



A! Aalto University
School of Electrical
Engineering

Model-Based Reinforcement Learning for Integrated Radar & Communications Systems

Dr. Petteri Pulkkinen
Data Scientist, Saab Finland

Co-authors: Dr. Majdoddin Esfandiari, Prof. H. Vincent Poor, and Prof. Visa Koivunen

The spectrum congestion problem

- The number of systems using radio frequencies increases all the time
- Most transmissions originate from radar and communications systems
- *One solution:* enable radar and communications systems to share spectrum resources
 - Coexist and co-operate

UNITED STATES FREQUENCY ALLOCATIONS

THE RADIO SPECTRUM

Radio Services



Key Wi-Fi Bands



Allocation Usage Designation

SERVICE	EXAMPLE	DESCRIPTION
Primary	FM/AM	Capital letters
Secondary	Mobile	Initial capital followed by lowercase letters

NTIA
U.S. DEPARTMENT OF COMMERCE
National Telecommunications and Information Administration
Office of Spectrum Management
September 2025
Anella Nath, Assistant Secretary of Commerce for Communications and Information

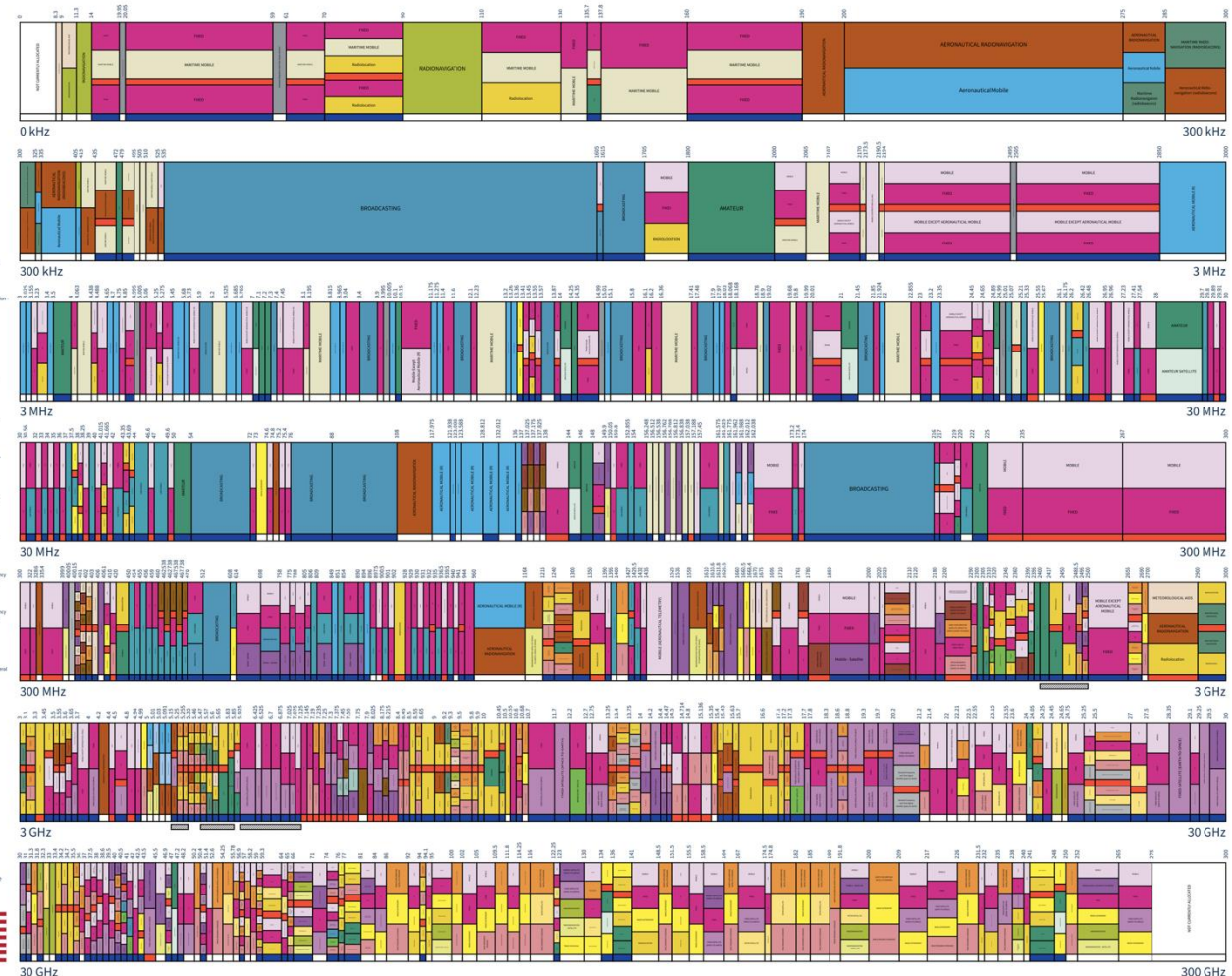
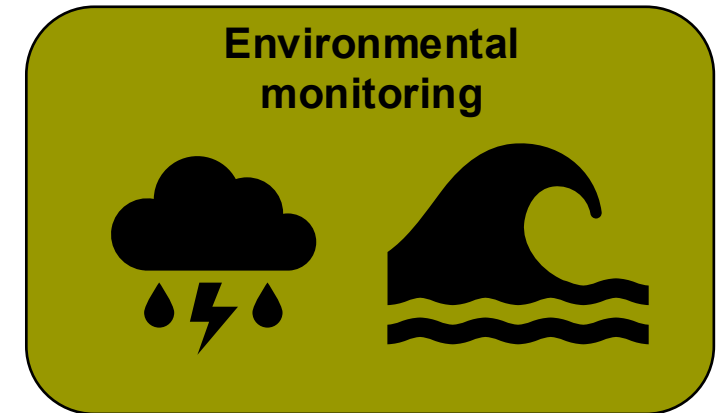
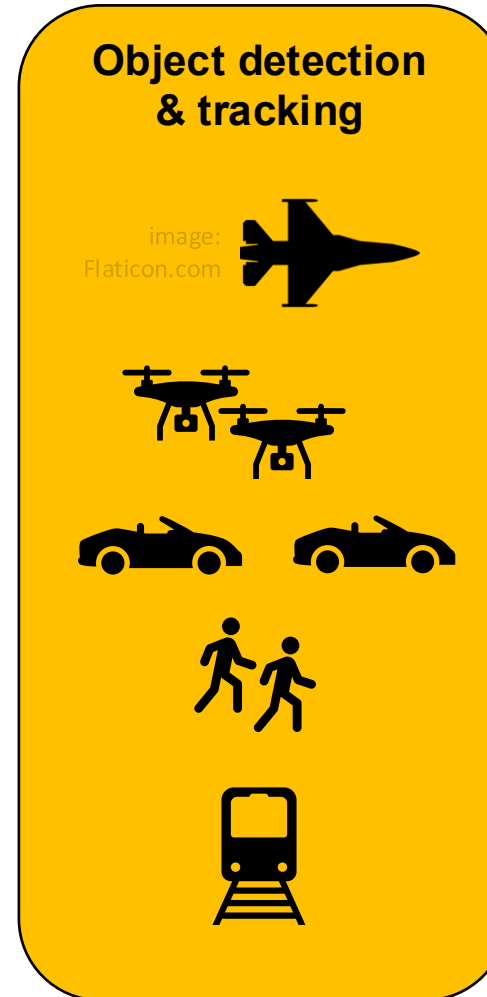


Image: <https://www.ntia.gov/page/united-states-frequency-allocation-chart>



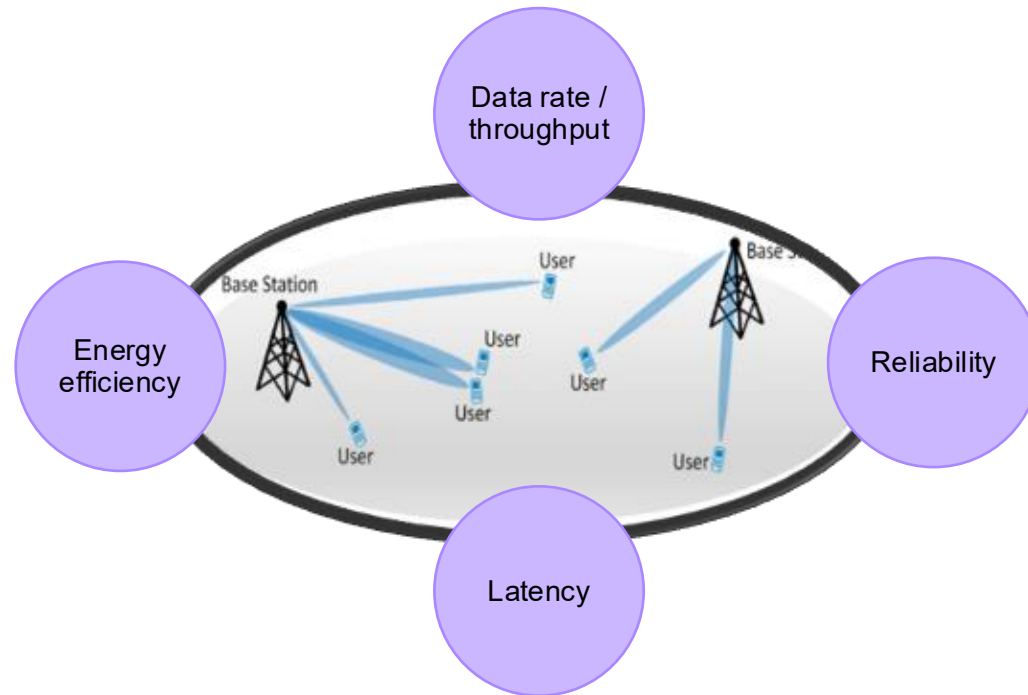
Integrated Sensing & Communications (ISAC)

- **Paradigm shift:** integration of radar and communications into unified ISAC systems
 - Both functions potentially operated on the same hardware, portion of the spectrum, and waveforms
 - Multifunction transceiver, circuit and antenna resources
- Reduce **the system cost** and **power consumption** while also **mitigating congestion** in the radio spectrum
- A key new technology envisioned for the emerging **6G wireless networks**

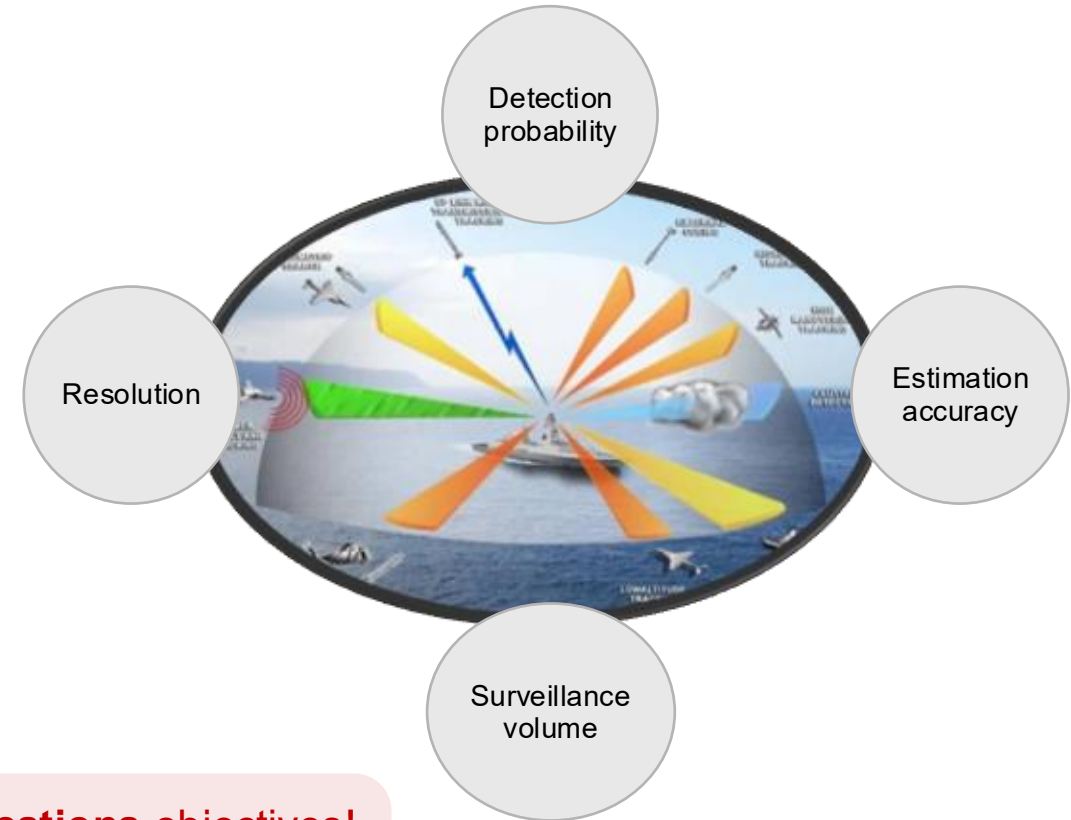


ISAC waveform design and resource management

Communications objectives



Radar objectives

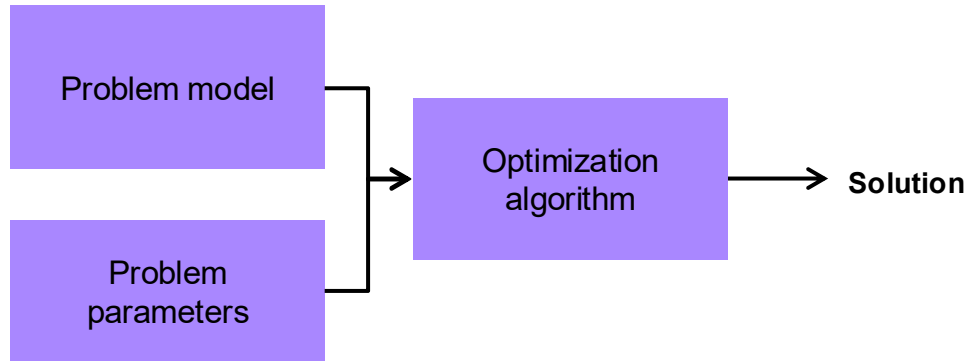


ISAC design should account both **radar** and **communications** objectives!

- Objectives may be conflicting → balancing trade-offs needed
- E.g., via multi-objective or constrained optimization

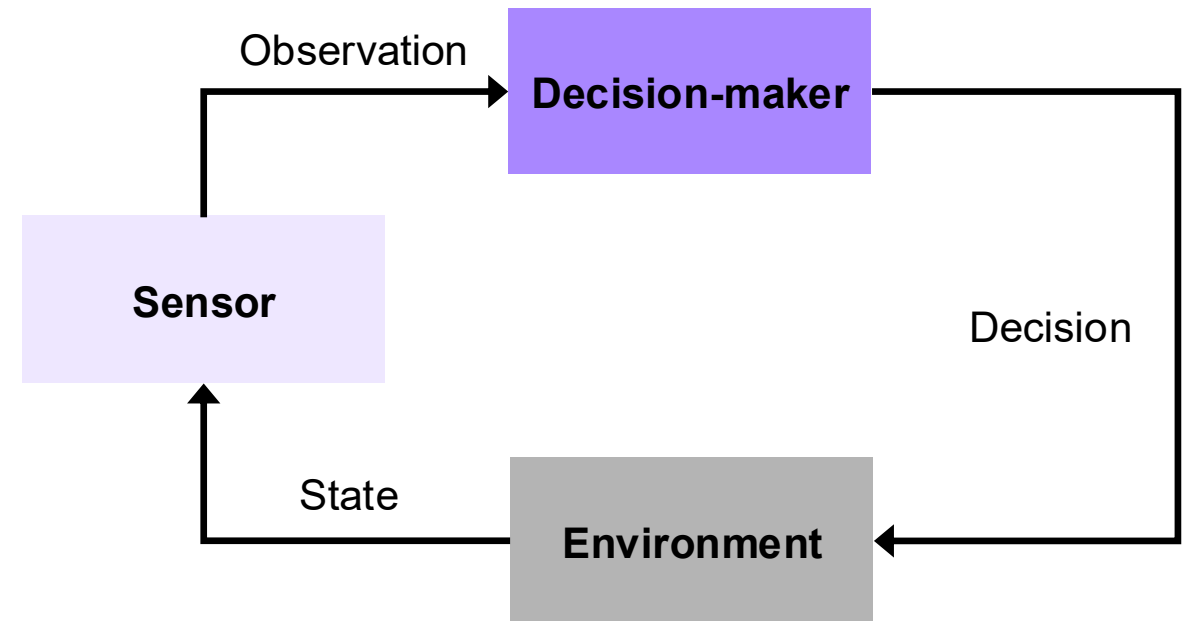
From per-slot optimization to sequential decision-making

Majority of literature treat resource allocation problems in radar and communications as **per-slot optimization** problems



Several weaknesses with this approach:

- May require accurate knowledge of the model parameters (e.g., perfect CSI and target locations)
- Cannot capture the inherent sequential and dynamic nature of the resource allocation tasks
- Cannot learn from past experiences
- May suffer from modeling deficits



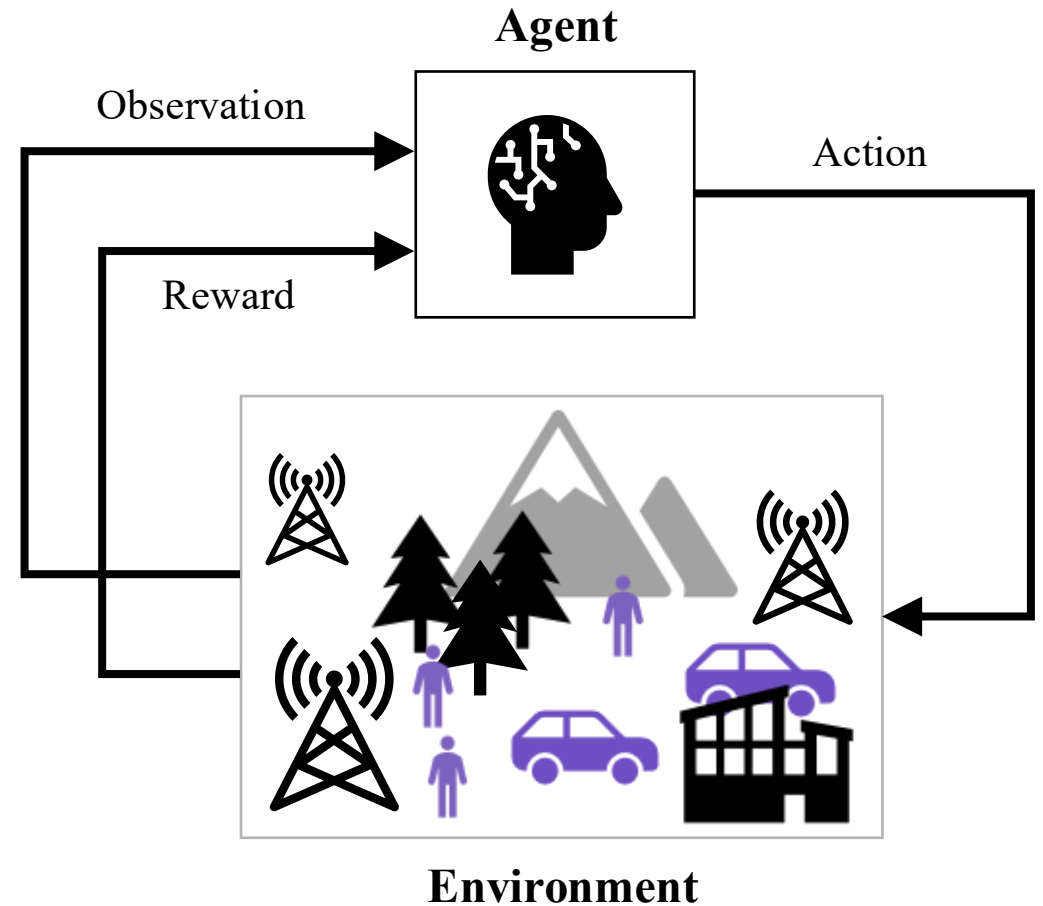
Modeling resource allocation as sequential decision-making allows developing algorithms that

- Consider long-term objectives
- Learn from experiences in real-time
- Handle the stochasticity in observations and environment dynamics in a rigorous manner

Reinforcement Learning (RL)

- **Reinforcement Learning (RL)** methods solve sequential decision-making problems via machine learning-based mechanisms
- Task is to maximize cumulative sum of rewards given the observed feedback from the environment
- Most RL methods do not require any knowledge about
 - Environment dynamics
 - Observations models
 - Rewards

RL methods could be used to address inherent complexity, modeling deficits, and highly dynamic nature of ISAC resource management problems



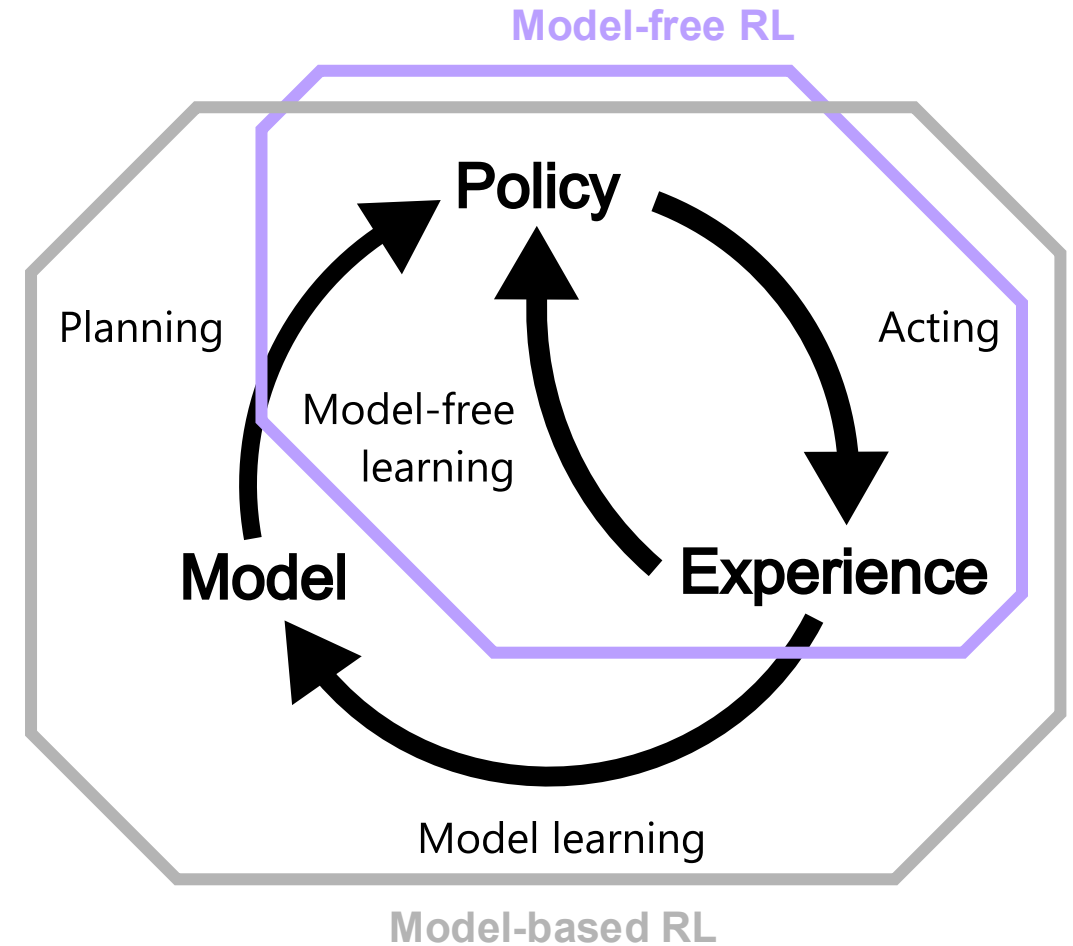
Model-based RL for ISAC resource allocation

Properties of typical ISAC resource allocation problems that make applying RL difficult:

- The decision spaces are very large and involve discrete and continuous variables
- Trial-and-error interactions with the environment are costly
- The environment state is not fully observable

Model-based RL:

- Related to the field of control theory
- Enables exploiting rich structural information about communications and radar systems
- Avoids tedious real-world trial-and-error interactions via predicting rewards and states using the model
- Use learning only where it gives added value!**

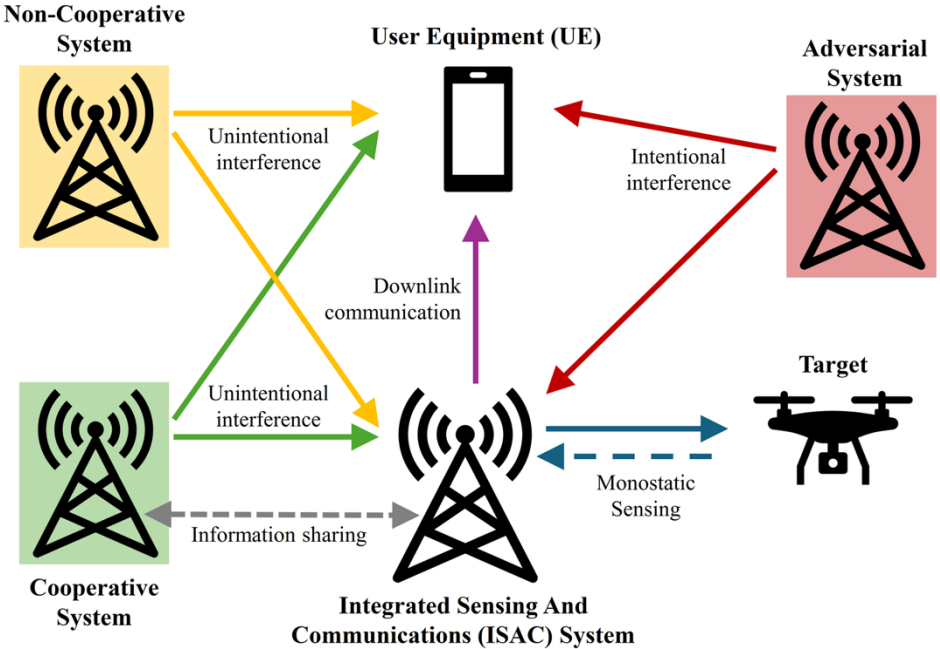


Practical examples of model-based RL in ISAC resource allocation

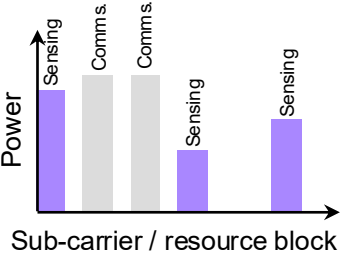
Frequency domain resource allocation under dynamic radio interference environments

Joint frequency & spatial domain resource allocation under dynamic radar target environments

Frequency resource allocation under dynamic interference



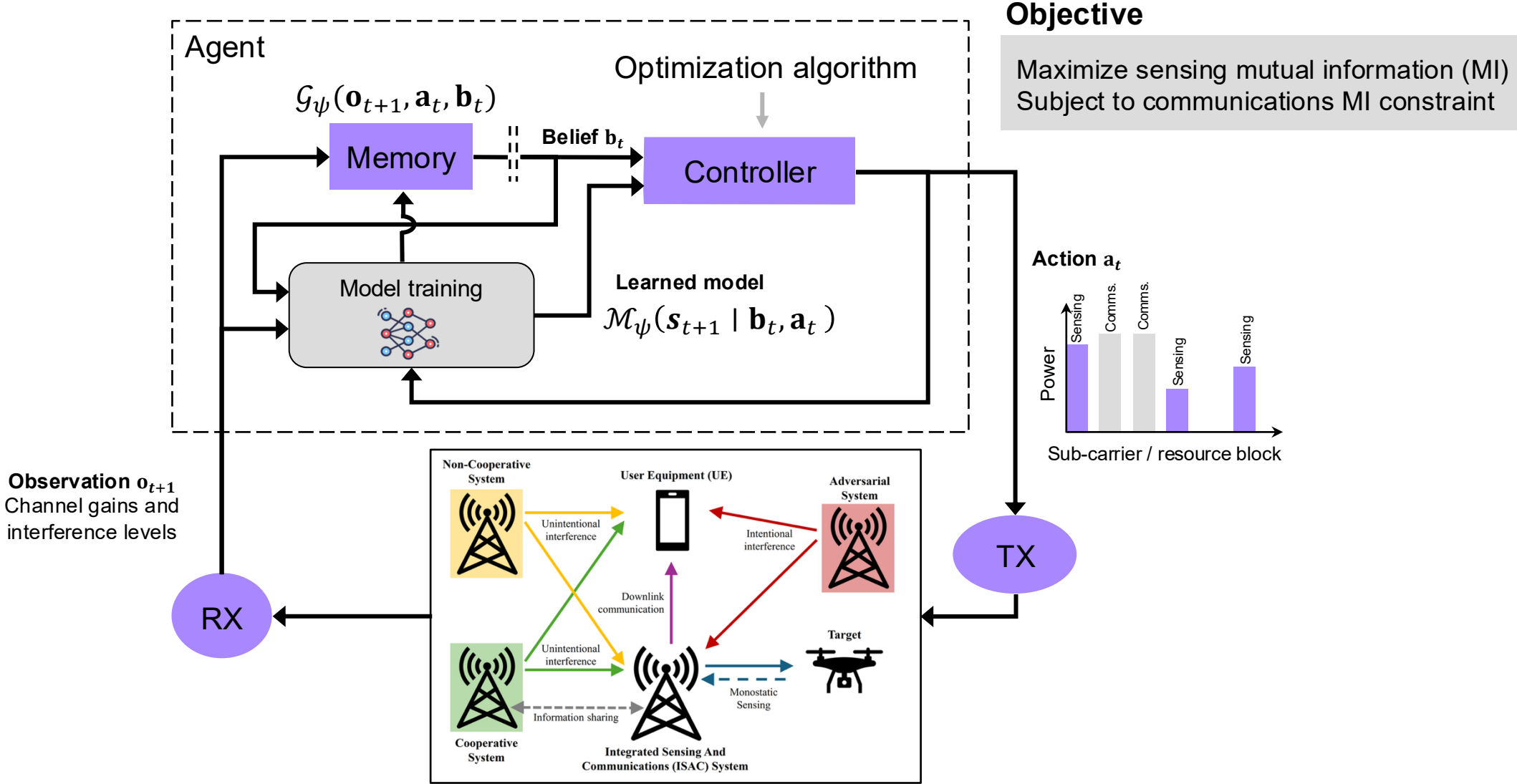
Resources to allocate



Objective

Maximize sensing mutual information (MI)
Subject to communications MI constraint

Frequency resource allocation under dynamic interference



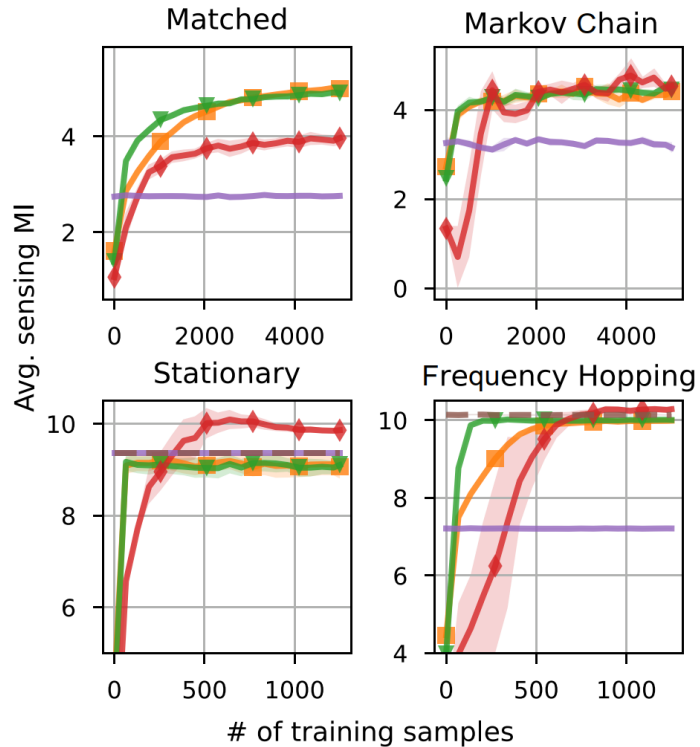
$$\mathbf{s}_{t+1} \sim \mathcal{M}(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t)$$



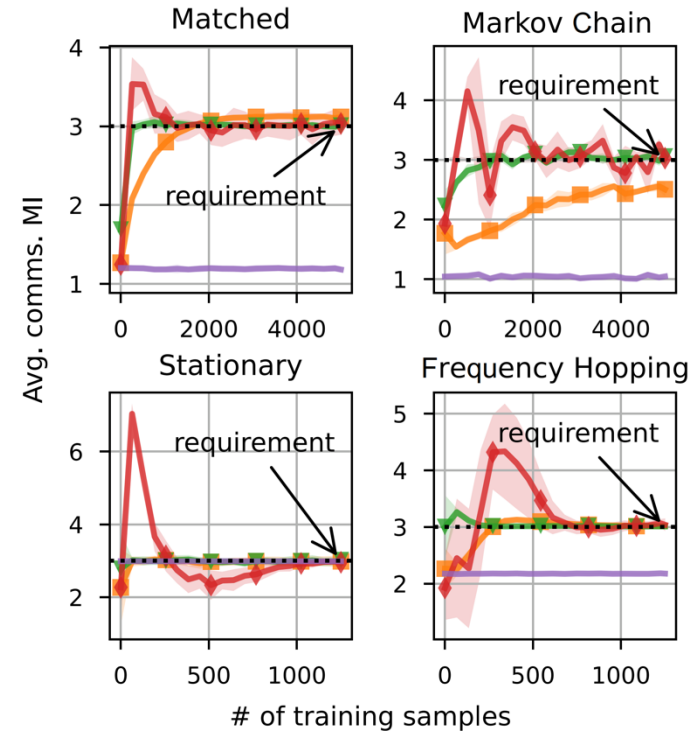
Frequency resource allocation under dynamic interference



MBRL methods, and sample efficient MFRL are proposed in our publications!



(a) Average sensing MI in different interference scenarios.



(b) Average communications MI in different interference scenarios.

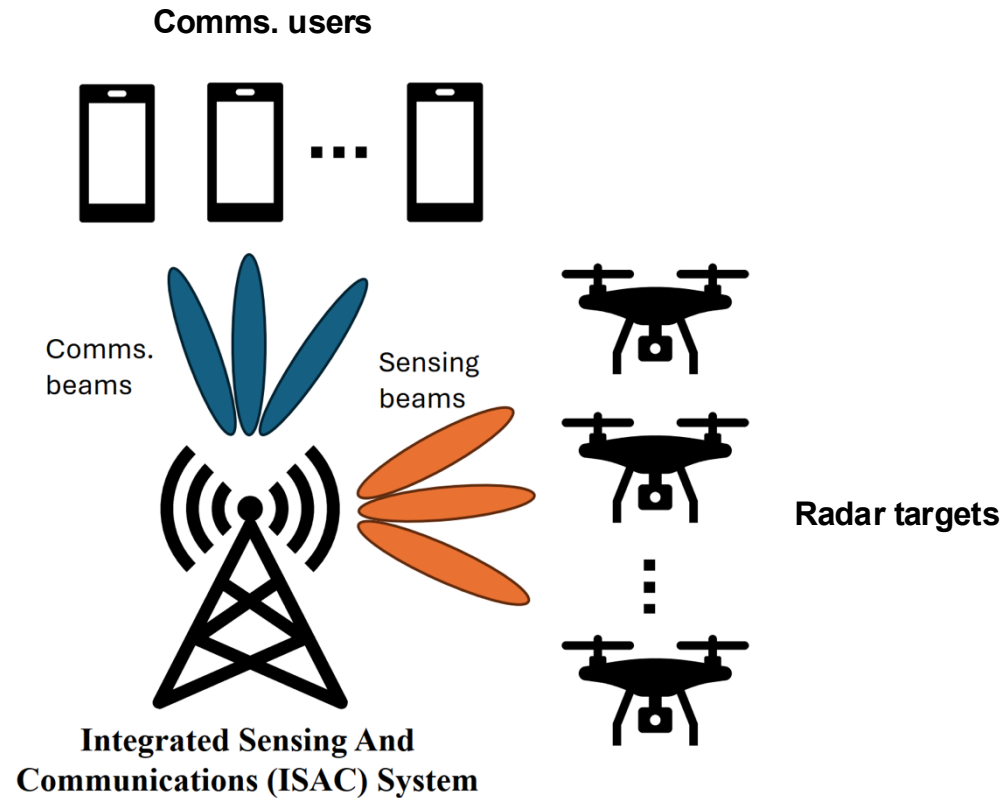
- 1) MBRL methods offer faster convergence speeds than the sample-efficient MFRL method
- 2) MFRL method may achieve better performance in static scenarios with large number of training samples

Practical examples of model-based RL in ISAC resource allocation

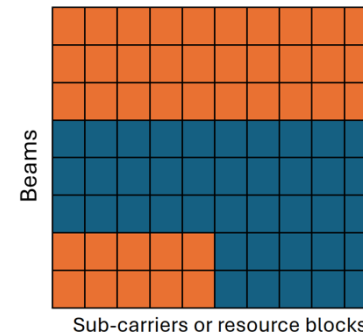
Frequency domain resource allocation under dynamic radio interference environments

Joint frequency & spatial domain resource allocation under dynamic radar target environments

Joint frequency–beam-space resource allocation



Resources to allocate



Orange cells: comms. allocations
Blue cells: sensing allocations

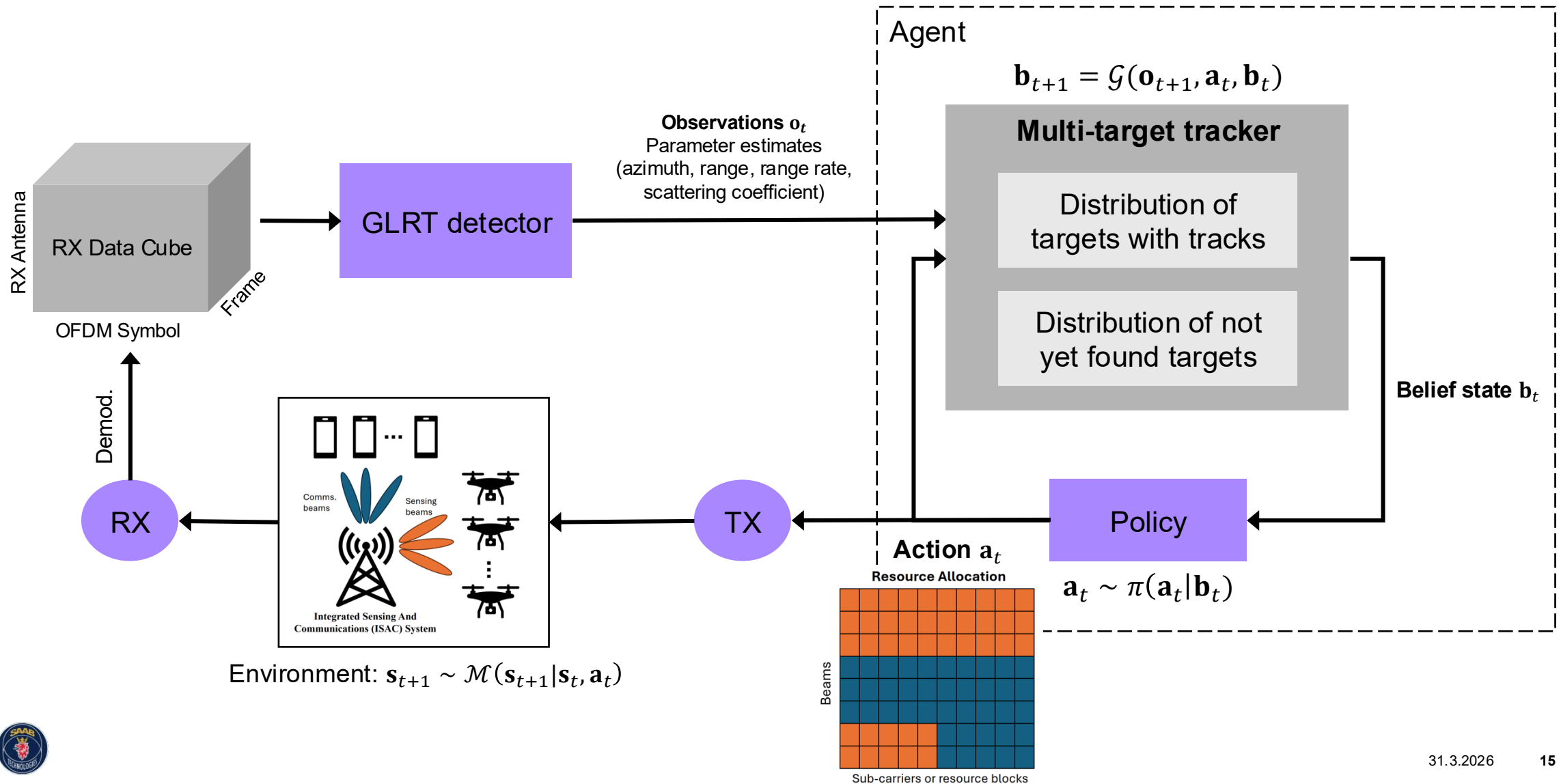
Each beam–frequency resource unit determines functionality (sensing/communication), transmit power, and user assignment for communications.

Objective

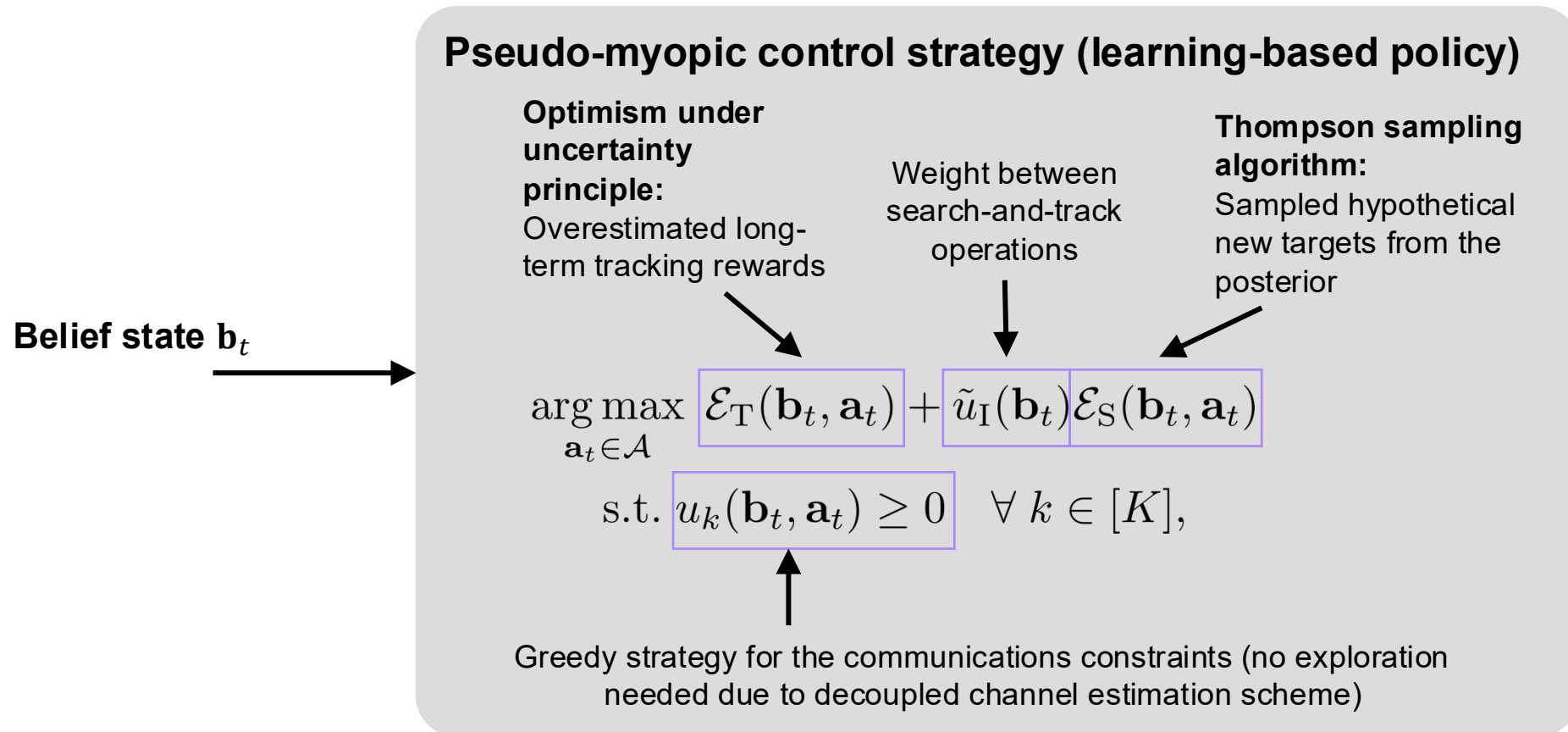
Maximize long-term radar objective supporting search-and-track operation

Subject to communications rate constraints for each user

Joint frequency–beam-space resource allocation

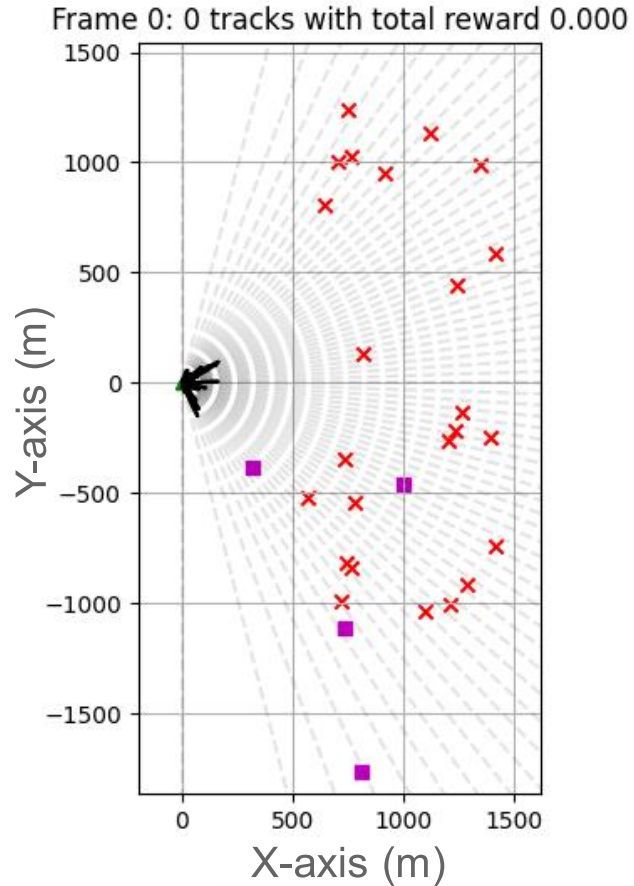


Joint frequency–beam-space resource allocation

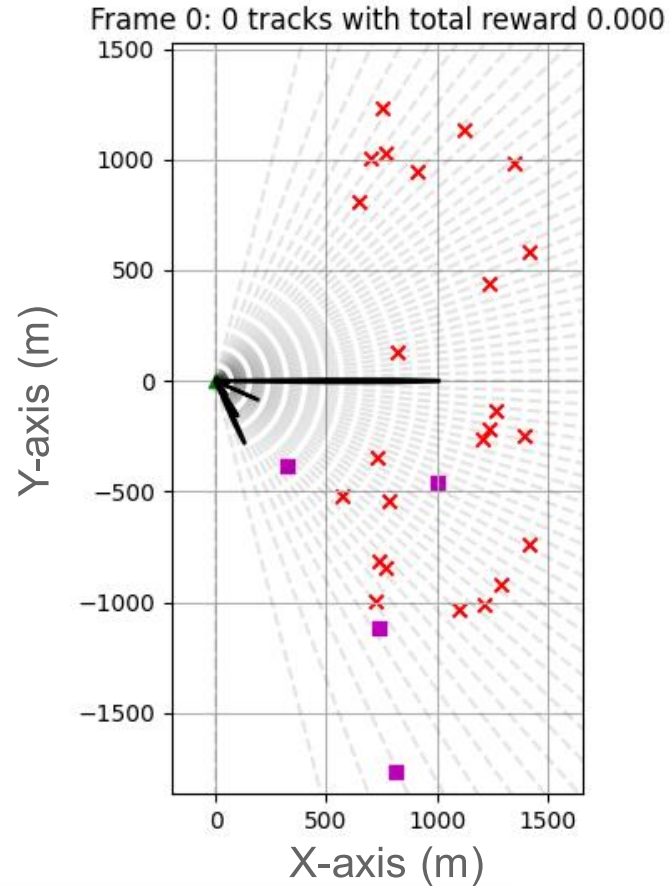


Joint allocation of beam and frequency resources

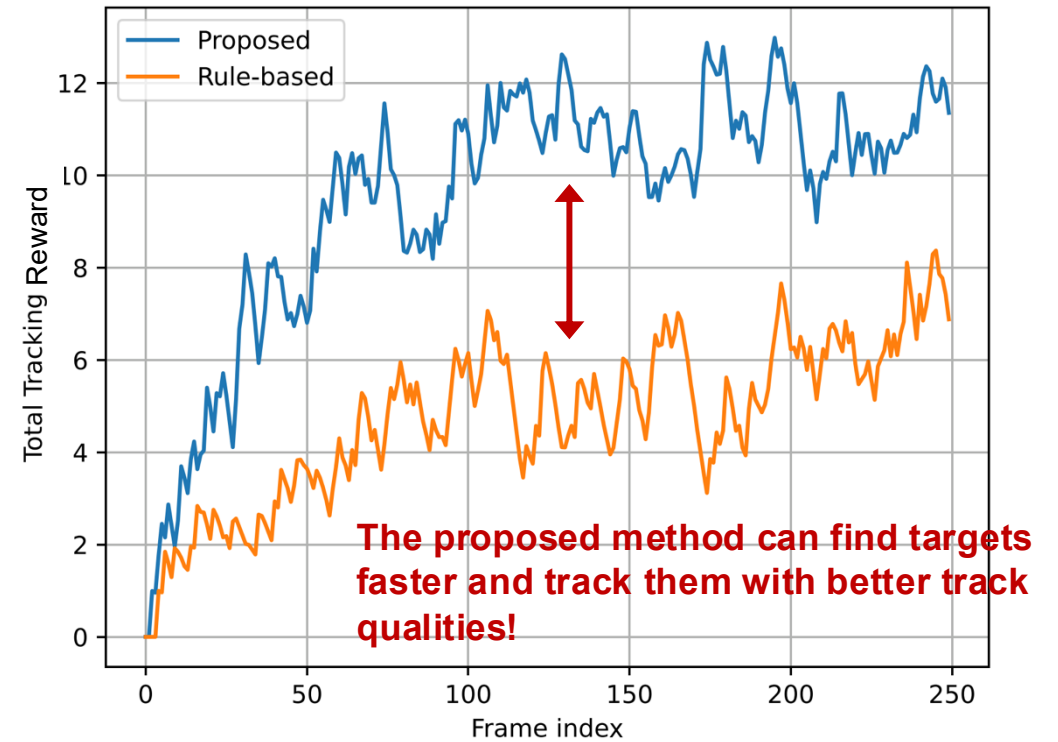
Proposed model-based RL method



Rule-based method



System parameter	Value
# of subcarriers	1200 (15 kHz spacing)
# of resource blocks	100 (12 subcarriers / RB)
# of beams	64 (DFT beams)



Conclusions

- Integrated radar sensing and communications system benefit from intelligent waveform design and resource management
 - Particularly useful in dynamic and congested spectrum environments
- Model-based RL enables
 - Learning from experiences in real-time for continuous and autonomous performance improvements
 - Exploiting rich knowledge on communications and radar systems as well as their operational environments
 - Avoiding tedious trial-and-error process required by the model-free RL
- Methods shown to be **data-efficient and practical** in multicarrier & multiantenna ISAC
- Potential future research directions:
 - Decision-making in distributed ISAC systems
 - Real-world implementations & demonstrations

Publications:

P. Pulkkinen and V. Koivunen. Model-Based Online Learning for Resource Sharing in Joint Radar-Communication Systems. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, pp. 4103-4107, May 2022.

P. Pulkkinen and V. Koivunen. Model-Based Online Learning for Joint Radar-Communication Systems Operating in Dynamic Interference. In *Proceedings of the 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, pp.992-996, Aug 2022.

P. Pulkkinen and V. Koivunen. Model-Free Online Learning for Waveform Optimization In Integrated Sensing And Communications. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, pp.1-5, Jun 2023.

P. Pulkkinen and V. Koivunen. Model-Based Online Learning for Active ISAC Waveform Optimization. *Journal of Selected Topics in Signal Processing*, vol. 18, no. 5, pp. 737-751, July 2024.

P. Pulkkinen and V. Koivunen. Partially Observable Model-Based Learning For ISAC Resource Allocation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, pp.12996-13000, Apr 2024.

P. Pulkkinen, M. Esfandiari and V. Koivunen. Cognitive BeamSpace Algorithm for Integrated Sensing and Communications. In *Proceedings of the IEEE Radar Conference (RadarConf)*, Denver, CO, USA, pp.1-6, May 2024.

P. Pulkkinen, M. Esfandiari and V. Koivunen. BeamSpace and Frequency Domain ISAC Resource Allocation using Reinforcement Learning. In *Proceedings of the 58th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 443-449, Oct 2024.

P. Pulkkinen, M. Esfandiari, H.V Poor and V. Koivunen. Multicarrier MIMO ISAC Resource Allocation Using Model-Based Reinforcement Learning. Submitted to *IEEE Transactions on Signal Processing*, May 2025.

P. Pulkkinen and V. Koivunen. Augmented Lagrangian controller design in model-based reinforcement learning for ISAC resource allocation. Accepted to 2026 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2026)*.

Thank you!

Model-based reinforcement learning for integrated radar and communications systems

Petteri Pulkkinen

A”
Aalto University

