Hochschule Konstanz
Fakultät Elektrotechnik
und Informationstechnik

Reinforcement Learning zur Regelung eines Batteriespeichersystems unter Unsicherheit

Masterprojekt von
Chin-I Feng

Hochschule Konstanz
Fakultät Elektrotechnik
und Informationstechnik

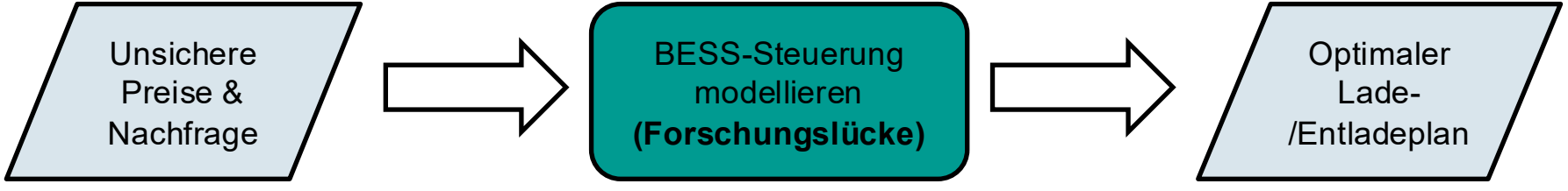
Agenda

- Einführung
- Simulationsumgebung
- Methoden
- Datengrundlage
- Ergebnisse
- Vergleich
- Fazit

Einführung

Problemstellung

- Hohe Preis- und Nachfragevolatilität erschwert die optimale Steuerung von Batteriespeichersystems (BESS).
- Klassische Ansätze können zukünftige Risiken nur unzureichend vorhersagen.
- Es fehlt ein Modell, das Preisunsicherheit, Nachfrageunsicherheit und Batteriedegradation gleichzeitig berücksichtigt.

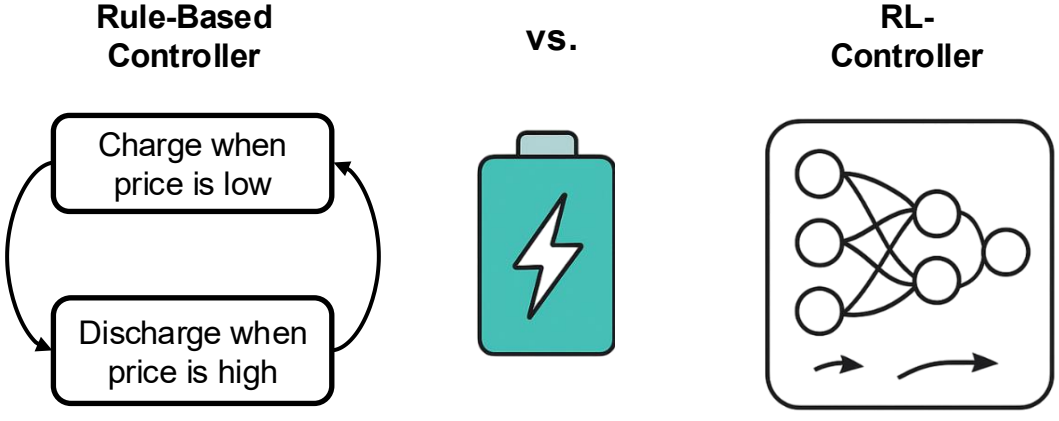


Einführung

Zielsetzung

- Entwicklung einer RL-basierten Simulationsumgebung für ein netzgekoppeltes BESS zur Analyse von Arbitrage und Peak-Shaving unter Marktunsicherheit.

- Vergleich von unsicherheitsbewussten RL (Distributional RL) mit klassischen und regelbasierten Strategien.



Simulationsumgebung

State, Action

State:

- SoC (State of Charge)
- SoH (State of Health)
- Temporal Features (Tageszeit, Jahreszeit)
- Letzte Action
- Preis-Features (aktueller Preis + 3h Prognose)
- Nachfrage-Features (aktuelle Nachfrage + 3h Prognose)

Action:

- Diskrete oder Kontinuierlich: Lade-/Entladeleistung $\in$ [-10 kW,+10kW]

Simulationsumgebung

Reward, Transition, Episode

Reward:

- Belohnung = Betriebsökonomischer Beitrag (Marktgewinn bzw. Energiekosten)
  − Degradationskosten
  − Lastüberschreitungsstrafe (nur im Peak-Shaving-Szenario)

Transition:

- SoC/SoH Modell (physikalisches Batteriemodell + zyklische Alterung)

Episode:

- Eine Woche Simulation (15-minütige Zeitschritte)
- Nach einer Woche wird die Episode beendet und das Environment zurückgesetzt.

Methoden

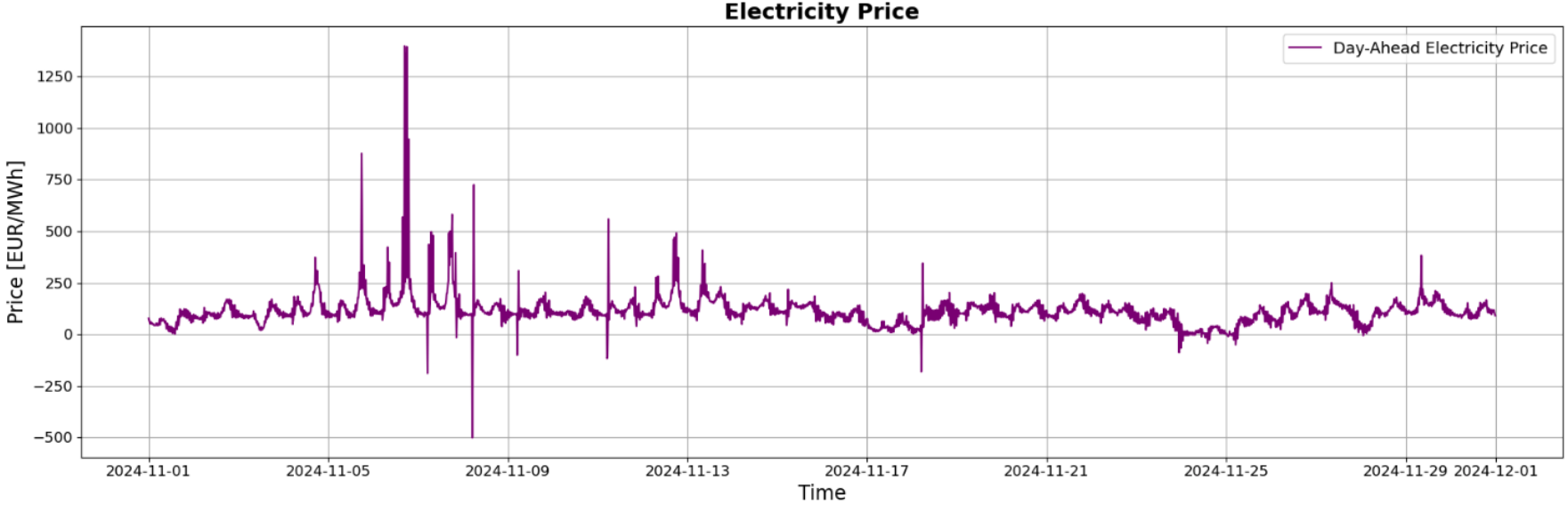
Methode	Eigenschaften
Rule-based	**Baseline** (kontinuierliche Aktionen), feste Entscheidungsregeln → if…else…
DQN	**Standard RL** (diskrete Aktionen), lernt den Erwartungswert der Returns → $\mathrm{E}[G_t]$
TD3	**Standard RL** (kontinuierliche Aktionen), lernt den Erwartungswert der Returns → $\mathrm{E}[G_t]$
QR-DQN	**Distributional RL** (diskrete Aktionen), lernt die Verteilung der Returns→ $\mathrm{Z}[G_t]$

Return G_t: gesamte zukünftige angesammelte Belohnung ab Zeitpunkt t

Hochschule Konstanz
Fakultät Elektrotechnik
und Informationstechnik

Datengrundlage

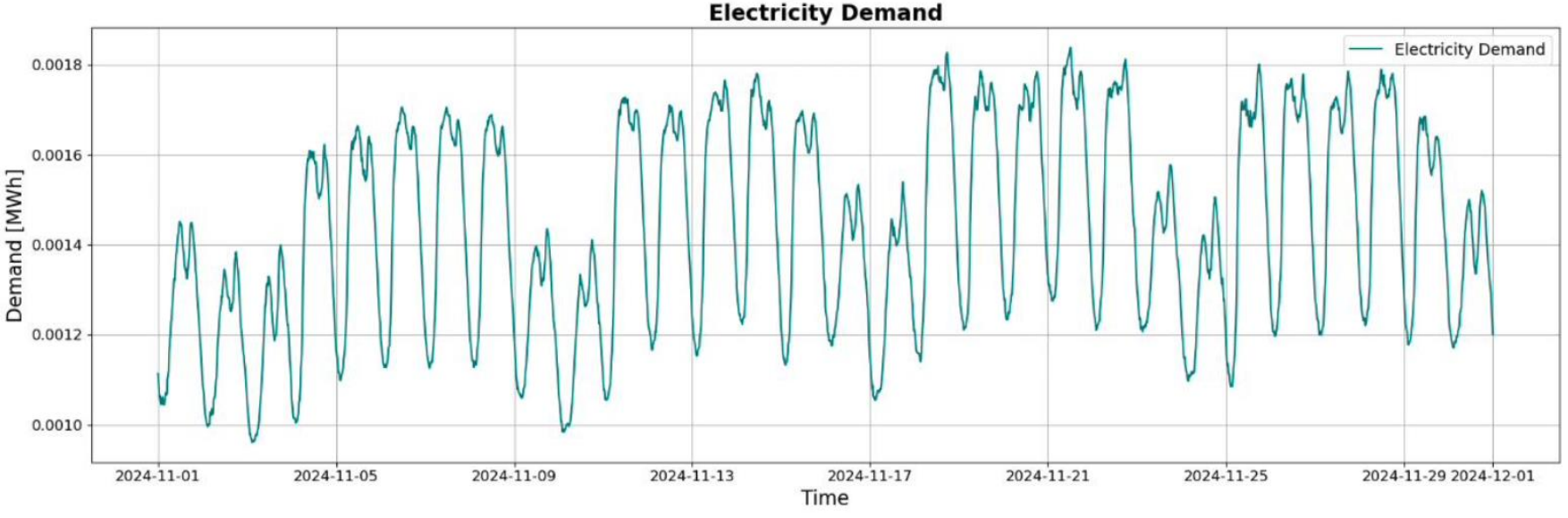
Strompreisdaten



Hochschule Konstanz
Fakultät Elektrotechnik
und Informationstechnik

Datengrundlage

Lastdaten

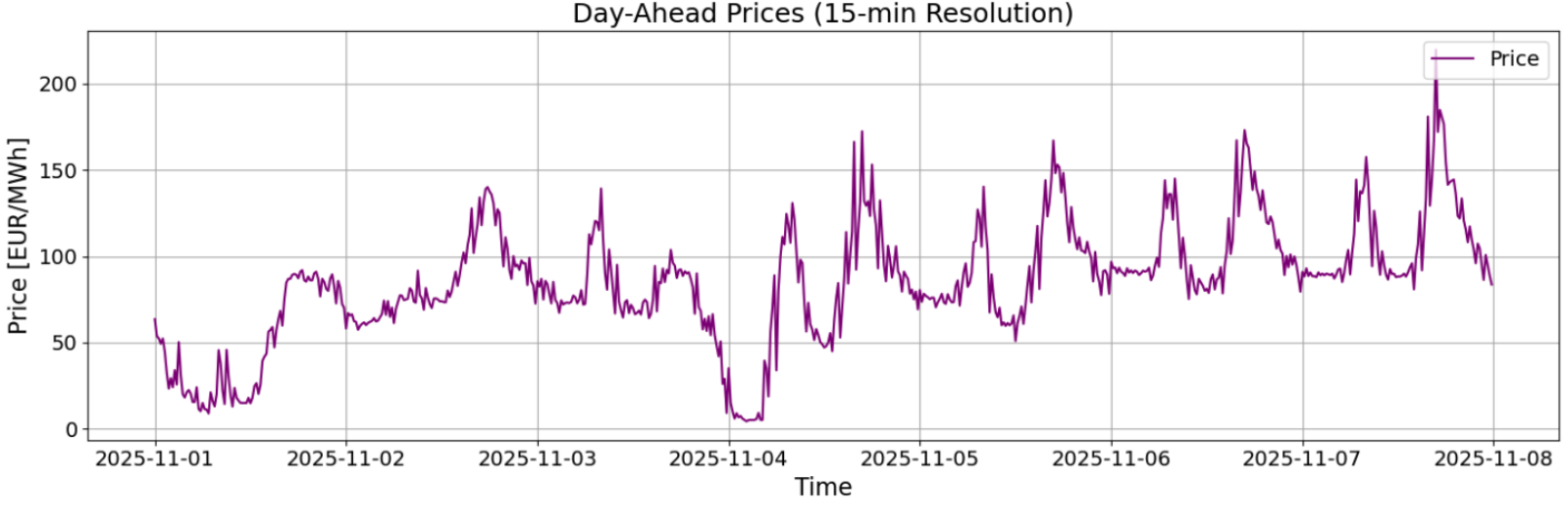


Datenquelle: SMARD (Bundesnetzagentur)

Datengrundlage

Training und Test

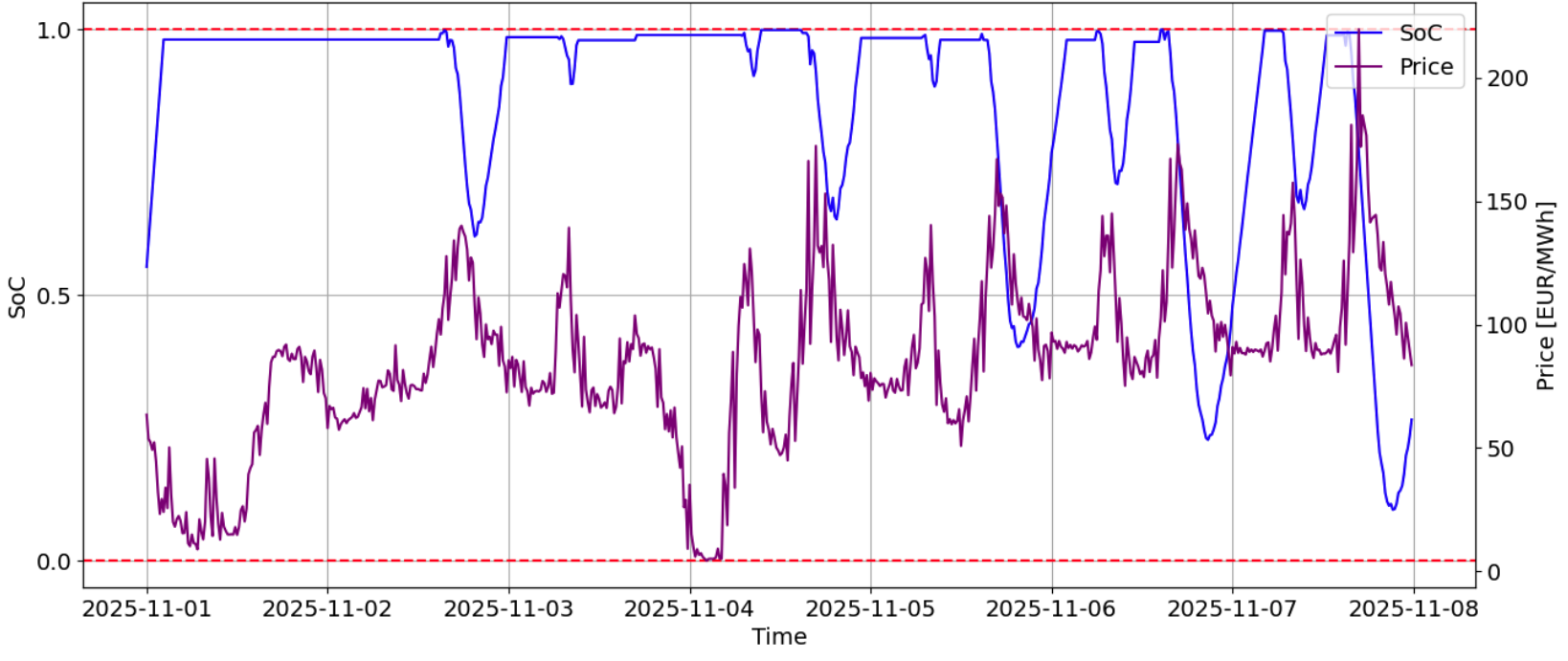
- **Trainingsdaten:** gesamter November 2024 (15-min Auflösung)
- **Testdaten:** erste Woche November 2025 (15-min Auflösung)
- **Ziel:** Evaluierung der Generalisierungsfähigkeit auf unbekannte Marktbedingungen



Datenquelle: Energy-Charts (Fraunhofer ISE)

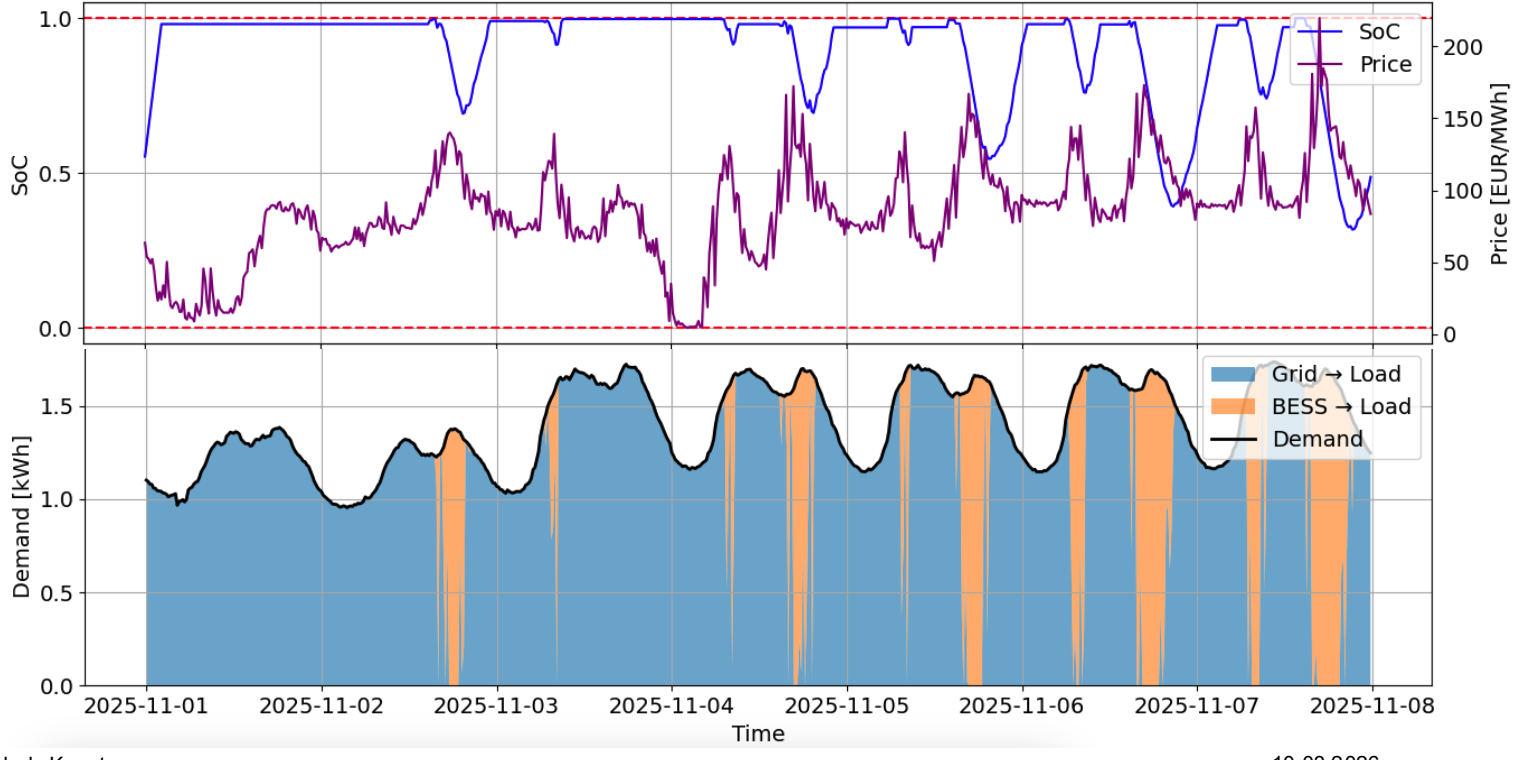
Ergebnisse

Rule-based (Arbitrage)



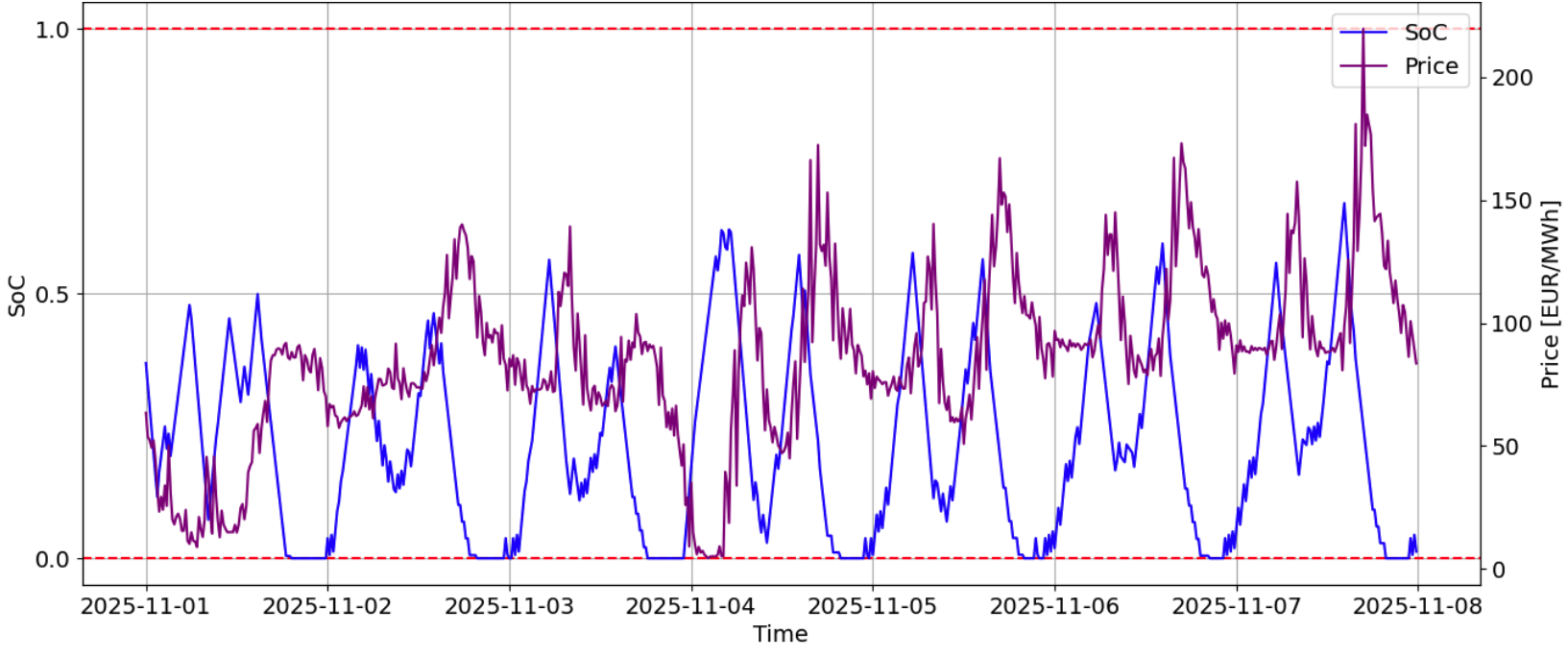
Ergebnisse

Rule-based (Peak-Shaving)



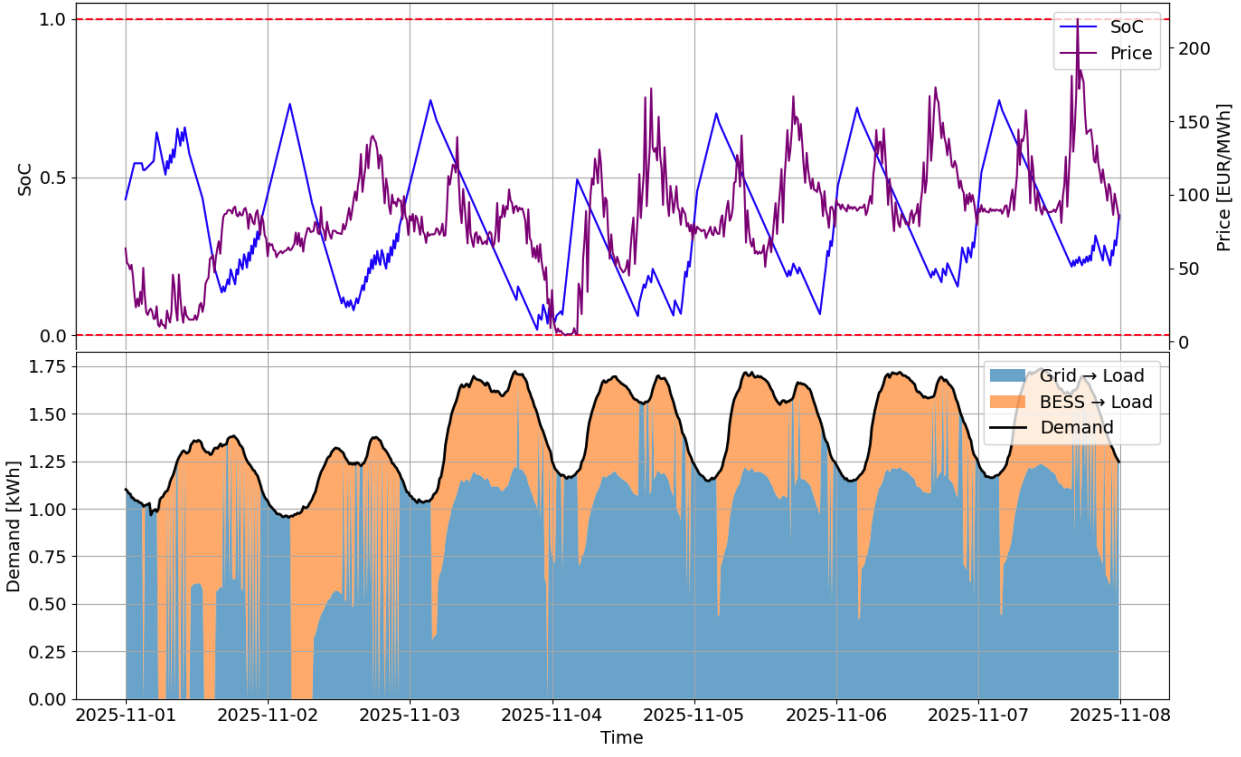
Ergebnisse

DQN (Arbitrage)



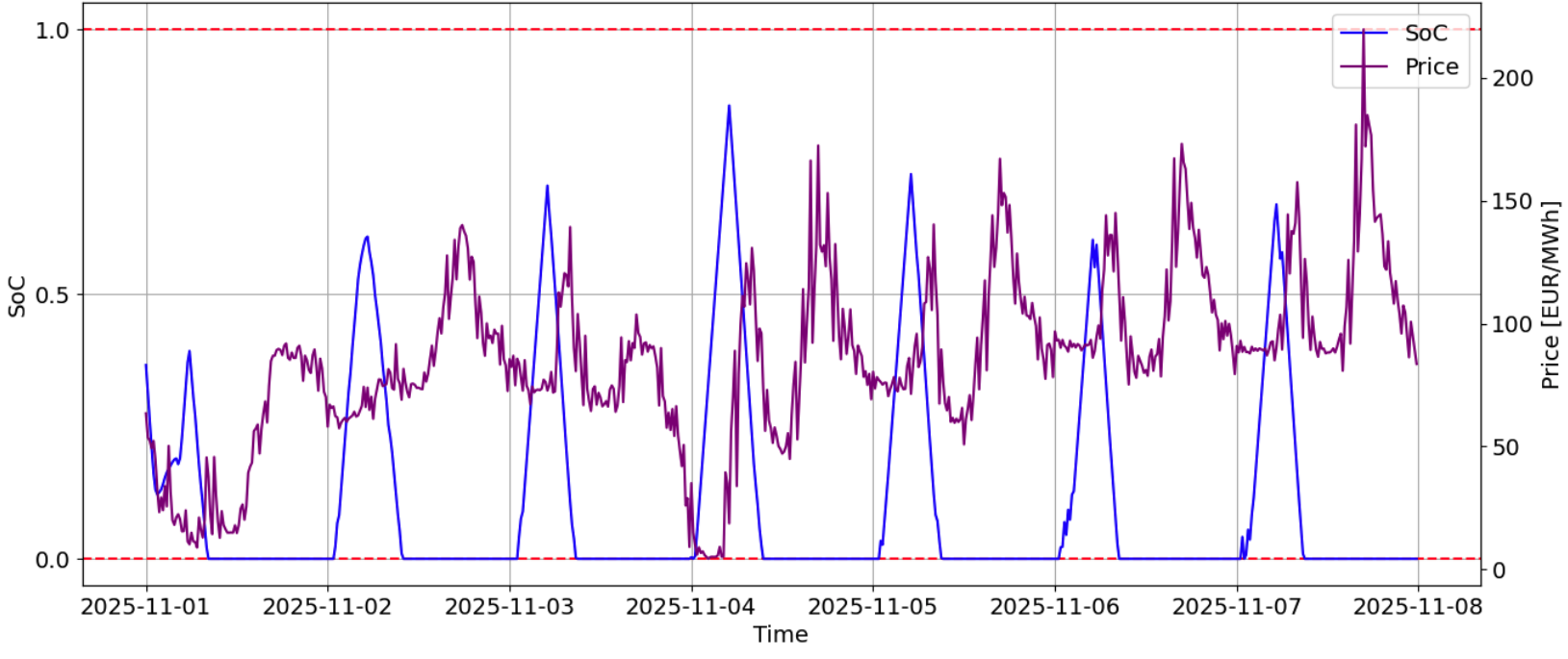
Ergebnisse

DQN (Peak-Shaving)



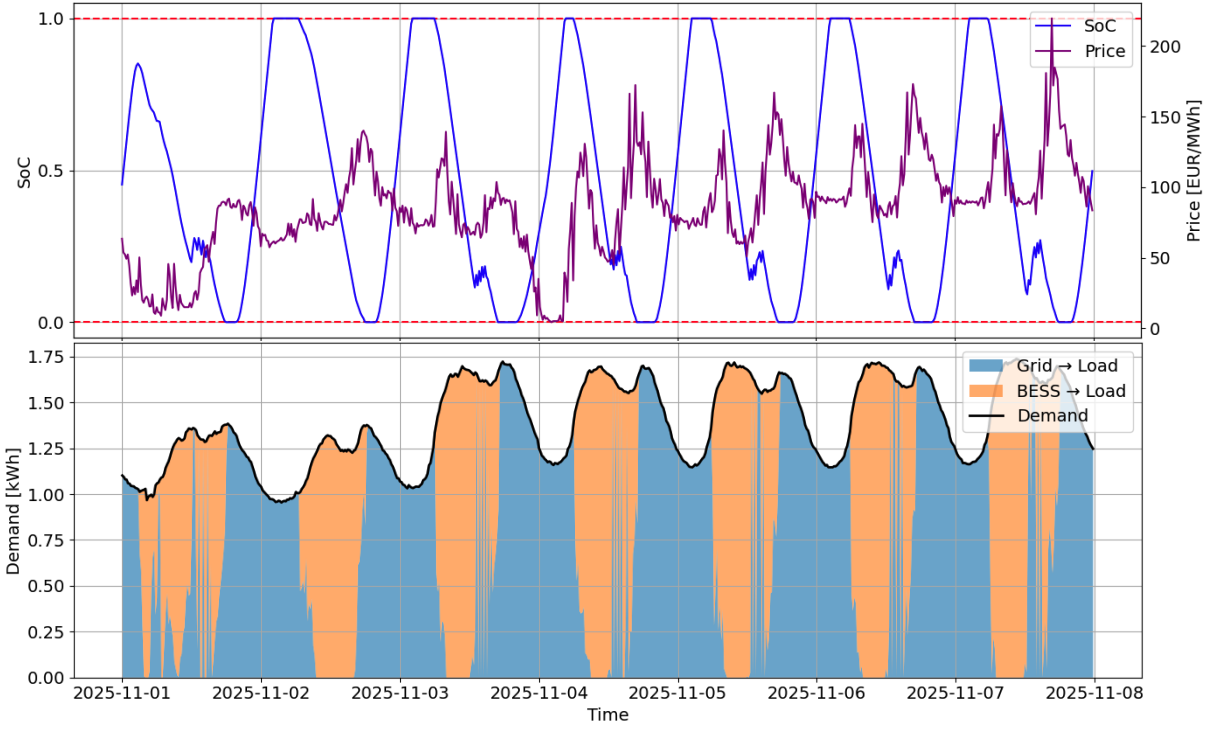
Ergebnisse

TD3 (Arbitrage)



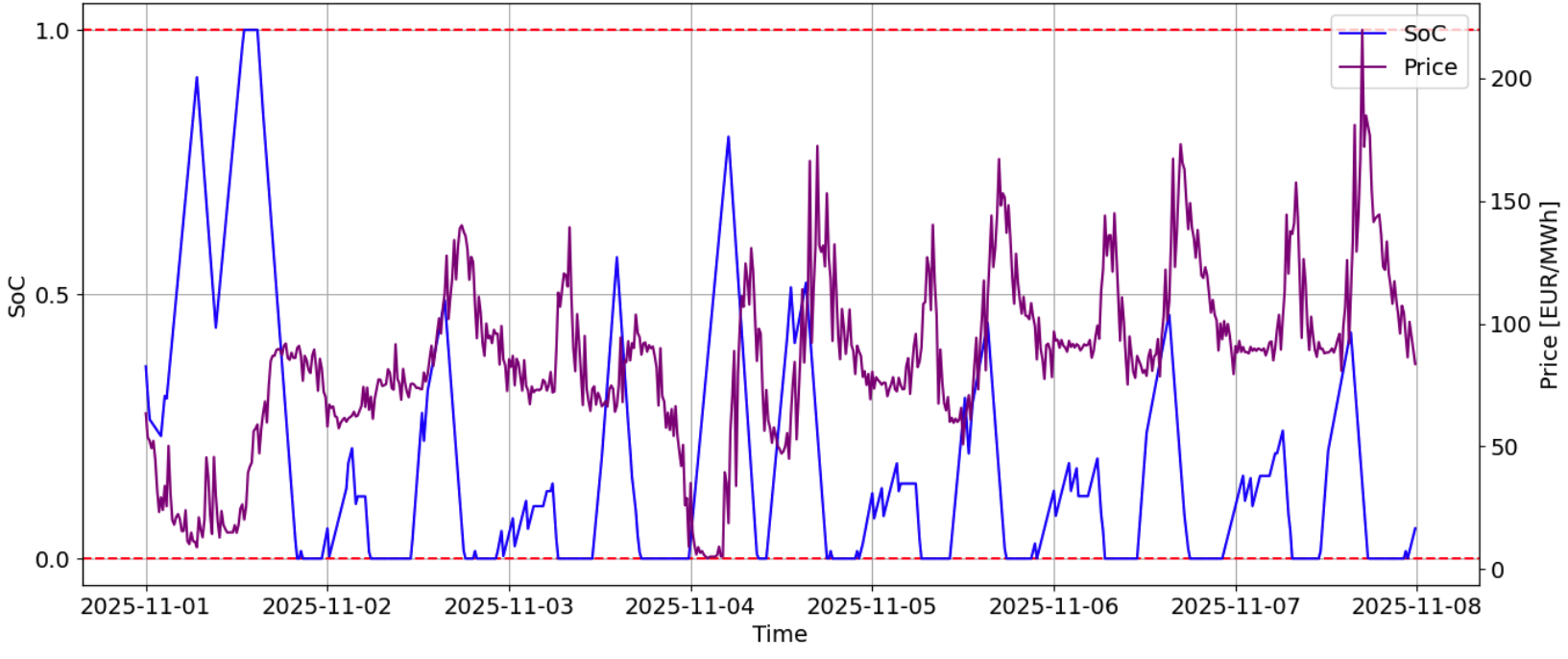
Ergebnisse

TD3 (Peak-Shaving)



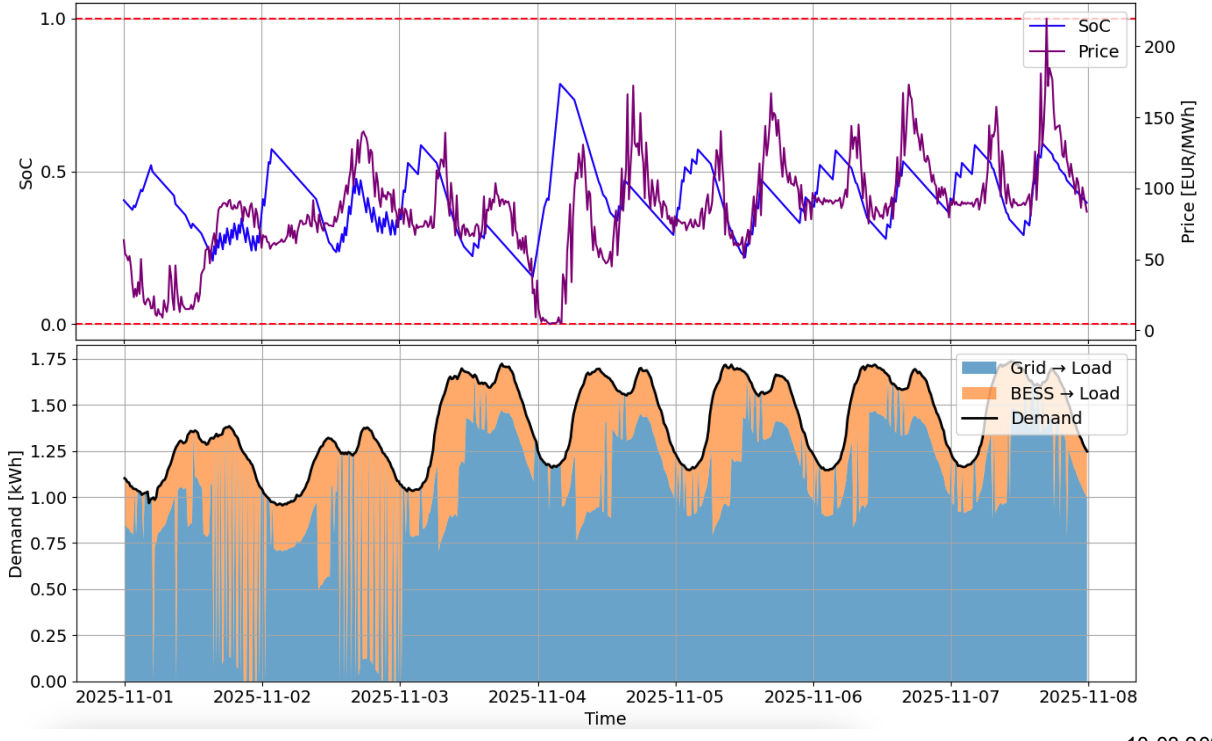
Ergebnisse

QR-DQN (Arbitrage)



Ergebnisse

QR-DQN (Peak-Shaving)



Vergleich

Arbitrage-Szenario

Controller	Erlös [EUR]	Deg. Kosten [EUR]	Finaler SoH [%]	Belohnung [EUR]
Rule-based	9.124	**0.196**	**98.0**	8.927
DQN	11.012	0.459	95.4	10.554
TD3	6.980	0.241	97.6	6.739
QR-DQN	**13.179**	0.355	96.4	**12.824**

Vergleich

Peak-Shaving-Szenario

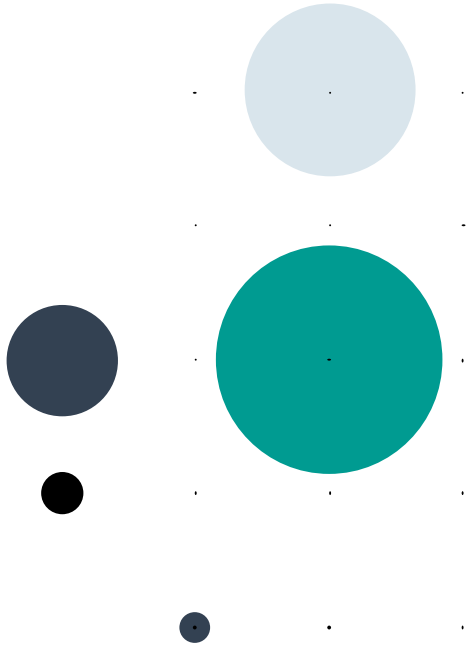
Controller	Kosten [EUR]	Deg. Kosten [EUR]	Finaler SoH [%]	Belohnung [EUR]
Rule-based	**76.395**	**0.160**	**98.4**	**-76.555**
DQN	83.602	0.302	97.0	-83.904
TD3	80.678	0.404	96.0	-81.082
QR-DQN	80.896	0.246	97.5	-81.142

Fazit

- Im Arbitrage-Szenario erzielt QR-DQN die höchste wirtschaftliche Performance unter Marktunsicherheit.
- Im Peak-Shaving-Szenario erreichen DQN und QR-DQN eine bessere Lastglättung als der Rule-Based-Controller und TD3.
- Die Modellierung der Return-Verteilung erhöht die Robustheit gegenüber Prognoseunsicherheiten.

Vielen Dank

für die Aufmerksamkeit



Einführung

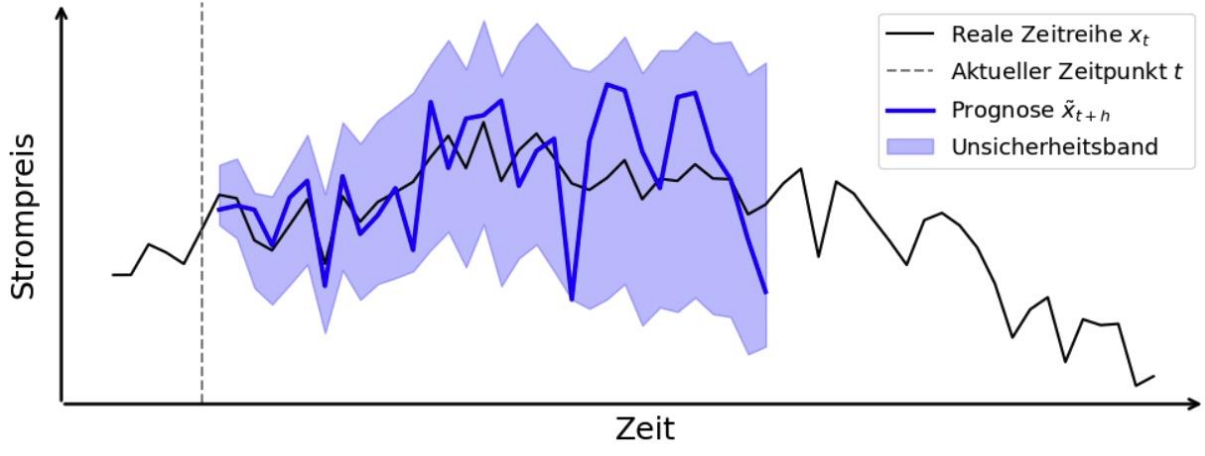
Forschungsfragen

- Wie effektiv sind RL-Regler für die Steuerung von BESS unter Preis- und Lastunsicherheit im Vergleich zu regelbasierten Ansätzen?
- Wie robust sind RL-Agenten gegenüber Prognoseunsicherheiten?
- Welche RL-Agenten zeigen die beste Performance bei Arbitrage und Peak-Shaving?

Prognoseunsicherheit

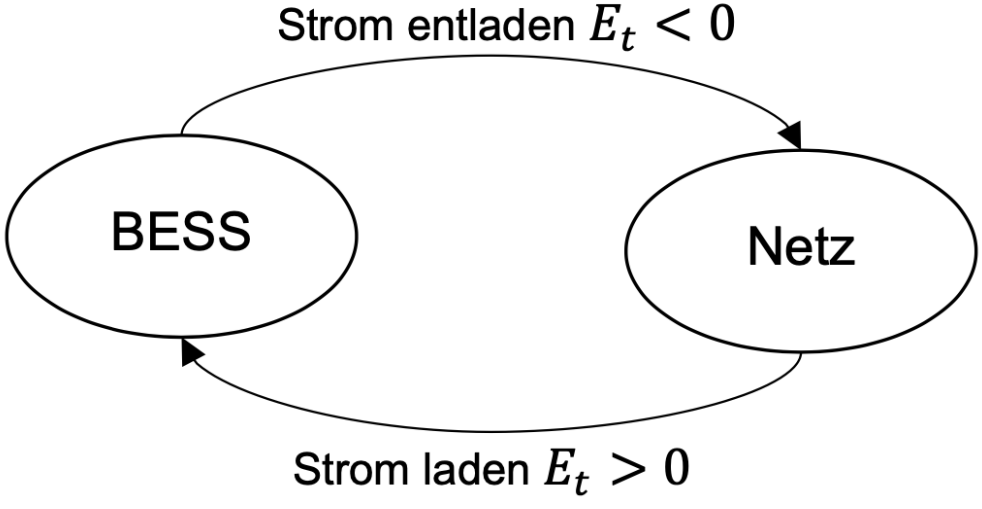
$$\tilde{x}_{t+h} = x_{t+h} + \sigma_h \varepsilon_{t,h} \quad \text{mit} \quad \varepsilon_{t,h} \sim \mathcal{N}(0,1)$$

- Prognosefehler gaußverteilt
- Varianz steigt mit Prognosehorizont



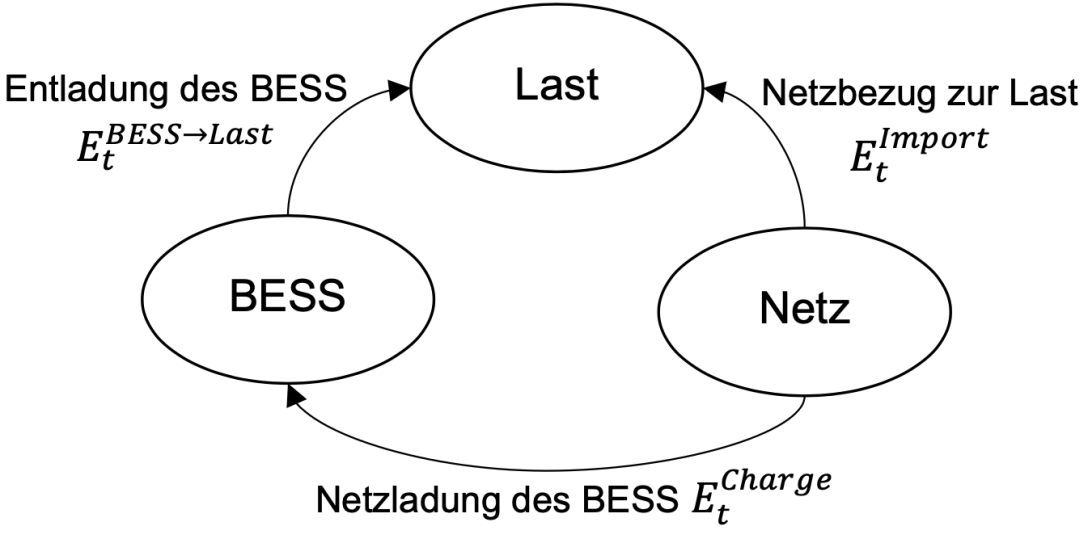
Arbitrage-Szenario

- Ziel ist die zeitlich optimierte Nutzung von Preisunterschieden durch Laden in Niedrigpreis- und Entladen in Hochpreisphasen.



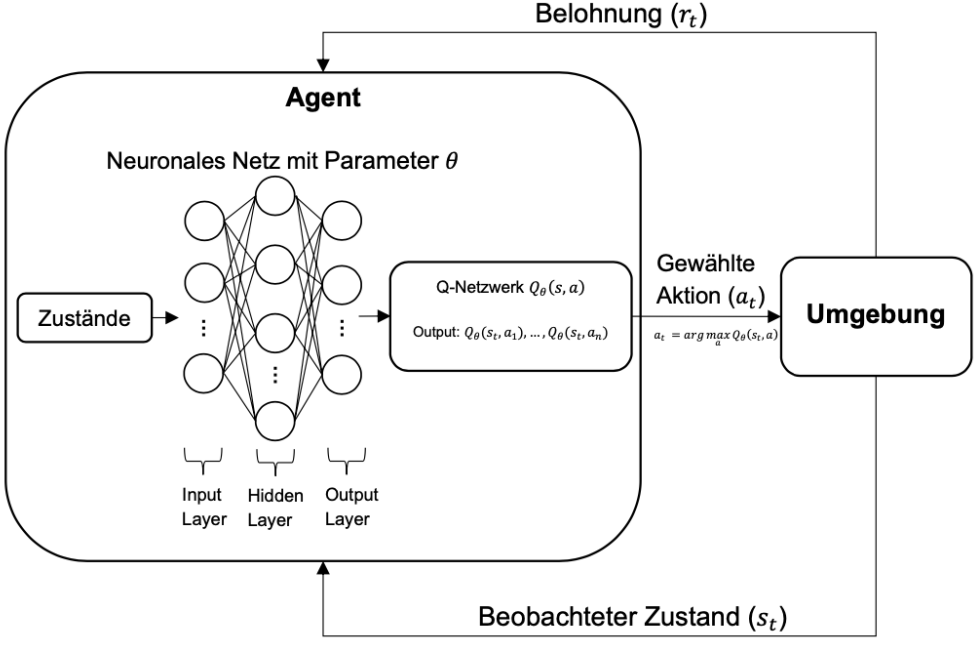
Peak-Shaving-Szenario

- Während der Peak-Glättung ist keine gleichzeitige Be- und Entladung des BESS zulässig.
- Bei Lastspitzen entlädt das BESS zur Reduktion der Netzlast; in lastarmen Phasen können Last und BESS Strom aus dem Netz beziehen.



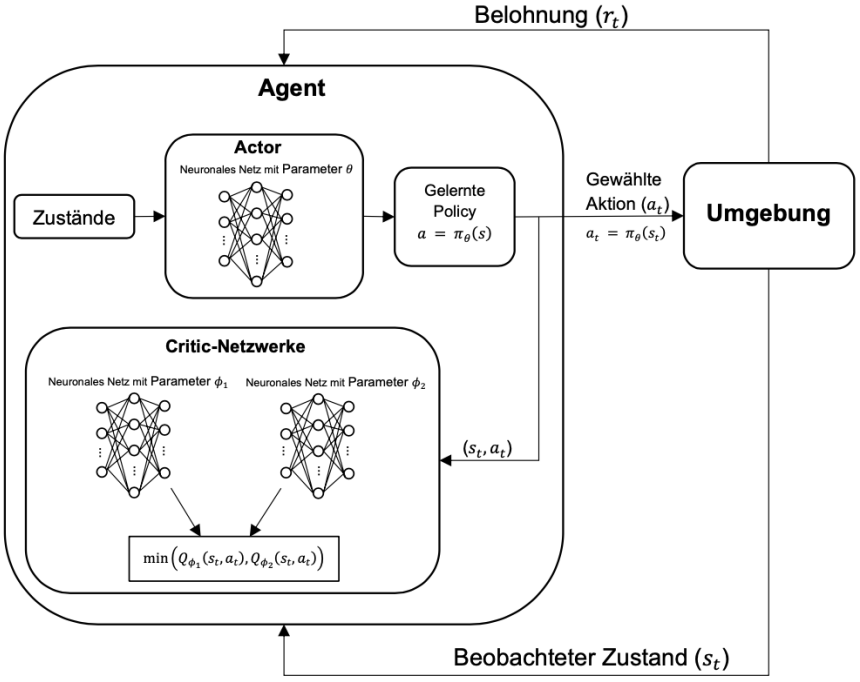
DQN

- Deep Q-Network
- Ein Deep Reinforcement Learning-Algorithmus, der die Aktionswertfunktion $Q(s, a)$ approximiert und auf Basis des erwarteten Returns $\mathrm{E}[G_t]$ optimale Aktionen auswählt.



TD3

- Delayed Deep Deterministic Policy Gradient
- Ein Deep Reinforcement Learning-Verfahren mit deterministischer Policy, das durch verzögerte Policy-Updates und doppelte Q-Schätzung die Trainingsstabilität verbessert.



QR-DQN

- Quantile Regression Deep Q-Network
- Im Gegensatz zu DQN approximiert QR-DQN nicht nur den Erwartungswert, sondern die gesamte Verteilung des Returns $\mathrm{Z}[G_t]$ mittels Quantilregression.