Verkehrssimulationen auf Binnenwasserstraßen mittels Reinforcement Learning

Bundesanstalt für Wasserbau
Technische Universität Dresden

BinSmart 2023

Fabian Hart
Ostap Okhrin

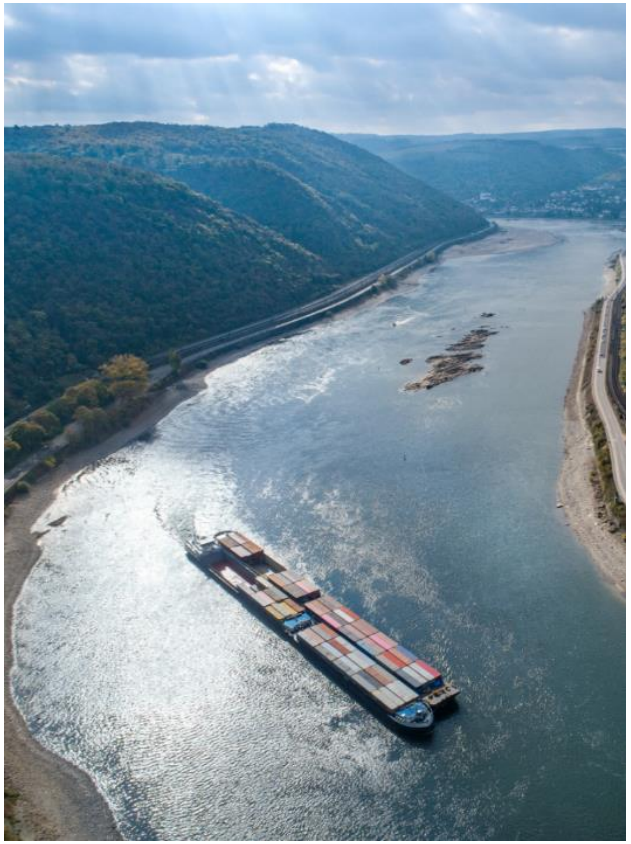
Verkehrssimulation

Projektkooperation BAW – TU Dresden

Entwicklung eines Verkehrssimulationsmodells auf Binnenwasserstraßen

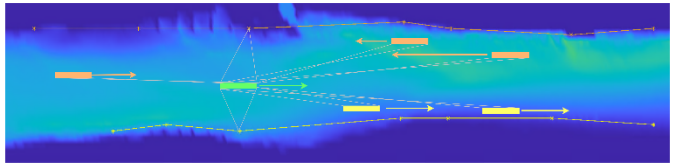
Ziel: wirtschaftliche und verkehrliche Analysen für Binnenwasserstraßen

- Engstellen
- Flottenstrukturen
- Verkehrskapazitäten
- Transportaufkommen
- Tendenz zu größeren Schiffen
- Klimawandel
- Konjunkturphasen



Quelle: BAW (Bundesanstalt für Wasserbau)

Simulator Architektur



Feature Detection

Controller:

Reinforcement Learning

vs

klassische Steuerungs- und Regelungstechnik

Vorteile RL:

- großes Potenzial bei komplexen, nicht-linearen Systemen
- Schiffsdynamik leicht austauschbar
- sehr effiziente Ausführung des gelernten Modells (einfache Matrixmultiplikationen)

rudder angle

engine power

Vessel Dynamics

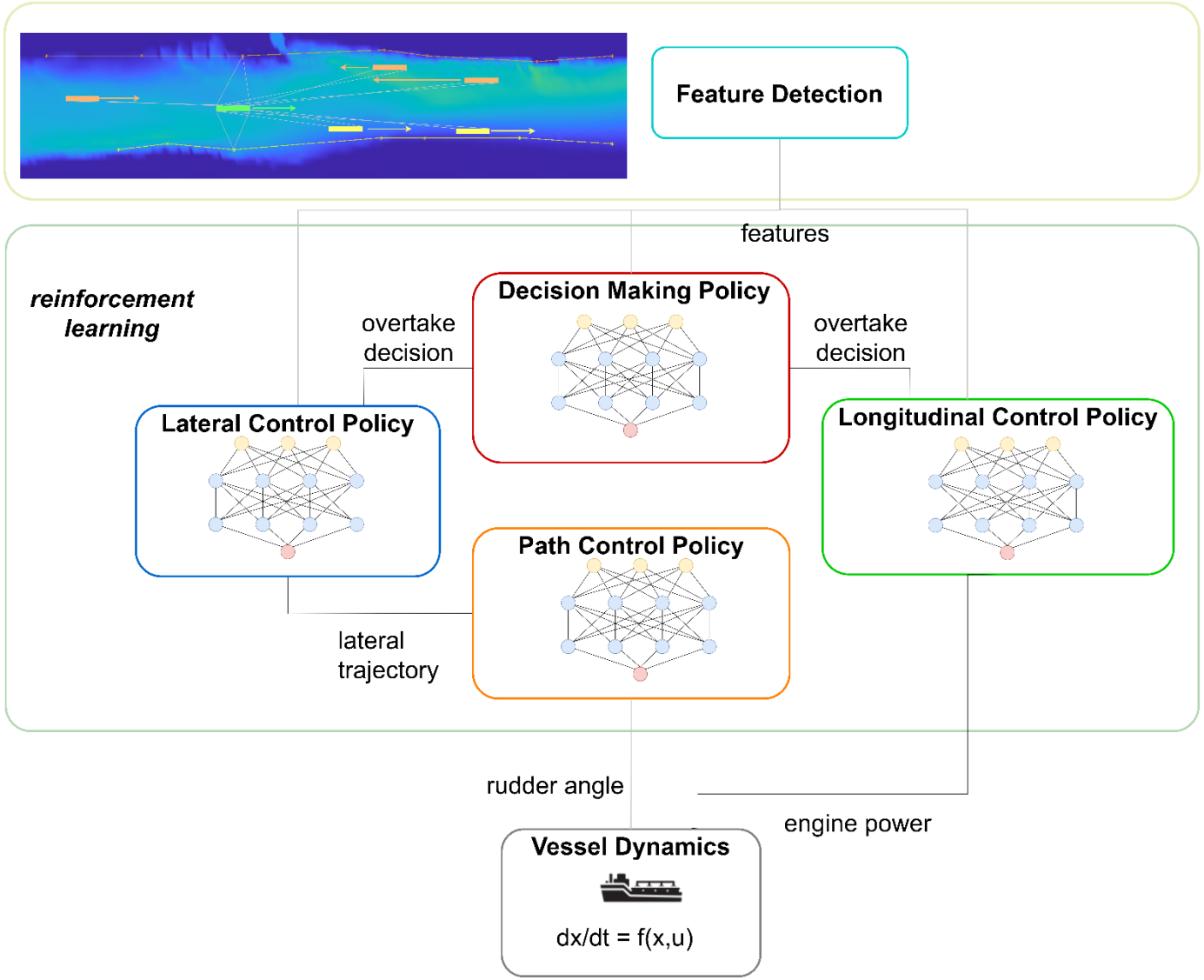
$dx/dt = f(x,u)$

TECHNISCHE UNIVERSITÄT DRESDEN

RL DRESDEN

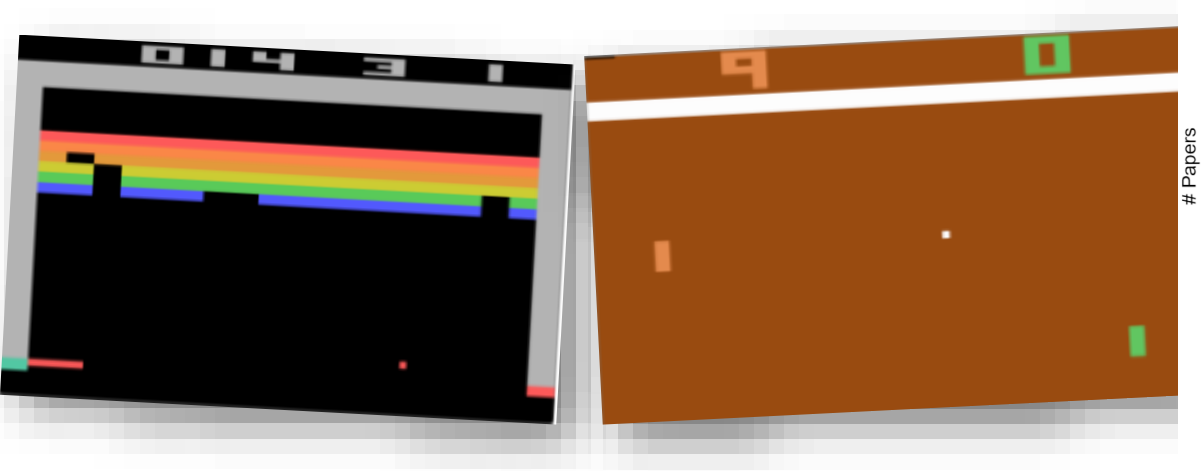
BAW Bundesanstalt für Wasserbau

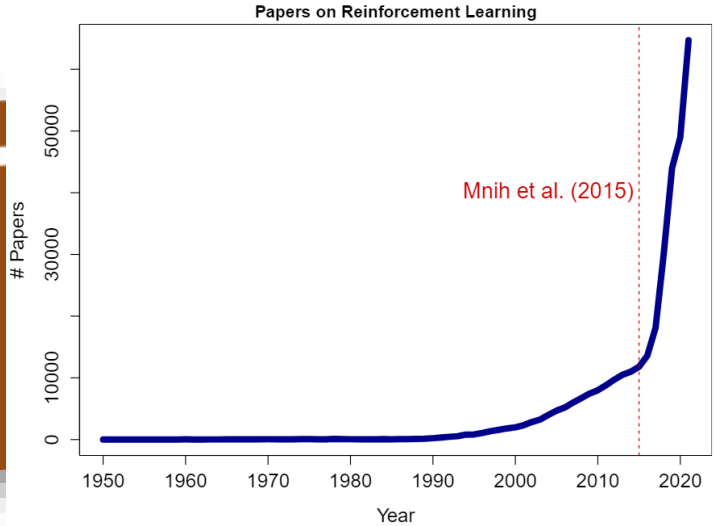
Simulator Architektur



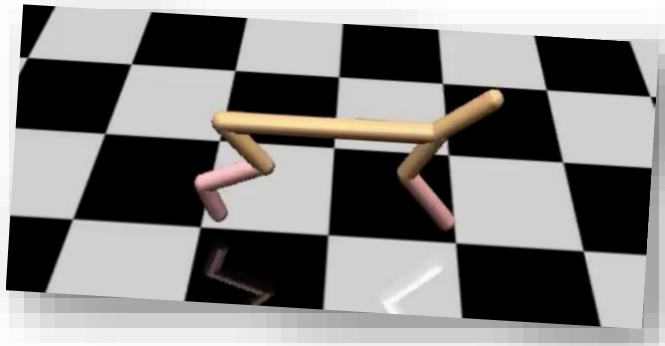
RL Basics

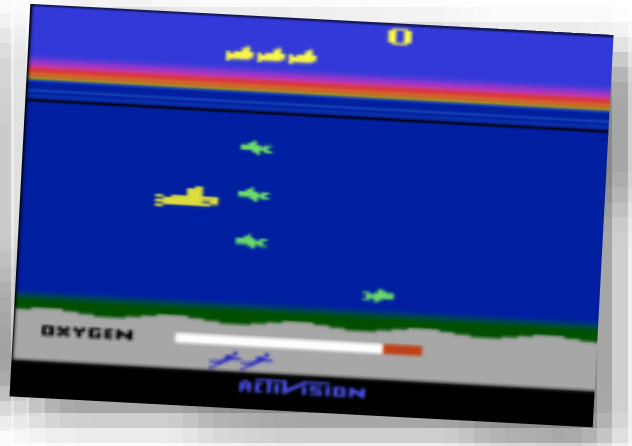
Lernen durch *Trial and Error*





Bellman, Richard. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.





Mnih, V., Kavukcuoglu, K., Silver, D., *et al*. **Human-level control through deep reinforcement learning**. *Nature*, 518, 529–533, 2015.

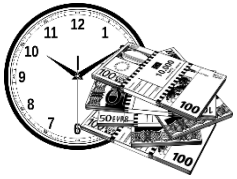
RL Basics

state space S

action space A

transition dynamics
$P(s_{t+1}|a_t, s_t)$

reward
$R(s_t, a_t, s_{t+1})$

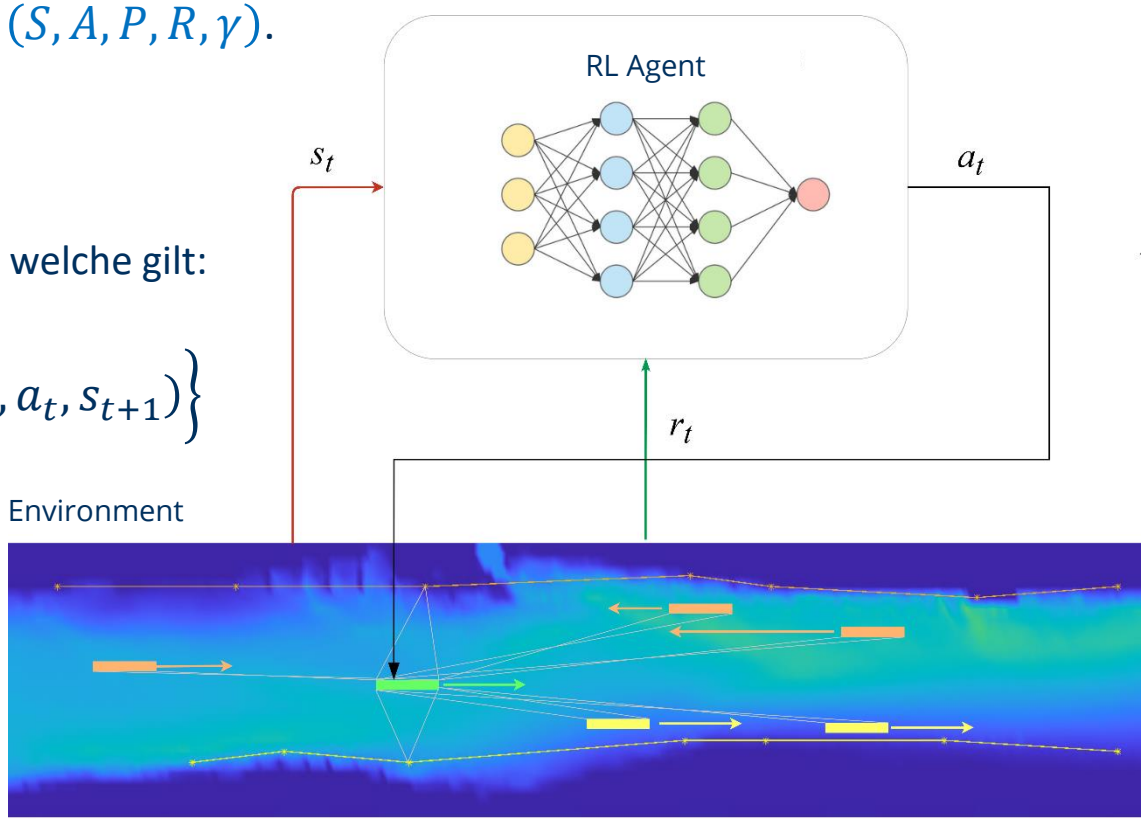


discount
$\gamma \in [0,1[$

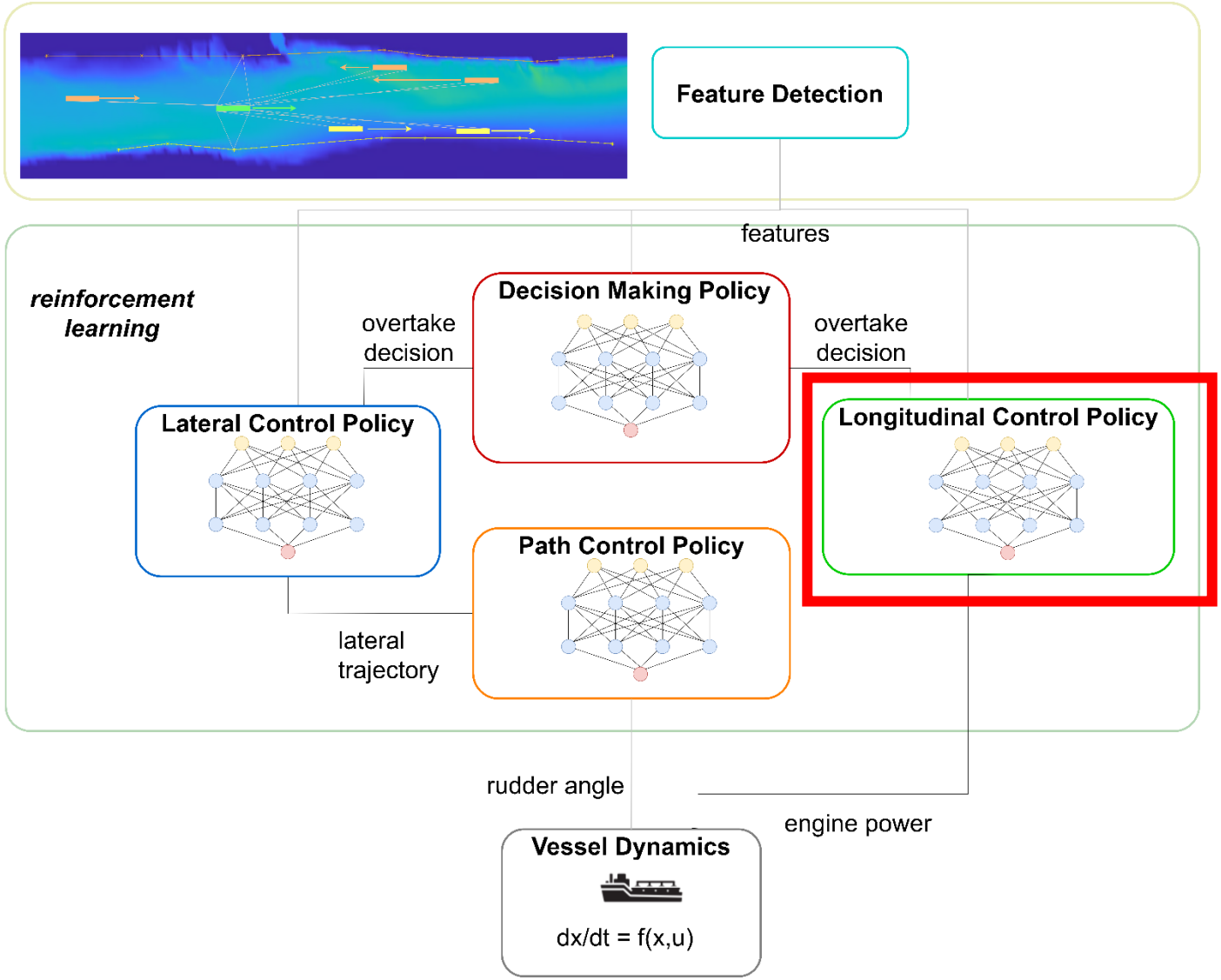
Annahme: Markov Decision Process (S, A, P, R, γ).

Ziel: Finden einer *policy* $a_t = \pi(s_t)$, für welche gilt:

$$\pi^* = argmax_\pi E_\pi \left\{ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1}) \right\}$$

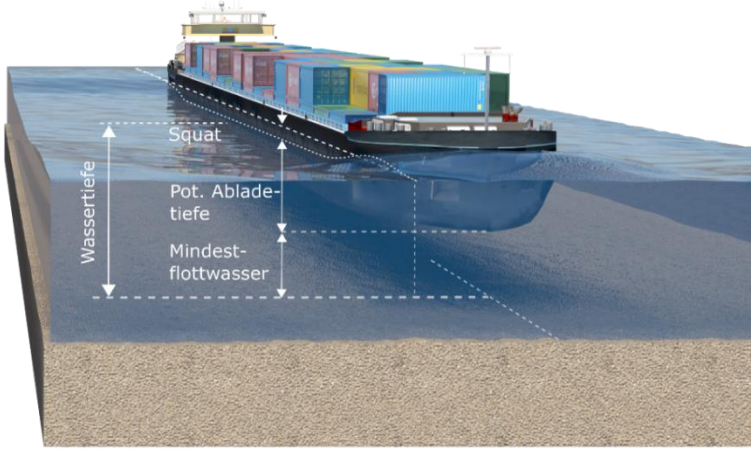


Fokus: Longitudinal Control Policy

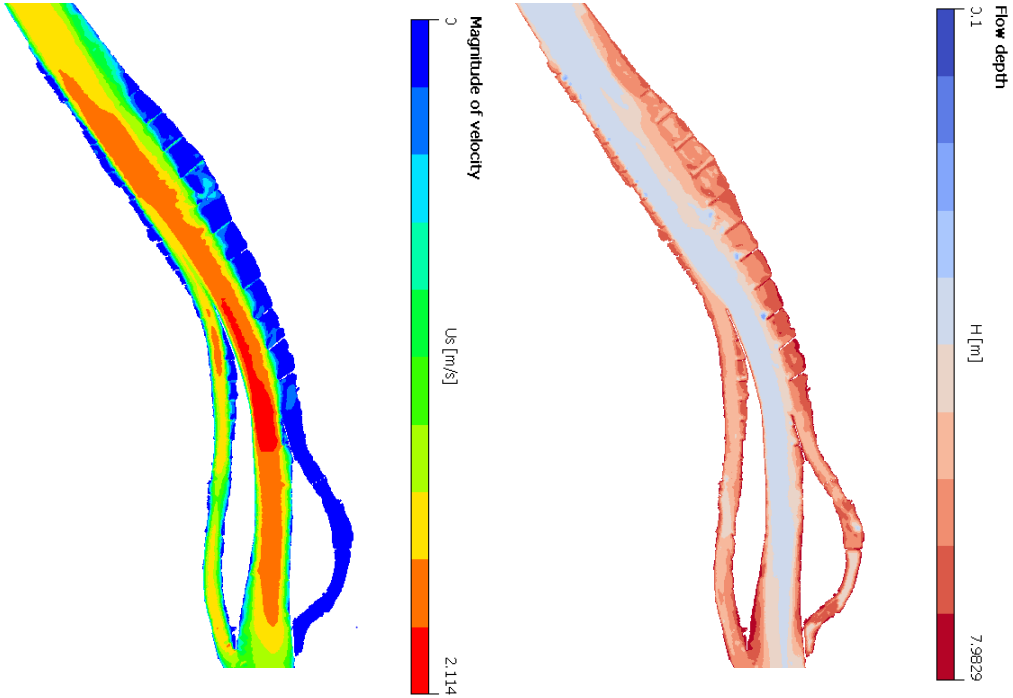


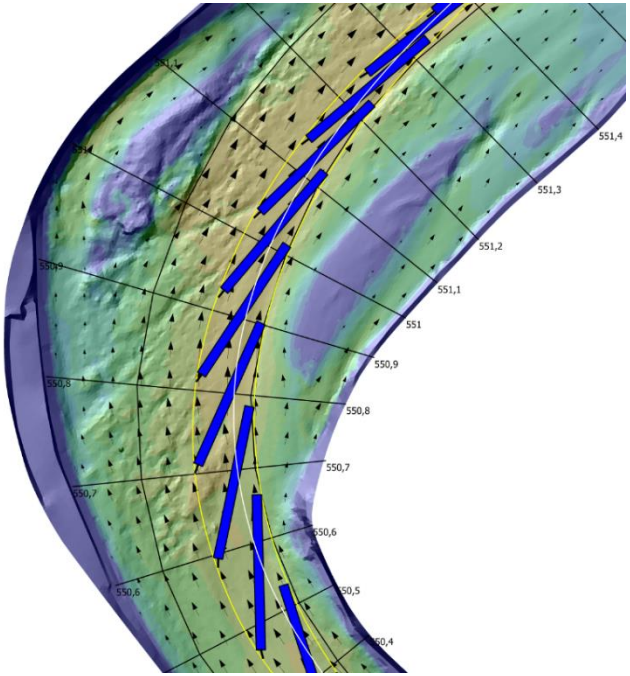
Herausforderungen

- Flussgeometrie
- Strömungen (2D)
- Bathymetrie
- Hydrodynamik
- Squat
- ...



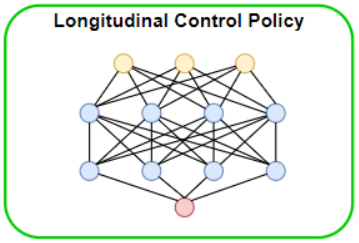
Quelle: BAW (Bundesanstalt für Wasserbau)





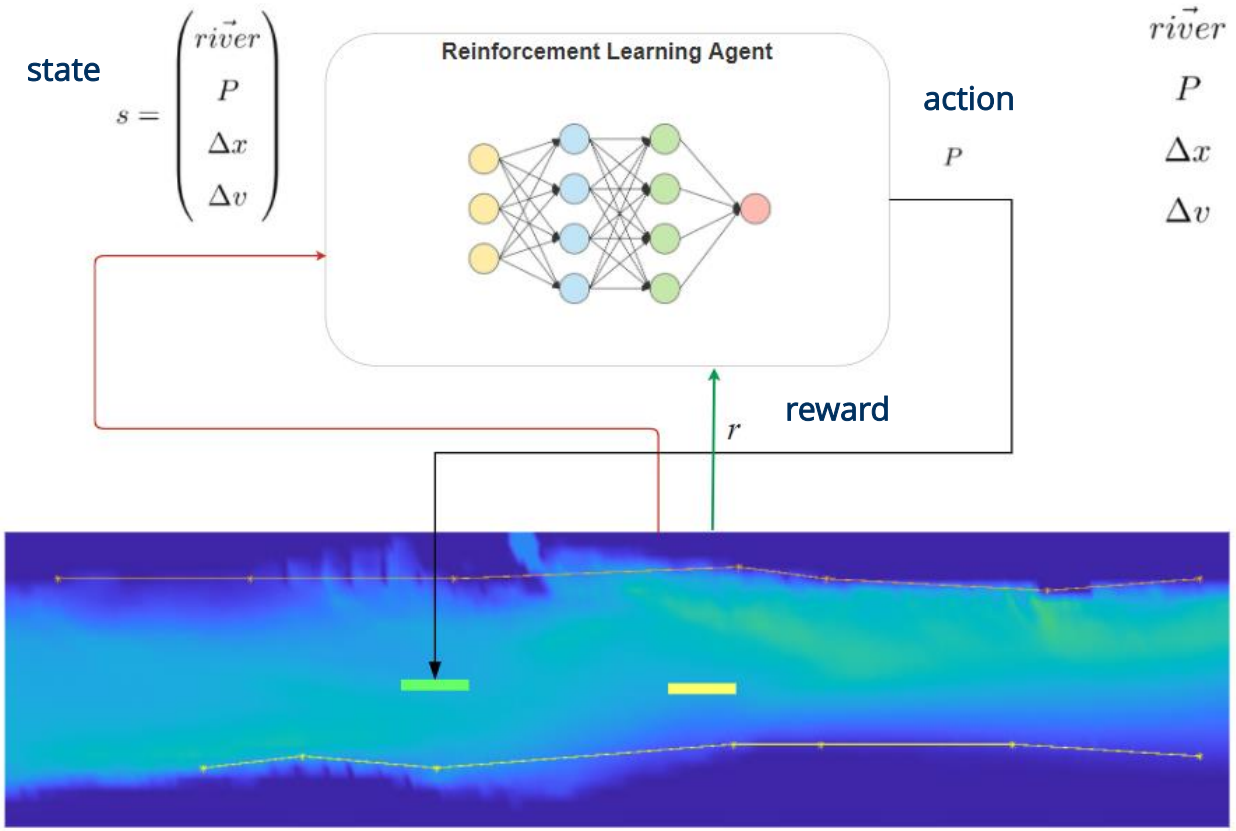
Quelle: BAW (Bundesanstalt für Wasserbau)

Longitudinal Control Policy



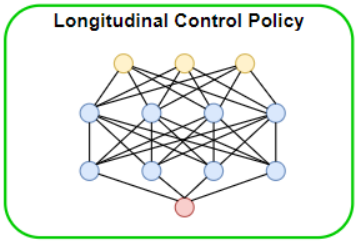
Ziel:

- Navigieren mit einer gewünschten Motorleistung
- Sicherheitsabstand zu evtl. vorausfahrendem Schiff



$\vec{river}$	Fluss Parameter
P	Motorleistung
Δx	Abstand
Δv	Relativgeschwindigkeit

Reward



- beschreibt quantitativ das gewünschte Verhalten, welches erlernt werden soll
- entscheidender Faktor in RL, häufig Fine-Tuning notwendig

Aber: *Wie verhält sich ein Schiffsführer in Schifffolgefahrt?*

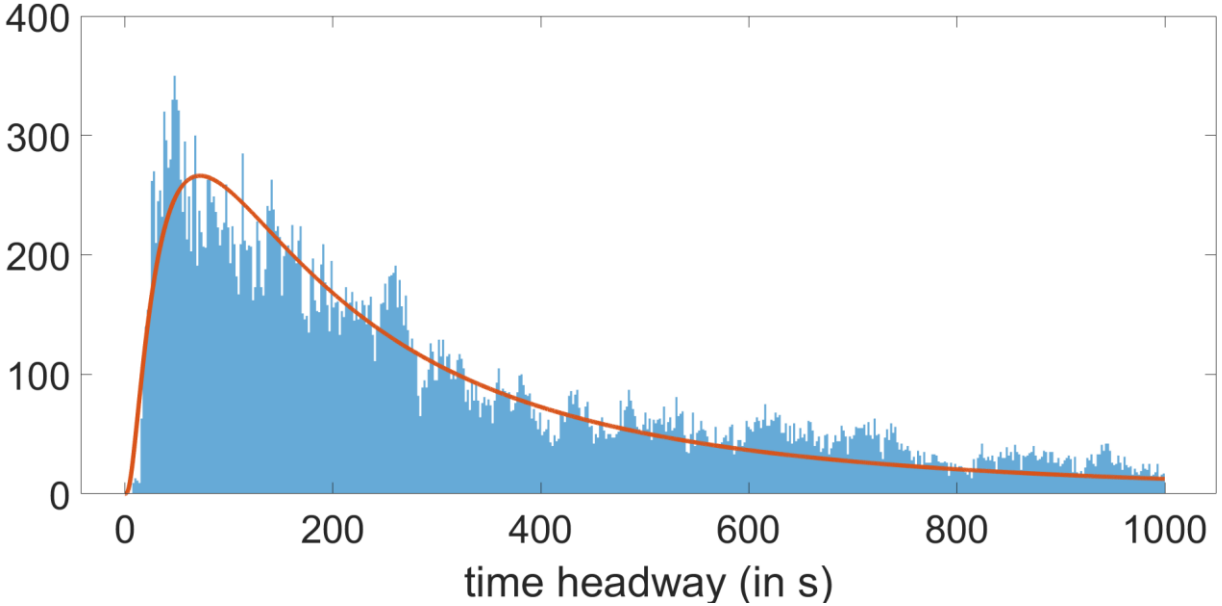
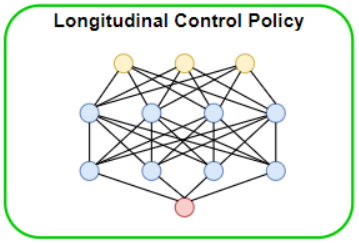


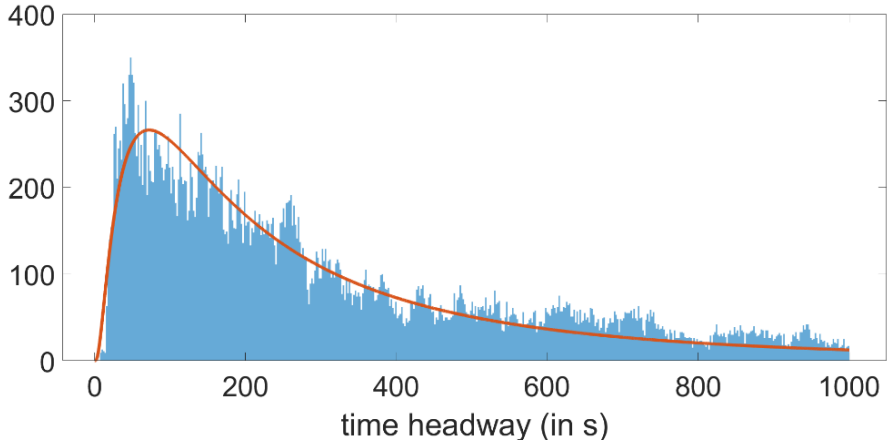
Abbildung: Time Headway für Schiffspaarungen aus AIS Daten des Mittelrheins

Reward



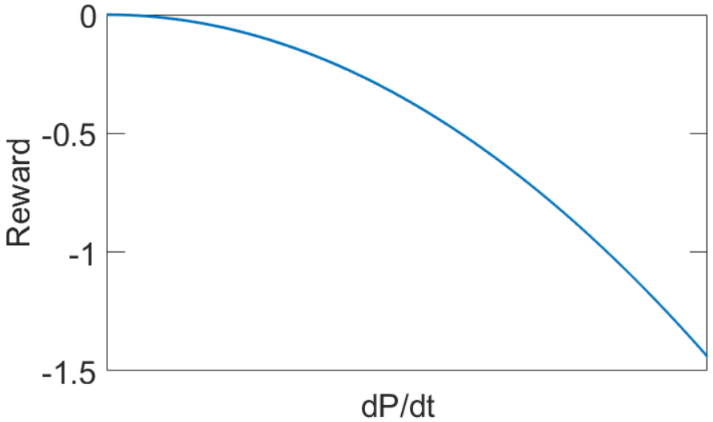
1. Sicherheit und Effizienz

- belohnt Fahren mit optimalem Time Headway



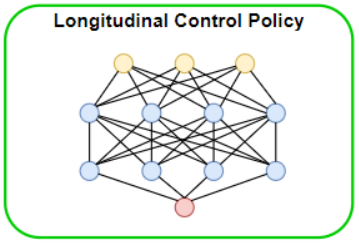
2. Komfort

- bestraft hohe Änderungsraten in Motorleistung



→ Gewichtung zwischen beiden Faktoren notwendig

Training Environment



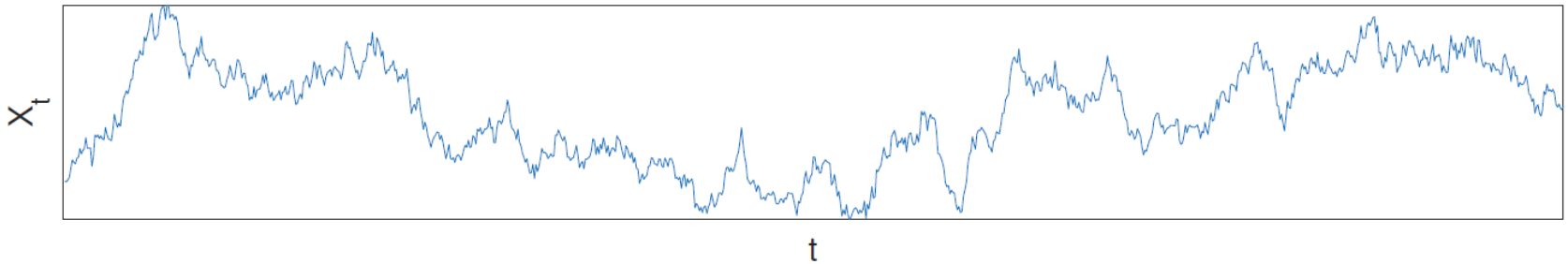
Ziel: fluss-unabhängig und so allgemein wie möglich

→ erhöht die Generalisierungsfähigkeiten des trainierten Modells

Ansatz: Umgebungsvariablen definiert als stochastische Prozesse

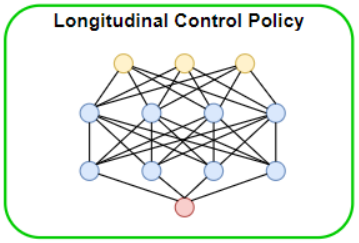
AR(1) Prozess:

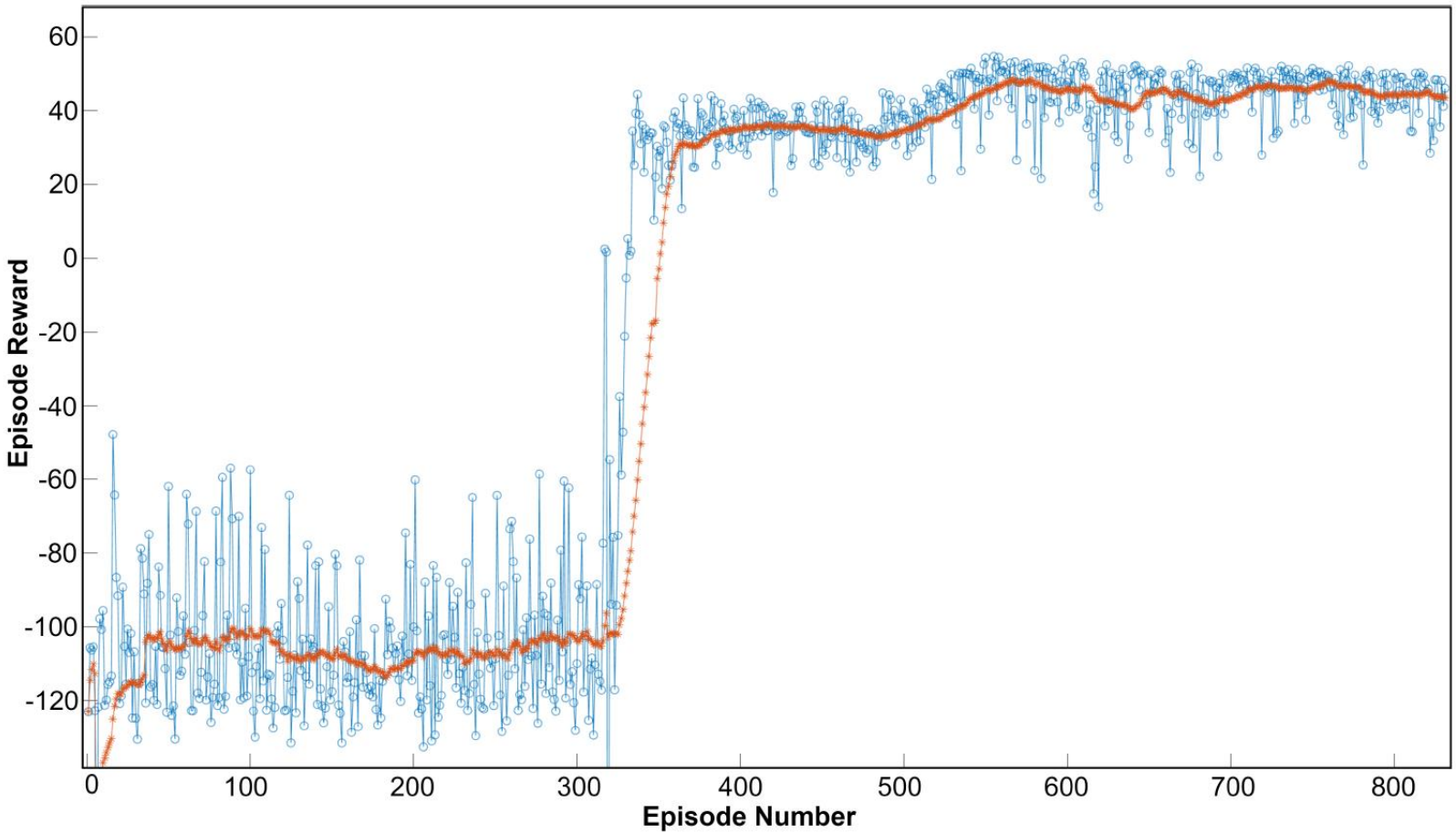
$$X_{t+1} = c_{AR} + \phi_{AR} X_t + u_t, \qquad \text{where } u_t \sim N(0, \sigma_{AR}{}^2)$$



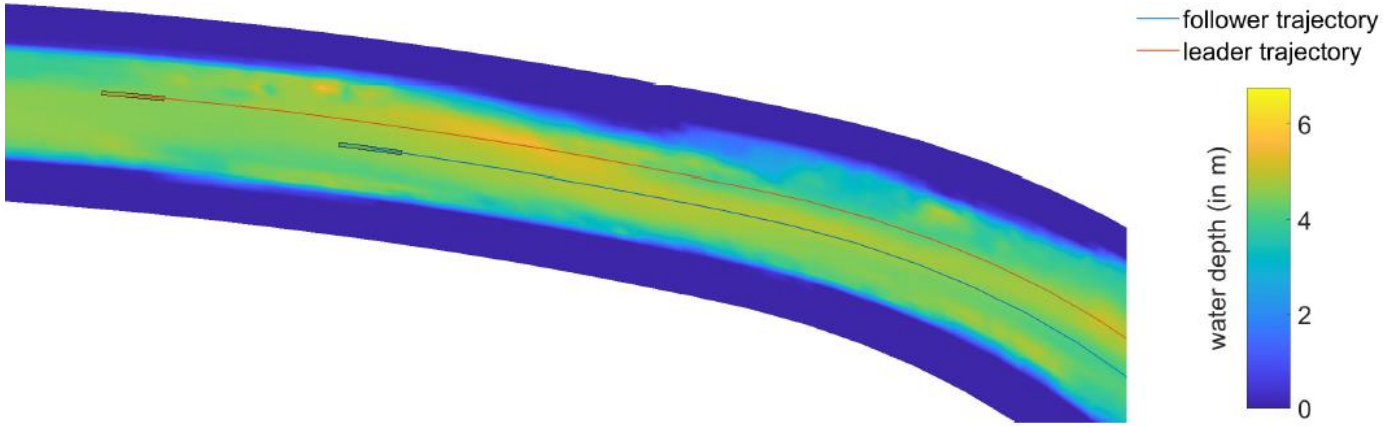
Umgebungsvariablen: Führungstrajektorie und Flussvariablen (Flusstiefe, Strömungsgeschwindigkeiten, Flussprofil)

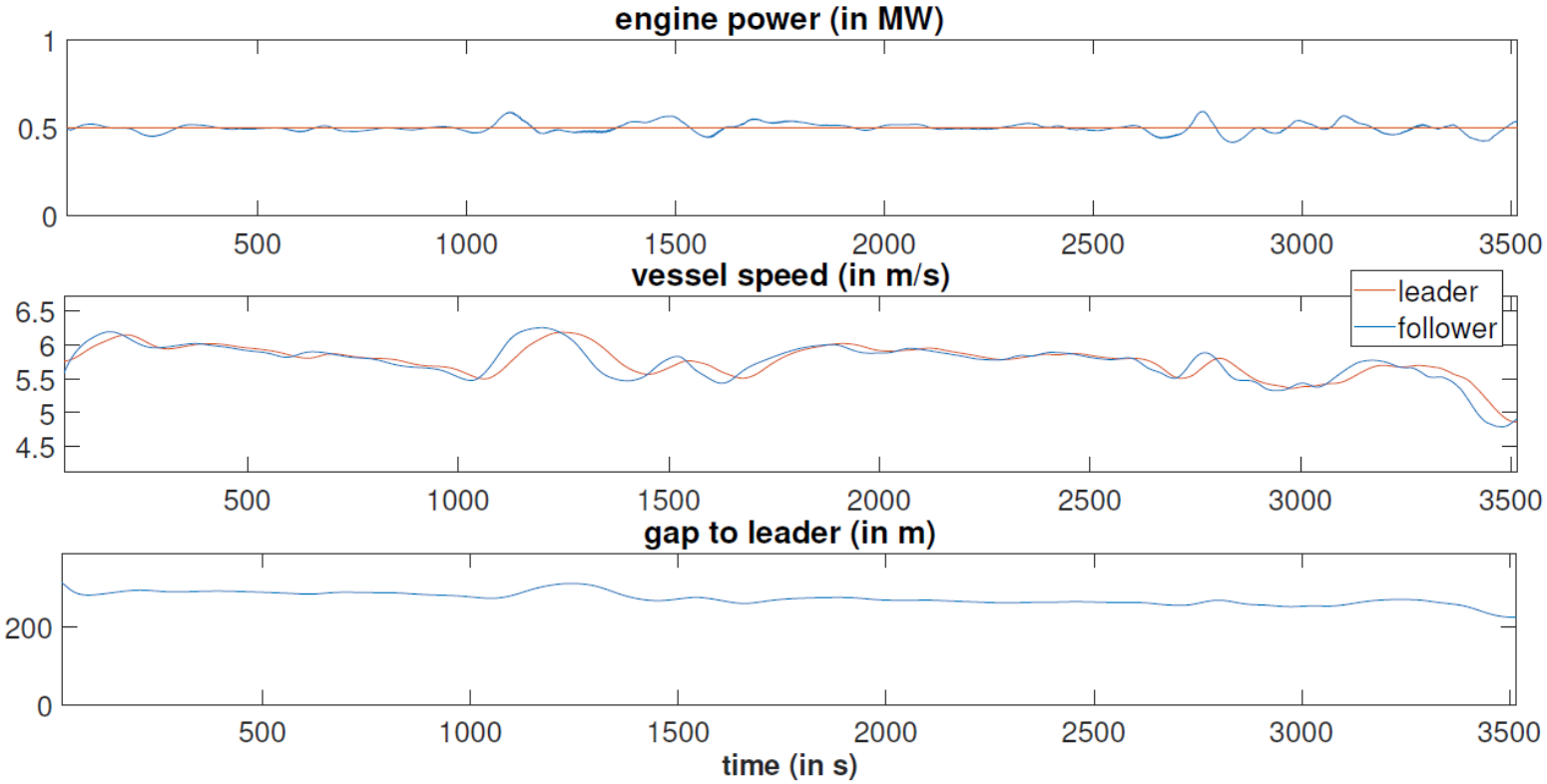
Training


Longitudinal Control Policy

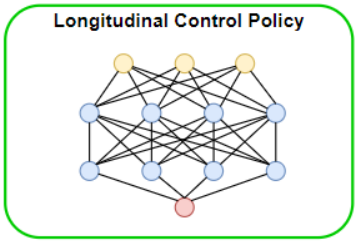

Episode Reward
60
40
20
0
-20
-40
-60
-80
-100
-120
0
100
200
300
400
500
600
700
800
Episode Number

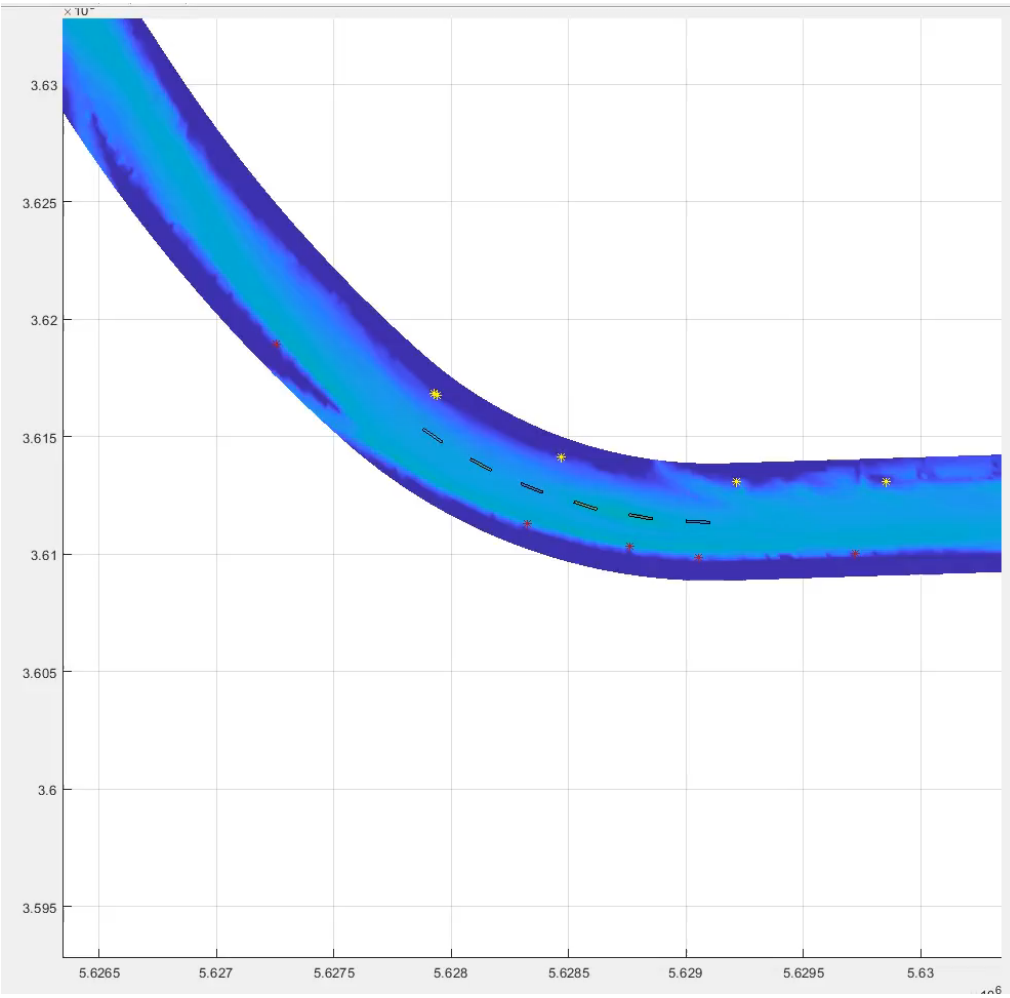
Validierung

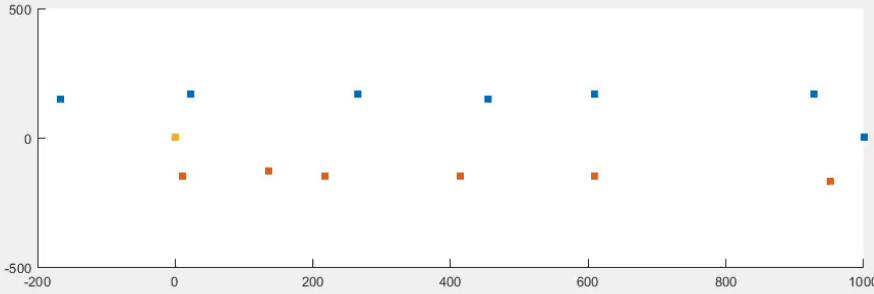


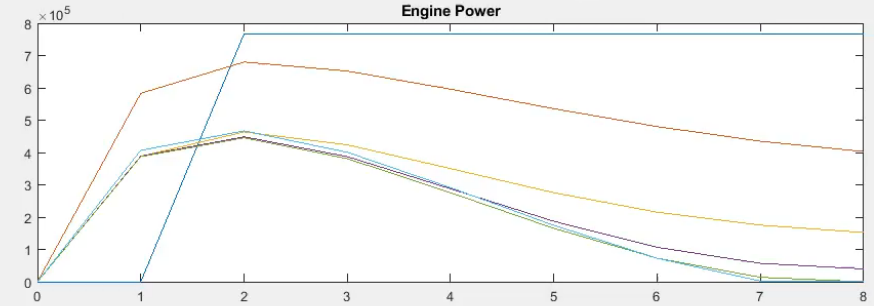


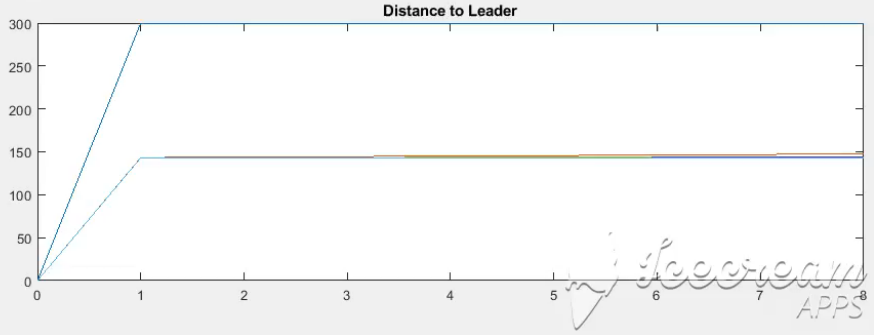
Validation


Longitudinal Control Policy

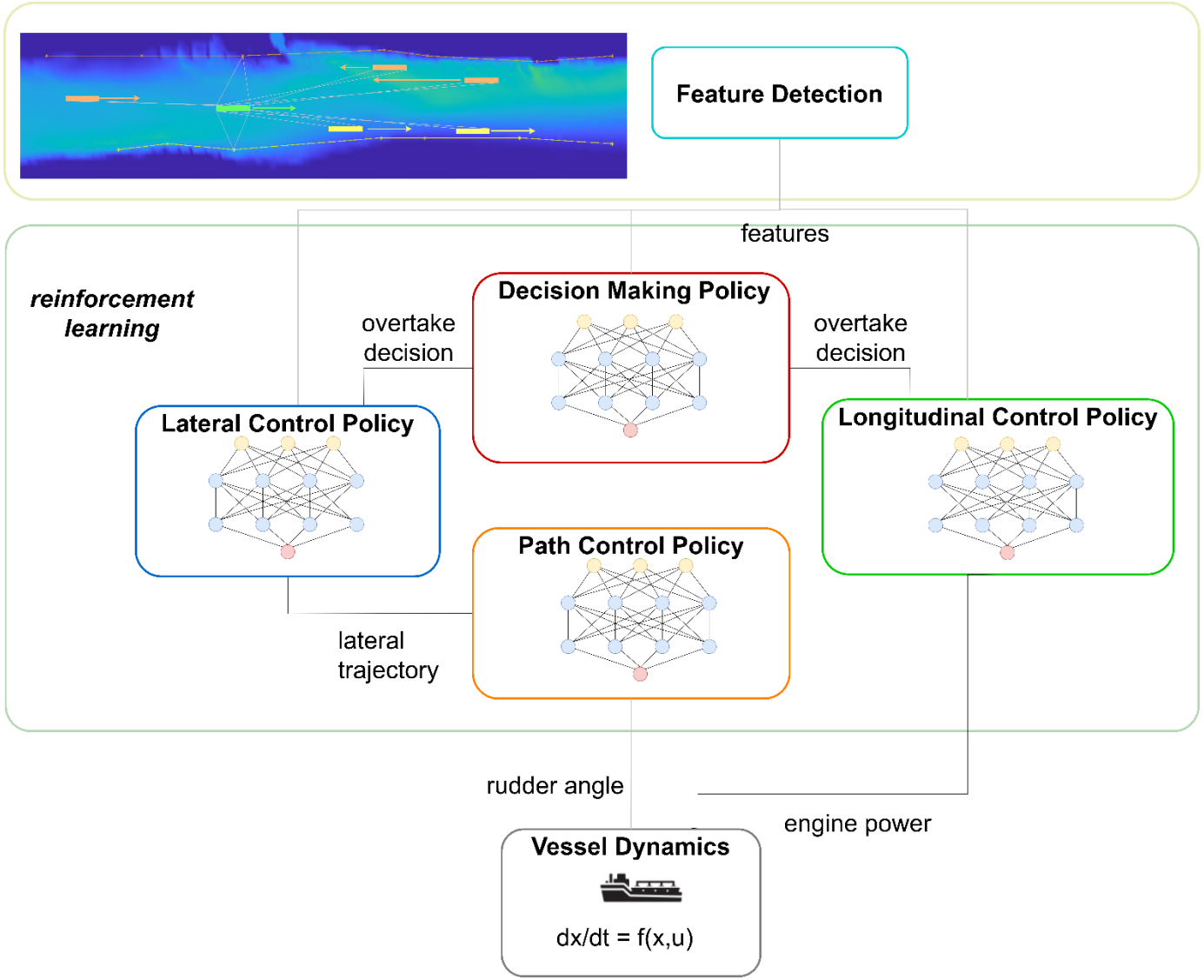

3.63
3.625
3.62
3.615
3.61
3.605
3.6
3.595
5.6265
5.627
5.6275
5.628
5.6285
5.629
5.6295
5.63


500
0
-500
-200
0
200
400
600
800
1000


Engine Power
× 10^5
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8


Distance to Leader
300
250
200
150
100
50
0
0
1
2
3
4
5
6
7
8

Simulator Architektur



»Wissen schafft Brücken.«