Towards Neural Charged Particle Tracking in Digital Tracking Calorimeters With Reinforcement Learning

Tobias Kortus, Ralf Keidel, and Nicolas R. Gauger,
Bergen pCT Collaboration

***Abstract*—We propose a novel technique for reconstructing charged particles in digital tracking calorimeters using reinforcement learning aiming to benefit from the rapid progress and success of neural network architectures without the dependency on simulated or manually-labeled data. Here we optimize by trial-and-error a behavior policy acting as an approximation to the full combinatorial optimization problem, maximizing the physical plausibility of sampled trajectories. In modern processing pipelines used in high energy physics and related applications, tracking plays an essential role allowing to identify and follow charged particle trajectories traversing particle detectors. Due to the high multiplicity of charged particles and their physical interactions, randomly deflecting the particles, the reconstruction is a challenging undertaking, requiring fast, accurate and robust algorithms. Our approach works on graph-structured data, capturing track hypotheses through edge connections between particles in the detector layers. We demonstrate in a comprehensive study on simulated data for a particle detector used for proton computed tomography, the high potential as well as the competitiveness of our approach compared to a heuristic search algorithm and a model trained on ground truth. Finally, we point out limitations of our approach, guiding towards a robust foundation for further development of reinforcement learning based tracking.**

***Index Terms*—Charged particle tracking, combinatorial optimization, proton imaging, reinforcement learning, track reconstruction.**

I. Introduction

THE rather recent introduction of deep neural network architectures, particularly graph neural networks, for the reconstruction of particle trajectories from discrete measurements in silicon detectors led to major progress in tracking performances, reducing problems associated with the combinatorial explosion arising from increased density of particle readouts [1], [2]. However, in contrast to many earlier approaches, suffering from the aforementioned phenomena [3], [4], [5], deep learning based methods require computationally costly simulated data containing ground-truth information. In this work we propose a novel track reconstruction scheme based on model-free reinforcement learning, inspired by applications in combinatorial optimization [6], [7], [8], [9], where we aim to find a policy, parametrized by a deep neural network, that functions as a heuristic approximation to the full combinatorial optimization problem. Therewith, our solution aims to provide a unified solution between iterative and deep reconstruction algorithms to reduce the additional complexity arising from the combinatorial explosion of the possible solution space, while being able to train on partial information without ground-truth labels. Our work aims primarily at applications in high energy physics such as high energy physics research, where the reconstruction of discrete detector readouts generated by particle trajectories in collision events is a central task in the data processing, providing indispensable information for further analysis, such as vertex finding [10], [11], [12], particle reconstruction [13], [14], [15] or jet flavor tagging [16], [17], [18]. Likewise, with the recent development of high granularity scanner prototypes [19] in medical physics applications such as proton computed tomography (pCT) or proton radiography (pRad), where residual energy and path of high energetic particles measured in particle detectors are used for imaging, track reconstruction becomes a key processing step, providing estimates of track parameters for image reconstruction [20], [21]. Both applications require sophisticated algorithms, capable of reconstructing the traversal path of particles with high purity and efficiency to maximize spatial resolution, while in medical applications also minimizing the radiological dose [21]. Using physical interaction models together with discrete action spaces, our work provides an alternative view on reinforcement learning based particle tracking as compared to the approach concurrently developed by *Våge* [22]. Further, in contrast to the model in [22], our approach can be used as a standalone module, providing competitive reconstruction performance. All source code together with training details, hyperparameters, data, and models are publicly available on GitHub[1] and Zenodo.[2] Our key contributions and conclusions in this paper summarize as follows:

- We propose a novel reconstruction scheme using model-free deep reinforcement learning building upon concepts from neural combinatorial optimization, allowing for the

Manuscript received 30 December 2022; revised 2 August 2023; accepted 10 August 2023. Date of publication 15 August 2023; date of current version 3 November 2023. This work was supported in part by German Federal State Rhineland-Palatinate (Forschungskolleg SIVERT), in part by the Research Council of Norway (Norges forskningsråd), and in part by the University of Bergen under Grant 250858. Recommended for acceptance by M. White. *(Corresponding author: Tobias Kortus.)*

Tobias Kortus and Ralf Keidel are with the Center for Technology and Transfer (ZTT), University of Applied Sciences Worms, 67549 Worms, Germany (e-mail: kortus@hs-worms.de; keidel@hs-worms.de).

Nicolas R. Gauger is with the Chair for Scientific Computing, TU Kaiserslautern, 67663 Kaiserslautern, Germany (e-mail: nicolas.gauger@scicomp.uni-kl.de).

Digital Object Identifier 10.1109/TPAMI.2023.3305027

[1] https://github.com/SIVERT-pCT/rl-tracking

[2] https://doi.org/10.5281/zenodo.7426388

ground-truth free optimization of a Pointer-Network [23] based architecture for particle tracking.

- We optimize the architecture with custom positional encoding for improved spatial inductive bias, allowing for a trainable selection of an area of interest, without any restrictions on the graph connectivity, requiring manual tuned priors.
- We demonstrate the efficiency of modeling the underlying effects of elastic nuclear interactions as an effective, yet easy to estimate quantity for a dense reward function performing approximately on par with supervised optimization.
- We exemplify the out-of-the-box generalization abilities of our approach to unseen particle densities, phantom geometries, track topologies and beam spot positions without requiring additional optimization.
- Finally, we demonstrate the competitiveness of the learned policy to a manual tuned heuristic search algorithm [5], [24].

II. Related Work and Motivation

Given the high demand for tracking algorithms in particle physics, various algorithms have been introduced in the last decades. In the following, we briefly point out key developments together with the motivation to deviate from existing approaches by combining the advantages of iterative reconstruction algorithms and deep learning in a unified reinforcement learning based approach. For a more comprehensive review of existing literature, we refer the reader to [25] and [1], [2].

Classical/iterative reconstruction methods: In the early development of reconstruction algorithms, classical and iterative algorithms, relying on rule based reconstruction or on physical models of particle interactions, dominated the landscape of research. Most prominent were approaches based on local track finding techniques such as Kalman filters [26] or cellular automata [3] together with global [27], [28] and combinatorial approaches [4], [5], [24]. However, with the increasing computational demands in modern high energy physics introduced by the combinatorial explosion caused by higher particle counts, these approaches are progressively replaced due to their scarce parallelization ability together with the often required manually tuned heuristics or costly evaluation of physical models during reconstruction.

Deep learning reconstruction methods: With the availability of modern deep learning architectures together with dedicated computing hardware, deep learning demonstrated impressive performance in particle tracking due to the ability to learn from raw data with no or minimal assumptions about the underlying system. Early approaches heavily utilized LSTM [29], [30], [31] and CNN [30], [31] architectures, and were later on superseded by graph neural networks [1], [2], sparsely capturing relations between particle hits. Here, most approaches rely on an edge classification scheme together with a final planning module extracting feasible track candidates leveraging predicted edge scores [29], [32], [33], [34].

With the usage of deep learning architectures, particle tracking becomes highly dependent on large amounts of computationally intensive Monte Carlo simulations, possibly introducing additional simulation-to-reality gaps (e.g., cluster sizes [35]), which might affect inference performance. Training fixed reconstruction policies on partial information using a reward signal, physical models used in iterative reconstruction can guide the training process of generalizable network architectures, while being able to reconstruct tracks independently of simulated data and costly evaluation of physical processes during inference.

III. Particle Interactions in Matter

In this section, we provide the reader a short lineup of predominant physical interaction mechanisms that can be observed and influence the path of charged particles traversing matter at energies relevant for pCT. For brevity, we intentionally leave out further interactions. For a full and in depth review of interaction mechanisms in high energy physics, the mindful reader is referred to Groom and Klein [36].

A. Ionizing Energy Loss

Charged particles passing through material of thickness dx lose a fraction of their initial energy, caused by repeated inelastic Coulomb interactions with atomic electrons. This relationship in terms of the mean energy loss per unit length $-dE/dx$, also referred to as linear stopping power (S), was first captured in a non-relativistic form by Bohr [37] and later for relativistic velocities by Bethe and Bloch [38]. Note that the linear stopping power is approximately inversely proportional to the particle velocity and thus residual energy, resulting in relatively low energy losses at high velocities and a maximum energy deposition right before the particle stops. This characteristic point in the energy loss curve is often referred to as Bragg-peak, providing the beneficial characteristics of precise energy deposition used in proton or ion therapy.

B. Multiple Coulomb Scattering

While interacting with atomic electrons, protons or heavier ions remain on their original path due to the relatively high mass opposed to the atomic electrons. However, in the case of Coulomb interactions with atomic nuclei, also referred to as multiple Coulomb scattering, heavy charged particles observe a repelling force causing a deflection of the particle from its original path [39]. The amount and direction of deflection is mostly random, following approximately a Gaussian distribution [40], [41], while being influenced by the radiation length X_0 of the material traversed and the particle momentum.

C. Inelastic Nuclear Interactions

Occasionally, high energetic charged particles directly pierce the Coulomb barrier of an atomic nucleus and collides with it. During this collision, the primary particle is absorbed, and secondary particles are created. Due to the chaotic nature of this interaction, obscuring information of the initial residual particle energy, particle tracks with inelastic interactions are

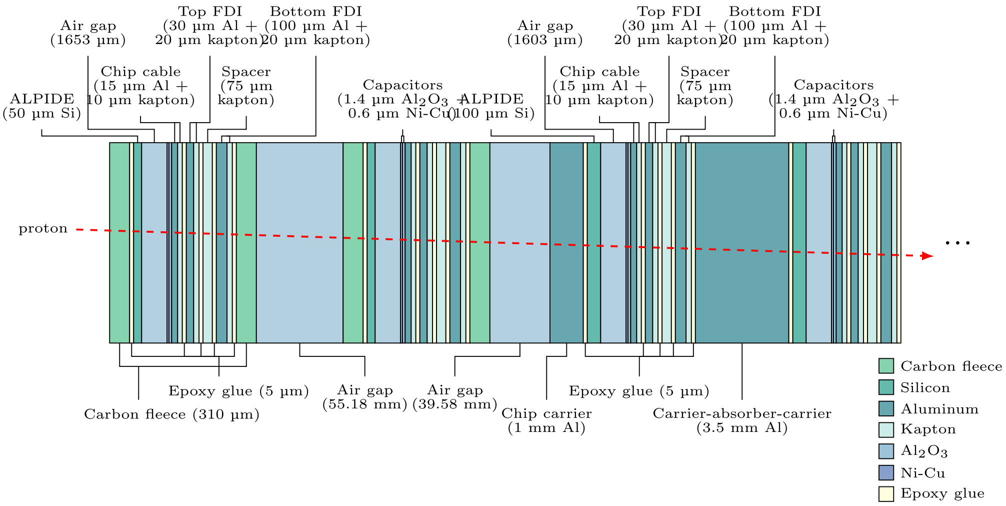

Air gap (1653 µm)
Top FDI (30 µm Al + 20 µm kapton)
Bottom FDI (100 µm Al + 20 µm kapton)
Chip cable (15 µm Al + 10 µm kapton)
Spacer (75 µm kapton)
ALPIDE (50 µm Si)
Capacitors (1.4 µm Al2O3 + 0.6 µm Ni-Cu)
ALPIDE (100 µm Si)
Air gap (1603 µm)
Top FDI (30 µm Al + 20 µm kapton)
Bottom FDI (100 µm Al + 20 µm kapton)
Chip cable (15 µm Al + 10 µm kapton)
Spacer (75 µm kapton)
Capacitors (1.4 µm Al2O3 + 0.6 µm Ni-Cu)
proton
...
Epoxy glue (5 µm)
Carbon fleece (310 µm)
Air gap (55.18 mm)
Air gap (39.58 mm)
Chip carrier (1 mm Al)
Epoxy glue (5 µm)
Carrier-absorber-carrier (3.5 mm Al)
Carbon fleece
Silicon
Aluminum
Kapton
Al2O3
Ni-Cu
Epoxy glue

Fig. 1. Schematic diagram of the material composition of the detector geometry consisting of, from left to right, the first two tracking layers and the first detector-absorber sandwich layer used for simulating proton tracks in GATE. Adapted from [19].

not relevant for pCT and thus are not further considered during reconstruction.

IV. The Digital Tracking Calorimeter

In this paper, we focus on reconstructing particle trajectories of protons measured during pCT and pRad, captured in a high granularity digital tracking calorimeter (DTC) proposed by the Bergen pCT collaboration. The Bergen pCT collaboration [19], established at the University of Bergen (Norway), focuses on the design of a novel prototype scanner for pCT using a multilayer structured high granularity DTC. The sensitive area of the proposed scanner is built of high resolution 1-bit ALPIDE silicon sensors [42], developed for the upgrade of the Inner Tracking System (ITS) of the ALICE experiment at CERN [43]. The shared structure of high granularity silicon sensors with other high energy experiments [44] makes this detector well suited for initial development of our algorithm, while reducing the overall complexity of the system. The pCT scanner, described in Alme et al. [19], consists in total of two downstream trackers in the frontal section of the detector, followed by a stack of 41 detector/absorber sandwich layers, each composed of a sensitive layer of silicon detectors followed by an 3.5 mm aluminum absorber plate. A detailed description of the material composition of the scanner is depicted in Fig. 1.

V. Tracking as a Markov Decision Process

For the representation of the particle reconstruction task, we consider a framework similar to existing approaches in combinatorial optimization [6], [7], [8], [9], [45], where we aim to approximately solve an optimization problem by finding a policy $\pi(a_t|s_t)$ capable of determining adequate sets of trajectories maximizing the physical plausibility of the undertaken transition. We therefore model the sequential task of following particles in subsequent high granularity silicon pixel layers as a Markov Decision Process (MDP) [46] on a graph defined by the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}, P, \mu_0 \rangle$. Here $\mathcal{S}$ is the set of possible states defining partial particle tracks, $\mathcal{A}$ is the set of possible actions defining all possible transitions between hit centroids in the graph, where $\mathcal{A}_t \subset \mathcal{A}$ is the set of feasible actions at time t defined by the neighborhood $\mathcal{N}$, and $\mathcal{R}$ is a scalar reward signal $\mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ defined by the underlying physics of particle interactions in matter. Further, let P be the (unknown) underlying state transition kernel and μ_0 the initial state distribution of the MDP. In our use case, we operate in an episodic setting where each episode starts with a randomly sampled state $s_0 \sim \mu_0$ in the detector and ends after $T \leq T_{max}$ time steps in a terminal state, upper bounded by a full traversal of all sensitive layers contained in the DTC, where a full particle track candidate is defined by a trajectory $\tau = (s_0, a_0, \ldots, s_T, a_T)$.

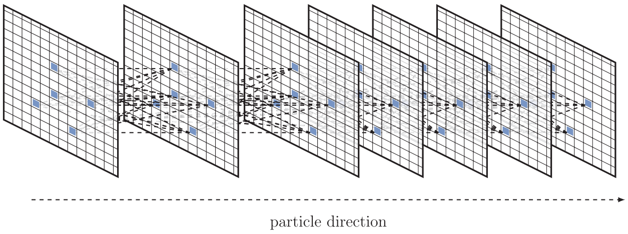


Fig. 2. Schematic representation of the proposed, fully connected, detector graph with multiple particle track hypotheses over two tracking layers and the first four detector-absorber sandwich layers.

A. Detector Graph

Following the initial problem definition in Section V, we now consider the definition of the graph-structure built upon the detector readouts. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed acyclic graph over the layered point cloud of detector readouts, where $\mathcal{V} = \{v_i\}_{i=1:N_v}$ denotes a set of graph vertices defined by proton hit centroids and $\mathcal{E} = \{e_k\}_{k=1:N_e}$ be the graph's edges, connecting existing neighboring readouts over subsequent layers, in a direction opposite to the particle traversal path (ref. Fig. 2).

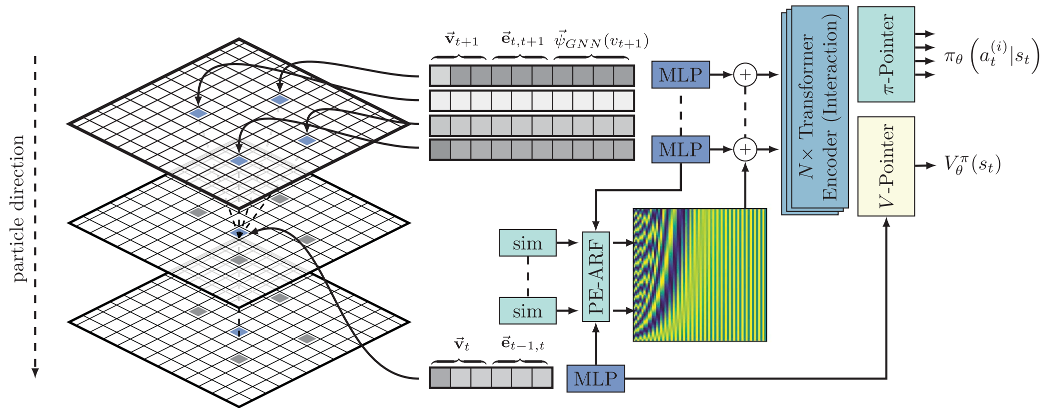

particle direction
$\vec{\mathbf{v}}_{t+1}$
$\vec{\mathbf{e}}_{t,t+1}$
$\vec{\psi}_{GNN}(v_{t+1})$
MLP
MLP
$N \times$ Transformer Encoder (Interaction)
π-Pointer
$\pi_\theta\left(a_t^{(i)}|s_t\right)$
V-Pointer
$V_\theta^\pi(s_t)$
sim
sim
PE-ARF
$\vec{\mathbf{v}}_t$
$\vec{\mathbf{e}}_{t-1,t}$
MLP

Fig. 3. Shared network architecture for policy and state-value estimates based on N stacked Transformer encoder, leveraging positional encoding with adaptive receptive field, together with individual network branches for V_θ^π and π_θ using additive attention.

In our case, each edge e_{ij} represents a possible hypothesis for a track segment between readouts v_i and v_j that need to be considered as a possible action for reconstruction. Here, we intentionally perform no prior filtering of hypotheses to avoid any assumptions, requiring detector dependent information. Further, we use inverted edge directions compared to the particle traversal direction to allow for a backward tracking scheme, starting in the distal section of the detector. This allows us to take advantage of the decreased particle density in this part of the detector. We parameterize each vertex v_i by a set of features $\vec{\mathbf{v}}_i = (\Delta E_i^C, x_i, y_i, \mathbb{1}_z(z_i))$, where ΔE_i^C is the measured energy deposition of the particle in terms of discrete cluster sizes, estimated according to [19], and x_i, y_i is the pixel position in the detector plane. Further, $\mathbb{1}_z(z_i)$ is a one-hot indicator function encoding the index of the detector plane in the DTC as a 43-dimensional vector. Each edge is then parameterized as $\vec{\mathbf{e}}_{ij} = (r_{ij}, \theta_{ij}, \phi_{ij})$, where r_{ij}, θ_{ij} and ϕ_{ij} are spherical coordinates describing the connection of a transition hypothesis. Finally, we normalize all features, where we additionally center the hit positions in the detector w.r.t the position of the pencil beam (modified during scanning), to provide translation invariant representations.

B. State Definition

To define a sufficient state representation in the detector graph which satisfies the Markov property w.r.t. a single track, we consider various components of the graph describing a sufficient statistic over the reconstructed track candidate. Given the independence of individual scattering events, we formulate a state in the MDP as a combination of a one-step history describing the last reconstructed track segment as well as all possible next segments, defined as

$$s_t = \{v_t, e_{t-1,t}\} \cup \bigcup_{i=0}^{\mathcal{N}(v_t)} \left\{v_{t+1}^{(i)}, e_{t,t+1}^{(i)}, \psi\left(v_{t+1}^{(i)}\right)\right\}. \quad (1)$$

Here, $\{v_t, e_{t-1,t}\}$ and $\{v_{t+1}^{(i)}, e_{t,t+1}^{(i)}\}$ describe the last reconstructed segment and the i-th possible next segment, defined by vertices and edges between two particle hits in subsequent sensitive layers, respectively. Further, $\psi(v_{t+1}^{(i)})$ is an additional representation, summarizing a statistic of an n-hop neighborhood in the graph. In Section VIII-K, Table V we demonstrate the equivalence of the MDP with the described state representation over a partial observable MDP (POMDP), where we assume partial observability of the system, introduced by essential information being lost in the unobserved history of the track candidate. We therefore consider an updated state representation $s_t = \{v_t, e_{t-1,t}, \Gamma(v_t)\} \cup \bigcup_{i=0}^{\mathcal{N}(v_t)}\{\cdot\}$ with an additional belief state $\Gamma(v_t)$, captured recursively over the preceding track segments by a LSTM.

VI. Preprocessing and Model Architecture

Working upon the state representation in Section V-B, we parametrize both policy $\pi_\theta(a_t|s_t)$ and state-value function $V_\theta^\pi(s_t)$ using a Pointer-Network style architecture (ref. Fig. 3) with encoder-decoder scheme [23], commonly utilized in neural combinatorial optimization [6], [7], [23]. To reduce the size of the policy and value network and allow parameter-sharing, we rely on a shared network trunk, containing the computationally demanding encoding task, combined with a novel adaptive positional encoding mechanism, while only separating the final decoders used for estimating π_θ and V_θ^π. According to the state definition in Section V-B, we provide the proposed network architecture two distinct set of features, in the following referred to as *action-* and *observation-features* with

$$\vec{\mathbf{h}}_{obs} = [\vec{\mathbf{v}}_t \| \vec{\mathbf{e}}_{t-1,t}] \quad \text{and} \quad (2)$$

$$\vec{\mathbf{h}}_{act,i} = \left[\vec{\mathbf{v}}_{t+1}^{(i)} \| \vec{\mathbf{e}}_{t,t+1}^{(i)} \| \vec{\psi}_{GNN}\left(v_{t+1}^{(i)}\right)\right], \quad (3)$$

Here, $\vec{\mathbf{v}}$ and $\vec{\mathbf{e}}$ are vertex and edge features with the concatenation operator $\|$, and $\vec{\psi}(v_{t+1}^{(i)})$ is a context vector generated for $v_{t+1}^{(i)}$ by a graph neural network (GNN). Further, we generate equally sized embeddings $\vec{\mathbf{h}}_{obs}^{emb}$ and $\vec{\mathbf{h}}_{act,i}^{emb}$ by transforming each set of features by a distinct multi-layer perceptron (MLP). In the following, we describe the individual components of the network.

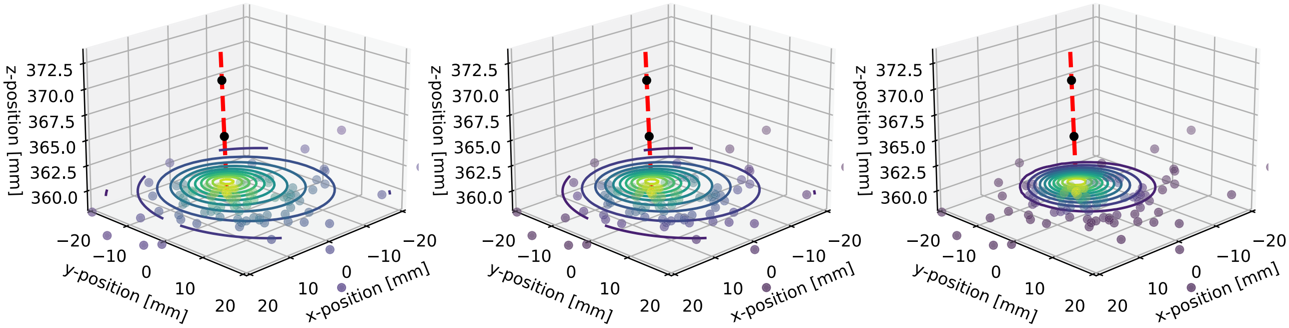


Fig. 4. Area of focus of adaptive receptive field defined by clip$\{\cdot, 0, 1\}$ with fixed rescaling factors of $\mathbf{\Phi}(\vec{h}^{emb}_{obs}, \vec{h}^{emb}_{act}) \in \{0.5, 0.4, 0.3\}$ together with previously reconstructed track segment with projected straight track and possible next hits in the subsequent detector layer.

A. Node Embedding

To capture structural information in the graph to find long-term dependencies over multiple detector layers, we add a statically calculated context representation $\vec{\psi}_{GNN}(v^{(i)}_{t+1})$ of the multi-hop neighborhood of vertex $v^{(i)}_{t+1}$ to the feature vector. We realize this with a graph neural network with inverted message flow (target-to-source), composed of multiple stacked graph attention (GAT) layers [47], [48] with added self-loops. This allows to efficiently capture relations in the graph geometry by calculating in each network layer $l > 0$ an updated node representation $\vec{\psi}^{(l)}_{i,GNN}$ as:

$$\vec{\psi}^{(l)}_{i,GNN} = \sigma\left\{\mathbf{W}_k\vec{\mathbf{h}}^{(l-1)}_i + \sum_{j=0}^{\mathcal{N}(i)} \alpha_{ij}\mathbf{W}_k\vec{\mathbf{h}}^{(l-1)}_j\right\}, \tag{4}$$

Here $\vec{\psi}^{(l)}_{i,GNN} \in \mathbb{R}^{d_E}$ is the d_E dimensional vector embedding, generated by the previous layer of the GNN, $\mathbf{W}_k \in \mathbb{R}^{d_E \times d_E}$ is a $d_E \times d_E$ dimensional parameter matrix and σ is a nonlinear activation function. Further, a_{ij} is an attention weight generated for each node in the neighborhood $\mathcal{N}(i)$ using

$$\alpha^{(l)}_{ik} = \frac{\exp\left\{\vec{\mathbf{v}}^T \mathrm{LReLU}\left(\mathbf{W}\left[\vec{\mathbf{h}}^{(l-1)}_i \middle\| \vec{\mathbf{h}}^{(l-1)}_j\right]\right)\right\}}{\sum_k^{\mathcal{N}(i)} \exp\left\{\vec{\mathbf{v}}^T \mathrm{LReLU}\left(\mathbf{W}\left[\vec{\mathbf{h}}^{(l-1)}_i \middle\| \vec{\mathbf{h}}^{(l-1)}_k\right]\right)\right\}}, \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^{d_E \times d_E}$ and $\vec{\mathbf{v}} \in \mathbb{R}^{d_E \times 1}$ are, again, network parameters and LReLU is the Leaky Rectified Linear Unit activation function.

B. Action Candidate Encoding

Aiming to transform the *action-features* ($\vec{\mathbf{h}}^{emb}_{act}$) to find existing relations between action candidates in $\mathcal{A}_t$, we use a multi-head attention (MHA) [49] based encoding mechanism similar to [8], [50]. We select this mechanism over the proposed LSTM layer in the original Pointer-Network architecture [23] to overcome shortcomings introduced by the lack of meaningful ordering of the input sequence. The encoding module is built according to (6) and 7 by stacking three encoder sub-layers, each composed by a multi-head attention block, with four heads with a dimensionality of $d_H = 32$ for each attention head, combined with a residual connection [51] and layer normalization (LN) [52]:

$$\vec{\mathbf{h}}^{(l)}_{act,i} = \mathrm{LN}\left(\vec{\mathbf{h}}^{(l-1)}_{act,i} + \mathrm{MHA}_i\left(\vec{\mathbf{h}}^{(l-1)}_{act,i}, \ldots, \vec{\mathbf{h}}^{(l-1)}_{act,N}\right)\right) \tag{6}$$

$$\vec{\mathbf{h}}^{(l)}_{act,i} = \mathrm{LN}\left(\vec{\mathbf{h}}^{(l)}_{act,i} + \mathrm{MLP}\left(\vec{\mathbf{h}}^{(l)}_{act,i}\right)\right) \tag{7}$$

Here, $\vec{\mathbf{h}}^{(l)}_{act,i}$ and $\vec{\mathbf{h}}^{(l-1)}_{act,i}$ are input and output features of network layer l, with $\vec{\mathbf{h}}^{(0)}_{act,i} = \vec{\mathbf{h}}^{emb}_{act,i}$. The final output of each sub-layer is then generated by transforming the output feature of the first step using a multilayer perceptron, once again combined with a residual connection and layer normalization. For a clear distinction of the transformed features, we denote the final output of the encoding layers in the following as $\vec{\mathbf{h}}^{attn}_{act,i}$.

C. Positional Encoding With Adaptive Receptive Field

To capture supplementary spatial information, we propose a modified form of positional encoding (PE-ARF) similar to [49], using cosine-similarities of track hypotheses as positional information augmented by an adaptive rescaling mechanism based on additive attention [53]. This allows us to re-allocate attention, as depicted in Fig. 4, to improve spatial resolution for each reconstruction step independently:

$$N^{(i)}_{\text{ARF}} = \alpha \cdot \mathrm{clip}\left\{\frac{0.5 \cdot \left(1 - \mathrm{sim}\left(e_{t-1,t}, e^{(i)}_{t,t+1}\right)\right)}{\mathbf{\Phi}(\vec{\mathbf{h}}^{emb}_{obs}, \vec{\mathbf{h}}^{emb}_{act,0:N})}, 0, 1\right\} \tag{8}$$

Here $0.5 \cdot (1 - \mathrm{sim}(e_{t-1,t}, e^{(i)}_{t,t+1}))$ is the cosine-similarity of the edges of a partial track hypothesis defined by the vertices $v_{t-1} \rightarrow v_t \rightarrow v_{t+1}$, rescaled to a range of $[0, 1]$, where 0 corresponds to straight edge connections. Further, clip$(\cdot, 0, 1)$ denotes a clipping function restricting the range of the rescaled similarity values to $[0, 1]$ removing unique encoding information from connections outside the normalization range, defined by the function mapping $\mathbf{\Phi}(\vec{\mathbf{h}}^{emb}_{obs}, \vec{\mathbf{h}}^{emb}_{act,0:N})$ with $\mathbb{R}^{d_M} \rightarrow \mathbb{R}$:

$$\mathbf{\Phi}(\cdot) = \mathbf{\Phi}\left\{\mathbf{W}^{\Phi}_1\vec{\mathbf{h}}^{emb}_{obs} + \sum_{j=0}^{\mathcal{N}(i)} \alpha_j\left(\mathbf{W}^{\Phi}_2\vec{\mathbf{h}}^{emb}_{act,j}\right)\right\} \tag{9}$$

Here, $\mathbf{W}^{\Phi}_1$ and $\mathbf{W}^{\Phi}_2$ are linear projections with $\mathbf{W}^{\Phi}_1, \mathbf{W}^{\Phi}_1 \in \mathbb{R}^{d_M \times d_M}$ and α_i is a learnable attention weight defined by the additive attention mechanism [53] with $\mathbf{W}^{\Phi}_1, \mathbf{W}^{\Phi}_1 \in \mathbb{R}^{d_M \times d_M}$

and $\vec{\mathbf{v}}^T \in \mathbb{R}^{d_M \times 1}$:

$$\alpha_i = \vec{\mathbf{v}}^T \tanh(\mathbf{W}_1^{\alpha} \vec{\mathbf{h}}_{act,i}^{emb} + \mathbf{W}_2^{\alpha} \vec{\mathbf{h}}_{act,j}^{emb}) \tag{10}$$

We demonstrate the positive effect of positional encoding with adaptive receptive field over regular positional encoding and embedding without encoding in Section VIII-K, Tables VI and VII.

D. Policy and State-Value Decoding

To obtain the final output describing the policy $\pi_\theta(a_t|s_t)$ and state-value estimate for a given input, we correlate the transformed *action-features* with the *observation-features* using the attention based decoder proposed in Vinyals et al. [23]. Here, we use separate branches for both state-value and policy estimate, calculating for each possible action $a_t^{(i)} \in \mathcal{A}_t$, two scalar weights α_i^{π} and α_i^{V} according to

$$\alpha_i = \vec{\mathbf{v}}^T \tanh(\mathbf{W}_1 \vec{\mathbf{h}}_{act,i}^{attn} + \mathbf{W}_2 \vec{\mathbf{h}}_{obs}^{emb} + \vec{\mathbf{b}}_{12}) + \vec{\mathbf{b}}_{\vec{\mathbf{v}}}. \tag{11}$$

We then leverage the information contained in each attention weight to calculate π_θ as

$$\pi_\theta\left(a_t^{(i)}|s_t, \theta\right) = \frac{\exp\left(\alpha_i^{\pi}\right)}{\sum_{j=0}^{\mathcal{N}(i)} \exp\left(\alpha_j^{\pi}\right)}, \tag{12}$$

where $\pi_\theta(a_t^{(i)}|s_t, \theta)$ represent the agent's policy in terms of softmax in action perferences, calculated for each possible action $a_t^{(i)} \in \mathcal{A}_t$ using the corresponding attention weight α_i^{π}. Similarly, the state-value function is estimated as the average over all attention weights a_i^V:

$$V_\pi^\theta(s_t) = \frac{1}{|\mathcal{N}(v_t)|} \sum_{i=0}^{\mathcal{N}(v_t)} \alpha_i^V \tag{13}$$

VII. Reward Design and Network Optimization

During the traversal of the detector-absorber sandwich layers, the proton trajectory is mainly influenced by the effects of multiple Coulomb scattering, where the magnitude of scattering observed in a thin slab follows approximately a Gaussian distribution [41]. Therefore, the joint probability for an observed trajectory τ, crossing multiple sensitive detector layers, factorizes according to

$$P(\tau) = \prod_{t=1}^{T-1} p\left(\Delta\theta(s_t : s_{t+1}) | X_0(s_t : s_{t+1}), R_0(\tau)\right), \tag{14}$$

where $p(\Delta\theta(s_t : s_{t+1})|\cdot)$ defines the probability of observing the angular deflection $\Delta\theta$ given the state transition $s_t \rightarrow s_{t+1}$ and the full trajectory τ under multiple Coulomb scattering. Here, $X_0(s_t : s_{t+1})$ denotes the radiation length of the detector section enclosed by $s_t : s_{t+1}$ and $R_0(\tau)$ denotes the estimated range of the track candidate τ. Intuitively, we aim to find a policy π_θ from the family of parametrized policies Π_θ, that maximizes the expected factorized probability $P(\tau)$:

$$\pi_\theta^* = \underset{\pi_\theta \in \Pi_\theta}{\arg\max} \;\; \mathbb{E}_{s_0 \sim \mu_0, a_t \sim \pi_\theta}\left[P(\tau)\right]. \tag{15}$$

We aim to indirectly optimize the described quantity of the factorized probability $P(\tau)$ by maximizing the return G_t obtained by repeated interactions in randomly sampled environments. We therefore define the immediate reward r_t for an action a_t as

$$r_t = \log p\left(\Delta\theta(s_t : s_{t+1}) | X_0(s_t : s_{t+1}), R_0(\tau)\right), \tag{16}$$

where the undiscounted return $G_t^{\gamma=0}$ directly corresponds to the factorized log-probability $\log P(\tau_{t:T})$ of the partial track $\tau_{t:T}$.

A. Sampling of Track Candidates During Training

During training, we sample multiple track hypotheses $\tau^{(i)} = (s_0, a_0, \ldots, s_T, a_T)$, from randomly selected environments, each capturing individual readout frames of the scanner. As we do not know the actual stopping point of tracks in the graph without stepwise solving all tracks starting from the last detector layer, we manually establish an initial state distribution μ_0, by sampling random starting positions from the pool of all vertices in the last five layers using a uniform distribution. By choosing a margin of five detector layers, we aim to account for varying track lengths in the detector, while still providing enough energy information for estimating the track's energy characteristic. Further, to be able to parametrize our initial state representation s_0, we combine our initial graph vertex v_0 with track seeding [54] for an estimation of initial track properties. For simplicity, we rely here on ground-truth track seeding to avoid measuring negative impacts of particular track seeding techniques on the RL approach. Doing so, we can obtain a possible upper bound on the performance in the later sections.

B. Reward Estimation Using MCS

For estimating the algebraically complicated theory for multiple Coulomb scattering described by Molière [39], we use a Gaussian approximation by Highland [41], [55], where the $2\sigma_{\theta_0}$ angle for each scattering transition in the sampled track hypothesis can be roughly estimated by

$$2\sigma_{\theta_0} = \frac{14.1\text{MeV}}{pv}\sqrt{\frac{z}{X_0}}\left[1 + \frac{1}{9}\log_{10}\left(\frac{z}{X_0}\right)\right]. \tag{17}$$

Here X_0 is the radiation length and z is the thickness of the target slab traversed. To avoid the additional complexity of composite materials, we only consider the predominant influence of aluminum absorbers and air gaps. We further calculate the kinetic energy (pv) of the particle respectively as [55]

$$pv = \frac{\tau + 2}{\tau + 1}E \quad \text{where} \quad \tau \equiv \frac{E}{mc^2}, \tag{18}$$

where E is the residual energy of the proton, m is the proton's mass and c is the speed of light. We estimate the residual energy $E(z)$ of the proton at each sensitive layer, given the range, based on the approximate mean energy loss curve defined by the Bragg-Kleeman rule [56] as

$$E(z) = \alpha^{-1/p}(R_0 - z)^{1/p}. \tag{19}$$

We find the range of the particle by performing a nonlinear least square fit on the energy depositions of a sampled trajectory τ using the differentiated Bragg-Kleeman equation −dE/dz [56], defined respectively as

$$-\frac{dE}{dz} = p^{-1}\alpha^{-1/p}(R_0 - z)^{1/p-1}. \tag{20}$$

Here $\alpha = 0.0262\,\text{MeV/mm}^{-1}$ and $p = 1.736$ are both parameters obtained from model fits to range-energy data performed for this particular detector by Pettersen et al. [57].

C. Data Correction Mechanisms

To counteract the significant imbalance in sampled experience and reward introduced by the unbalanced frequency of state transitions observed for tracker and calorimeter layers, we introduce two correction mechanisms:

- *Reward Normalization:* We observed degraded performance of the model due to different orders of magnitude in reward estimates of layers with different material budget (Section VIII-K, Table VIII). To counteract this phenomenon, we employ an adapted version of reward normalization as proposed by Van Hasselt et al. [58], where we normalize rewards to zero mean and unit variance depending on the traversed material budget.
- *Data resampling:* To stabilize training, we further balance the amount of sampled state transitions $\langle s_t, a_t, s_{t+1}\rangle$, by, resampling the gathered experience from tracking and calorimeter transitions during each optimization step to a 1:1 ratio.

D. Policy and State-Value Updates

For the optimization of the proposed agent architecture, we rely on proximal policy optimization (PPO-CLIP) [59], an actor-critic algorithm providing state-of-the art results in many application domains by allowing larger policy updates while avoiding costly operations of KL-constrained objectives [59]. We specifically rely on an actor-critic method in order to apply bootstrapping for decreased gradient variance while still being able to use on-policy optimization. We prefer on-policy over off-policy experience due to the substantial dependence of the chosen reward function on the particular sampled trajectory, conditioned by the required range fitting in Section VII-B. For each update step, we calculate the policy loss according to Schulman et al. as

$$\mathcal{L}^P = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right], \tag{21}$$

where r_t is the probability ratio between old and new policy, denoted by $r_t = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$. Further, $\hat{A}_t$ is an estimate of the advantage function, estimated using generalized advantage estimation [60]:

$$\hat{A}_t^{GAE} = \sum_{l=0}^{\infty}(\gamma\lambda)^l \delta_{t+l}^V, \tag{22}$$

Here, both γ and λ are tunable hyperparameters controlling discount and bias-variance-tradeoff for the advantage estimate, and δ_t^V denotes the temporal difference (TD) residual

$$\delta_t^V = r_t + \gamma V(s_{t+1}) - V(s_t). \tag{23}$$

Similar to the policy loss, we calculate the value-loss as a clipped mean squared error objective, as defined in Engstrom et al. [61]. For the combined update of the shared network architecture for state-value and policy function, we determined the combined loss function according to Schulman et al. [59] as

$$\mathcal{L}(\theta) = \hat{\mathbb{E}}\left[\mathcal{L}_t^P(\theta) + c_1\mathcal{L}_t^V(\theta) + c_2 S\left[\pi_\theta\right](s_t)\right], \tag{24}$$

where $\mathcal{L}^P$ and $\mathcal{L}^V$ are policy- and value-loss, $S[\pi_\theta]$ is an additional entropy regularization term and c_1, c_2 are weighting factors for $\mathcal{L}_t^P$ and $S[\pi_\theta]$ respectively. We use orthogonal initialization for all network layers, with varying scaling for each layer [61], [62].

VIII. Experimental Results

In this section, we demonstrate and analyze the performance of the proposed sequential reconstruction approach from different aspects, including various phantom geometries, phantom material composites and particle densities. We particularly focus on the reconstruction and generalization performance of particle trajectories recovered from particle spots, measured for a radiograph of an inhomogeneous head phantom with various particle configurations. We then extend the analysis of reconstruction performance to individually analyze the impact of phantom thickness, particle density as well as beam spot positioning to decompose existing sources of error. We finally compare the results with a supervised trained model and heuristic search algorithm [5], [24], demonstrating the competitive performance of the reinforcement learning based approach.

A. Generation of Track Candidates During Inference

Unlike trajectory generation during training, during inference we aim to extract all valid tracks from a given readout frame, while following the full length of a track. We therefore deviate from the candidate generation described in Section VII-A. During inference, we add, starting from the last layers, all vertices not covered by track candidates to the reconstruction queue and reconstruct all candidates, following the learned policy greedily, until all candidates reach their respective terminal state in the first tracking layer (ref. Fig. 5).

B. Track Filtering

Due to the occurring physical interactions, a fraction of particles undergoes unrecoverable inelastic nuclear interactions, either observable as large angle scattering or abrupt stops. Further, some reconstructed particles do not match the required physical properties. To remove those particle trajectories from evaluation, we apply a cut based filtering, limiting the scattering angle in calorimeter and tracking layers to $\Delta\theta_{\max} = 271$ mrad, corresponding to a 2σ upper bound for particles in the last layer before stopping using extrapolated values from the PSTAR database [63]. Further, all tracks are required to show the characteristic high energy deposition of a Bragg peak, which we identify by an energy filter in the last layer of $\Delta E_{\min} = 2.5\,\text{keV}/\mu\text{m}$ [5], [64].

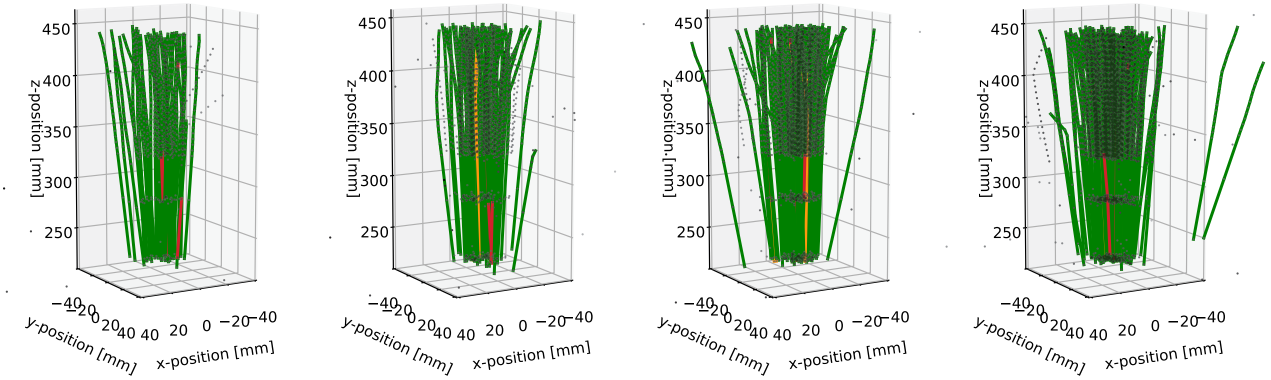


Fig. 5. Reconstructed particle trajectories from randomly sampled environments of the 100 mm water phantom dataset, described in Section VIII-D with, from left to right, 50, 100, 150, and 200 p^+/F. Green: correctly reconstructed track segment, red: incorrect reconstructed track segment, orange: correct reconstruction following the wrong primary particle.

C. Performance Scores

To evaluate and quantify the performance of the proposed tracking scheme, we calculate purity (p) and efficiency (ϵ) of reconstructed particle tracks based on the ground truth of particle tracks. We calculate the purity of reconstructed tracks, according to (25) as the fraction of correctly reconstructed tracks after filtering and the total number of reconstructed tracks after filtering:

$$p = \frac{N^{filt}_{rec,+}}{N^{filt}_{rec,+/-}}, \quad \epsilon = \frac{N^{filt}_{rec,+}}{N^{prim}_{total}}, \tag{25}$$

Here we rely on the strict *tight-match* definition of a correctly reconstructed track, where all reconstructed vertices have to correspond, in addition to matching the criteria in Section VIII-B, to the same primary particle. Similarly, we calculate the efficiency (ϵ) of reconstructed particle tracks as the number of correctly reconstructed tracks after filtering divided by the number of total primary tracks present in the readout frame.

D. Simulations and Data Generation

As the there is currently no working prototype of the proposed scanner setup described in Alme et al. [19], we rely on Monte Carlo (MC) simulated data [65] using the Gate 9.2 simulation toolkit [66] built upon Geant4 [67]. We therefore use a model of the detector geometry, following the material budgets, described in Fig. 1. We simulate the proton source, used in pCT and pRad for pencil beam scanning, as a mono-energetic 230 MeV proton beam with $\sigma_{xy} = 2$ mm. In order to study the quality of reconstructed particle tracks in detail, we consider the following two phantom types, placed in between the particle beam and detector:

- *Pediatric head phantom:* To provide realistic reconstruction results for inhomogeneous material composites, we rely on spot scanning data with 7 mm spot spacing, generated for a pediatric head phantom described in Giacometti et al. [68]. To reduce the overall runtime of the evaluation, we only consider beam spots contained inside the patient with 2,000 primaries per frame each. We therefore remove all frames from the simulation, where the proton beam center misses the head phantom.
- *Water phantoms:* We additionally provide simulated data with proton beams degraded in homogeneous water phantoms with 100, 150 and 200 mm thickness, to specifically investigate the effect of particle densities and phantom thickness on the reconstruction performance.

E. Model Training

In all following experiments, we use simulated training data of a particle beam (10,000 primaries), directly penetrating the detector without any degrading phantom material placed in between beam and detector. This allows us to remove any dependence on phantom geometry during training phase, avoiding any kind of overfitting to particular test configurations. We then optimize fifteen independent models (to reduce the overall runtime, we only use the first five models for evaluating the spot scanning performances) on randomly sampled environments for 1,000 iterations, to provide a stable estimate of the model performance together with inter-run uncertainties ($\pm 1\sigma$; ± 1 standard error of the mean (SEM) for model comparisons). We further provide statistical confidences in terms of p-values for all model comparisons using a two-sided Welch's t-test [69], [70]. Each environment is generated using particle tracks from the training dataset, each constructing a graph geometry with 100 primaries each frame (p^+/F).

F. Reconstruction Performance: Head Phantom

To validate the overall reconstruction performance of the proposed approach, we analyze the purity (p) and efficiency (ϵ) of reconstructed particle trajectories for the spot scanning dataset generated for a pediatric head phantom, as described in Section VIII-D. Here, we focus particularly on reconstruction results for particle densities of 50, 100 and 150 p^+/F, which corresponds closest to the particle densities to be expected in the Bergen pCT detector, given the readout capabilities of the system. In Fig. 6, Fig. 7 and Table I we visualize and compare the individual reconstruction performance for all beam spot positions directly penetrating the head phantom geometry. As depicted in Table I, we report average median purities in the

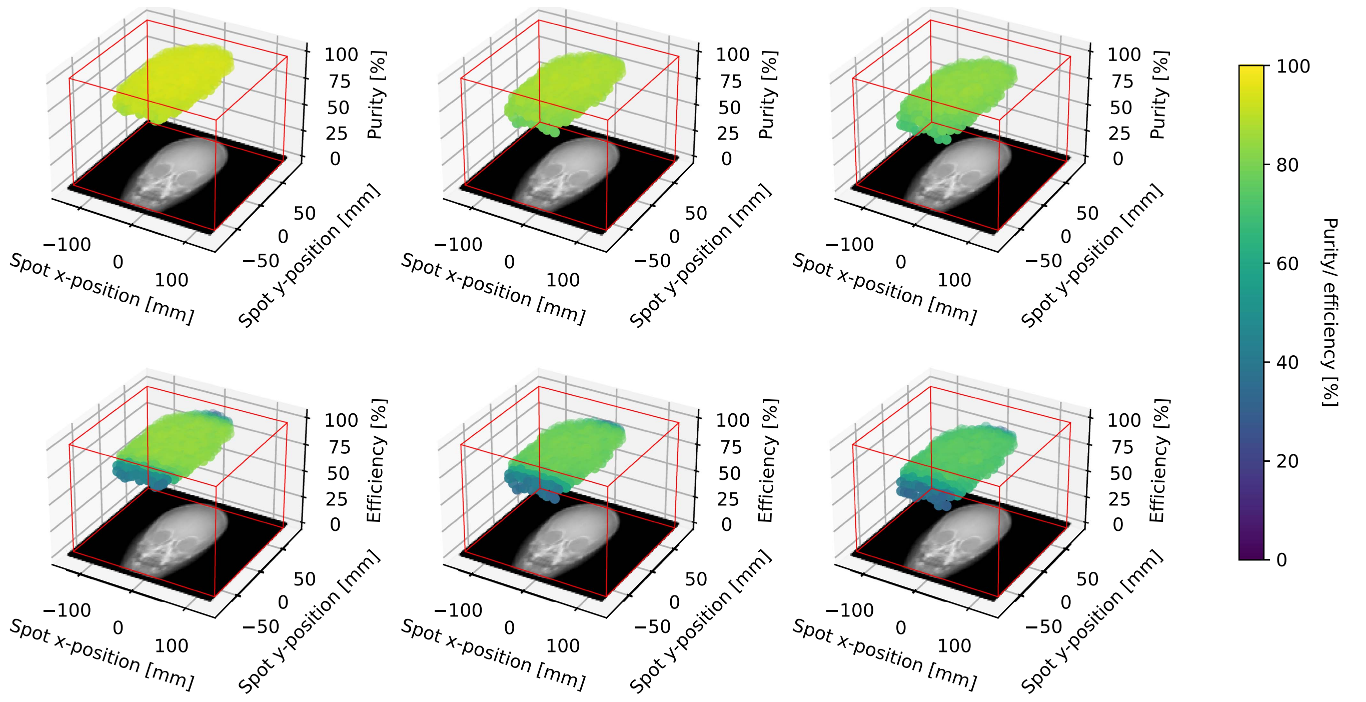


Fig. 6. Reconstruction performance, including purity p (bottom) and efficiency ϵ (top) on pediatric head phantom with 5 mm spot spacing and $\sigma_{xy} = 2$ mm, used for reconstruction of coronal pRad. Left to right: $50p^+/F$, $100p^+/F$, and $150p^+/F$. Marked in red is the projected sensitive detector area with 270 mm × 164 mm.

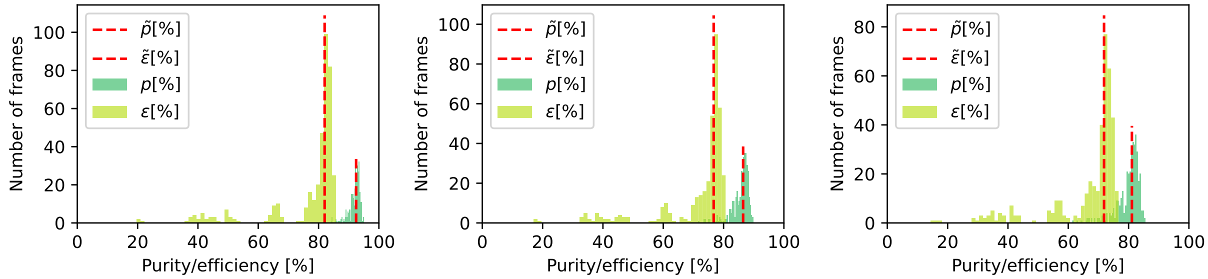


Fig. 7. Distribution of reconstruction performance including purity p and efficiency ϵ on pediatric head phantom with 7 mm spot spacing and $\sigma_{xy} = 2$ mm used for reconstruction of coronal pRad, left to right: $50p^+/F$, $100p^+/F$, and $150p^+/F$. Marked in red are the median values.

TABLE I
MEDIAN $\tilde{x}$ AND AVERAGE $\overline{x}$ PURITY (p) AND EFFICIENCY (ϵ) OF RECONSTRUCTED TRACKS FOR HEAD PHANTOM WITH DENSITIES IN BETWEEN 10 AND 100 p^+/F

	50 p$^+$/F		**100 p$^+$/F**		**150 p$^+$/F**	
Metric	p [%]	ϵ [%]	p [%]	ϵ [%]	p [%]	ϵ [%]
$\tilde{x}$	92.8±0.1	80.7±0.2	86.3±0.2	75.4±0.4	80.8±0.3	70.8±0.5
$\overline{x}$	92.1±0.1	77.8±0.2	85.7±0.2	72.3±0.4	80.1±0.3	67.6±0.5

range of 80.8±0.3% up to 92.8±0.1% with efficiencies between 70.8±0.5% and 80.7±0.2%. Further, we find decreased average values, compared to the median, due to the long tail of outliers, as depicted in Fig. 7, present in the beam spot positions in the upper head and lower neck sections. We further analyze this effect in Section VIII-H.

Otherwise, the overall reconstruction performance for all beam spots is fairly consistent. Most of the beam spots contained in the center of the detector yield similar results, with some Gaussian noise around the median reconstruction performance.

G. Phantom Thickness and Particle Density

To decompose the impact of phantom thickness and particle density on both purity and efficiency, we further analyze the reconstruction performance on homogeneous water phantoms of various thickness, as described in Section VIII-D. We choose phantom thicknesses of 100, 150 and 200 mm as realistic equivalences for human tissue and evaluate the results on a wide variety of particle densities (p^+/F) in the range of 10 to 200 primaries per frame. As depicted in Table II, we achieve average purities in the range of 75.3±0.6% up to 98.8±0.1% and efficiencies in between 66.6±0.6% and 88.4±1.6% for

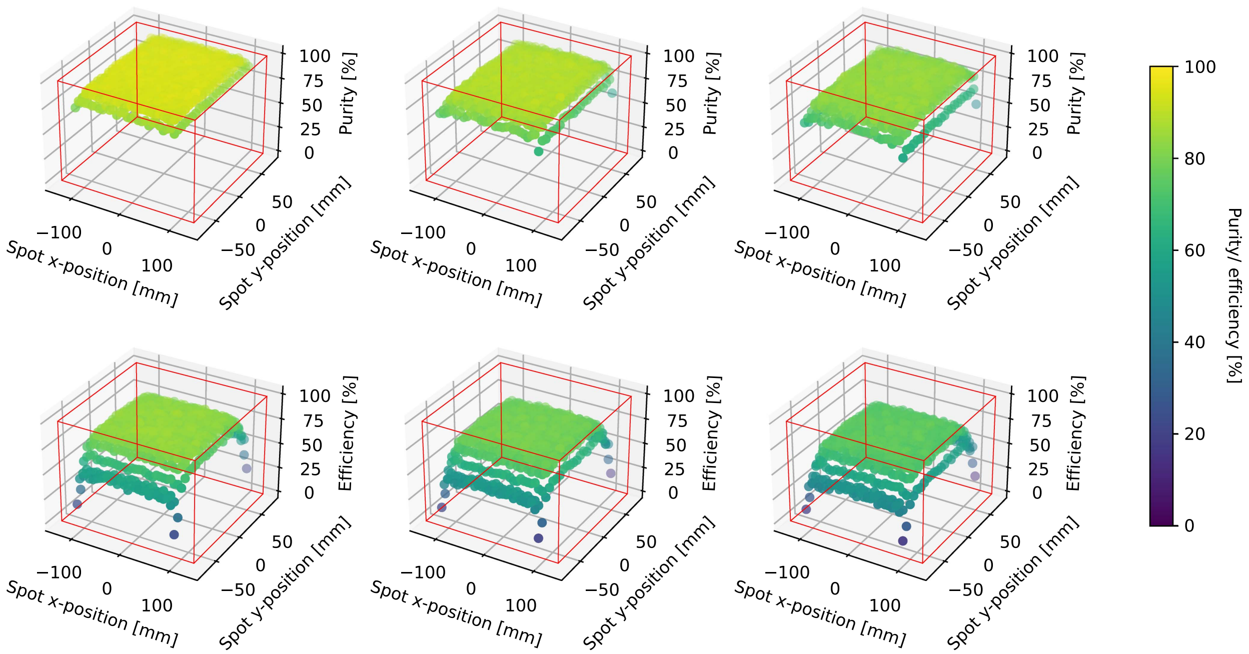


Fig. 8. Reconstruction performance including purity p (top) and efficiency ϵ (bottom) for various beam spot positions on a homogeneous 150 mm water phantom with 7 mm spot spacing and $\sigma_{xy} = 2$ mm, left to right: $50p^+/F$, $100p^+/F$, and $150p^+/F$. Marked in red is the projected sensitive detector area with 270 mm × 164 mm.

TABLE II
PURITY (p) AND EFFICIENCY (ϵ) OF RECONSTRUCTED TRACKS FOR 100, 150, AND 200 MM WATER PHANTOMS WITH DENSITIES IN BETWEEN 10 AND 100 p^+/F

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	p [%]	ϵ [%]	p [%]	ϵ [%]	p [%]	ϵ [%]
10	98.3±0.1	87.6±1.1	98.9±0.1	88.0±1.3	98.8±0.1	88.4±1.6
20	96.5±0.1	85.7±0.4	97.5±0.1	87.4±0.5	97.7±0.1	88.0±0.8
30	95.1±0.2	84.3±0.3	96.1±0.2	86.3±0.4	96.5±0.1	87.2±0.5
40	93.5±0.2	82.5±0.4	95.1±0.2	85.3±0.4	95.4±0.2	86.3±0.4
50	92.5±0.2	81.5±0.3	93.7±0.2	84.0±0.4	94.4±0.2	85.4±0.4
100	85.6±0.3	75.2±0.5	88.8±0.5	79.0±0.5	89.5±0.4	80.8±0.5
150	80.5±0.4	70.8±0.6	83.8±0.7	74.4±0.6	85.3±0.6	76.9±0.5
200	75.3±0.6	66.6±0.6	80.0±0.8	70.9±0.6	81.7±0.6	73.8±0.5

all particle densities and phantom geometries. Further, we can observe better performances for higher phantom thicknesses, which can be explained by the decreased number of layers to be reconstructed, due to the lower residual energy. Similarly, we achieve higher performances for lower particle densities and observe increasing deterioration in both p and ϵ with higher particle densities.

H. Beam Spot Positioning and Detector Boundaries

We further analyze the impact of particular beam spot positions on homogeneous phantom geometries to identify systematic biases in the reconstruction performance. We therefore compare both purity (p) and efficiency (ϵ) for various beam spot positions contained in an area of 196 mm × 154 mm, with a spacing between beam spots of 7 mm in x and y direction. We further add homogeneous water phantoms of 100, 150 and 200 mm thickness in between particle beam and detector. Fig. 8 shows the obtained results for the 150 mm water phantom with 50, 100 and 150 p^+/F. For brevity, we omit the detailed results for the remaining water phantoms, as the overall results are directly comparable. Fig. 8 depicts the performance result of reconstructed particle trajectories for different beam spot positions throughout the area of detector aperture (marked by a wireframe cube in red). Here, the overall reconstruction performance inside the center area of the detector remains stable (with expected random Gaussian noise around the average reconstruction performance) for all particle densities, demonstrating the translation invariance introduced by the beam spot dependent normalization of features. However, we observe a drastic drop in reconstruction efficiency in the outer sections, increasing exponentially with respect to the distance to the detector boundary. In contrast, the purity of reconstructions remains significantly more stable, while the overall tendency of decreased performance is still observable. We argue, based on the direct correlation of reconstruction performance and distance to the detector boundary, that the drop in efficiency can be fully explained by particles leaving the detector which, thus, cannot be fully reconstructed. This claim is further confirmed by the fact that the efficiency degrades with much faster speed than the purity, due to the strict filtering of particle trajectories without a Bragg peak (Section VIII-B). The same effect of decreased reconstruction performance for particles leaving the detector can be found in Section VIII-F Fig. 6, generating multiple outliers which can be observed in Fig. 7.

I. Performance Gap on Heuristic Search

To ensure the competitiveness of our approach, we compare performance of our learned policy with a manually tuned heuristic search algorithm, that has been developed for this

TABLE III
PERFORMANCE GAP Δp AND $\Delta\epsilon$ ON HEURISTIC SEARCH FOR 100, 150, AND 200 MM WATER PHANTOMS WITH DENSITIES BETWEEN 10 AND 200 p^+/F

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta\epsilon$ [pp]	Δp [pp]	$\Delta\epsilon$ [pp]	Δp [pp]	$\Delta\epsilon$ [pp]
10	6.0±0.1	3.6±1.1	5.2±0.1	1.8±1.3	4.4±0.1	1.7±1.6
20	5.3±0.1	2.8±0.4	4.8±0.1	2.2±0.5	4.3±0.1	2.2±0.8
30	5.1±0.2	2.6±0.3	4.3±0.2	1.7±0.4	3.9±0.1	2.1±0.5
40	4.4±0.2	1.8±0.4	4.1±0.2	1.7±0.4	3.4±0.2	1.9±0.4
50	4.4±0.2	1.9±0.3	3.5±0.2	1.2±0.4	3.2±0.2	1.6±0.4
100	2.6±0.3	0.6±0.5	2.3±0.5	0.0±0.5	2.1±0.4	0.5±0.5
150	1.4±0.4	-0.1±0.6	0.5±0.7	-1.3±0.6	0.6±0.6	-0.8±0.5
200	-0.2±0.6	-0.8±0.6	-0.2±0.8	-2.0±0.6	0.1±0.6	-1.3±0.5

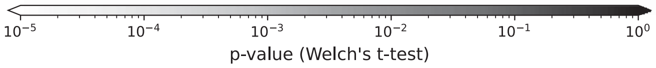
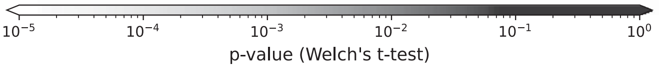


TABLE IV
PERFORMANCE GAP PURITY (Δp) AND EFFICIENCY ($\Delta\epsilon$) OF PARTIAL INFORMATION FOR 100, 150, AND 200 MM WATER PHANTOMS WITH DENSITIES BETWEEN 10 AND 200 p^+/F

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta\epsilon$ [pp]	Δp [pp]	$\Delta\epsilon$ [pp]	Δp [pp]	$\Delta\epsilon$ [pp]
10	-0.0±0.0	-0.5±0.3	-0.0±0.0	-0.6±0.3	0.0±0.0	-0.7±0.4
20	0.0±0.0	-0.2±0.1	0.0±0.0	-0.3±0.1	-0.0±0.0	-0.4±0.2
30	-0.1±0.0	-0.2±0.1	0.0±0.1	-0.4±0.1	-0.1±0.0	-0.2±0.1
40	0.0±0.1	-0.1±0.1	0.0±0.1	-0.2±0.1	-0.0±0.0	-0.2±0.1
50	-0.0±0.1	-0.1±0.1	0.0±0.1	-0.2±0.1	-0.1±0.1	-0.3±0.1
100	0.0±0.1	-0.0±0.1	-0.1±0.1	-0.2±0.2	-0.2±0.1	-0.2±0.2
150	-0.2±0.1	-0.2±0.2	-0.0±0.2	-0.1±0.2	-0.1±0.2	-0.1±0.2
200	0.6±0.4	0.6±0.3	0.2±0.3	0.2±0.3	-0.1±0.2	-0.0±0.2



particular detector by Pettersen et al. [5], based on a previous approach described in Amrouche et al. [24]. Here, the possible solution space is restricted by $S_n < S_{\max}$, where $S_n = (\sum_{i=1}^{N} \Delta\theta_i^2)^{1/2}$ is the square root of the sum of squared angular deflection terms and $S_{\max} = 278$ mrad defines a manually optimized upper limit using ground truth based on MC simulated data [5]. Following an initial seed pair, feasible candidates are identified recursively based on the aforementioned metric. To further limit the search cone, new candidates are only added to the queue if the respective S_n values are within 15% of each other. Finally, candidates with $\Delta\theta < 50$ mrad are always added independently of the aforementioned restrictions. After all feasible solutions are explored, the optimal track candidate is selected based on the lowest S_n score [5]. To make the results comparable, we also replace the process of finding suitable seed pairs with a ground truth seeding, restricting the search tree only to solutions including the correct seed. As depicted in Table III, our method is able to achieve comparable results to the algorithm in Pettersen et al. [5] with better or mostly on par performances. Particularly for densities up to 100 p^+/F, the learned policy outperforms the manually tuned heuristic, with improvements in purity up to 6.0±0.1 percentage points (pp). However, we can observe rare cases of decreased performance, especially for high particle densities, demonstrating the limits of the learned heuristic.

For particle densities of 150 and 200 p^+/F the results vary between phantom thicknesses. In most cases, we still are able to obtain marginally improved purities while observing different amounts of decrease in the reconstruction efficiencies. In spite of that, all efficiencies remain in a range of approximately two percentage points (-2.0±0.6 pp) compared to our proposed approach.

J. Performance Gap of Partial Information

Finally, we analyze the performance gap of our approach, as opposed to an equivalent model trained in a supervised manner. We therefore compare the baseline model described in the preceding Sections VIII-F, VIII-G, and VIII-H with a second model, sharing the same model architecture, by minimizing the negative log likelihood of random undertaken track transitions (x) of primary particle tracks given the ground truth (y) provided by MC simulations ($\mathcal{D}$) according to:

$$\theta^* = \underset{\theta \in \Theta}{\arg\max} \quad \mathbb{E}_{x,y\sim\mathcal{D}} \left[P(y|x) \right]. \tag{26}$$

Here, the parametrization of x follows the feature description provided in Section VI. To limit the scope of this work, we intentionally do not compare the approach to state-of-the-art graph neural network architectures used for particle tracking, as we only intend to identify the limitations introduced by learning in a partial information setting using the proposed reward signal. As shown in Table IV, we report comparable results with a non-significant worst case performance gap of -0.2 ± 0.1pp purity and -0.7 ± 0.4pp efficiency. In all cases, the proposed algorithm stays in less than a percentage point range compared to the comparable supervised algorithm, demonstrating the strong performance of the proposed algorithm and the suitability of the chosen reward function, given the proposed state representation. Further, based on the comparable performance of all three algorithms (Sections VIII-G and VIII-I) we argue that the parametrization of only a single particle trajectory, while providing a good amount of information, limits the possible reconstruction performance of the approach especially in complex scenarios in high density sections of the particle detector with multiple comparable track hypotheses. This lack of information, might be adequately resolved by also considering the preferences of track hypotheses of similar tracks in the direct neighborhood.

K. Ablation Studies

To better understand and verify the effectiveness of the contributions to the proposed architecture, we ablate key components that represent substantial changes to existing architectures and approaches in traditional neural combinatorial optimization. In the following tables, we present the results in terms of performance gaps. In Table V we demonstrate the sufficiency of the proposed state representation in Section V considering only a history of the particle track over a single layer as opposed to the more complicated representation capturing the full particle history in a hidden representation learned using a LSTM network layer.

TABLE V
PERFORMANCE GAP OF PURITY (Δp) AND EFFICIENCY ($\Delta \epsilon$) FOR MODELS WITH MDP AND POMDP STATE DEFINITION

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]
10	0.0±0.0	-0.2±0.3	0.0±0.1	-0.2±0.4	0.1±0.1	-0.2±0.4
20	0.0±0.0	-0.0±0.1	0.0±0.1	-0.1±0.1	0.0±0.1	-0.0±0.2
30	-0.0±0.1	-0.0±0.1	0.0±0.1	-0.1±0.1	-0.0±0.1	-0.0±0.1
40	0.1±0.1	-0.0±0.1	0.0±0.1	-0.1±0.1	-0.0±0.1	-0.0±0.1
50	-0.0±0.1	-0.1±0.1	0.0±0.1	-0.1±0.1	-0.1±0.1	-0.1±0.1
100	0.1±0.1	-0.1±0.2	0.2±0.2	-0.0±0.2	0.0±0.2	-0.0±0.2
150	0.2±0.2	0.0±0.2	0.2±0.3	0.0±0.2	0.2±0.3	0.1±0.2
200	0.2±0.2	-0.1±0.2	0.1±0.3	-0.0±0.2	0.1±0.3	0.1±0.3

TABLE VI
PERFORMANCE GAP Δp AND $\Delta \epsilon$ FOR BASELINE MODEL AND MODEL WITHOUT DEFAULT POSITIONAL ENCODING MECHANISM

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]
10	14.3±2.1	13.6±2.1	12.8±1.9	12.2±2.1	12.7±2.1	12.2±2.8
20	22.9±2.6	21.6±2.3	20.3±2.3	20.1±2.3	19.6±2.1	19.8±2.5
30	30.1±2.6	27.6±2.3	26.6±2.4	25.6±2.2	25.7±2.2	25.2±2.2
40	34.9±2.6	31.7±2.2	32.1±2.4	30.2±2.1	31.4±2.2	29.9±2.0
50	39.3±2.4	35.4±2.1	36.7±2.2	34.2±2.0	35.9±2.1	33.6±1.9
100	50.6±1.7	44.7±1.5	49.9±1.6	44.9±1.4	49.2±1.5	44.9±1.4
150	55.2±1.3	48.6±1.1	54.9±1.2	48.9±1.1	54.9±1.1	49.8±1.0
200	56.1±0.9	49.5±0.8	57.7±0.9	51.2±0.8	58.0±0.9	52.4±0.8

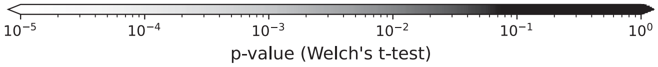
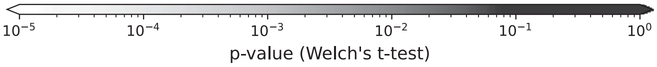


TABLE VII
PERFORMANCE GAP Δp AND $\Delta \epsilon$ FOR BASELINE MODEL AND MODEL WITH POSITIONAL ENCODING WITHOUT ADAPTIVE RECEPTIVE FIELD

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]
10	0.4±0.2	-0.0±0.3	0.2±0.1	-0.1±0.4	0.2±0.1	-0.3±0.4
20	0.8±0.3	0.5±0.3	0.5±0.2	0.2±0.2	0.4±0.1	0.1±0.3
30	1.2±0.4	0.9±0.4	0.7±0.3	0.4±0.3	0.7±0.2	0.4±0.2
40	1.4±0.5	1.0±0.5	0.9±0.3	0.6±0.3	0.8±0.2	0.5±0.2
50	1.9±0.6	1.4±0.6	1.2±0.4	0.8±0.4	0.9±0.3	0.6±0.3
100	2.6±1.0	2.2±0.9	2.1±0.6	1.5±0.6	1.5±0.4	1.1±0.4
150	3.4±1.1	2.7±1.0	2.5±0.8	2.0±0.7	2.1±0.6	1.5±0.5
200	3.9±1.3	3.1±1.2	3.1±1.0	2.6±0.9	2.5±0.7	2.0±0.6

TABLE VIII
PERFORMANCE GAP Δp AND $\Delta \epsilon$ FOR BASELINE MODEL TRAINED WITH AND WITHOUT REWARD NORMALIZATION

	WPT 100 mm		WPT 150 mm		WPT 200 mm	
p^+/F	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]	Δp [pp]	$\Delta \epsilon$ [pp]
10	1.4±0.9	1.2±0.9	1.1±0.7	0.6±0.7	0.7±0.4	0.1±0.6
20	2.4±1.6	2.4±1.4	1.7±1.1	1.7±1.1	1.2±0.7	1.0±0.7
30	3.2±2.0	3.2±1.8	2.3±1.5	2.3±1.3	1.8±1.0	1.7±1.0
40	3.9±2.4	3.9±2.1	2.9±1.8	2.9±1.6	2.2±1.3	2.2±1.2
50	4.4±2.6	4.3±2.3	3.5±2.1	3.3±1.8	2.6±1.5	2.5±1.4
100	6.3±3.3	5.9±2.9	5.3±2.7	4.9±2.4	4.2±2.1	4.0±1.9
150	7.5±3.6	6.9±3.1	6.8±3.0	6.1±2.7	5.4±2.4	5.0±2.2
200	8.4±3.6	7.7±3.1	7.8±3.2	7.0±2.8	6.5±2.7	6.0±2.4



We further present the ablation results comparing the default attention mechanism without positional encoding with the proposed adaptive positional encoding mechanism described in Section VI-C and the non-modified version proposed for Transformer Architectures in [49].

We observe a big improvement in both purity and efficiency introduced by positional encoding (up to 58.0±0.9 pp in p and 52.4±0.8 pp in ϵ; ref. Table VI) in combination with a significant improvement in inter-run convergence quality. By introducing an adaptive rescaling mechanism, allowing to reduce the receptive area of the encoding mechanism, we can further improve the reconstruction results, particularly in high density particle configurations. Finally, we demonstrate the improvement in reconstruction quality by introducing reward normalization, independently handling gained rewards of tracking and calorimeter layer in order to handle the different order of scales in reward signals. We particularly observe improved purities and efficiencies for high particle densities (ref. Table VIII).

IX. CONCLUSION

In this paper, we introduce a novel reconstruction scheme for particle tracking in high energy physics applications using model-free reinforcement learning on graph structured data, maximizing the physical plausibility of undertaken state transitions. With this approach, we take a step towards a unified solution combining the advantages of ground-truth-free iterative reconstruction algorithms with the power of deep neural network architectures. Our approach generalizes well to various, previously unseen, phantom geometries, particle densities and beam spot positions. We demonstrate on simulated data that our approach is able to consistently learn good policies, able to reconstruct trajectories better, or on par with the performance of a comparable heuristic search algorithm. We further show that we are able to achieve similar results to an equivalent supervised trained model minimizing the negative log likelihood of undertaken transitions, given the correct label. We argue, based on the strong performance compared to the heuristic search algorithm and the equivalent supervised trained model, that the current state definition, which only satisfies the Markov property w.r.t. a single track, is unlikely sufficient to resolve complicated track reconstruction errors due to confusion of particles in the dense Gaussian core of the particle beam or large angle scattering, limiting the possible reconstruction performance of the approach. Instead, a richer representation of the whole system considering a full parametrization of surrounding tracks might be essential to resolve this kind of reconstruction conflicts. Finally, further work is still required to locate the performance of the proposed optimization scheme in the existing literature, using both simulated and real detector data, when available.

MEMBERS OF THE BERGEN PCT COLLABORATION

Max Aehle[a], Johan Alme[b], Gergely Gábor Barnaföldi[c], Tea Bodova[b], Vyacheslav Borshchov[d], Anthony van den Brink[e], Mamdouh Chaar[b], Viljar Eikeland[b], Gregory Feofilov[f],

Christoph Garth[g], Nicolas R. Gauger[a], Georgi Genov[b], Ola Grøttvik[b], Håvard Helstrup[h], Sergey Igolkin[f], Ralf Keidel[i], Chinorat Kobdaj[j], Tobias Kortus[i], Viktor Leonhardt[g], Shruti Mehendale[b], Raju Ningappa Mulawade[i], Odd Harald Odland[k, b], George O'Neill[b], GÃ¡bor Papp[l], Thomas Peitzmann[e], Helge Egil Seime Pettersen[k], Pierluigi Piersimoni[b,m], Maksym Protsenko[d], Max Rauch[b], Attiq Ur Rehman[b], Matthias Richter[n], Dieter Röhrich[b], Joshua Santana[i], Alexander Schilling[i], Joao Seco[o, p], Arnon Songmoolnak[b, j], Ákos Sudár[c, q], Jarle Rambo SÃ¸lie[r], Ganesh Tambave[s], Ihor Tymchuk[d], Kjetil Ullaland[b], Monika Varga-Kofarago[c], Lennart Volz[t, u], Boris Wagner[b], Steffen Wendzel[i], Alexander Wiebel[i], RenZheng Xiao[b, v], Shiming Yang[b], Hiroki Yokoyama[e], Sebastian Zillien[i]

a) Chair for Scientific Computing, TU Kaiserslautern, 67663 Kaiserslautern, Germany b) Department of Physics and Technology, University of Bergen, 5007 Bergen, Norway; c) Wigner Research Centre for Physics, Budapest, Hungary; d) Research and Production Enterprise "LTU" (RPELTU), Kharkiv, Ukraine; e) Institute for Subatomic Physics, Utrecht University/Nikhef, Utrecht, Netherlands; f) St. Petersburg University, St. Petersburg, Russia; g) Scientific Visualization Lab, TU Kaiserslautern, 67663 Kaiserslautern, Germany; h) Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, 5020 Bergen, Norway; i) Center for Technology and Transfer (ZTT), University of Applied Sciences Worms, Worms, Germany; j) Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand; k) Department of Oncology and Medical Physics, Haukeland University Hospital, 5021 Bergen, Norway; l) Institute for Physics, Eötvös Loránd University, 1/A Pázmány P. Sétány, H-1117 Budapest, Hungary; m) UniCamillus – Saint Camillus International University of Health Sciences, Rome, Italy; n) Department of Physics, University of Oslo, 0371 Oslo, Norway; o) Department of Biomedical Physics in Radiation Oncology, DKFZ–German Cancer Research Center, Heidelberg, Germany; p) Department of Physics and Astronomy, Heidelberg University, Heidelberg, Germany; q) Budapest University of Technology and Economics, Budapest, Hungary; r) Department of Diagnostic Physics, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway; s) Center for Medical and Radiation Physics (CMRP), National Institute of Science Education and Research (NISER), Bhubaneswar, India; t) Biophysics, GSI Helmholtz Center for Heavy Ion Research GmbH, Darmstadt, Germany; u) Department of Medical Physics and Biomedical Engineering, University College London, London, U.K.; v) College of Mechanical & Power Engineering, China Three Gorges University, Yichang, People's Republic of China.

Acknowledgment

The simulations were partly executed on the high performance cluster "Elwetritsch" at the TU Kaiserslautern, which is part of the "Alliance of High Performance Computing Rheinland-Pfalz" (AHRP). We kindly acknowledge the support of the regional university computing center (RHRK). The ALPIDE chip was developed by the ALICE collaboration at CERN. The open access publication was supported by the publications fund of the University of Applied Sciences Worms.

References

[1] J. Duarte and J. R. Vlimant, "Graph neural networks for particle tracking and reconstruction," 2020, *arXiv: 2012.01249*.

[2] S. Thais et al., "Graph neural networks in particle physics: Implementations, innovations, and challenges," 2022. [Online]. Available: http://arxiv.org/abs/2203.12852

[3] A. Glazov et al., "Filtering tracks in discrete detectors using a cellular automaton," *Nucl. Inst. Methods Phys. Res. A*, vol. 329, no. 1/2, pp. 262–268, 1993.

[4] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system," *Nucl. Instruments Methods Phys. Res. Sect. A: Accelerators Spectrometers Detectors Assoc. Equip.*, vol. 395, no. 2, pp. 169–184, 1997.

[5] H. E. Pettersen et al., "Proton tracking algorithm in a pixel-based range telescope for proton computed tomography," 2020, *arXiv: 2006.09751*.

[6] I. Bello et al., "Neural combinatorial optimization with reinforcement learning," in *Proc. 5th Int. Conf. Learn. Representations*, 2019, pp. 1–15.

[7] Q. Ma et al., "Combinatorial optimization by graph pointer networks and hierarchical reinforcement learning," in *Proc. AAAI Workshop Deep Learn. Graphs: Methodol. Appl.*, 2020.

[8] W. Kool, H. Van Hoof, and M. Welling, "Attention, learn to solve routing problems!," in *Proc. 7th Int. Conf. Learn. Representations*, 2019, pp. 1–25.

[9] H. Dai et al., "Learning combinatorial optimization algorithms over graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6349–6359.

[10] S. Boutle et al., "Primary vertex reconstruction at the ATLAS experiment," *J. Physics: Conf. Ser.*, vol. 119, 2017, Art. no. 117.

[11] W. Erdmann, "Vertex reconstruction at the CMS experiment," *J. Physics: Conf. Ser.*, vol. 110, 2008, Art. no. 092009.

[12] J. Mašísk, "Vertex and track reconstruction in ATLAS and CMS," in *Proc. Sci.*, vol. 13, 2013, Art. no. C02035.

[13] H. Bakhshiansohi et al., "Particle-flow reconstruction and global event description with the CMS detector," *J. Instrum.*, vol. 12, 2017, Art. no. P10003.

[14] I. Belikov, "Event reconstruction and particle identification in the ALICE experiment at the LHC," in *Proc. EPJ Web Conf.*, 2014, pp. 1–9.

[15] M. Aaboud et al., "Jet reconstruction and performance using particle flow with the ATLAS detector," *Eur. Phys. J. C*, vol. 77, no. 7, 2017, Art. no. 466.

[16] M. Centonze, "Jet flavour tagging for the ATLAS experiment," in *Proc. Sci.*, vol. 380, 2022, pp. 1–5.

[17] L. Feldkamp, "Study of b-jet tagging performance in ALICE," *J. Physics: Conf. Ser.*, vol. 509, no. 1, pp. 10–12, 2014.

[18] M. Verzetti, "Machine learning techniques for jet flavour identification at CMS," in *Proc. EPJ Web Conf.*, 2019, Art. no. 06010.

[19] J. Alme et al., "A high-granularity digital tracking calorimeter optimized for proton CT," *Front. Phys.*, vol. 8, pp. 1–20, 2020.

[20] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications," *J. Appl. Phys.*, vol. 34, no. 9, pp. 2722–2727, 1963.

[21] R. P. Johnson, "Review of medical radiography and tomography with proton beams," *Rep. Prog. Phys.*, vol. 81, no. 1, 2018, Art. no. 016701.

[22] L. H. Våge, "Reinforcement learning for charged-particle tracking reinforcement learning," in *Proc. Connecting Dots Workshop*, 2022, pp. 1–8.

[23] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2692–2700.

[24] S. Amrouche et al., "Track reconstruction at LHC as a collaborative data challenge use case with RAMP," in *Proc. EPJ Web Conf.*, 2017, pp. 1–12.

[25] A. Strandlie and R. Frühwirth, "Track and vertex reconstruction: From classical to adaptive methods," *Rev. Modern Phys.*, vol. 82, no. 2, pp. 1419–1458, 2010.

[26] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl. Instruments Methods Phys. Res. Sect. A: Accelerators Spectrometers Detectors Assoc. Equip.*, vol. 262, no. 2, pp. 444–450, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0168900287908874

[27] P. V. C. Hough, "Machine analysis of bubble chamber pictures," in *Proc. 2nd Int. Conf. High-Energy Accelerators Instrum.*, 1959, pp. 554–558.

[28] D. Primor et al., "A novel approach to track finding in a drift tube chamber," *J. Instrum.*, vol. 2, 2007, Art. no. P01009.

[29] S. Farrell et al., "Novel deep learning methods for track reconstruction," 2018. [Online]. Available: http://arxiv.org/abs/1810.06111

[30] D. Baranov et al., “The particle track reconstruction based on deep neural networks,” in *Proc. EPJ Web Conf.*, 2019, Art. no. 06018.

[31] P. Goncharov, G. Ososkov, and D. Baranov, “Particle track reconstruction with the TrackNETv2,” in *Proc. AIP Conf.*, 2019, pp. 1–5.

[32] G. DeZoort et al., “Charged particle tracking via edge-classifying interaction networks,” *Comput. Softw. Big Sci.*, vol. 5, no. 1, pp. 1–13, 2021, doi: 10.1007/s41781–021-00073-z.

[33] J. Cremer et al., “Equivariant graph neural networks for toxicity prediction,” 2023. [Online]. Available: https://chemrxiv.org/engage/chemrxiv/article-details/63e236b75c37ece322b6f162

[34] X. Ju et al., “Graph neural networks for particle reconstruction in high energy physics detectors,” 2019, *arXiv: 2003.11603*.

[35] G. Tambave et al., “Characterization of monolithic CMOS pixel sensor chip with ion beams for application in particle computed tomography,” *Nucl. Instruments Methods Phys. Research, Sect. A: Accelerators Spectrometers Detectors Assoc. Equip.*, vol. 958, 2020, Art. no. 162626, doi: 10.1016/j.nima.2019.162626.

[36] D. Groom and S. Klein, “Passage of particles through matter,” *Eur. Phys. J. C. - EUR PHYS J. C*, vol. 15, pp. 163–173, 2000.

[37] N. Bohr, “II. On the theory of the decrease of velocity of moving electrified particles on passing through matter,” *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 25, no. 145, pp. 10–31, 1913.

[38] F. Bloch, “On the deceleration of fast-moving particles passing through matter,” *Annalen der Physik*, (in German), vol. 408, no. 3, pp. 285–320, 1933.

[39] G. Moliere, “Theory of the scattering of fast charged particles. II. Repeated and multiple scattering,” *Zeitschrift fur Naturforschung - Section A: J. Phys. Sci.*, (in German), vol. 3, no. 2, pp. 78–97, 1948.

[40] G. R. Lynch and O. I. Dahl, “Approximations to multiple coulomb scattering,” *Nucl. Inst. Methods Phys. Res. B*, vol. 58, no. 1, pp. 6–10, 1991.

[41] V. L. Highland, “Some practical remarks on multiple scattering,” *Nucl. Instruments Methods*, vol. 129, no. 2, pp. 497–499, 1975.

[42] M. Mager, “ALPIDE, the monolithic active pixel sensor for the ALICE ITS upgrade,” *Nucl. Instruments Methods Phys. Res. Sect. A: Accelerators Spectrometers Detectors Assoc. Equip.*, vol. 824, no. 2016, pp. 434–438, 2016.

[43] K. Aamodt et al., “The ALICE experiment at the CERN LHC,” *J. Instrum.*, vol. 3, no. 08, pp. S08 002–S08 002, Aug. 2008, doi: 10.1088/1748–0221/3/08/s08002.

[44] A. P. De Haas et al., “The FoCal prototype - an extremely fine-grained electromagnetic calorimeter using CMOS pixel sensors,” *J. Instrum.*, vol. 13, no. 1, 2018, Art. no. P01014.

[45] N. Mazyavkina et al., “Reinforcement learning for combinatorial optimization: A survey,” *Comput. Operations Res.*, vol. 134, 2021, Art. no. 105400.

[46] R. Bellman, “A Markovian decision process,” *J. Math. Mechanics*, vol. 6, no. 5, pp. 679–684, 1957. [Online]. Available: http://www.jstor.org/stable/24900506

[47] P. Veličković et al., “Graph attention networks,” in *Proc. 6th Int. Conf. Learn. Representations*, 2018, pp. 1–12.

[48] S. Brody, U. Alon, and E. Yahav, “How attentive are graph attention networks?,” 2021. [Online]. Available: http://arxiv.org/abs/2105.14491

[49] A. Vaswani et al., “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5999–6009.

[50] M. Deudon et al., “Learning heuristics for the TSP by policy gradient,” in *Proc. Int. Conf. Integration Constraint Program. Artif. Intell. Operations Res.*, 2018, pp. 170–181.

[51] K. He et al., “Deep residual learning for image recognition,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[52] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” 2016. [Online]. Available: http://arxiv.org/abs/1607.06450

[53] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[54] X. Ai et al., “A common tracking software project,” *Comput. Softw. Big Sci.*, vol. 6, no. 1, 2022, Art. no. 8.

[55] B. Gottschalk, “On the scattering power of radiotherapy protons,” *Med. Phys.*, vol. 37, no. 1, pp. 352–367, 2010.

[56] T. Bortfeld and W. Schlegel, “An analytical approximation of depth-dose distributions for therapeutic proton beams,” *Phys. Med. Biol.*, vol. 41, pp. 1331–1339, 1996.

[57] H. E. Pettersen et al., “Accuracy of parameterized proton range models; A comparison,” *Radiat. Phys. Chem.*, vol. 144, pp. 295–297, 2018.

[58] H. Van Hasselt et al., “Learning values across many orders of magnitude,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4294–4302.

[59] J. Schulman et al., “Proximal policy optimization algorithms,” 2017. [Online]. Available: http://arxiv.org/abs/1707.06347

[60] J. Schulman et al., “High-dimensional continuous control using generalized advantage estimation,” in *Proc. 4th Int. Conf. Learn. Representations*, 2016, pp. 1–14.

[61] L. Engstrom et al., “Implementation matters in deep policy gradients: A case study on PPO and TRPO,”2020. [Online]. Available: http://arxiv.org/abs/2005.12729

[62] W. Hu, L. Xiao, and J. Pennington, “Provable benefit of orthogonal initialization in optimizing deep linear networks,” 2020. [Online]. Available: http://arxiv.org/abs/2001.05992

[63] M. J. Berger, J. S. Coursey, M. A. Zucker, and J. Chang, “ESTAR, PSTAR, and ASTAR: Computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions (Version 1.2.3),” Nat. Inst. Standards Technol., Gaithersburg, 2005.

[64] H. E. S. Pettersen et al., “Investigating particle track topology for range telescopes in particle radiography using convolutional neural networks,” *Acta Oncologica*, vol. 60, pp. 1413–1418, 2021.

[65] T. Kortus et al., “Particle tracking data: Bergen DTC prototype,” Dec. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7426388

[66] S. Jan et al., “GATE -Geant4 application for tomographic emission: A simulation toolkit for PET and SPECT,” *Phys Med Biol. Phys Med Biol*, vol. 49, no. 19, pp. 4543–4561, 2004.

[67] S. Agostinelli et al., “GEANT4 - a. simulation toolkit,” *Nucl. Instruments Methods Phys. Research, Sect. A: Accelerators Spectrometers Detectors Assoc. Equip.*, vol. 506, no. 3, pp. 250–303, 2003.

[68] V. Giacometti et al., “Development of a high resolution voxelised head phantom for medical physics applications,” *Physica Medica*, vol. 33, pp. 182–188, 2017.

[69] B. L. Welch, “The generalisation of student’s problems when several different population variances are involved,” *Biometrika*, vol. 34, no. 1–2, pp. 28–35, 1947.

[70] C. Colas, O. Sigaud, and P. Y. Oudeyer, “A hitchhiker’s guide to statistical comparisons of reinforcement learning algorithms,” in *Proc. RML@ICLR 2019 Workshop - Reproducibility Mach. Learn.*, 2019, pp. 1–23.

Tobias Kortus received the BSc degree in medical engineering from the University of Applied Sciences Furtwangen, University Campus Tuttlingen, in 2019, and the MSc degree in applied computer science from the University of Applied Sciences Esslingen, in 2021. He is currently working towards the PhD degree with the University of Applied Sciences Worms. His research interests include machine learning and reinforcement learning, with focus on applications in high energy, and medical physics.

Ralf Keidel is senior professor with the University of Applied Sciences Worms and Principal Investigator of the SIVERT research training group dealing with the algorithmic part of the proton Computed Tomography (pCT) project of the Bergen pCT Collaboration. He is member of the ALICE collaboration board and the Inter-experimental Machine Learning Working Group with CERN, Geneva. His research interests include pCT, machine learning, and optimization techniques.

Nicolas R. Gauger is a full professor and chairholder for scientific computing and director with the Computing Center (RHRK), University of Kaiserslautern as well as Principal Investigator of the SIVERT research training group dealing with the algorithmic part of the proton Computed Tomography (pCT) project of the Bergen pCT Collaboration. His research interests include numerical optimization, high-performance computing, machine learning, and pCT amongst other fields of application.