Multi-use Energy Management Concept for PV Battery Storage Systems Based on Reinforcement Learning

Florus Härtel(✉) and Thilo Bocklisch

Technische Universität Dresden, Chair of Energy Storage Systems, Dresden, Germany
florus.haertel@tu-dresden.de

Abstract. This contribution introduces an energy management concept for multi-use applications of PV battery storage systems based on reinforcement learning (RL). The approach uses the state-of-the-art Proximal Policy Optimization algorithm in combination with recurrent Long Short-Term Memory networks to derive locally optimal energy management policies from a data-driven, simulation-based training procedure. For this purpose, an AC-coupled residential PV battery storage system is modelled and parametrized. Qualitative advantages of the RL-based approach compared to the commonly used model predictive control (MPC) approaches with regard to multi-use energy management applications, such as the ability to optimize a control policy over an infinite, discounted time horizon, are highlighted. From a large-scale training run of over 200 hyperparameter configurations, the five best energy management policies are selected and evaluated against state-of-the-art MPC and rule-based energy management concepts. In the evaluation over one year it is shown, that the energy management learned by the RL algorithm reduces curtailment losses from 5.70% to 4.78%, specific energy cost from 7.16 Cent kWh^{-1} to 7.09 Cent kWh^{-1} and increase the share of PV energy fed into the grid under a fixed feed-in limit from 49.95% to 50.99% compared to the MPC energy management, which is the second best one.

Keywords: PV Battery Storage System (PVBSS) · Energy Management · Multi-Use · Reinforcement Learning (RL) · Artificial Neural Networks (ANN) · Long Short-Term Memory (LSTM) · Proximal Policy Optimization (PPO)

1 Introduction

The rapid expansion of photovoltaic and wind power to meet the climate protection targets legally defined by the Federal Republic of Germany [1] is also driving the demand for stationary battery storage systems to compensate the volatility of renewable electricity sources and thereby ensure the stability of the grid [2, 3]. Intelligent energy management concepts are needed in order to take into account the techno-economic optimization criteria for the deployment of a battery storage system, such as low operating costs or a short amortisation period.

P. Schossig et al. (Eds.): IRES 2022, AHE 16, pp. 206–214, 2023.
https://doi.org/10.2991/978-94-6463-156-2_15

Reinforcement learning (RL) is a model-free method that is used to optimize a control policy by interacting with an environment – in this case the energy management is interacting with the PV battery storage system - and thereby receiving and maximizing rewards [4, 5]. RL offers a number of advantages over the widely used model predictive control (MPC) concepts [6–11], such as the ability to implicitly learn the system dynamics, without any prior knowledge of the system at hand. Furthermore, the policy can be optimized over an infinite time horizon a via the learned, state value function. No restrictions need to be imposed on the modelling of the environment, i.e. the PV battery storage system, or the target reward function as they are regarded as black-box functions by the RL algorithm. Practical limitations of the RL method have been a challenge for its application in real-world problem like the energy management of PV battery storage systems. These limitations include the non-guaranteed convergence of the learning algorithm to an optimal policy, the possible need for retraining if the change in system parameters becomes too large compared to the system used for training the policy or the sensibility to the hyperparameters of the respective algorithm [12, 13]. Advancements both in the field of RL methodology and machine learning hardware over last decade have led to the successful application of RL to a diversity of problems [14, 15] and therefore make it an interesting candidate to solve complex energy management problems, such as multi-use energy storage applications.

2 Reference System

The considered reference system is an AC-coupled PV battery storage system, consisting of a PV generator, cumulative loads, a battery storage and a connection to the power grid. Both the PV generator and the battery storage have their own inverter and they are coupled via the common AC grid of the building (s. Fig. 1).

In the simulation needed for the training and evaluation of the RL-based energy, the loads as well as the solar irradiation are simulated from measured time series provided by the HTW Berlin [16] and the Chair of Meteorology of the TU Dresden at the Tharandt weather station respectively. The PV battery storage system is simulated in a temporal resolution of one minute with control intervals of fifteen minutes. The battery storage itself is modelled with a constant conversion efficiency of 0.92 and a maximum inverter power of 4 kW for both the charging and discharging direction as well as a storage capacity of 7 kWh. An installed PV peak power of 5 kWp, an average yearly consumption of 4500 kWh, resulting from the input time series, and a feed-in limitation of 50% of the peak PV power is assumed [17]. The cost for energy drawn from the grid is set to 0.32 € kWh^{-1} and the feed-in tariff is 0.08 € kWh^{-1}.

3 Reinforcement Learning Based Energy Management Concept

3.1 Basics of Reinforcement Learning

RL optimizes a policy by maximizing the rewards received from the environment. This is achieved by interaction of the RL agent with the environment following the current policy and subsequent parameter optimization of the policy over the set of collected state transitions. This interaction loop is the basis of all RL algorithms (s. Fig. 2).

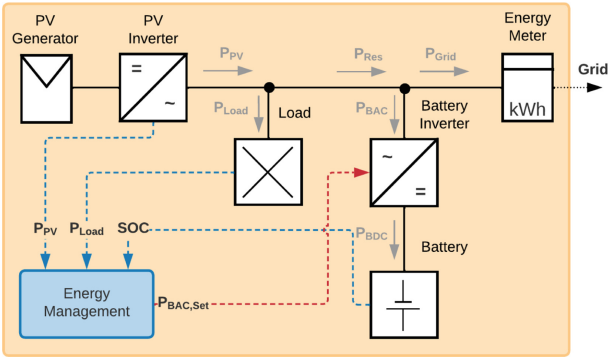

Fig. 1. Modelled PV battery storage system with energy management, input values for the energy management (blue) and output values (red).

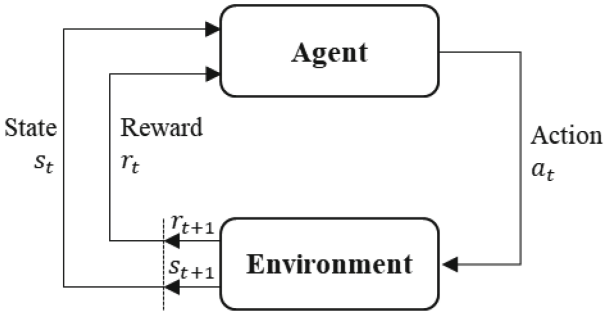

Fig. 2. Interaction loop between the RL agent and the environment [4].

The agent is the controllable part of a RL problem, optimized in the process: The policy of the energy management $\pi(s_t)$ as well as the auxiliary state value function $V(s_t)$, needed for the RL algorithm. The environment, on the other hand, is the part of the RL problem whose system dynamics and reward structure are to be exploited, both of which are regarded as black box by the RL algorithm.

3.2 State, Action and Reward

In order to formulate a RL problem from the given energy management task, three core values need to be defined first: the state vector s_t describing the state of the PV battery system at each time step, the action vector a_t containing all the setpoints to be executed by the system in the next time step and the reward scalar r_t used as optimization criterion for the RL algorithm.

In case of the PV battery system under consideration the state vector s_t contains the state of charge of the battery storage SOC_t, the energy generated by the PV generator

$E_{PV,t}$ and the energy consumed by the loads in the previous time step $E_{L,t}$.

$$s_t = \begin{bmatrix} SOC_t \\ E_{Load,t} \\ E_{PV,t} \end{bmatrix} \tag{1}$$

The action vector a_t contains the reference value for the grid power of the PV battery storage system, which the energy management is then trying to hold by charging and discharging the battery.

$$a_t = \begin{bmatrix} \frac{P_{BAC,set,t}}{P_{BAC,max}} \end{bmatrix} \tag{2}$$

The reward r_t is the sum of the income from energy fed into the grid minus the costs of energy drawn from the grid.

$$r_t = E_{Feed,t} p_{Feed,t} - E_{Draw,t} p_{Draw,t} \tag{3}$$

Since the reward depends on the input time series of load and solar irradiation and the RL agent has no knowledge about these system dynamics, they are regarded as stochastic processes and therefore the reward is regarded as stochastic function of the current policy, too.

$$r_t \sim R(\cdot|\pi) \tag{4}$$

3.3 Algorithm and Neural Network Topology

While iterating the RL interaction loop (s. Fig. 2) the generated transition tuples (s_t, a_t, r_t, s_{t+1}) are stored. A multitude of different RL algorithms exist to optimize the parameterized policy $\pi(s_t)$ from these collected transitions. The state-of-the-art Proximal Policy Optimization (PPO) algorithm is chosen as RL algorithm, which has been demonstrated to reach faster convergence and produce better policies than other RL algorithms in a variety of continuous control problems [18, 19].

Artificial neural networks (ANN) are utilized as parametrized functions for the policy $\pi(s_t)$, which maps states s_t to actionsa_t, and the auxiliary state value function$V(s_t)$, which is needed for the PPO algorithm in order to estimate the value of a given states_t. To exploit temporal information from past load and PV power data for the energy management, a recurrent Long Short-Term Memory (LSTM) network [20] is shared between the policy $\pi(s_t)$ the state value function $V(s_t)$ (s. Fig. 3). This so-called state encoder reduces the temporal dimension of a series of past state observations $o_{\leq t}$ to a vector of useful features – the encoded states_t^*.

The prediction targets of the state value function – the returns R_t - are defined as the discounted sum of rewards over an infinite time horizon.

$$R_t = \sum_{i=0}^{\infty} r_{t+i} \gamma^i \tag{5}$$

This infinite geometric series can be truncated by recursively estimating the value of the terminal state of the collected transitions $V(s_T)$.

$$R_t = \sum_{i=0}^{T-1} r_{t+i} \gamma^i + V(s_T) \tag{6}$$

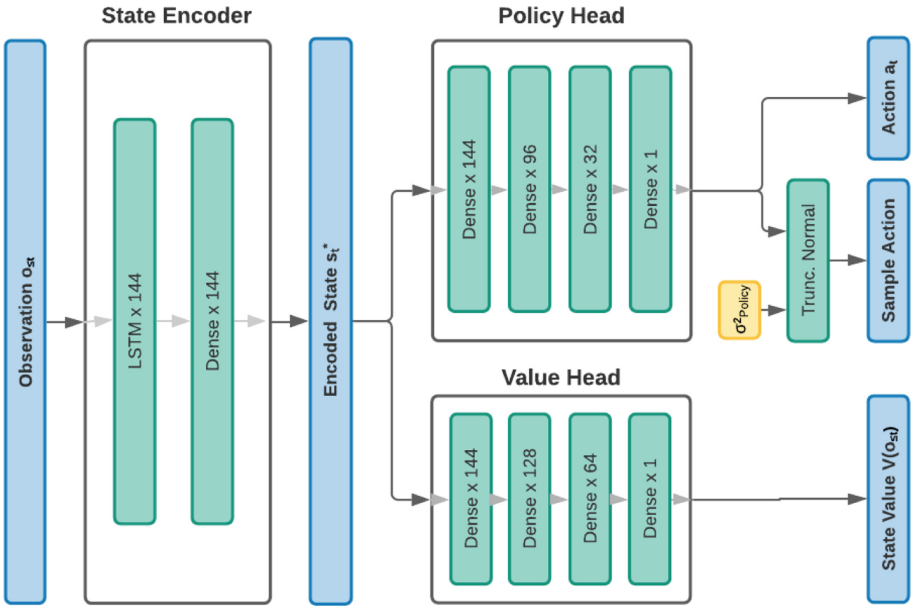

Fig. 3. Neural network topology of the policy $\pi(s_t)$ and state value function $V(s_t)$. Both share the LSTM state encoder in their computation path, which is used to reduce the temporal dimension and extract meaningful features from a series of past states.

The discounting factor $\gamma \in [0, 1)$ controls how quick the exponential weights are decaying and thus how far future rewards are regarded for the policy optimization. This ability to estimate the value of states over an infinite time horizon sets apart the RL-based concept from MPC-based concepts: It allows the energy management policy to be optimized for reward signals that may occur many time steps in the future. In multi-use energy management applications this might be used to minimize power demand fees, which are only accounted a few times per year, or for minimizing battery aging.

4 Evaluation Results

After the training of the policy is finished, the converged policy can be used for evaluation. For this purpose, separate time series of the load and solar radiation are held back from the training dataset and form separate evaluation episodes with a length of one year. A simple priority-based (PRIO) [21, 22] and a MPC-based peak shaving energy management (MPC-PS) [17, 22] are used as reference energy managements for the evaluation. The amounts of energy of the load E_{Load}, the PV generator E_{PV}, the PV curtailment losses E_{CL}, energy drawn from the grid E_{Draw} and fed into the grid E_{Feed} for each time step are summed over the whole year.

$$E_x = \sum_t E_{x,t} \tag{7}$$

The degree of self-sufficiency k_{SS}, the self-consumption k_{SC}, the relative curtailment losses k_{CL}, the share of PV energy fed into the grid k_{FI} and the specific energy costs k_{SEC} are used as evaluation criteria.

$$k_{SS} = \frac{E_{Load} - E_{Draw}}{E_{Load}} \tag{8}$$

$$k_{SC} = \frac{E_{PV} - E_{Feed} - E_{CL}}{E_{PV}} \tag{9}$$

$$k_{CL} = \frac{E_{CL}}{E_{PV}} \tag{10}$$

$$k_{FI} = \frac{E_{Feed}}{E_{PV}} \tag{11}$$

$$k_{SEC} = \frac{p_{Draw} E_{Draw} - p_{Feed} E_{Feed}}{E_{Load}} \tag{12}$$

In a large-scale training run prior to the evaluation, over 200 policies were learned with different hyperparameters [23], which serve as tuning parameters for the RL algorithm. The five best performing policies from this training RL- $<$ 1–5 $>$ are evaluated against the references MPC-PS and PRIO.

The analysis of evaluation episode 1 (s. Table 1), which has a cumulative energy consumption E_{Load} of 4210 kWh and PV generation E_{PV} of 6104 kWh, shows the RL-based energy management RL-3 beats both PRIO and MPC-PS in terms of reduced curtailment losses k_{CL} and increased share of PV energy fed into the grid k_{FI} under the fixed feed-in limit. However, the degree of self-sufficiency k_{SS} and self-consumption k_{SC} is slightly worse, where PRIO performs best, as expected. The increased energy fed into the grid however causes the specific energy price k_{SEC} to be the lowest with RL-3. The evaluation suggests, that the RL-based concept is able to optimize for the multi-use target of maximizing self-sufficiency, while minimizing curtailment losses. It is important to point out that all of this is possible only by defining the energy cost as reward and thus as sole signal for the policy optimization. The RL agent does not have any knowledge about the dynamics of the PV battery storage or the reward structure prior to training.

5 Summary and Outlook

In this contribution an energy management concept for PV battery storage systems in multi-use applications based on RL was introduced. It can provide a number of qualitative advantages over the commonly used MPC concepts, such as learning a locally optimal policy without any prior assumptions about the system dynamics or the ability to optimize over an infinite time horizon via the infinite geometric series of discounted rewards approximated by the state value function.

An AC-coupled PV battery storage system was modelled and a residential use case was defined. The evaluation results suggest that the RL-based concept exceeds the performance of the MPC-PS and PRIO reference energy managements in terms of reduced curtailment losses k_{CL}, specific energy costs k_{SEC} and increased share of PV energy fed into the grid k_{FI} under a fixed feed-in limit.

Table 1. Comparative metrics for the evaluation of episode 1.

Energy Management	k_{SS} (%)	k_{SC} (%)	k_{CL} (%)	k_{FI} (%)	k_{SEC} $\left(\frac{\text{Cent}}{\text{kWh}}\right)$
PRIO	**59.98**	**44.73**	9.87	45.39	7.54
MPC-PS	59.51	44.35	5.70	49.95	7.16
RL-1	59.20	44.10	5.08	50.82	7.16
RL-2	59.18	44.08	5.09	50.82	7.17
RL-3	59.36	44.23	**4.78**	**50.99**	**7.09**
RL-4	59.43	44.28	5.18	50.54	7.12
RL-5	59.57	44.40	5.69	49.91	7.15

The ability of regarding a very long time horizon for the policy optimization via the auxiliary state value function makes RL a promising candidate for multi-use energy management applications with long-term optimization objectives, such as minimization of power demand fees or battery aging. The inclusion of these shares of cost as well as time-variable energy tariffs are currently investigated. Another interesting extension is the application of this concept for hybrid energy storage systems [24] with more than one degree of freedom, which has also been conceptualized.

Acknowledgement. The presented paper is based on the results of the research project "HYBAT – Hybrid lithium-ion battery storage solution with 1500 V system technology, innovative thermal management and optimized system management", supported by the Federal Ministry of economic affairs and Climate action (funding code: 03EI3009C).

The authors want to thank the Centre for Information Services and High Performance Computing [Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH)] Technische Universität Dresden for providing its facilities for high throughput calculations. Also we want to thank the Chair of Meteorology Technische Universität Dresden for providing irradiation data for our simulations.

References

1. Deutscher Bundestag, Bundes-Klimaschutzgesetz: KSG, 2021.
2. M. Sterner, I. Stadler (Eds.), Energiespeicher - Bedarf, Technologien, Integration, 2nd ed., Springer Berlin Heidelberg; Imprint; Springer Vieweg, Berlin, Heidelberg, 2017. ISBN: 978–3–662–48893–5.
3. P. Sterchele, J. Brandes, J. Heilig, D. Wrede, C. Kost, T. Schlegl, A. Bett, H.-M. Henning, Wege zu einem klimaneutralen Energiesystem: Die deutsche Energiewende im Kontext gesellschaftlicher Verhaltensweisen –Update für ein CO2-Reduktionziel von 65% in 2030 und 100% in 2050 (2020).
4. R.S. Sutton, A.G. Barto, Reinforcement learning: an introduction, Second edition, The MIT Press, Cambridge, Massachusetts, 2018. ISBN: 978-0-262-03924-6
5. I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT Press, Cambridge (EE. UU.), London, 2016. ISBN: 9780262035613.
6. A. Nottrott, J. Kleissl, B. Washom, Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems, Renewable Energy 55 (2013) 230–240. https://doi.org/https://doi.org/10.1016/j.renene.2012.12.036.
7. R. Hanna, J. Kleissl, A. Nottrott, M. Ferry, Energy dispatch schedule optimization for demand charge reduction using a photovoltaic-battery storage system with solar forecasting, Solar Energy 103 (2014) 269–287. https://doi.org/https://doi.org/10.1016/j.solener.2014.02.020.
8. J. Moshövel, K.-P. Kairies, D. Magnor, M. Leuthold, M. Bost, S. Gährs, E. Szczechowicz, M. Cramer, D.U. Sauer, Analysis of the maximal possible grid relief from PV-peak-power impacts by using storage systems for increased self-consumption, Applied Energy 137 (2015) 567–575. https://doi.org/https://doi.org/10.1016/j.apenergy.2014.07.021.
9. T. Bocklisch, M. Paulitschke, M. Böttiger, Investigation of Energy Management Concepts for a Hybrid Battery Storage and Power-To-Heat Device for Renewable Energy Applications, TU Dresden, 2016, p. 1.
10. R. Gelleschus, M. Böttiger, T. Bocklisch, Optimization-Based Control Concept with Feed-in and Demand Peak Shaving for a PV Battery Heat Pump Heat Storage System, Energies 12 (2019) 2098. https://doi.org/https://doi.org/10.3390/en12112098.
11. M. Böttiger, M. Paulitschke, T. Bocklisch, Innovative Reactive Energy Management for a Photovoltaic Battery System, Energy Procedia 99 (2016) 341–349. https://doi.org/https://doi.org/10.1016/j.egypro.2016.10.124.
12. M. McCloskey, N.J. Cohen, Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem, in: G.H. Bower (Ed.), Psychology of Learning and Motivation, Academic Press, 1989, pp. 109–165.
13. K. Khetarpal, M. Riemer, I. Rish, D. Precup, Towards Continual Reinforcement Learning: A Review and Perspectives (2020).
14. Z. Zhang, D. Zhang, R. C. Qiu, Deep reinforcement learning for power system: An overview, CSEE Journal of Power and Energy Systems 6 (2019) 213–225. https://doi.org/10.17775/CSEEJPES.2019.00920.
15. Y. Li, Deep Reinforcement Learning: An Overview (2017). https://doi.org/10.48550/arXiv.1701.07274.
16. Tjarko Tjaden, Joseph Bergner, Johannes Weniger, Volker Quaschning, Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis (2015). https://doi.org/10.13140/RG.2.1.5112.0080/1.
17. J. Bergner, J. Weniger, T. Tjaden, V. Quaschning, Feed-in Power Limitation of Grid-Connected PV Battery Systems with Autonomous Forecast-Based Operation Strategies, 29th European Photovoltaic Solar Energy Conference and Exhibition; 2363–2370 (2014) 8 pages, 4237 kb. https://doi.org/10.4229/EUPVSEC20142014-5CO.15.1.

18. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, arXiv:1707.06347 [cs] (2017).
19. M. Hessel, J. Modayil, H. van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, D. Silver, Rainbow: Combining Improvements in Deep Reinforcement Learning (2017).
20. S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9 (1997) 1735–1780. https://doi.org/https://doi.org/10.1162/neco.1997.9.8.1735.
21. T. Weitzel, C.H. Glock, Energy management for stationary electric energy storage systems: A systematic literature review, European Journal of Operational Research 264 (2018) 582–606. https://doi.org/https://doi.org/10.1016/j.ejor.2017.06.052.
22. M. Böttiger, Multikriteriell optimierendes Betriebsführungsverfahren für PV-Batteriespeichersysteme. Dissertation, Dresden, 2019.
23. M. Andrychowicz, A. Raichuk, P. Stańczyk, M. Orsini, S. Girgin, R. Marinier, L. Hussenot, M. Geist, O. Pietquin, M. Michalski, S. Gelly, O. Bachem, What Matters In On-Policy Reinforcement Learning? A Large-Scale Empirical Study (2020). https://doi.org/10.48550/arXiv.2006.05990.
24. T. Bocklisch, Hybrid energy storage approach for renewable energy applications, Journal of Energy Storage 8 (2016) 311–319. https://doi.org/https://doi.org/10.1016/j.est.2016.01.004.