Tracking of Proton Traces in a Digital Tracking Calorimeter using Reinforcement Learning

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, TU Kaiserslautern

On behalf of the Bergen pCT collaboration and the SIVERT research training group

May 13, 2022

Proton Computed Tomography & The Bergen pCT Detector [1]

- **Proton CT**: Alternative imaging technique to conventional computed tomography → promises reduced uncertainties for proton/hadron therapy treatment planning.
- Bergen (Norway) pCT collaboration develops novel pCT scanner completely based and the ALPIDE pixel sensor.
- Consists of 41 detector absorber layers and 2 tracking layers.

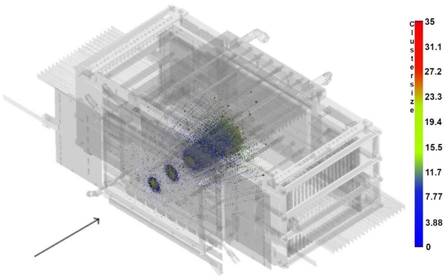
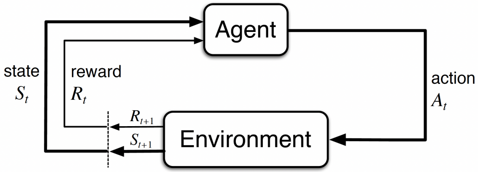


Image courtesy of Alexander Wiebel

[1] Alme et al. A High-Granularity Digital Tracking Calorimeter Optimized for Proton CT. Frontiers in Physics.

Reinforcement Learning



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

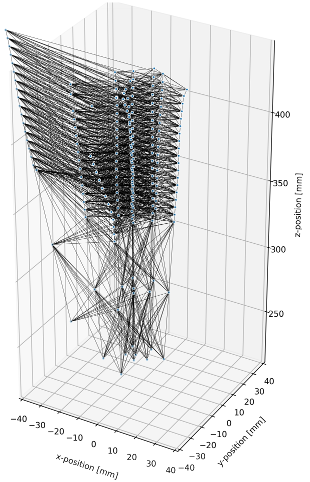
- **Basic idea**: Learn a policy from raw data that optimizes the physical plausibility of the reconstruction.

Detector Graph Generation

- All possible combinations of proton tracks as a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with:
- $v \in \mathcal{V}$: Particle hits in the detector.
- $e \in \mathcal{E}$ Possible track segments connecting two hits of adjacent layers (reversed $\rightarrow$ backward tracking).
- Edge and node features $(\vec{v}, \vec{e})$:

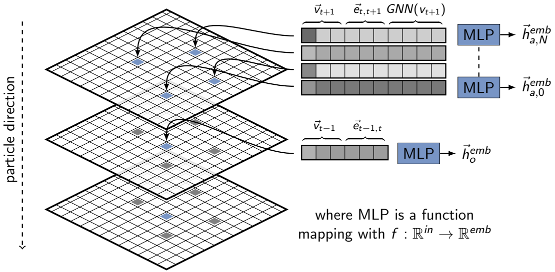
$$\vec{v_i} = (edep, x_i, y_i, z_i) \tag{1.1}$$

$$\vec{e_{ij}} = (r_{ij}, \theta_{ij}, \phi_{ij}) \tag{1.2}$$

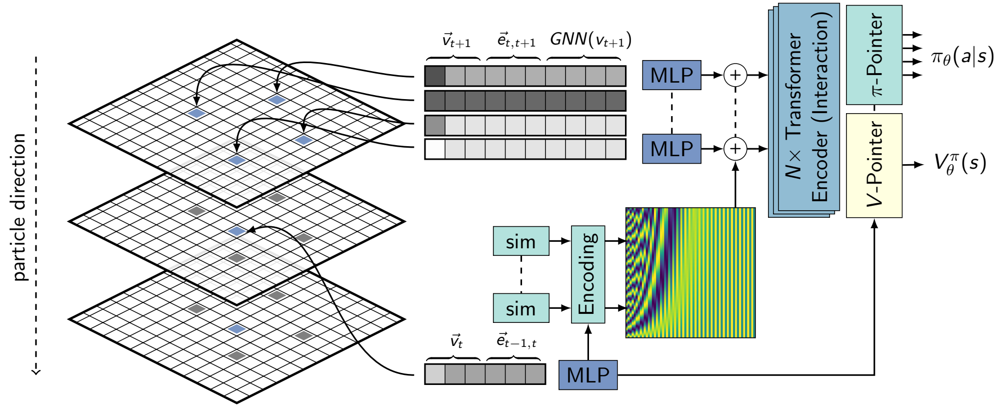


Extraction of Observation- and Action Features

- Select features to provide sufficient history (w.r.t single track).
- Independence of scattering events → considering only a one-step history is sufficient.
- Two different set of features:
  1. *observation-features*: History over last segment.
  2. *action-features*: Collection of possible next segments (correspond to actions).



Network Architecture - Interaction & Pointer Modules

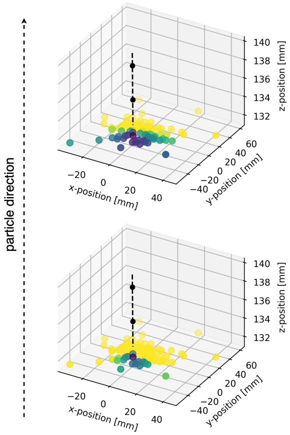


Inductive Bias using Positional Encoding & Dynamic Receptive Fields

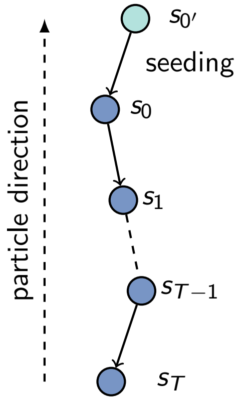
- **Goal:** Encode positional information of different orders of magnitude with fixed spatial resolution → Employ controllable re-scaling using idea of dynamic receptive fields:

$$N_{DRF}(s_{t-1:t}, s_{t:t+1}) = clip\left(\frac{0.5 \cdot (1 - sim)}{\Phi_{clip}(\vec{h}_o^{emb})}, 0, 1\right) \cdot \alpha_{scale} \qquad (2)$$

- where sim denotes the cosine similarity $sim(e_{t-1,t}, e_{t,t+1})$ and $\Phi_{clip} : \mathbb{R}^d \to \mathbb{R}$ denotes a MLP with $clip(\Phi(\vec{h}_o^{emb}), \epsilon, 1)$.



Policy/Value Optimization



- For every training iteration:
  - Initial **"pre-state" sampled from uniform distribution** over last N layers. State definition requires a transition in the detector to be fully parametrized $\rightarrow$ **track seeding** (currently using ground truth).
  - Sample stepwise **multiple track candidates** over all layers $a_t \sim \pi_{\theta_k}(a_t|s_t)$ from environment following the current behavior policy.
  - **Reward & advantage calculations** based on physical likelihood of observing the sampled trajectory (multiple Coulomb scattering).
  - **Multiple optimization steps** for π_θ and V_θ^π using PPO-CLIP.

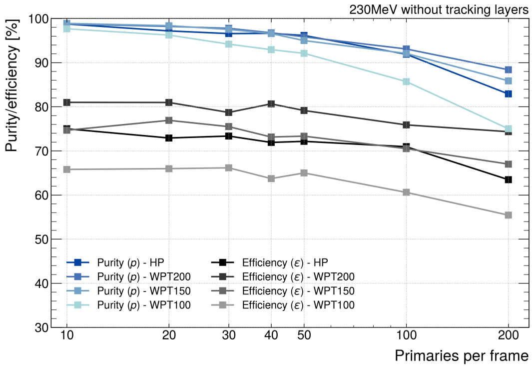
Preliminary Results: Setup

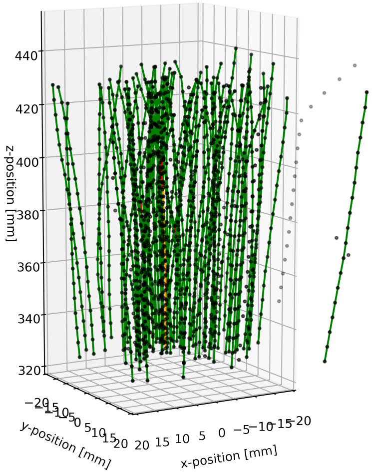
- **Phantoms**: Head phantom [2], water phantoms $t \in \{100, 150, 200\}$ mm with $1e^4$ primaries.
- **Dataset**: Split into N reinforcement learning environments (**tracking layers were removed**) with $M \in \{10, 20, 30, 40, 50, 100, 200\}$ primaries per frame (80/20 train test split).
- **Training**: Train 500 steps on environments with 100 primaries per frame ($\approx$ 15 min)
- **Track filtering**: Thresholds for scattering angle and energy deposition in last layer$\rightarrow$ remove secondaries and tracks leaving the detector.
- **Metrics**: Purity (p) and Efficiency (ϵ) $\rightarrow$ results averaged over 5 runs

$$p = \frac{N_{rec,+}}{N_{rec,+/-}}, \quad \epsilon = \frac{N_{rec,+}}{N_{total}}, \tag{3}$$

[2] Giacometti et al.. Development of a high resolution voxelised head phantom for medical physics applications. Phys Med. 2017 Jan;33:182-188.

Preliminary Results





head phantom, 100 primaries

Conclusion and Outlook

- 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 particle densities**.
- Still some difficulties with optimizing inhomogeneous detector geometries $\rightarrow$ symmetries in the transitions are the main factor of success.

Future Work

- Stabilize training with tracking layers.
- When reconstructing a single the system remains still partial observable (influence of other tracks). $\rightarrow$ **Multi-Agent Reinforcement Learning (MARL)**.

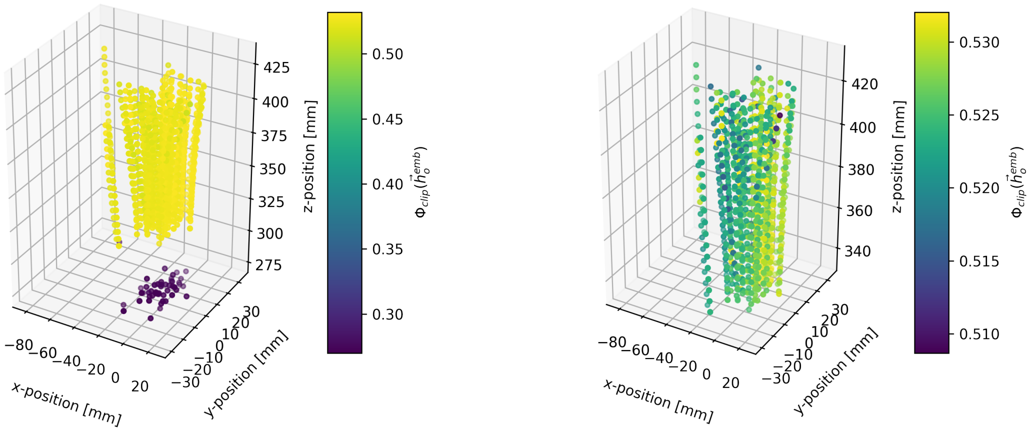
The Bergen pCT Collaboration and SIVERT Research Training Group

- 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
- Technical University TU Kaiserslautern, Germany

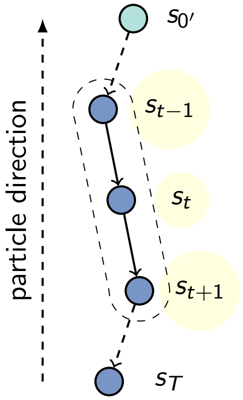
Contact: kortus@ztt.hs-worms.de

Backup Slides

Backup Slides - Learned $\Phi_{clip}(\vec{h}_o^{emb})$ Values: (100 primaries, head phantom)



Backup Slides - Reward Calculation



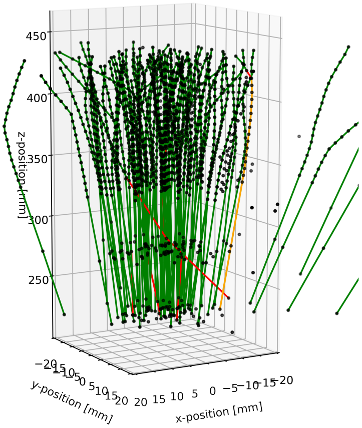
- Reward r_t for time step t is based on the state triplet $\langle s_{t+1}, s_t, s_{t-1} \rangle$:

$$r_t = \log P_{Highland}(\theta_{s_t:s_{t-1}} | \theta_{s_{t+1}:s_t}) \tag{4}$$

- Where $P_{Highland}$ is a normal distribution with zero mean and θ_0 with

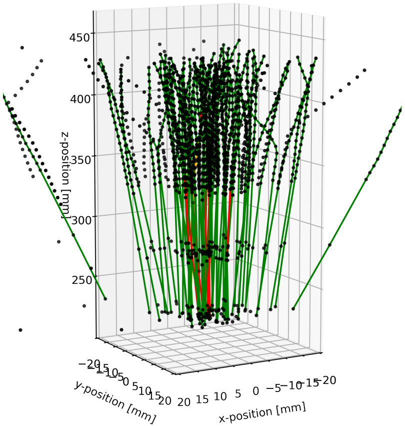
$$\theta_0 = \frac{14.1\text{MeV}}{pv} \sqrt{\frac{x}{X_0}} \left[1 + \frac{1}{9} \log_{10} \left(\frac{x}{X_0} \right) \right] . \tag{5}$$

- **Modifications**:
  1. Decrease carrier thickness of first detector layer to match carbon carrier of tracking layers → symmetry of material budget.
  2. Increase number of training iterations to 2000
  3. Independent reward normalization for detector → detector and detector/tracker → tracker transitions.



head phantom, 100 primaries

- **Modifications**:
  1. Increase number of training iterations to 2000
  2. Independent reward normalization for detector → detector, detector → tracker and tracker → tracker transitions.



head phantom, 100 primaries