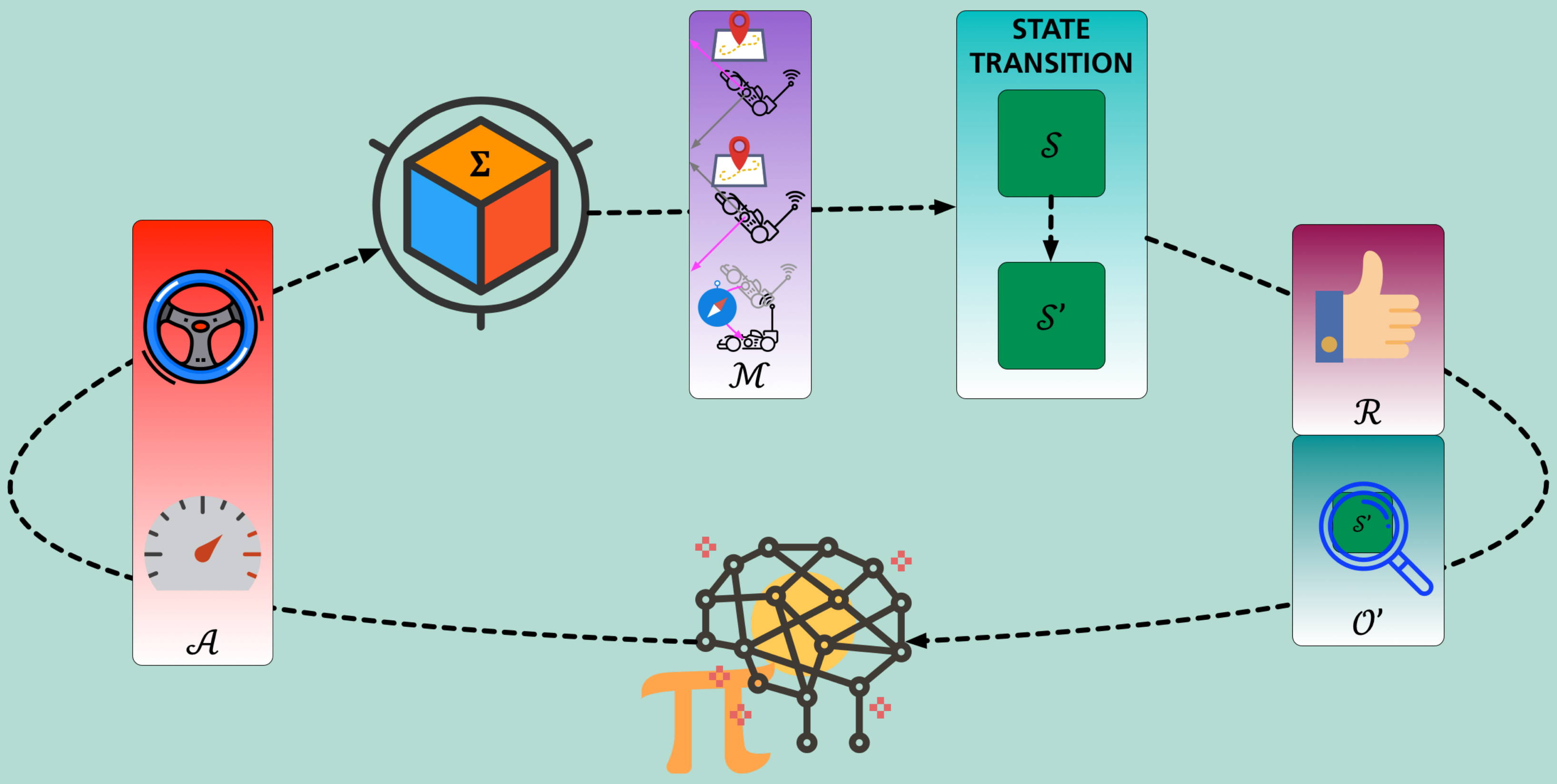
COST-EFFICIENT DEEP REINFORCEMENT LEARNING FOR SIM2REAL CAR DRIVING

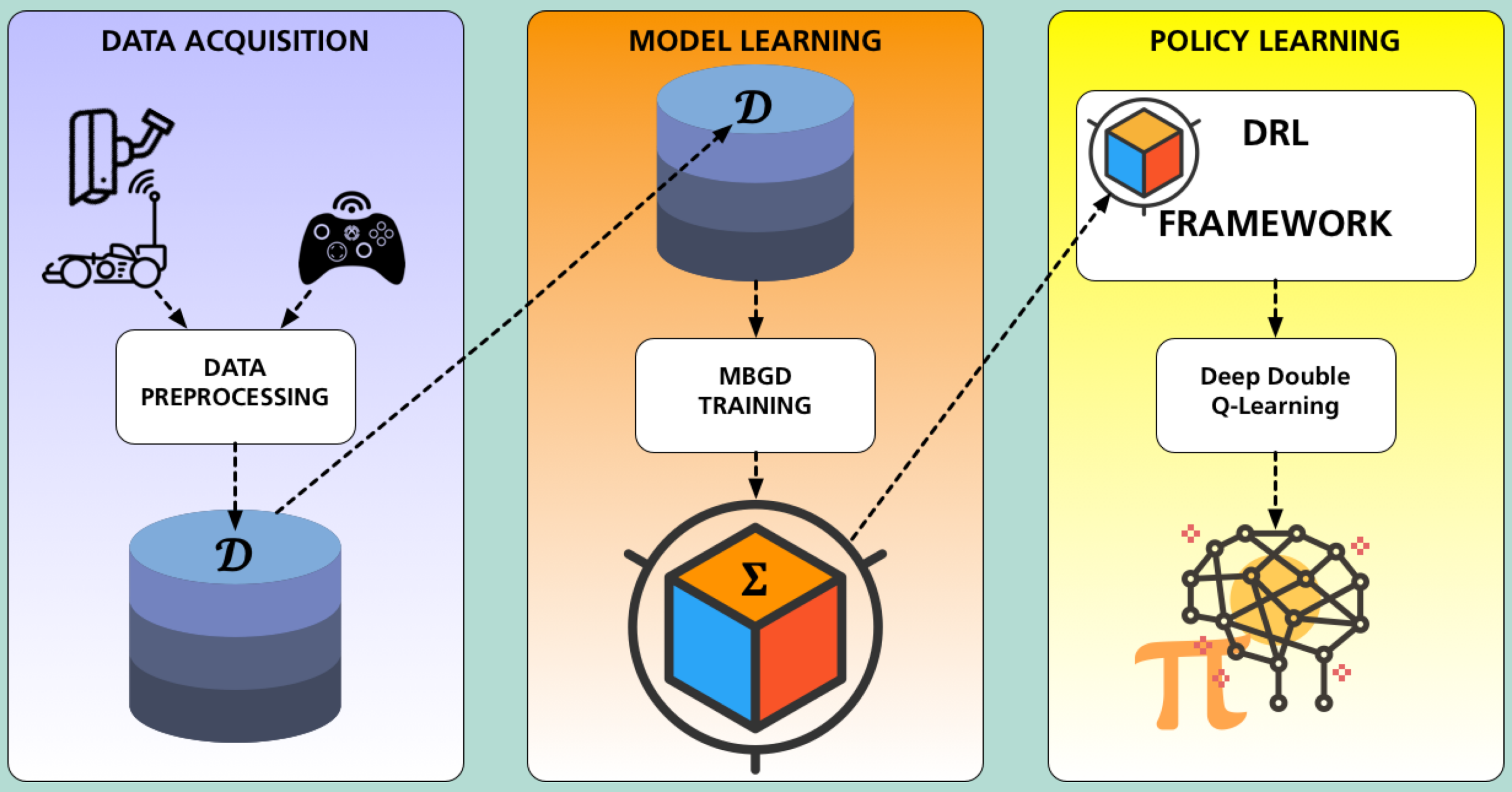
Tim Nisslbeck, Thorsten Edelhäußer, Christopher Mutschler
Fraunhofer-Institut für Integrierte Schaltungen IIS, Machine Learning and Information Fusion Group
{nisslbtm | thorsten.edelhaeusser | christopher.mutschler}@iis.fraunhofer.de

MOTIVATION

- *Controllers* for autonomous robots are learned by means of ***machine learning*** rather than engineered by hand.
- Lately, ***Deep reinforcement learning (DRL)*** methods learned controllers demonstrating *intelligent* behavior in various application domains (video games, robotics, ...).
- Learning controllers for real-world applications is associated with ***high costs*** (interaction with real world, wearout, ...).



IDEA



1. **Data Aquisition:**
   Sample the real world to get training data ($\mathcal{D}$).
2. **Model Learning:**
   Learn a model (Σ) of the real world.
3. **Policy Learning:**
   Learn a controller / policy (π)
   via a DRL framework using a fixed (converged) Σ.
4. **Policy and Model Evaluation:**
   Apply π in the real world to evaluate both π and Σ.

IMPLEMENTATION

- Application is to learn policies for **RC car driving**.
- Each step of the pipeline is implemented as follows:

1. **Data Aquisition:**
   Manually drive an RC car in a field covered by a motion capture system (ART). Collect both control inputs (***actions***) and ART data (***states***). Pre-process raw data to appropriate training data ($\mathcal{D}$).
2. **Model Learning:**
   Learn a ***vehicle dynamics model*** (Σ) from $\mathcal{D}$. We represent Σ with a ***Long Short-Term Memory*** (LSTM) Recurrent Neural Network (RNN). We train the LSTM network with mini-batch gradient descent (MBGD).
3. **Policy Learning:**
   We implement a DRL framework that uses Σ for state-transition prediction (as pictured above) to learn π with Deep Double Q-Learning.

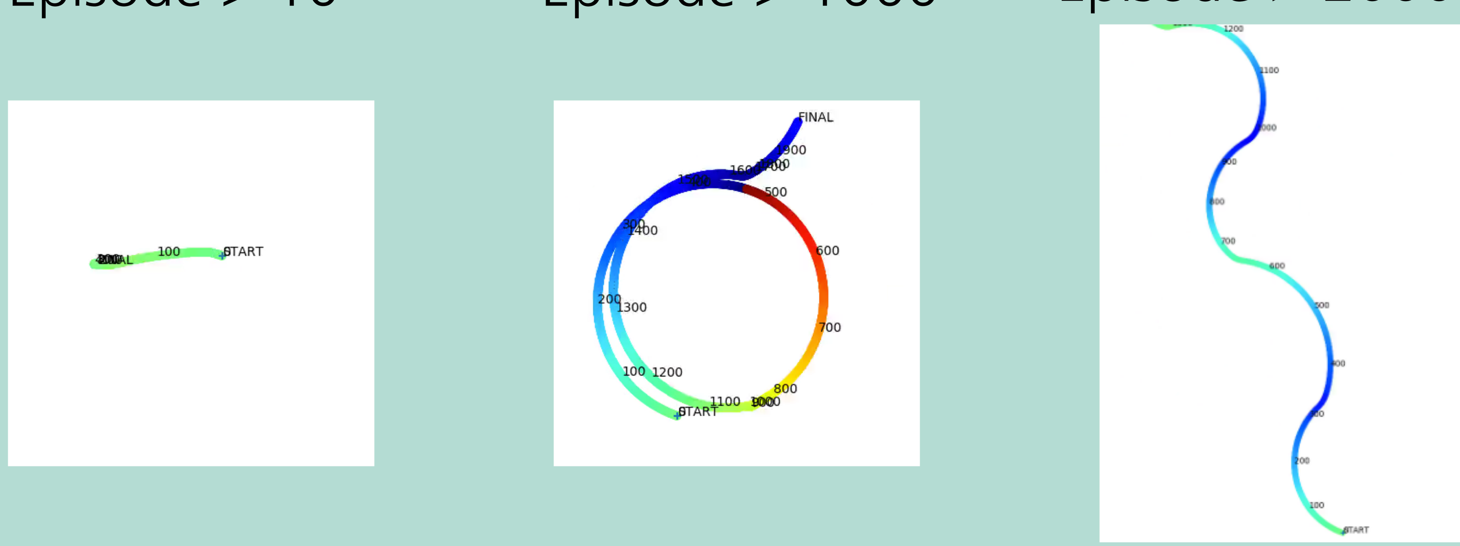
EVALUATION METHOD

- For different *(non-)linear* RC car maneuvers m learn π_m in implemented framework and execute π_m in real world.
- If each π_m performs well in the real world, then:
  - Σ provides reliable vehicle dynamics, and
  - Σ generalizes to *non-linear* real-world dynamics.

EVALUATION RESULTS

- Agent learns efficient π_m in simulation *and* the real world.
- → Predictions of Σ are accurate enough to learn π_m that are applicable in the real world.

Episode > 10 | Episode > 1000 | Episode > 2000



Policy performance in simulation for car driving maneuver "Slalom".

CONCLUSION

Efficient vehicle dynamics models and car driving policies can be learned using the proposed method, without requiring much human overhead. Further evaluation is necessary to compare the performance of learned policies with other approaches.

Fraunhofer
IIS