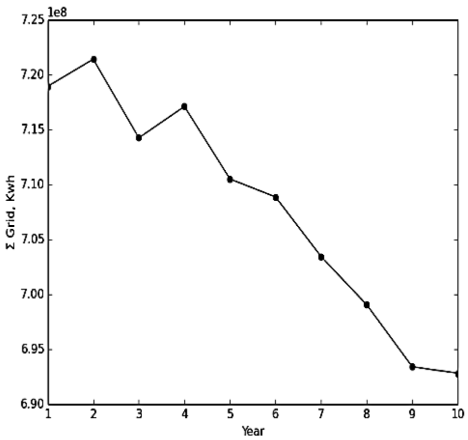


Fig.13. Power from grid with Individual QL

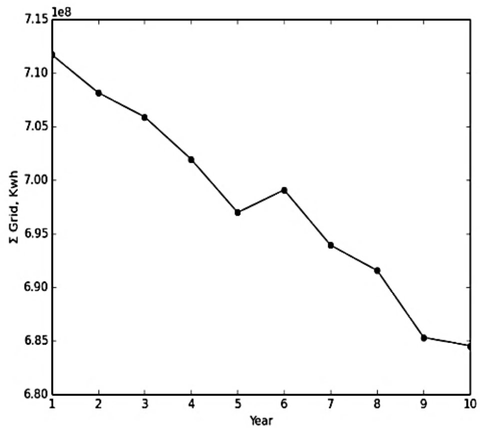


Fig. 14. Power from grid with CQL

## 11. Conclusion

The optimization of distributed energy management of solar micro-grids is done with a Multi-Agent Reinforcement Learning approach, namely Coordinated Q learning (CQ Learning) and the performance is compared with conventional distributed optimization method and single agent learning independently using Q learning. A simulation model was developed using python programming to prove that CQ learning achieves the lowest possible cost of power generation under intermittency of solar PV system and randomness of load. Thus the dynamic interactions between the consumer agent and its environment is carried out for autonomous optimization of battery scheduling to increase the utility of the battery, utility of solar power, thereby reduce the power consumption from grid in the long run. Uncertainties in the solar power generator due to the stochastic nature of the irradiance and temperature are accounted for. The proposed framework gives the intelligent consumer the ability to explore and understand the stochastic environment and reuse this experience for selecting the optimal energy management actions to reduce dependency on the grid in a distributed environment. Future work will focus on extension to multiple agents integrating diverse renewable generators (solar and wind) using Multi-Agent Reinforcement Learning with several intelligent consumers with conflicting requirements. Also, neural network can be used for generalisation of state space to reduce the complexity of the problem.

## References

1. Ross Guttromson, Steve Glover. The advanced micro-grid integration and interoperability. *Sandia National Laboratories, Sandia report*, March 2014.
2. Hatziargyriou ND. *European transactions on electrical power*. Special issue: Micro-grid and energy management, December 2010, p. 1139-141.
3. Reddy P P, Veloso M. Strategy learning for autonomous agents in smart grid market. *In Twenty-second international joint conference on artificial intelligence*; 2005, p. 1446-51.
4. Chen C, Duan S, Cai T, Liu B, Hu G. Smart energy management system for optimal micro-grid economic operation. *Renewable Power Generation*, IET 5(3); 2011, p. 258-67.
5. Mohamed F A, Koivo H N. System modelling and online optimal management of Micro-grid with battery storage. *International Journal on Electrical Power and Energy Systems*, 32(5); 2010, p. 398-407.
6. Colson C M, Nehrir M H, Pourmousavi S A. Towards real-time micro-grid power management using computational intelligence methods. *IEEE*; 2010, p. 1-8.
7. Abdirahman M Abdilahi, M W Mustafa, G Aliyu, J Usman. Autonomous Integrated Micro-grid (AIMG) System. *International Journal of Education and Research*, Vol.2, no.1; 2014, p.77-82.
8. Jun Z, Junfeng L, Jie W, and Ng H. A multi-agent solution to energy management in hybrid renewable energy generation system. *Renewable Energy*, vol. 36, no. 5; 2011, p.1352-63.
9. Aymen Chaouachi, Rashad M, Kamel M, Ridha Andoulsi, and Ken Nagasaka. Multi objective Intelligent Energy Management for a Micro-grid. IEEE Transactions, *Industrial Electronics,* 60(4); 2013, p. 1688-99.
10. Kuznetsova E, Li Y F, Ruiz C, Zio E, Ault G, and Bel L. K. Reinforcement learning for micro-grid energy management. *Energy Journal*, 59; 2013, p. 133-46.
11. Sutton R S, Barto A G. *Reinforcement Learning: An introduction*. London, England: MIT Press; 1998.
12. De Hauwere, Y.M., Vrancx, P, Now´e. Learning Multi-Agent State Space Representations. *In Proceedings of the 9th International Conference on Autonomous Agents and Multi-Agent Systems*. Toronto, Canada; 2010, p. 715–722.
13. Engin Karatepe, Mutlu Boztepe, Metin Cola. Neural network based solar cell model. *Energy Conversion and Management*, 47; 2006, p.1159-78.
14. Yujin Lim, Hak-Man Kim. Strategic bidding using reinforcement learning for load shedding in micro-grid. *Computers and Electrical Engineering*, Elsevier; 2014.
15. Marco Wiering , Martijn van Otterlo Ma. Reinforcement Learning: State- of-the- Art. New York: Springer; 2012.