Intelligent Power Management in Modern Automotive Power Distribution Systems Based on Reinforcement Learning

Michael Gerten, On-board Systems Lab, TU Dortmund University, Dortmund, michael.gerten@tu-dortmund.de
Kilian Medger, On-board Systems Lab, TU Dortmund University, Dortmund, kilian.medger@tu-dortmund.de
Stephan Frei, On-board Systems Lab, TU Dortmund University, Dortmund, stephan.frei@tu-dortmund.de

Abstract

To ensure functional safety in highly automated vehicles, critical undervoltages of safety-relevant components must be prevented. Undervoltages can occur because of voltage drops resulting from large currents within the distribution system. While larger cross-section areas of the wiring harness could prevent critical undervoltages, they are undesirable because of the added weight and cost. In this contribution, intelligent power management is presented as an alternative method to stabilize the supply voltage in modern reconfigurable distribution systems. By using reinforcement learning, a neural network learns to actively control the power flow within the system. This way, temporary undervoltages can successfully be prevented.

1 Introduction

In highly automated vehicles, voltage stability of the power distribution system is increasingly important. Functional safety must be ensured at all times, critical components shall not experience undervoltage. [1]
Significant undervoltages usually occur in case of high load currents within the system that result in a large voltage drop across the supply wires. In the worst case, such voltage drops may temporarily violate the required minimum supply voltage of a critical component, as exemplarily depicted in Figure 1. To improve the voltage stability of the system and prevent these critical voltage drops, wire cross-section areas can be increased [2]. However, the additional material results in increased cost and weight of the wiring harness, which is undesirable. Stabilizing the voltage during these critical time spans could alternatively be achieved by power management. This might include temporarily disabling comfort loads or redirecting the power flow in meshed or ring topologies.

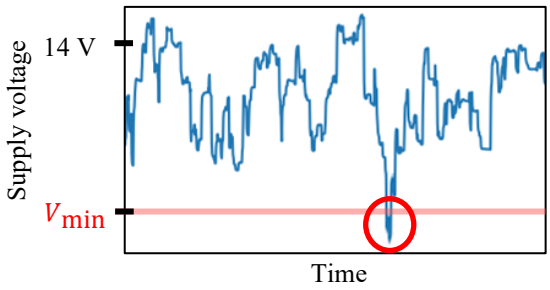

Figure 1: Exemplary supply voltage curve with temporary undervoltage

Commonly, simple rule-based power management schemes are proposed. For example, a non-safety-relevant load is programmed to limit its power consumption if its supply voltage drops below a specific threshold [3]. In [4], loads are assigned to different power reduction levels; if the battery is discharged excessively, the power consumption is limited according to these reduction levels. As the goal here was to prevent a depletion of the battery, load voltage stability was not directly considered.
Such static thresholds, however, might require large safety margins and are not trivial to define. Also, because they do not consider the actual impact on the voltage stability and interactions between different components, they can be ineffective [5].
A more intelligent, system-oriented operating strategy could be realized using machine learning. Currently, machine learning is rarely used in this context. [6] proposes a reinforcement learning approach to optimize the state of charge of the low-voltage battery. Many works also focus on machine learning-based energy management in hybrid electric vehicles (for example [7]), but they focus on efficiency and range of the drive system.
In this contribution, a power management scheme is developed using machine learning that optimizes the voltage stability. As depicted in Figure 2, the developed approach uses the reconfigurability of modern automotive distribution systems. Instead of conventional melting fuses, electronic semiconductor fuses (eFuses) are increasingly used for protection. These offer an active switching and current measurement capability. To improve the voltage stability, uncritical loads can be selectively switched off by the eFuse if necessary, or the power flow can be redirected, for example by opening a segment of a ring structure (for redundant supply). An artificial neural network shall interpret the current state of the distribution system (represented by the measured eFuse currents) and chose an appropriate switching configuration of the system. Reinforcement learning is especially suited for such a task, because it doesn't require labelled training data and learns an operating strategy by simply interacting with a simulation model of the distribution system.
The paper is structured as follows: In Section 2, the fundamentals of reinforcement learning and the used proximal policy algorithm are explained. The developed power management approach is presented in Section 3. It is applied to an exemplary supply system in Section 4 and the results are analysed. Section 5 concludes the paper.

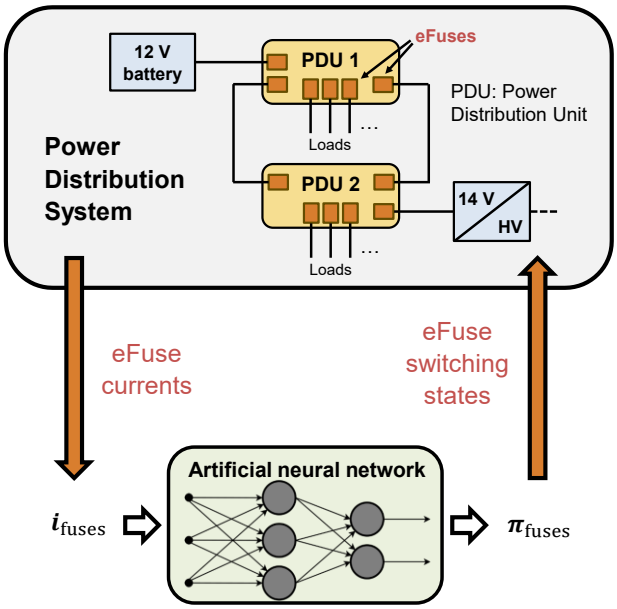

Figure 2: Management of an automotive power distribution system using an artificial neural network

2 Reinforcement Learning

In this section, the fundamentals of reinforcement learning are introduced and the specific algorithm used in this contribution is explained.

2.1 Basic Concepts

The basic idea of reinforcement learning is depicted in Figure 3. During training, an agent interacts with the environment (for example an automotive power distribution system) to learn an optimized operating strategy or policy π. Specifically, the agent performs an action a on the environment. As a result of this action, the environment transitions to the state s. Additionally, a reward r is calculated that evaluates the current state. This way, desirable actions can be encouraged (positive reward) and undesirable actions can be discouraged (negative reward). Based on this feedback, the agent updates the policy to perform better actions in the future. The policy $\pi(a|s)$ can for example be represented by an artificial neural network that determines the next action. In this case, the input of the network is the current state s, and the output is the possibility that action a is performed next. [8]

While different reinforcement learning algorithms exist, they share some basic concepts. To not only focus on the immediate reward of the next step, the return is defined as the sum of all future rewards. Furthermore, the discounted return R weights the future rewards. With the discount factor $\gamma < 1$, rewards are decreasingly taken into account the further they lie in the future [8]:

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots \quad (1)$$

The index denotes the time step, with t being the current time step.

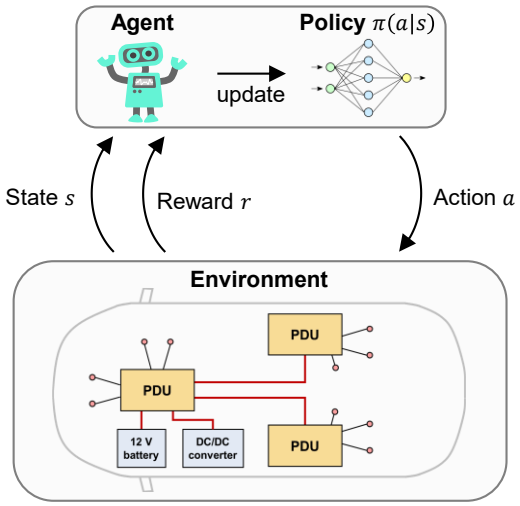

Figure 3: General overview of reinforcement learning

Furthermore, the following quantities / functions are defined to evaluate states and actions [8]:

- Value function $V(s)$: the value describes the discounted return that is to be expected when using the current policy from state s. In other words, it describes how "good" the current state s is.
- Action value function $Q(s|a)$: the action value describes the expected discounted return, if the environment is in state s and the action a is performed. In other words: how good is the action a?
- Advantage function $A(s|a) = Q(s|a) - V(s)$: the advantage quantifies, how much *additional* (positive or negative) return a specific action a yields in state s compared to the current policy.

2.2 Value-based vs. policy-based methods

Different reinforcement learning methods can generally be categorized as value-based or policy-based. Value-based methods, e.g., Deep Q-learning [9], use the neural network to approximate the action value Q. After training, a greedy policy is used: in each state, the action with the highest approximated Q value is performed.

Policy-based approaches, e.g., proximal policy optimization [10], the policy $\pi(a|s)$ is learned directly by the neural network. With the input being the current state, the outputs of the neural network are the probabilities of each possible action. This allows continuous actions and multiple actions to be performed simultaneously. [8]

For the investigated problem of intelligent power management, the action space needs to address the switching state of all N eFuses of the system independently. Therefore, 2^N discrete actions are needed to allow all possible switching configurations. As policy-based methods can perform multiple actions simultaneously, only N output neurons are necessary, reducing the complexity of the neural network. As a state-of-the-art policy-based reinforcement learning method, proximal policy optimization (PPO) [10] is used.

2.3 Proximal Policy Optimization

In PPO, two neural networks are used during training. One is called the actor network and outputs the policy, i.e., the probability of each action a. The other network is called the critic and approximates the value function $V(s)$, which is only used during the training process. While having separate outputs, both networks can share common layers. [8]
At the beginning of a training iteration, a batch of T steps is performed in the environment using the current policy. The parameters θ of the neural networks are then updated depending on the observed feedback. For this optimization, cost functions are defined. The cost function of the actor network is given in equation (2) and shall be maximized.

$$J_\theta^{\text{actor}} = \mathrm{E}_t\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} A_t\right] \tag{2}$$

$\pi_\theta(a_t|s_t)$ denotes the probability of action a_t with the new network parameters and $\pi_{\theta_{\text{old}}}(a_t|s_t)$ the probability with the old ones. The advantage A_t is estimated based on the actually received return R_t and the estimated value $V(s)$ of the critic network. The E denotes the expectation that averages over multiple acquired state transitions. In other words, by maximizing J_θ^{actor}, actions with a positive advantage are made more likely, while actions with a negative reward are made less likely. To prevent too large parameter changes within one optimization step, however, the probability ration $\pi_\theta/\pi_{\theta_{\text{old}}}$ is clipped within the interval $[1-\epsilon, 1+\epsilon]$. [10]
For the loss function of the value net (critic), a quadratic error between the predicted value $V_\theta(s_t)$ and the actually received return R_t can be used:

$$J_\theta^{\text{critic}} = \mathrm{E}_t[(V_\theta(s_t) - R_t)^2] \tag{3}$$

Finally, the network parameters are optimized. partial derivatives of the cost functions with regard to the network parameters are calculated using backpropagation, and an optimization method like gradient descent is applied to update the parameters. [8]

3 Intelligent Power Management

PPO is applied to learn an intelligent operating strategy. As presented, the voltage stability shall be optimized by identifying a suitable eFuse switching configuration.
As an environment for training and testing, a simulation model is used, which is built from existing modelling approaches [11]. For simplicity, only dynamic load behaviour is considered, with time constants in the milliseconds range. Higher frequency effects in the microsecond range or below, like switching transients or faults, are neglected. As machine learning frameworks, PyTorch and the reinforcement learning extension TorchRL [12] are used within Python. The simulation model is implemented in MATLAB/Simscape, compiled into a shared library and then directly called from Python. Equidistant time steps are used in the simulation.
The state vector $\boldsymbol{s}_t$ that is returned by the environment and processed by the neural network consists of the current and the last sample of all the N eFuse currents. The load current sums $i_{\text{sum,PDU}}$ of all M PDUs are included as an additional feature:

$$\begin{aligned}\boldsymbol{s}_t = \boldsymbol{i}_{\text{fuses}} = \big[&i_{\text{fuse},1}(t), \dots, i_{\text{fuse},N}(t),\\ &i_{\text{fuse},1}(t-1), \dots, i_{\text{fuse},N}(t-1),\\ &i_{\text{sum,PDU},1}(t), \dots, i_{\text{sum,PDU},M}(t)\big]\end{aligned} \tag{4}$$

The actor network maps this input data to the action probabilities $\boldsymbol{\pi}_{\text{fuses}}$ (refer to equation (5)**)**. For example, $a_{\text{fuse},1}$ denotes the action of switching fuse 1. If the network output $\pi(a_{\text{fuse},1})$ is larger than 0.5, fuse 1 is switched on. If $\pi(a_{\text{fuse},1}) \leq 0.5$, the fuse is switched off.

$$\boldsymbol{\pi}_{\text{fuses}} = \big[\pi(a_{\text{fuse},1}), \dots, \pi(a_{\text{fuse},N})\big] \tag{5}$$

The structure of the neural network may be adapted depending of the size of the distribution system and, hence, the complexity of the problem. For this contribution, one hidden layer with a total of 256 neurons is used.
For the quality of the learned strategy, the training procedure and the received reward are pivotal. After each time step, a reward is returned that incentivizes stable supply voltages and penalizes undervoltages and the deactivation of individual loads. To achieve this, the total reward r consists of several parts. First, a bonus B is rewarded, if all safety-critical loads experience a sufficient supply voltage:

$$B = \begin{cases} 2, & n_{\text{crit,uv}} = 0 \\ 0, & \text{else} \end{cases} \tag{6}$$

$n_{\text{crit,uv}}$ denotes the number of safety-critical loads experiencing an undervoltage. If undervoltages occur, an undervoltage penalty P_{uv} is determined. This penalty scales with the amount of the undervoltage:

$$P_{\text{uv}} = \sum_{i=1}^{n_{\text{crit,uv}}} 1 + 2 \cdot V_i^{\text{crit,uv}} + \sum_{j=1}^{n_{\text{comfort,uv}}} V_j^{\text{comfort,uv}} \tag{7}$$

$n_{\text{comfort,uv}}$ denotes the number of comfort loads experiencing an undervoltage. The vectors $\boldsymbol{V}^{\text{crit,uv}}$ and $\boldsymbol{V}^{\text{comfort,uv}}$ contain the voltages by which the undervoltage thresholds are exceeded for the critical and the comfort loads, respectively. Furthermore, a penalty is given if eFuses were actively switched off:

$$P_{\text{switch}} = n_{\text{switch}} \tag{8}$$

Here, n_{switch} is the number of loads switched-off during the last action. Finally, a penalty is given based on the total number of deactivated loads n_{off}:

$$P_{\text{off}} = 0.2 \cdot n_{\text{off}} \tag{9}$$

n_{off} being the number of all currently switched-off loads. All in all, this leads to the total reward r:

$$r = B - P_{\text{uv}} - P_{\text{switch}} - P_{\text{off}} \tag{10}$$

This way, the agent receives direct feedback after every step and the current reward only depends on the previous two actions and the state of the loads. To focus on this immediate feedback during training, a small discount parameter γ is beneficial, e.g., 0.2.
For the training procedure, a curriculum learning approach [13] is used. The general idea of curriculum learning is that

the neural network shall first be confronted with an easier task that progressively gets harder.
In this case, the network is first trained on a small number of static load configurations. This way, the algorithm has enough time to learn the basic operating principles. Afterwards, dynamic scenarios with realistic load behaviour are introduced.

4 Demonstration

The proposed method is now demonstrated on an exemplary power distribution system. First, this system is introduced, then the PPO training procedure is presented and finally, the results are examined.

4.1 Investigated system

For demonstration, the simplified distribution system depicted in Figure 4 is used. It consists of two power distribution units (PDUs) in a zonal architecture, which is increasingly used and discussed for modern distribution systems [14, 15]. The PDUs are connected in a redundant ring topology and are supplied by two sources, i.e., a battery and a DC/DC converter. The supply wires and all loads are protected and actively switchable by individual eFuses.
A total of ten loads is considered for each PDU. For simplification, the functions of the loads are not further specified, nominal currents are given in Table 1. As a voltage stability criterion, a minimum supply voltage of 9 V is assumed for all loads. For PDU 1, all supplied loads with even numbers are considered safety-relevant, while for PDU 2 the odd-numbered loads are safety-relevant.
Based on the given nominal currents, a load dynamic is assumed. For simplicity, all loads are considered to have equal dynamic characteristics according to the probability density depicted in Figure 5. Currents smaller than or equal to the nominal current I_{nom} have the largest probabilities. Larger currents up to $5 \cdot I_{\mathrm{nom}}$ occur less frequently with a linearly decreasing probability.
An exemplary load current profile resulting from this behaviour is depicted in Figure 6 for a load with $I_{\mathrm{nom}} = 50$ A.

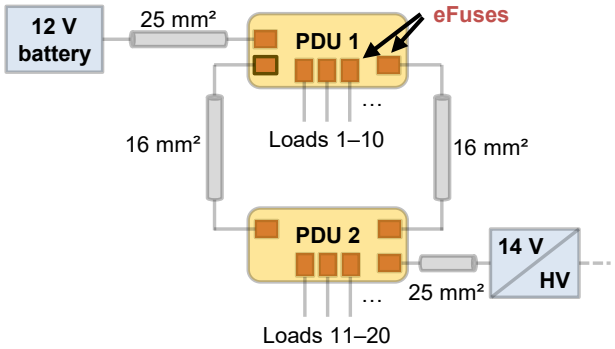

Figure 4: Exemplary power distribution system

The load current is randomly changed between piecewise constant values according to the presented probability density. A new constant current value is approached exponentially, as also shown in Figure 6 (bottom). This low-pass behaviour represents the resistive-capacitive input characteristic of a typical load because of its stabilizing capacitance [16].

Table 1: Load parameterization of exemplary system

Load	Nominal current	Load wire cross section
1, 11	50 A	10 mm²
2, 12	40 A	6 mm²
3, 13	30 A	4 mm²
4, 14	20 A	2.5 mm²
5, 15	10 A	1 mm²
6, 16	5 A	0.5 mm²
7, 17	1 A	0.35 mm²
8, 18	0.5 A	0.35 mm²
9, 19	0.1 A	0.35 mm²
10, 20	0.1 A	0.35 mm²

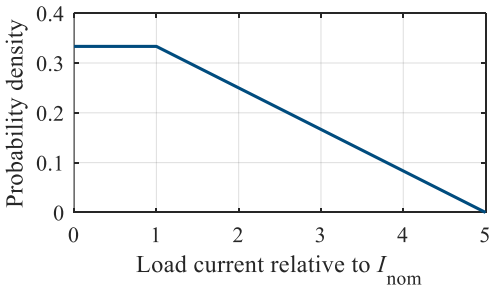

Figure 5: Assumed probability density for load current distribution

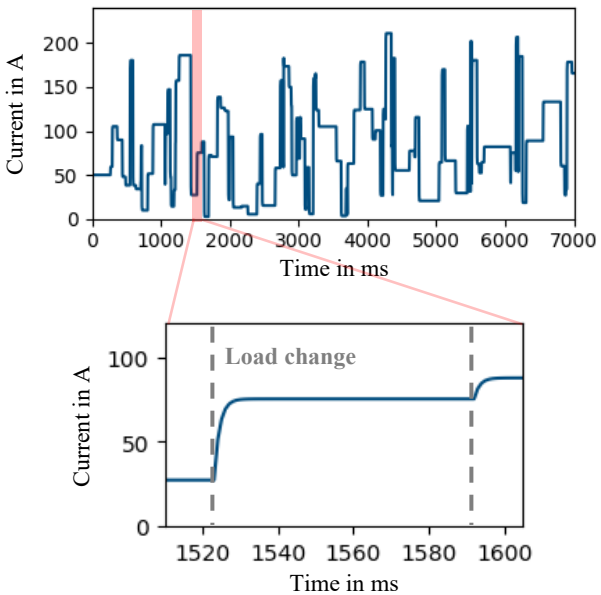

Figure 6: Current profile of exemplary load with low-pass behaviour during load current change

First, the described distribution system is simulated without any power management. Here, and in the following investigations, a simulation time step of 1 ms is used. The resulting supply voltages of safety-relevant loads over a span of 1 million simulation steps are depicted in Figure 7. Without power management, during about 2,4 % of the

time safety-relevant loads experience a critical undervoltage. To confirm that the system is theoretically stabilizable by only operating the eFuses, another simulation is performed without any comfort loads for the same number of steps. As Figure 7 (b) shows, this increases the supply voltage of the safety-relevant loads to uncritical values above 9 V. An intelligent operating strategy should now find an optimum between these two extremes; safety-relevant loads need to be stabilized, but comfort functions should only be deactivated if unavoidable.

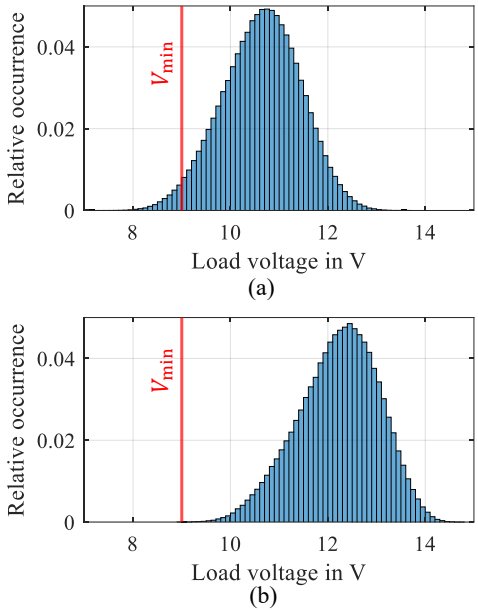

Figure 7: Supply voltage distribution of safety-relevant loads in case of (a) no power management and (b) without comfort loads

4.2 Training

The training is performed with the described PPO algorithm. The used training parameter are depicted in Table 2. As explained, the network is first trained on static system configurations without dynamic load behaviour. Additionally, critical scenarios are emphasised during training, so the network learns how to intervene to optimize the voltage stability. After 50 episodes, load dynamic is introduced. The resulting reward during this training procedure is depicted in Figure 8. Starting with a negative average reward due to the randomly initialized neural network weights, it quickly rises and then slowly converges to almost 2 at the end of the training.

Table 2: PPO training parameters

Parameter	Value
Discount factor γ	0.2
GAE parameter λ	0.6
Entropy factor	0.05
Initial learning rate	0.0003
Batch size	10.000
Sub-batch size	1.000
No. epochs	6
Total no. of steps	10.000.000

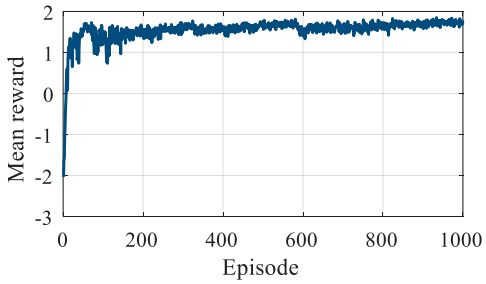

Figure 8: Average reward after each training batch

4.3 Results

To demonstrate the trained neural network, the environment is now nominally operated as described in Section 4.1. In Figure 9, an exemplary interval is depicted that shows the learned intelligent operating strategy. All loads that change their current consumption during the depicted time period are shown. The voltage of the safety-critical load 6 fluctuates as a result of the load current changes. At about 350 ms, load 1 significantly increases its current consumption. Without load management, this would result in a critical undervoltage of load 6 (dotted black line). However, the neural network intervenes accordingly and switches off the comfort load 12, as also depicted in Figure 9. This action stabilized the supply voltage of load 6. All other loads remain switched on. Figure 10 depicts all supply voltages of safety-relevant loads for a larger interval of 1.5 s. Without the neural network (blue), several critical undervoltages occur during this interval, with activated neural network, those scenarios are stabilized (orange).
All in all, the network successfully learned to efficiently stabilize the distribution system. This can also be seen in the resulting supply voltage histogram, depicted in Figure 11. The critical undervoltages are prevented, while a comfort load availability of 97.7 % could be achieved using the intelligent operating strategy.

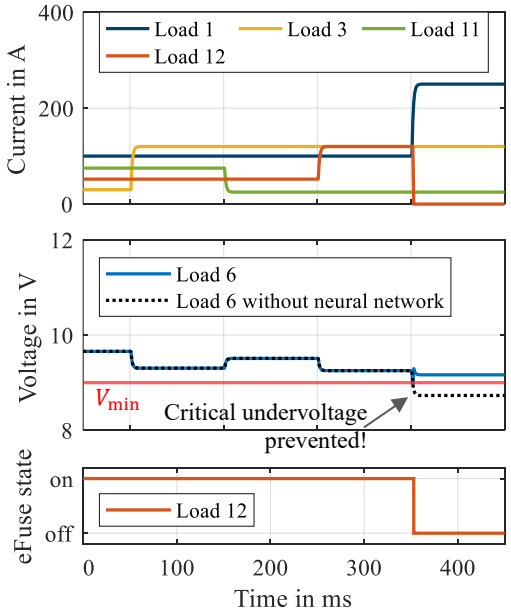

Figure 9: Exemplary prevention of critical undervoltage by learned operating strategy. Selected currents (top), voltage of safety-relevant load 6 with and without neural network (middle) and fuse state of comfort load 12 (bottom)

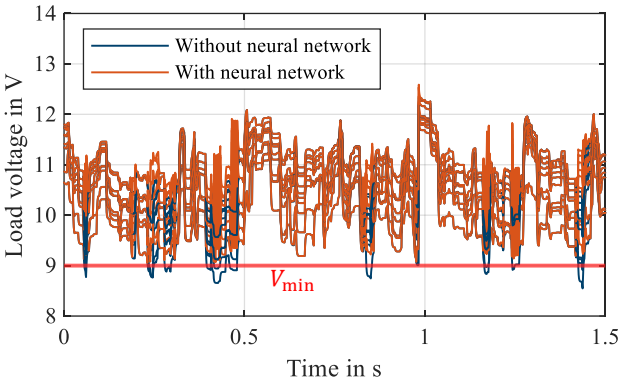

Figure 10: Exemplary interval of supply voltages of all safety-relevant loads with and without neural network

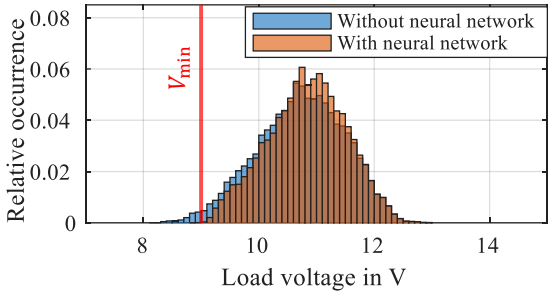

Figure 11: Supply voltage distribution of safety-relevant loads with and without neural network control

5 Conclusion

In this paper, a power management method has been developed that stabilizes the supply voltage of safety-relevant components in modern automotive distribution systems. The approach utilizes the reconfigurability of eFuse-based power systems and actively prevents critical undervoltages by selective temporary deactivation of comfort loads.

This operating strategy is realized by a neural network that processes the measured eFuse currents and then decides if one or more eFuses need to be switched. This behaviour is learned through reinforcement learning, where the algorithm interacts with a simulation model of the distribution system. In a simulative demonstration, the approach effectively and selectively prevents undervoltages of safety-critical loads, while also maximizing the availability of the comfort functions. Larger wire cross-section areas are therefore not needed to stabilize the supply system. In future work, the method can be further optimized and applied to a real vehicular system. Optimizations could include prioritisation of individual comfort loads during training and integration of real load current profiles.

Acknowledgement

This work presented in this paper was supported by the German Federal Ministry of Research, Technology and Space (BMFTR) as part of the project KI4BoardNet under Grant 16ME0779. The responsibility for this publication is held by the authors only.

References

[1] P. Kilian *et al.,* "Principle Guidelines for Safe Power Supply Systems Development," *IEEE Access*, vol. 9, pp. 107751–107766, 2021, doi: 10.1109/ACCESS.2021.3100711.

[2] F. Ruf *et al.,* "Topology and Design Optimization of a 14 V Automotive Power Net Using a Modified Discrete PSO in a Physical Simulation," in *2013 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Beijing, China, 2013.

[3] J. Fröschl and O. Sirch, *Bordnetze und E/E-Architektur: Eine Einführung in die Zusammenhänge zwischen Elektrik/Elektronik-Architektur und Energiebordnetz im Automobil,* 1st ed. Tübingen: Expert Verlag, 2023.

[4] R. M. Fabis, *Beitrag zum Energiemanagement in Kfz-Bordnetzen*: Dissertation, Technische Universität Berlin, 2006.

[5] T. P. Kohler, R. Gehring, J. Froeschl, D. Buecherl, and H.-G. Herzog, "Voltage Stability Analysis of Automotive Power Nets based on Modeling and Experimental Results," in *New Trends and Developments in Automotive System Engineering*: IntechOpen, 2011.

[6] Andreas Heimrath, Joachim Froeschl, Razieh Rezaei, Martin Lamprecht, and Uwe Baumgarten, "Reflex-Augmented Reinforcement Learning for Operating Strategies in Automotive Electrical Energy Management," in *2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, London, UK, 2019.

[7] T. Liu, Y. Zou, D. Liu, and F. Sun, "Reinforcement Learning of Adaptive Energy Management With Transition Probability for a Hybrid Electric Tracked Vehicle," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7837–7846, 2015, doi: 10.1109/TIE.2015.2475419.

[8] M. Lapan, *Deep Reinforcement Learning Hands-On,* 3rd ed. Birmingham: Packt Publishing Limited, 2024.

[9] V. Mnih *et al.,* "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[10] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Jul. 2017. [Online]. Available: http://arxiv.org/pdf/1707.06347v2

[11] M. Gerten, M. Rübartsch, and S. Frei, "Models of Automotive Power Supply Components for the Transient Analysis of Switching Events and Faults," in *AmE - Automotive meets Electronics 2022*, Dortmund, Germany, 2022.

[12] A. Bou *et al.,* "TorchRL: A data-driven decision-making library for PyTorch," Jun. 2023. [Online]. Available: http://arxiv.org/pdf/2306.00577v2

[13] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey," *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1–50, 2020.

[14] T. Liebetrau, "E/E Architecture Transformation How it impacts value chain and networking technologies," in vol. 104, *AmE 2022 - Automotive meets Electronics; 13. GMM-Symposium*, Dortmund, Germany, 2022.

[15] V. Bandur, G. Selim, V. Pantelic, and M. Lawford, "Making the Case for Centralized Automotive E/E Architectures," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1230–1245, 2021, doi: 10.1109/TVT.2021.3054934.

[16] M. S. Bechteler, C. M. Schessl, and T. F. Bechteler, "Electrical Power Net Systems in Cars—Impedance Modeling and Measurement," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1148–1155, 2010, doi: 10.1109/TVT.2009.2037886.