Reinforcement Learning Algorithms for Charged Particle Tracking with Applications in Proton Computed Tomography

Tobias Kortus [1] Ralf Keidel[1] Nicolas R. Gauger[2]

[1] Center for Technology and Transfer, University of Applied Sciences Worms
[2] Chair for Scientific Computing, University of Kaiserslautern-Landau (RPTU)

on behalf of the Bergen pCT collaboration

6th IML Workshop – January 30, 2024

Proton Computed Tomography and Particle Tracking

- **Goal**: Reconstruct path of protons in multi-layer detector after patient to obtain sufficient information (energy, direction) required for image reconstruction.
- Bergen pCT detector prototype [1]:
  - 2 tracking layer
  - 41 detector-absorber layer

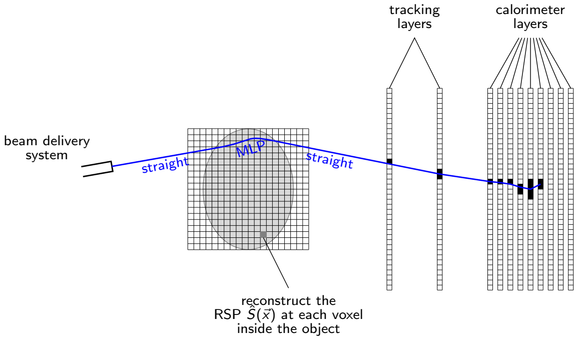
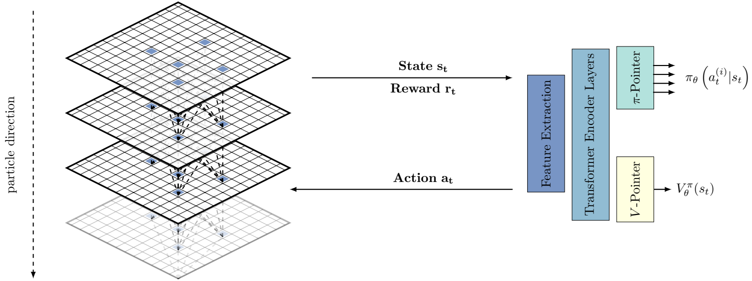


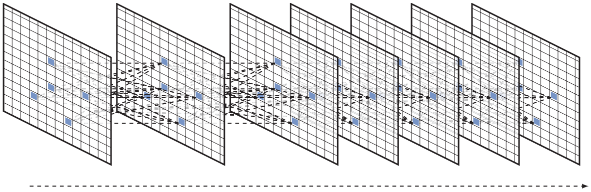
Image courtesy: Aehle et al., 2023 (`https://doi.org/10.1088/1361-6560/ad0bdd`)

Reinforcement Learning for Particle Tracking



- **Goal**: Find a good reconstrution policy π^* by interacting with the environment.
- **Policy**: Decision strategy of the agent for each given state.
- **Value**: How good is a state in the long run (expected discounted future reward).

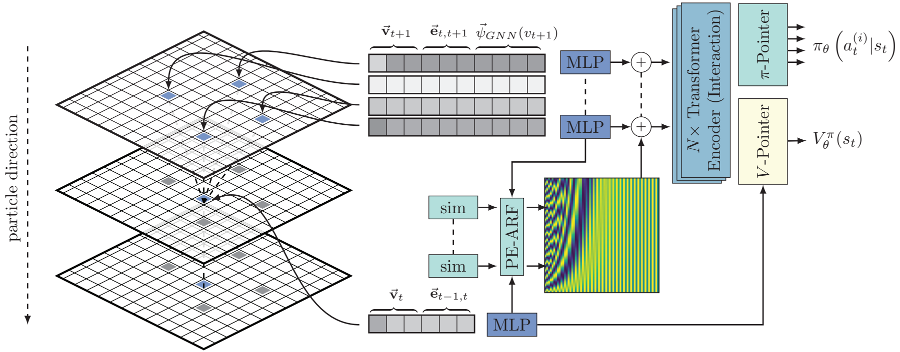
Representation as a Directed Acyclic Graph



- Directed acylic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - $\mathcal{V} = \{v_i\}_{\{i=1,N\}}$: Particle hit centroids
  - $\mathcal{E} = \{e_{ij}\}_{\{i=1,M\}}$: Possible track segments (actions) in opposite direction to the traversal direction.

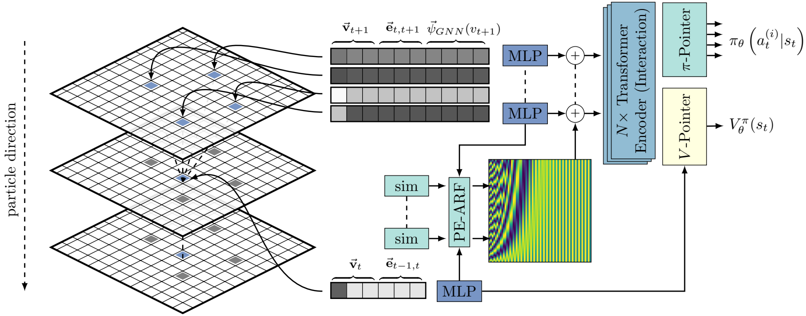
- Parametrization of vertices and edges as
  - $\vec{v} = \{\Delta E, x, y, z\}$
  - $\vec{e} = \{r, \theta, \phi\}$
- Environment dynamics defined by the edges between particle hits.

Single-Agent Network Architecture



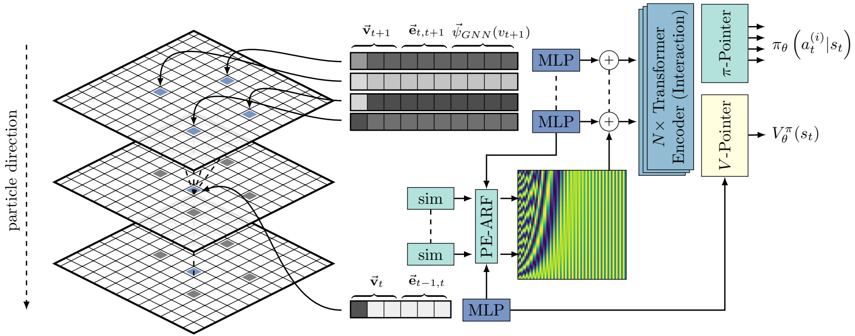
Single-Agent Network Architecture

Features from node, edge features and node embedding (using GNN):
→ last reconstructed segment: h_{obs}
→ possible next segments: $h_{act}^{(i)}$

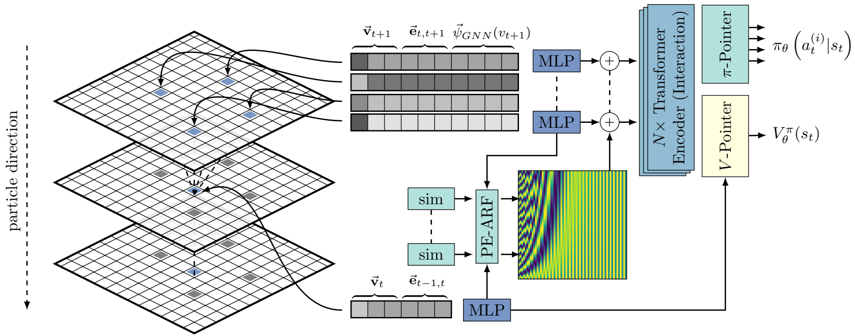


Single-Agent Network Architecture

Exchange information between action candidates:
→ updated h_{act} using dot-product attention mechanism [2]
→ operable with dynamic number of actions
→ permutation invariant operation



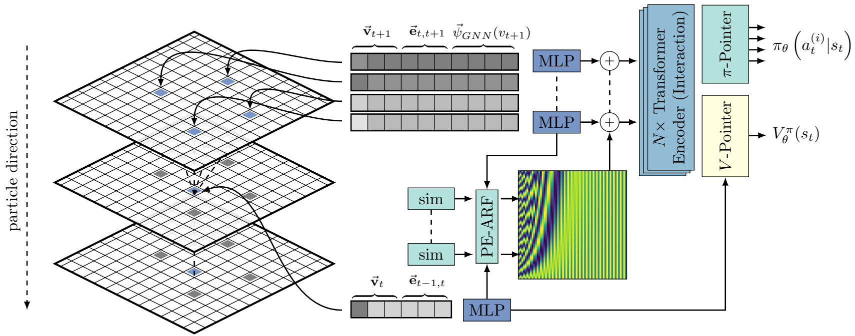
Single-Agent Network Architecture



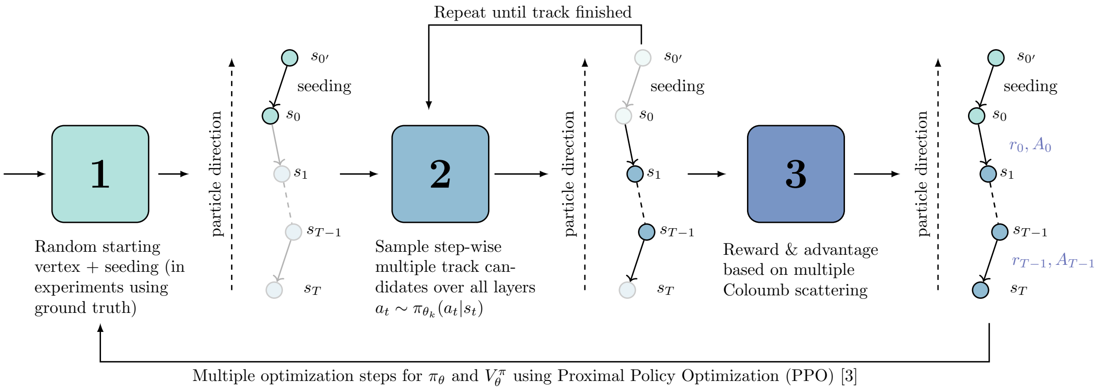
PE-ARF: positional information based on similarity + adaptive "focus" depending on input features

Single-Agent Network Architecture

Pointer network: $v^T \tan(W_1 h_{obs} + W_2 h_{act}^{(i)})$ [6]:
$\rightarrow$ scalar output for each $h_{act}^{(i)}$, h_{obs} combination



Single-agent Optimization of Behavior Policy



Limitations of Single-Agent Reinforcement Learning

Limitations of Single-Agent Reinforcement Learning

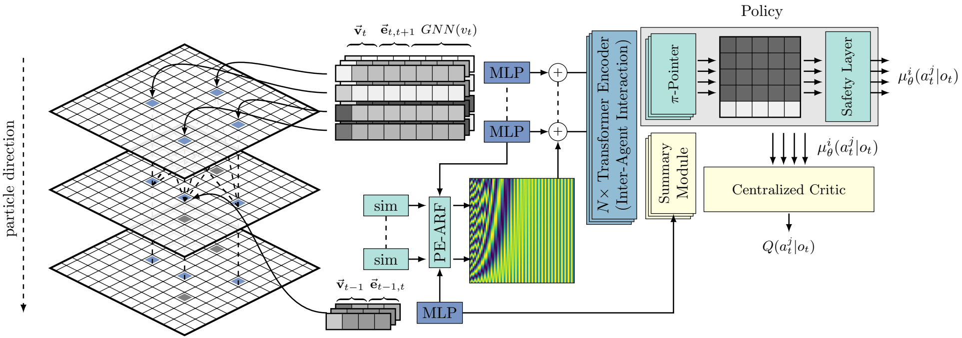
- **Partial observability:** reconstruction w.r.t. entire readout frame remains still partial observable (other tracks are not taken into consideration)
- **Ambiguities in assignments:** Conflicts in reconstruction can assign the same particle to multiple tracks $\rightarrow$ generation of implausible tracks.

Multi-Agent Reinforcement Learning (Work in Progress)

Design considerations for training MARL agents for particle tracking:

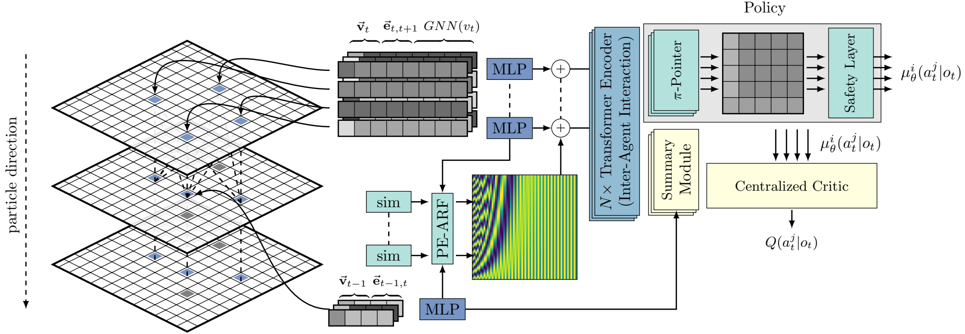
1. **Dec-POMDP:** Consider multiple decentralized agents (similar to single-agent) with only local observations per agent, only limited communication → minimal performance impact by avoiding global information or complex communication protocols.
2. **CTDE:** Use information during training that would be unavailable during inference (centralized critic) → Better training performance (reduces instationarity).
3. **Constraints**. Enforce agreement between agents → unique particle assignment (constraint satisfaction by designing a safety layer [4], [5]).

Multi-Agent Network Architecture (Work in Progress)



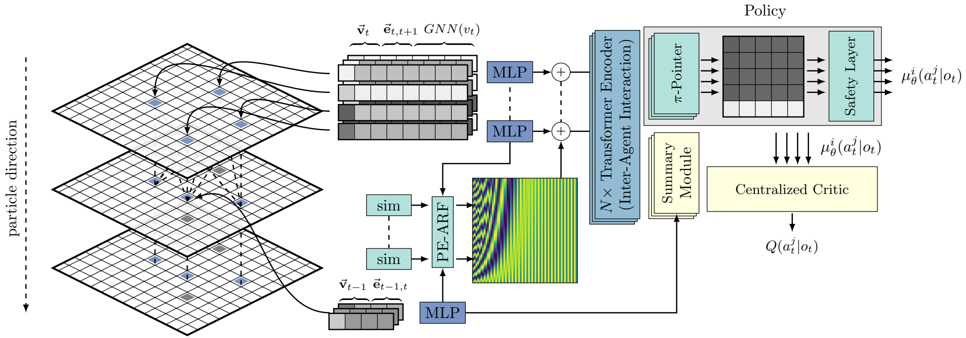
Multi-Agent Network Architecture (Work in Progress)

$$c_{ij} = \begin{cases} d\left(\mu_\theta^i \| \mathbb{1}(a^j)\right) & \text{if } a^j \text{ not used for seeding} \\ \infty & \text{otherwise} \end{cases}$$

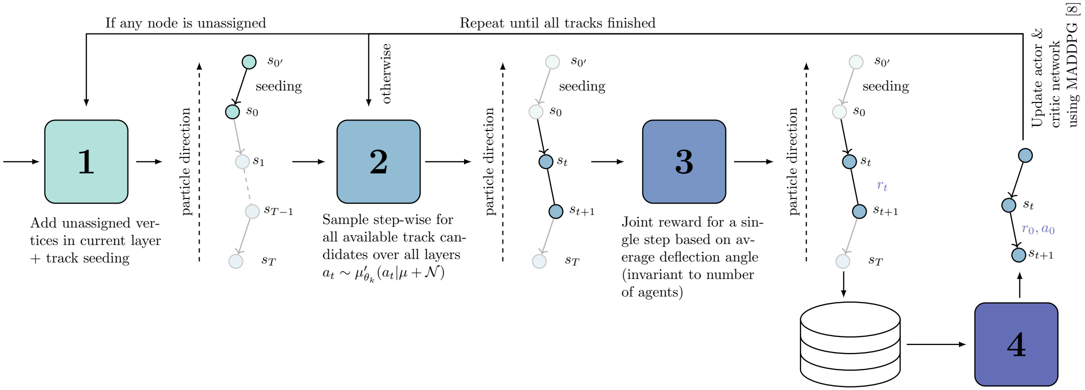


Multi-Agent Network Architecture (Work in Progress)

$\min_{M'} \quad \sum_{(i,j)\in\mathcal{E}} \mu'_{ij} c_{ij}, \quad \text{where} \quad \mu'_{ij} \in \{0,1\}$
with implicitly interpolated gradients using [7]



Multi-agent Optimization of Behavior Policy



Add unassigned vertices in current layer + track seeding

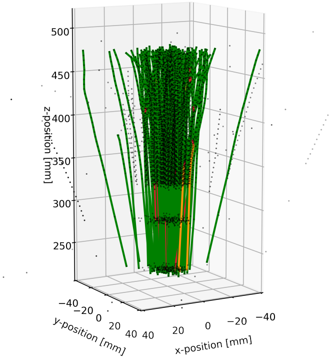
Sample step-wise for all available track candidates over all layers $a_t \sim \mu'_{\theta_k}(a_t|\mu + \mathcal{N})$

Joint reward for a single step based on average deflection angle (invariant to number of agents)

Sample random transitions $\langle\cdot\rangle \sim \mathcal{D}$ from replay buffer.

Preliminary Results (Work in Progress)

		100 mm Water		150 mm Water		200 mm Water	
Density	Algorithm	p [%]	ϵ [%]	p [%]	ϵ [%]	p [%]	ϵ [%]
100	Search [9]	83.0±0.0	74.6±0.0	86.5±0.0	79.0±0.0	87.4±0.0	80.3±0.0
	PPO [10]	85.6±0.3	75.2±0.5	88.8±0.5	79.0±0.5	89.5±0.4	80.8±0.5
	MADDPG*	**90.8**±–.–	**75.7**±–.–	**92.8**±–.–	**79.1**±–.–	**93.1**±–.–	**81.0**±–.–
150	Search [9]	79.1±0.0	70.9±0.0	83.2±0.0	**75.7±0.0**	84.7±0.0	**77.7±0.0**
	PPO [10]	80.5±0.4	70.8±0.3	83.8±0.7	74.4±0.6	85.3±0.6	76.9±0.5
	MADDPG*	**87.3**±–.–	**71.0**±–.–	**89.1**±–.–	74.2±–.–	**90.7**±–.–	73.6±–.–
200	Search [9]	75.4±0.0	**67.4±0.0**	80.1±0.0	**72.9±0.5**	81.6±0.4	**75.0±0.0**
	PPO [10]	75.3±0.6	66.6±0.6	80.0±0.8	70.9±0.6	81.7±0.6	73.8±0.5
	MADDPG*	**83.0**±–.–	65.6±–.–	**86.9**±–.–	71.1±–.–	**87.7**±–.–	73.6±–.–



Reconstruction performance, measured in terms of purity p and efficiency ϵ for different configurations. Results marked with * are cherry-picked runs due to instability of training. Results for *Search* and *PPO* are taken from [10].

Conclusion and Current/Future Work

Conclusion

- Reinforcement learning proves to be a promising optimization technique for track reconstruction **leveraging deep neural networks** while **requiring no manual supervision**.
- Architecture allows for **generalization to previously unseen phantom geometries and particle densities**.
- More information & results: `10.1109/tpami.2023.3305027`

Current/Future Work

- When reconstructing a single, the system remains still partial observable (influence of other tracks). $\rightarrow$ **Multi-Agent Reinforcement Learning (MARL)**.
- First promising results, further work required to stabilize training.

The Bergen pCT Collaboration

- University of Bergen, Norway
- Helse Bergen, Norway
- Western Norway University of Applied Science, Bergen, Norway
- Wigner Research Center for Physics, Budapest, Hungary
- DKFZ, Heidelberg, Germany
- Saint Petersburg State University, Saint Petersburg, Russia
- Utrecht University, Netherlands
- RPE LTU, Kharkiv, Ukraine
- Suranaree University of Technology, Nakhon Ratchasima, Thailand
- China Three Gorges University, Yichang, China
- University of Applied Sciences Worms, Germany
- University of Oslo, Norway
- Eötvös Loránd University, Budapest, Hungary
- University of Kaiserslautern Landau, Germany

Contact: kortus@hs-worms.de

References I

[1] J. Alme *et al.*, "A high-granularity digital tracking calorimeter optimized for proton ct," *Frontiers in Physics*, vol. 8, pp. 1–20, October 2020, ISSN: 2296424X. DOI: 10.3389/fphy.2020.568243.

[2] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 5999–6009, Nips 2017, ISSN: 10495258.

[3] J. Schulman *et al.*, "Proximal policy optimization algorithms,", pp. 1–12, 2017. [Online]. Available: http://arxiv.org/abs/1707.06347.

[4] G. Dalal *et al.*, "Safe exploration in continuous action spaces,", 2018. [Online]. Available: http://arxiv.org/abs/1801.08757.

[5] Z. Sheebaelhamd *et al.*, "Safe deep reinforcement learning for multi-agent systems with continuous action spaces,", 2021. [Online]. Available: http://arxiv.org/abs/2108.03952.

References II

[6] O. Vinyals *et al.*, "Pointer networks," *Advances in Neural Information Processing Systems*, vol. 2015-Janua, pp. 2692–2700, 2015, ISSN: 10495258.

[7] M. Vlastelica *et al.*, "Differentiation of blackbox combinatorial solvers," *8th International Conference on Learning Representations, ICLR 2020*, pp. 1–19, 2020.

[8] R. Lowe *et al.*, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 6380–6391, 2017, MADDPG, ISSN: 10495258.

[9] H. E. Pettersen *et al.*, "Proton tracking algorithm in a pixel-based range telescope for proton computed tomography," *arXiv*, 2020, ISSN: 23318422.

[10] T. Kortus *et al.*, "Towards neural charged particle tracking in digital tracking calorimeters with reinforcement learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, ISSN: 19393539. DOI: `10.1109/TPAMI.2023.3305027`.