A Two-Layer Reinforcement Learning Solution for Energy Harvesting Data Dissemination Scenarios

TECHNISCHE UNIVERSITÄT DARMSTADT

Andrea Ortiz and Anja Klein
Technische Universität Darmstadt, Germany

Tobias Weber
Universität Rostock, Germany

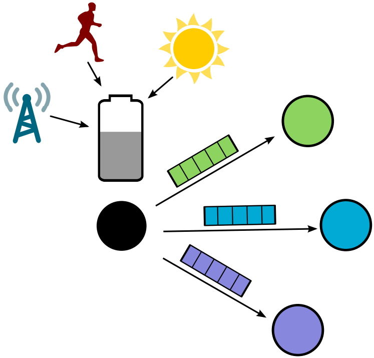
This work was funded by the LOEWE Priority Program NICER

2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

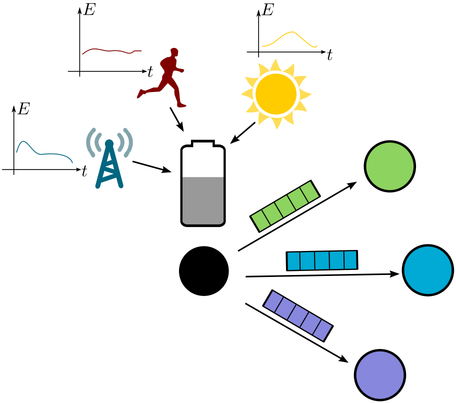
- Energy harvesting (EH) nodes are able to collect energy from the environment
- EH holds the promise of self-sustainability and perpetual operation
- Complete non-causal information is required for optimal power allocation
- What can be done if only causal information is available?



Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

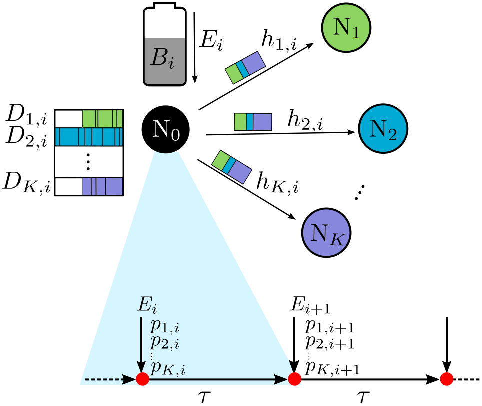
- Energy harvesting (EH) nodes are able to collect energy from the environment
- EH holds the promise of self-sustainability and perpetual operation
- Complete non-causal information is required for optimal power allocation
- What can be done if only causal information is available?



Outline

System Model

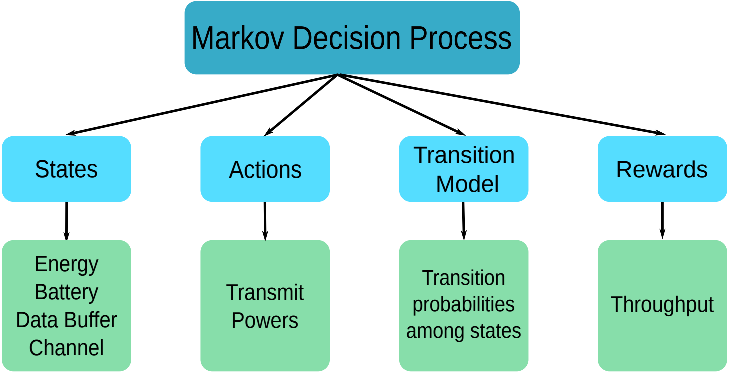
- N_0 harvests energy and uses it exclusively to transmit data to K receivers
- Constant time interval τ between two consecutive EH instants
- Only causal information available at the transmitter N_0
- Constant transmit powers during each time interval
- Each receiver treats the non-intended received data as interference



Problem Formulation

TECHNISCHE UNIVERSITÄT DARMSTADT

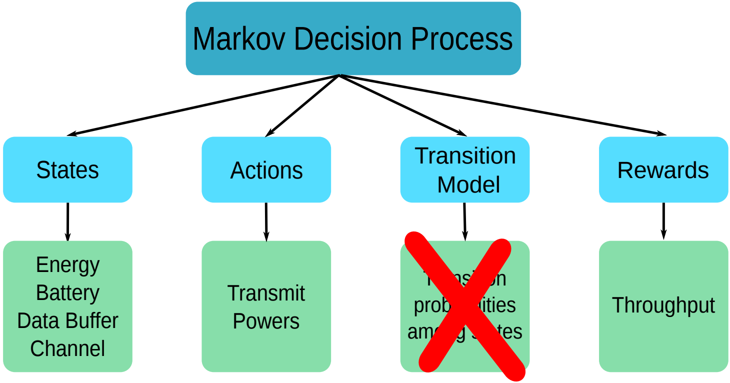
Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

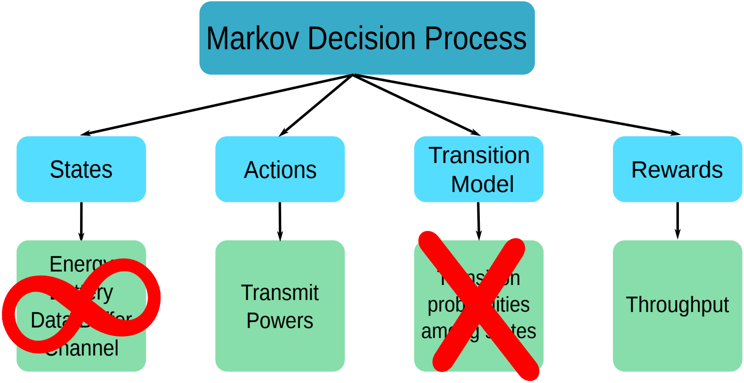
TECHNISCHE UNIVERSITÄT DARMSTADT

Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

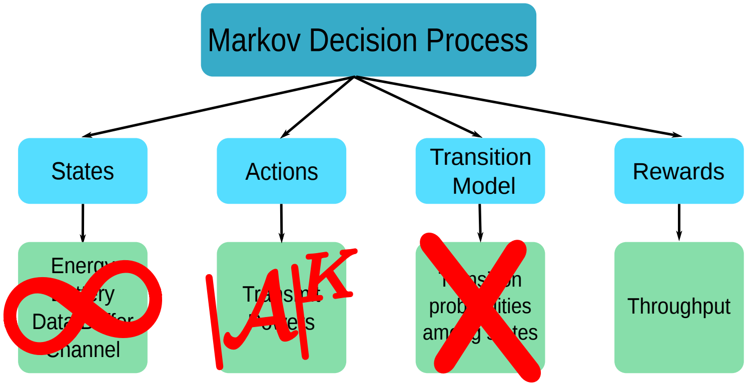
Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

TECHNISCHE UNIVERSITÄT DARMSTADT

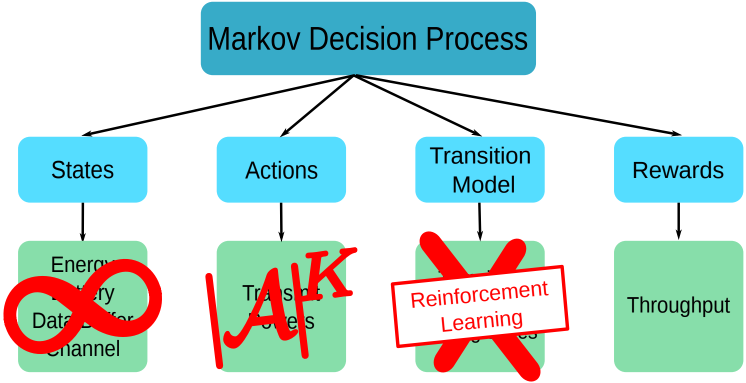
Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

TECHNISCHE UNIVERSITÄT DARMSTADT

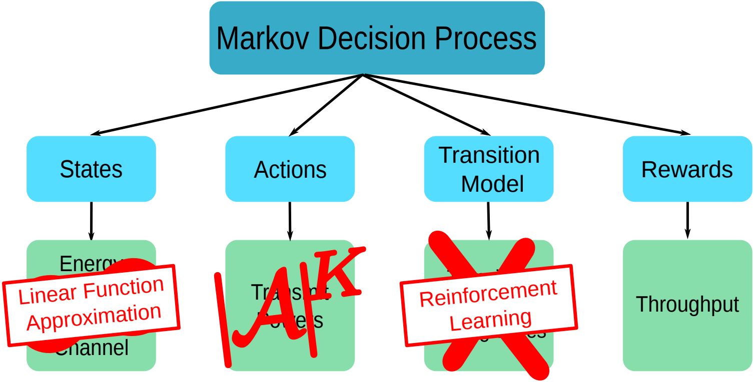
Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

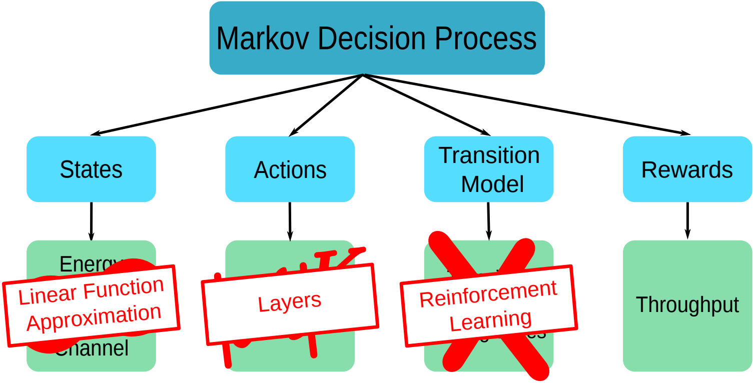
TECHNISCHE
UNIVERSITÄT
DARMSTADT

Goal: Power allocation policy to maximize the throughput having only causal information



Problem Formulation

Goal: Power allocation policy to maximize the throughput having only causal information

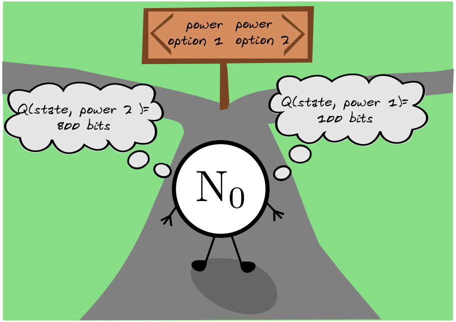


Reinforcement Learning (RL): The key idea

TECHNISCHE UNIVERSITÄT DARMSTADT

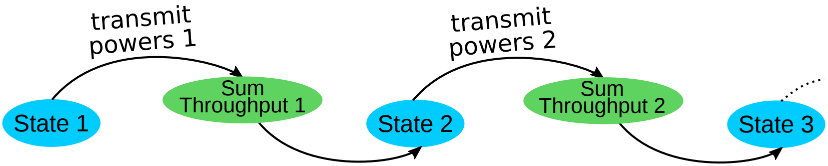
- The transmitter learns the power allocation policy considering only its causal information
- The policy is evaluated using the Q-function
- The optimal Q-function leads to the optimal policy

The Q-function is the expected throughput to be achieved by following a policy given a state and a transmit power



Reinforcement Learning (RL): The State-Action-Reward-State-Action (SARSA) algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT



- SARSA builds an estimate of the Q-function based on the states that are visited and the obtained throughput
- The Q-function is updated in every iteration
- The transmit power values are selected according to ε-greedy policy

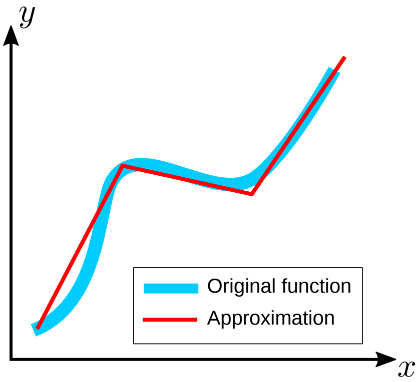
Linear Function Approximation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Only a limited number of Q-values can be stored
- The Q-function is approximated as a linear combination of feature functions

$$Q(\text{state, power}) = \mathbf{f}^{\mathsf{T}}\mathbf{w}$$

- The features **f** correspond to natural attributes of the EH problem
  - Different feature functions for each layer
- The weights **w** indicate the contribution of each feature function



Proposed Two-Layer Approach

TECHNISCHE
UNIVERSITÄT
DARMSTADT

RL for power allocation

Selection of the total transmit power for the time interval	Distribution among the data streams to be transmitted

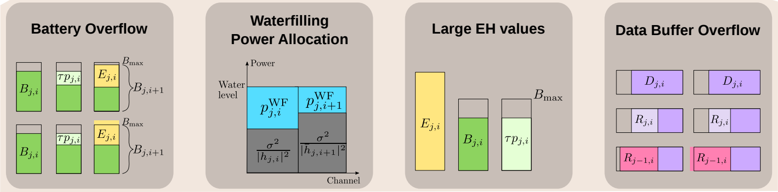
Proposed Two-Layer Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

RL for power allocation

Selection of the total transmit power for the time interval

Distribution among the data streams to be transmitted



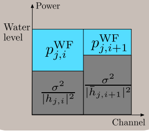
Proposed Two-Layer Approach

RL for power allocation

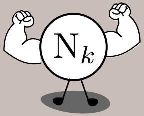
Selection of the total transmit power for the time interval

Distribution among the data streams to be transmitted

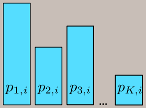
Waterfilling Power Allocation



Maximum Rate

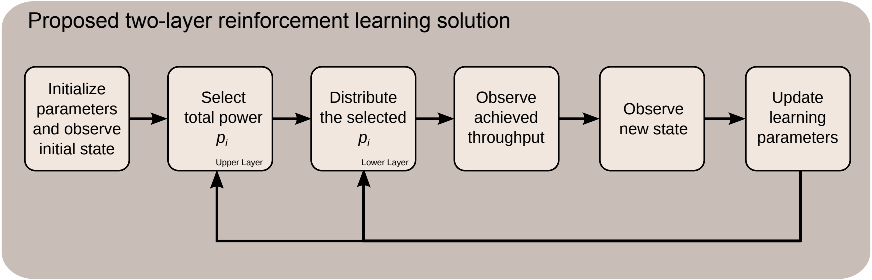


Proportional Fairness



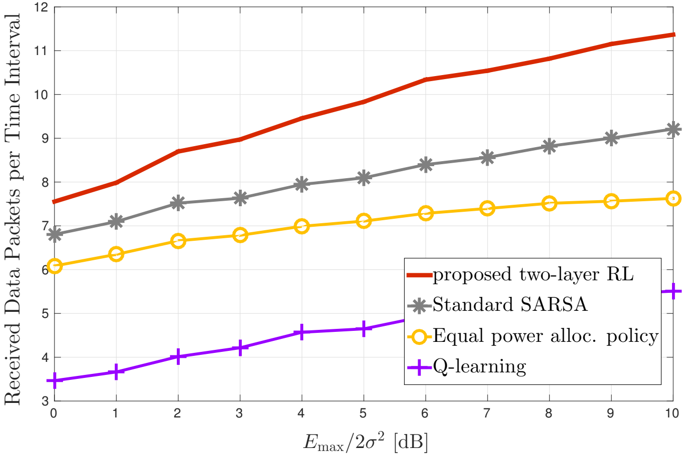
Summary

TECHNISCHE
UNIVERSITÄT
DARMSTADT



Performance Results: Received Data Packets vs. $E_{\mathbf{max}}/2\sigma^2$

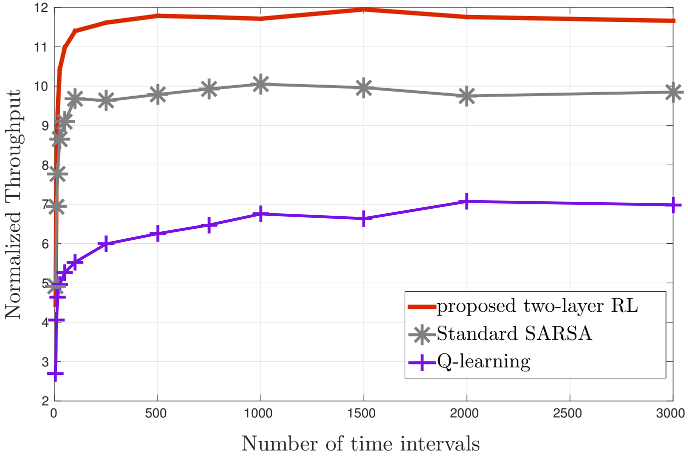
I = 1000, $E_i \in \mathcal{U}(0, E_{\mathbf{max}})$, K = 3, d = 200 **kbits**



Performance Results: Throughput vs. Number of Time Intervals

TECHNISCHE UNIVERSITÄT DARMSTADT

$E_i \in \mathcal{U}(0, E_{\mathbf{max}})$, $E_{\mathbf{max}}/2\sigma^2$ = 10**dB**, K = 3



Conclusions

TECHNISCHE UNIVERSITÄT DARMSTADT

- A two-layer reinforcement learning algorithm with linear function approximation was proposed to solve the power allocation problem in a data dissemination scenario.
  - Only local causal information available
  - No discretization required for the energy, battery level, data buffer level or channel values
- The proposed feature functions take into account the characteristics of the EH problem
- Better performance compared to standard learning techniques

Conclusions

TECHNISCHE UNIVERSITÄT DARMSTADT

- A two-layer reinforcement learning algorithm with linear function approximation was proposed to solve the power allocation problem in a data dissemination scenario.
  - Only local causal information available
  - No discretization required for the energy, battery level, data buffer level or channel values
- The proposed feature functions take into account the characteristics of the EH problem
- Better performance compared to standard learning techniques

Thank you for your attention!