

# A high-order deterministic dynamical low-rank method for proton transport in heterogeneous media

Pia Stammer<sup>a,\*</sup>, Niklas Wahl<sup>b,c</sup>, Jonas Kusch<sup>d,1</sup>, Danny Lathouwers<sup>a,1</sup>

<sup>a</sup> Delft University of Technology, Dept. of Radiation Science and Technology, Netherlands

<sup>b</sup> Division of Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Germany

<sup>c</sup> National Center for Radiation Research in Oncology (NCRO), Heidelberg Institute for Radiation Oncology (HIRO), Germany

<sup>d</sup> Dept. of Data Science, Norwegian University of Life Sciences, Norway

## ARTICLE INFO

### Keywords:

Deterministic transport  
Proton  
Boltzmann  
Fokker-Planck  
Model order reduction  
Dynamical low-rank approximation

## ABSTRACT

Dose calculations in proton therapy require the fast and accurate solution of a high-dimensional transport equation for a large number of (pencil) beams with different energies and directions. Deterministically solving this transport problem at a sufficient resolution can however be prohibitively expensive, especially due to highly forward peaked scattering of the protons. We propose using a model order reduction approach, the dynamical low-rank approximation (DLRA), which evolves the solution on the manifold of low-rank matrices in (pseudo-)time. For this, we compare a collided-uncollided split of the linear Boltzmann equation and its Fokker-Planck approximation. We treat the uncollided part using a ray-tracer and combine high-order phase space discretizations and a mixture model for materials with DLRA for the collided equation. Our method reproduces the results of a full-rank reference code at significantly lower rank, and thus computational cost and memory, and further makes computations feasible at much higher resolutions. At higher resolutions, we also achieve good accuracy with respect to TOPAS MC in homogeneous as well as heterogeneous materials. Finally, we demonstrate that several beam sources with different angles can be computed with little cost increase compared to individual beams.

## 1. Introduction

Accurate modeling of charged particle transport, particularly proton transport, is essential for applications such as radiation therapy, where dose deposition must be carefully controlled. The linear Boltzmann equation provides a physically accurate model for charged particle transport. However, solving it numerically is challenging due to multiscale effects, a high-dimensional phase space (space, angle, and energy), and strongly forward-peaked scattering. The latter necessitates a finely resolved phase space discretization or specialized methodological modifications, thus often leading to prohibitive memory requirements and computational costs. These computational requirements limit the feasibility of using the full Boltzmann model, particularly in settings that require many simulations, such as uncertainty quantification and treatment plan optimization in intensity-modulated proton therapy.

For this reason, Monte Carlo methods, which can statistically solve the full physical model at high precision by simulating large numbers of particles, require acceleration techniques and strong computational resources to be practical [1,2]. Deterministic solutions

\* Corresponding author.

E-mail address: [p.k.stammer@tudelft.nl](mailto:p.k.stammer@tudelft.nl) (P. Stammer).

<sup>1</sup> Equal contributions, order alphabetical.

to the Boltzmann equation remove the statistical noise and slow convergence of Monte Carlo methods and further allow access to full angle- and energy-resolved particle densities, thus better supporting further computations (e.g. uncertainty quantification) [1,3]. Although the accuracy of such solvers can compete with that of Monte Carlo methods given equivalent physical modeling [1,4], they are not as wide-spread due to their high algorithmic complexity and challenges in efficient implementation. Rather, simplified models such as pencil beam algorithms, which apply dose kernels precomputed from Fermi-Eyges theory [5] or Gaussian fits to measurements and Monte Carlo simulations, are frequently used in practice [6–9]. While these models offer a computationally efficient alternative, they are based on simplifying assumptions which limit precision especially in heterogeneous media [9,10]. Alternatively, deterministic approaches have been developed to efficiently resolve forward-peaked scattering, mainly based on multi-grid [11,12] and adaptive grid refinement [13,14]. While such refinement strategies can effectively resolve angular dependencies, they still require a sufficiently fine spatial resolution [15]. Another approach in computational electron therapy [16] reduces the memory footprint and computational costs by leveraging dynamical low-rank approximation (DLRA) [17]. This approach represents and evolves the solution to the linear Boltzmann equation in a low-rank format, thereby significantly reducing the number of degrees of freedom. Since DLRA is a dimensionality reduction method for time-dependent problems, a central idea in [16] is to interpret the energy as a pseudo-time by using a continuous slowing down approximation [18]. This interpretation treats energy loss in the medium as a surrogate for time evolution, enabling the use of time-dependent model reduction techniques such as DLRA.

A central challenge of DLRA is the inherent stiffness of the evolution equations of low-rank factors [19]. Therefore, robust time integrators that are not affected by this stiffness have been developed. The main idea of these integrators is to move only in flat subspaces of the low-rank manifold, thereby avoiding directions with a high curvature that result in stiff evolution equations [20]. Perhaps the most popular class of robust time integrators is the (augmented) basis-update & Galerkin (BUG) integrator [21,22], which has since its introduction been modified to enable parallel update computations [23] and to achieve high-order accuracy [24,25]. Robust integrators for DLRA have been used extensively in kinetic theory [26], ranging from applications in plasma physics [27,28] to neutron criticality [29,30]. The use of DLRA in such problems relies heavily on the derivation of efficient spatial and temporal discretizations that preserve underlying structures.

While the method proposed in [16] allows for highly efficient numerical simulations, it is currently limited to electron therapy, water-equivalent materials, as well as first-order temporal and spatial discretizations. An extension to proton simulations is challenging, primarily due to highly peaked scattering coefficients and the requirement of straggling terms in the continuous slowing down approximation. The latter makes the representation of energy as a pseudo-time more challenging and thus complicates the use of DLRA as a time-dependent model reduction technique.

In this work, we propose a novel extension of DLRA to proton transport that addresses these challenges through the following contributions: (1) a ray-tracing model for uncollided particles to handle straggling, (2) high-order phase-space discretizations, and (3) a low-rank material decomposition. The raytracing model exploits the small number of relevant angular directions of uncollided particles to achieve low computational costs while being able to efficiently incorporate straggling effects. Collided particles are then advanced without straggling, thus enabling the use of dynamical low-rank approximation. Moreover, sharp gradients in proton transport require high-order phase-space discretizations. To mitigate computational costs imposed by limiters, we propose a nodal second-order upwind stencil combined with a fourth-order Runge–Kutta method in pseudo-time, i.e. energy, for numerical stability. We further use a modal spherical harmonics method ( $P_N$ ) to discretize the directional domain. Lastly, to allow the incorporation of different materials, we propose a low-rank material decomposition similar to [30], which enables efficient evaluation of the DLRA evolution equations.

The remaining manuscript is structured as follows: First we give a short introduction to proton transport and the dynamical low-rank approximation in Section 2. The former relies on the continuous slowing down approximation of the Boltzmann or Fokker-Planck equation, including a transformation that enables the use of energy as pseudo-time in the presence of material-dependent parameters. Then, the main method is introduced in Section 3, including the ray tracer for the uncollided and spatial and angular discretizations as well as DLRA evolution equations based on the rank-adaptive BUG integrator for the time/energy update of the collided part of the equation. Finally, in Section 4, we present a variety of numerical experiments and in Section 5, we discuss the computational complexity and memory reduction. We provide a discussion and conclusion in Sections 6 and 7.

## 2. Background

### 2.1. Transport model for proton therapy

We model the transport of particles through a medium using the continuous slowing down approximation of the steady-state linear Boltzmann equation

$$\Omega \cdot \nabla_{\mathbf{r}} \psi(E, \mathbf{r}, \Omega) - \frac{\partial S(E, \mathbf{r}) \psi(E, \mathbf{r}, \Omega)}{\partial E} - \frac{1}{2} \frac{\partial^2 T(E, \mathbf{r}) \psi(E, \mathbf{r}, \Omega)}{\partial E^2} = \Gamma \psi(E, \mathbf{r}, \Omega), \quad (1)$$

where the phase space of the particle density  $\psi$  consists of energy  $E \in [E_{\min}, E_{\max}] \subset \mathbb{R}_+$ , space  $\mathbf{r} \in \mathbb{R}^3$  and direction of flight  $\Omega \in \mathbb{S}^2$ . The stopping power  $S$  describes the average rate of energy loss and straggling  $T$  represents the statistical spread in energy loss due to stochastic interactions. The collision operator  $\Gamma$  on the right-hand side depends on the differential and total scattering cross sections  $\Sigma_s$  and  $\Sigma_t$  and is defined as

$$\Gamma \psi = \int_{\mathbb{S}^2} \Sigma_s(E, \mathbf{r}, \Omega' \cdot \Omega) \psi(E, \mathbf{r}, \Omega') d\Omega' - \Sigma_t(E, \mathbf{r}) \psi(E, \mathbf{r}, \Omega) \quad (2)$$

in case of the Boltzmann operator and as

$$\Gamma\psi = \frac{\xi_1}{2} L\psi \quad (3)$$

for the Fokker-Planck operator, where  $\xi_1$  is the first transport coefficient defined as

$$\xi_1 = 2\pi \int_{-1}^1 (1 - \mu_0) \Sigma_s(E, \mathbf{r}, \mu_0) d\mu_0, \quad \mu_0 = \cos(\theta) = \boldsymbol{\Omega}' \cdot \boldsymbol{\Omega} \quad (4)$$

and  $L$  is the spherical Laplace–Beltrami operator

$$L = \left[ \frac{\partial}{\partial \mu} (1 - \mu^2) \frac{\partial}{\partial \mu} + \frac{1}{1 - \mu^2} \frac{\partial^2}{\partial \Phi^2} \right], \quad (5)$$

where the spherical coordinates  $\mu \in [-1, 1]$  and  $\Phi \in [0, 2\pi)$  represent the cosine of the polar angle and the azimuthal angle, respectively. Note, that, under the assumption of highly forward-peaked differential cross sections, the Fokker-Planck operator has been shown to be the  $\mathcal{O}(1)$  asymptotic approximation of the Boltzmann scattering operator in the small-angle parameter  $\epsilon$  [31].

In order to later use energy as pseudo-time and separate angular and spatial dependency in the collision operator we apply the transformation  $\tilde{\psi}(E, \mathbf{r}, \boldsymbol{\Omega}) := S(E, \mathbf{r})\psi(E, \mathbf{r}, \boldsymbol{\Omega})$  and assume that the tissue at each point in space is composed of a mixture of the same set  $\mathbb{A} = \{\text{H, C, N, O, Na, Mg, P, S, Cl, Ar, K, Ca}\}$  of base materials with differing weights  $w_i(\mathbf{r}), i \in \mathbb{A}$  [32]. Then the stopping power and cross sections can also be computed as weighted mixtures of the respective parameters for the individual elements using the Bragg additivity rule [33]. Note that this model for material composition is much more accurate [33,34] than assuming water-equivalence as in [16], while still enabling a separation of spatial and angular dependence which is instrumental for the construction of an efficient DLRA algorithm. Then scattering cross sections  $\Sigma_s$  are computed as

$$\Sigma_s(E, \mathbf{r}, \mu_0) = \sum_{i \in \mathbb{A}} w_i(\mathbf{r}) \Sigma_{s,i}(E, \mu_0), \quad (6)$$

where  $\Sigma_{s,i}$  is the scattering cross-section of material  $i$ . The total scattering cross section  $\Sigma_t(E, \mathbf{r})$  is computed analogously, where for material  $i$  we have  $\Sigma_{t,i}(E) = 2\pi \int_{-1}^1 \Sigma_{s,i}(E, \mu_0) d\mu_0$ . The resulting transport equation then takes the following form

$$-\frac{\partial \tilde{\psi}(E, \mathbf{r}, \boldsymbol{\Omega})}{\partial E} = -\boldsymbol{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\tilde{\psi}(E, \mathbf{r}, \boldsymbol{\Omega})}{S(E, \mathbf{r})} + \frac{1}{2} \frac{\partial^2 T(E, \mathbf{r}) \tilde{\psi}(E, \mathbf{r}, \boldsymbol{\Omega})}{\partial E^2 S(E, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i \Gamma_i \frac{\tilde{\psi}(E, \mathbf{r}, \boldsymbol{\Omega})}{S(E, \mathbf{r})}, \quad (7)$$

where for Boltzmann

$$\Gamma_i \psi = \int_{\mathbb{S}^2} \Sigma_{s,i}(E, \mathbf{r}, \boldsymbol{\Omega}' \cdot \boldsymbol{\Omega}) \psi(E, \mathbf{r}, \boldsymbol{\Omega}') d\boldsymbol{\Omega}' - \Sigma_{t,i}(E, \mathbf{r}) \psi(E, \mathbf{r}, \boldsymbol{\Omega}) \quad (8)$$

and for Fokker-Planck

$$\Gamma_i \psi = \frac{\xi_{1,i}}{2} L, \quad \xi_{1,i} = 2\pi \int_{-1}^1 (1 - \mu_0) \Sigma_{s,i}(E, \mathbf{r}, \mu_0) d\mu_0. \quad (9)$$

### 2.1.1. Modeling of physical quantities

The particle and material dependent quantities in the transport equation include scattering cross sections, stopping power and straggling coefficients. The accuracy of our model strongly depends on a good modeling of these quantities. We determine the values for the set  $\mathbb{A}$  containing 12 elements which are used to constitute tissue in the human body and mix these according to their proportions at each point in space. For comparability, we extract the stopping power for each material and different energies from the Monte Carlo code TOPAS MC [35]. Such data is not as easily obtained for straggling and differential cross sections, therefore we determine these based on physics models.

We use William's model [36] to compute the modified straggling coefficient:

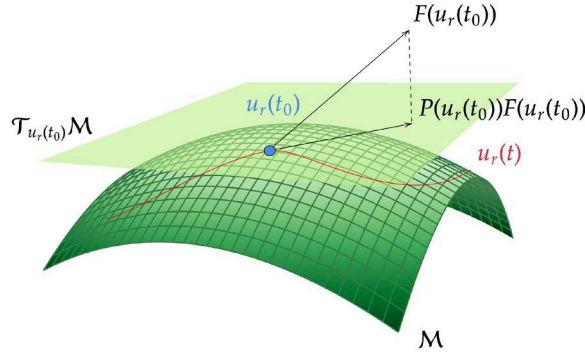
$$T(\mathbf{r}, E) = \sum_{i \in \mathbb{A}} \frac{1}{(4\pi\epsilon_0)^2} N_i(\mathbf{r}) 4\pi e^4 Z_i \left( \frac{4I_i}{3m_e v_p(E)^2} \ln \frac{2m_e v_p(E)^2}{I_i} \right), \quad (10)$$

where  $\epsilon_0$  is the vacuum permittivity,  $N_i(\mathbf{r})$  are the atomic densities of atom  $i$  within the material at position  $\mathbf{r}$ ,  $Z_i$  the respective atomic numbers,  $I_i$  the ionization energies,  $m_e$  the electron rest mass and  $v_p(E)$  the velocity of the incident proton.

For the differential elastic scattering, we use Molière's model [37,38] which has been found in [39] to achieve good agreement with TOPAS MC and improved results compared to the small-angle first Born approximation. Here, the macroscopic differential cross section of the mixture material is given by

$$\Sigma_s(E, \mathbf{r}, \mu_0) = \sum_{i \in \mathbb{A}} \tau_{i,\text{lab}}(\mu_0) \frac{4N_i(\mathbf{r})\alpha^2}{k(E)^2(1 - \mu_0 + \chi_{\alpha,i}(E))}, \quad (11)$$

where  $\chi_{\alpha,i}(E) = \chi_{0,i}(E)^2(1.13 + 3.76a_i(E)^2)$  is the Molière corrected screening parameter according to [37,38] with  $\chi_{0,i}(E) = \frac{1.13\alpha Z_i^{1/3} m_e c}{p(E)}$  and  $a_i = \frac{Z_i \alpha}{\beta(E)}$  representing the deviation from the Born approximation [40]. Further,  $\beta = \frac{v_p(E)}{c}$  is the relative velocity,  $\alpha$  the fine structure constant and  $k(E) = \frac{p(E)}{\hbar}$  the reduced wave number with  $p(E)$  the momentum and  $\hbar$  the reduced Planck constant.



**Fig. 1.** Graphical illustration of dynamical low-rank approximation. The solution at time  $t_0$  lies on the manifold of rank  $r$  matrices  $\mathcal{M}$ . It will not stay on  $\mathcal{M}$  if it moves according to the original dynamics (i.e., into direction  $\mathbf{F}(\mathbf{u}_r(t_0))$ ), but remains of rank  $r$  if the original dynamics are projected onto the tangent space  $\mathcal{T}_{\mathbf{u}_r(t_0)}\mathcal{M}$  via the projector  $\mathbf{P}(\mathbf{u}_r(t))$ .

Note, that Eq. (11) contains a factor to convert the cross sections from the center of mass to the laboratory frame of reference, which can be computed as

$$\tau_{i;\text{lab}}(\mu_0) = \frac{\left(1 + 2\mu_0 \cdot \frac{m_p}{m_i} + \left(\frac{m_p}{m_i}\right)^2\right)^{\frac{3}{2}}}{1 + \mu_0 \cdot \frac{m_p}{m_i}},$$

where  $m_p$  is the proton rest mass and  $m_i$  the rest mass of the target nucleus [32]. Further, since protons and neutrons have almost equal mass,  $\frac{m_p}{m_i} = A_i$  can be used to simplify the computation, where  $A_i$  is the atomic mass number.

For simplicity, we neglect nuclear scattering (including production of secondary neutrons and recoil nuclei) and absorption as their effect on the deposited energy is relatively small [41,42] and they do not affect the numerical method presented here.

## 2.2. Dynamical low-rank approximation

Discretizing the transport equations presented in the previous section will result in a large matrix ordinary differential equation

$$\dot{\mathbf{u}}(t) = \mathbf{F}(t, \mathbf{u}(t)), \quad \mathbf{u}(t_0) = \mathbf{u}_0, \quad (12)$$

where  $\mathbf{u}(t) \in \mathbb{R}^{n \times m}$  and energy will act as a pseudo-time. Here,  $n$  refers to the number of spatial cells and  $m$  is the number of moments of the modal angular discretization.

A rank  $r$  approximation to  $\mathbf{u}(t)$  is given by  $\mathbf{u}(t) \approx \mathbf{X}(t)\mathbf{S}(t)\mathbf{V}(t)^\top$  where  $\mathbf{X}(t) \in \mathbb{R}^{n \times r}$  and  $\mathbf{V}(t) \in \mathbb{R}^{m \times r}$  are orthonormal basis matrices and  $\mathbf{S}(t) \in \mathbb{R}^{r \times r}$  is a not necessarily diagonal coefficient matrix. Here, the rank  $r$  matrices form a smooth manifold, which we denote by  $\mathcal{M}$ . The core idea of dynamical low-rank approximation is to evolve the low-rank factors instead of the full solution in time [17]. This is achieved by projecting the dynamics of (12) onto the tangent space of  $\mathcal{M}$  at the current solution. The projection restricts the dynamics of the solution to the low-rank manifold  $\mathcal{M}$ , thus ensuring a memory-efficient solution representation during the entire simulation. The projected equations then read

$$\dot{\mathbf{u}}_r(t) = \mathbf{P}(\mathbf{u}_r(t))\mathbf{F}(t, \mathbf{u}_r(t)), \quad (13)$$

where for  $\mathbf{Z} = \mathbf{X}\mathbf{S}\mathbf{V}^\top$ , the projector onto the tangent space of  $\mathcal{M}$  at  $\mathbf{Z}$  takes the form

$$\mathbf{P}(\mathbf{Z})\mathbf{G} = \mathbf{X}\mathbf{X}^\top\mathbf{G}(\mathbf{I} - \mathbf{V}\mathbf{V}^\top) + \mathbf{G}\mathbf{V}\mathbf{V}^\top.$$

A graphical illustration of dynamical low-rank approximation can be found in Fig. 1. Since this manifold can exhibit high curvature, conventional time integration methods require prohibitively small step sizes. Therefore, careful construction of new time integration methods that take the geometry of  $\mathcal{M}$  into account when solving (13) is required [19]. Robust time integration methods include the projector–splitting integrator [20] and basis-update & Galerkin (BUG) integrators [21–23]. The main idea of these integrators is to move on flat subspaces of the low-rank manifold, thereby avoiding directions with high curvature. The augmented BUG integrator [22], which we use in this work, first updates the basis matrices  $\mathbf{U}$  and  $\mathbf{V}$  in parallel, constraining the dynamics to

$$\{\mathbf{K}\mathbf{V}(t_0)^\top \mid \mathbf{K} \in \mathbb{R}^{n \times r}\} \subset \mathcal{M} \text{ and } \{\mathbf{U}(t_0)\mathbf{L}^\top \mid \mathbf{L} \in \mathbb{R}^{m \times r}\} \subset \mathcal{M}. \quad (14)$$

This leads to the differential equations

$$\dot{\mathbf{K}}(t) = \mathbf{F}(t, \mathbf{K}(t)\mathbf{V}_0^\top)\mathbf{V}_0, \quad \mathbf{K}(t_0) = \mathbf{U}_0\mathbf{S}_0, \quad (15a)$$

$$\dot{\mathbf{L}}(t) = \mathbf{F}(t, \mathbf{U}_0\mathbf{L}(t)^\top)^\top\mathbf{U}_0, \quad \mathbf{L}(t_0) = \mathbf{V}_0\mathbf{S}_0^\top. \quad (15b)$$

The time updated basis matrices at time  $t_1 = t_0 + \Delta t$ , which we denote by  $\hat{\mathbf{U}}$  and  $\hat{\mathbf{V}}$  can then be retrieved by a QR decomposition of  $[\mathbf{K}(t_1), \mathbf{U}_0] = \hat{\mathbf{U}}\mathbf{S}_K$  and  $[\mathbf{L}(t_1), \mathbf{V}_0] = \hat{\mathbf{V}}\mathbf{S}_L$ , where  $[\mathbf{A}, \mathbf{B}] \in \mathbb{R}^{m \times 2r}$  denotes the concatenation of two matrices  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times r}$ . Lastly, to update the coefficient matrix  $\mathbf{S}$ , the integrator constrains the dynamics to  $\{\hat{\mathbf{U}}\mathbf{S}\hat{\mathbf{V}}^\top \mid \mathbf{S} \in \mathbb{R}^{2r \times 2r}\} \subset \mathcal{M}_{2r}$  by solving the Galerkin system

$$\dot{\mathbf{S}}(t) = \hat{\mathbf{U}}^\top \mathbf{F}(t, \hat{\mathbf{U}}\mathbf{S}(t)\hat{\mathbf{V}}^\top)\hat{\mathbf{V}}, \quad \mathbf{S}(t_0) = (\hat{\mathbf{U}}^\top \mathbf{U}_0)\mathbf{S}_0(\mathbf{V}_0^\top \hat{\mathbf{V}}). \quad (15c)$$

The time-updated coefficient matrix  $\hat{\mathbf{S}}$  is retrieved from  $\mathbf{S}(t_1)$ . Since the time-updated solution  $\hat{\mathbf{Y}} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top$  is of rank  $2r$ , a new rank can be determined by a truncated singular value decomposition of  $\hat{\mathbf{S}}$ . Given  $\hat{\mathbf{S}} = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^\top$ , where  $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_{2r})$  collects the singular values of  $\hat{\mathbf{S}}$ , the singular value decomposition is truncated at a new rank  $r_1$  such that  $\sum_{i=r_1+1}^{2r} \sigma_i \leq \vartheta$  with a user-determined truncation threshold  $\vartheta$ , see [22]. This yields a new rank  $r_1$  factorization  $\mathbf{Y}_1 = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^\top$ . The process of evolving the solution along (15) with subsequent truncation is repeated until the factorized solution at a given final time  $t_k$  is reached after  $k$  steps. The Eqs. (15) inherit the stiffness of the original problem, thus, no further stiffness is added from the underlying geometry of  $\mathcal{M}$ . Therefore, classical time integration methods can be used to evolve the factorized solution. For a detailed discussion of BUG integrators, their inherent parallelism, and extensions to higher order, see [21–25].

### 3. Methods

In this section we detail the numerical method used to solve the problem described in Section 2.1. We loosely follow the strategy in [16] of splitting the problem in collided and uncollided as well as scattering and streaming parts and using a  $\mathbf{P}_N$ -based dynamical low-rank approach for the collided equations. In contrast to [16], we however use a ray tracer including energy straggling for the uncollided part and the derivation for the collided equations differs due to the material dependency in cross sections and stopping power as well as the higher order spatial discretization. Together with the efficient incorporation of multiple materials as detailed in Section 2.1.1, this enables the derivation of an efficient and accurate numerical solver to simulate proton radiation in a realistic setting.

#### 3.1. Collided-uncollided split

We start by splitting Eq. (7) into a collided and uncollided part using a first collision source method. This will facilitate the inclusion of straggling and avoid challenging boundary conditions in the DLRA part of the method. Note that the transformation used in Eq. (7) is only applied to the equation representing collided particles here since the uncollided flux equation will be treated with a raytracer which does not require pseudo-time dependence.

We now separate the angular flux into collided particles  $\psi_c$  and uncollided particles  $\psi_u$  such that  $\psi = \psi_u + \psi_c$ . For collided particles, we assume that straggling can be neglected, that is,

$$\frac{\partial^2 T \psi_c}{\partial E^2} \approx 0. \quad (16)$$

We choose this simplification because the dominant contribution to range fluctuations is expected to arise from uncollided particles, while the collided flux primarily influences lateral spreading, making the contribution of straggling effects to the collided equation relatively small [15,33,43,44]. Then, we obtain

$$\mathbf{\Omega} \cdot \nabla \psi_u = \frac{\partial S \psi_u}{\partial E} + \frac{1}{2} \frac{\partial^2 T \psi_u}{\partial E^2} + \sum_{i \in \mathbb{A}} w_i \Gamma_{\text{out};i} \psi_u(E, \mathbf{r}, \mathbf{\Omega}), \quad (17a)$$

$$-\partial_E \tilde{\psi}_c(E, \mathbf{r}) = -\mathbf{\Omega} \cdot \nabla \frac{\tilde{\psi}_c(E, \mathbf{r})}{S(E, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i \Gamma_i \frac{\tilde{\psi}_c(E, \mathbf{r})}{S(E, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i \Gamma_{\text{in};i} \psi_u(E, \mathbf{r}, \mathbf{\Omega}), \quad (17b)$$

where  $\Gamma_{\text{out};i}$  and  $\Gamma_{\text{in};i}$  are the in- and outscattering parts of the collision operator, respectively. So,

$$\Gamma_{\text{in};i} \psi = \int_{\mathbb{S}^2} \Sigma_{s;i}(E, \mathbf{r}, \mathbf{\Omega}' \cdot \mathbf{\Omega}) \psi(E, \mathbf{r}, \mathbf{\Omega}') d\mathbf{\Omega}', \quad (18)$$

$$\Gamma_{\text{out};i} \psi = \Sigma_{t;i}(E, \mathbf{r}) \psi(E, \mathbf{r}, \mathbf{\Omega}) \quad (19)$$

for the Boltzmann operator and

$$\Gamma_{\text{in};i} \psi = \Gamma_i \psi(E, \mathbf{r}, \mathbf{\Omega}), \quad \Gamma_{\text{out};i} \psi = 0 \quad (20)$$

for the Fokker-Planck operator.

#### 3.2. Uncollided part: ray tracer

For the uncollided particles, we use an in-house analytical ray tracer implemented in fortran [see also 15,32].

Having split off the singular part as described above, the PDE along the ray for  $\psi_u$  reads

$$\frac{\partial \psi_u(E, z)}{\partial z} + \sum_{i \in \mathbb{A}} w_i(z) \Gamma_{\text{out};i} \psi_u(E, z) = \frac{\partial S^*(E, z) \psi_u(E, z)}{\partial E} + \frac{1}{2} \frac{\partial}{\partial E} T(E, z) \frac{\partial \psi_u(E, z)}{\partial E}, \quad (21)$$

where  $z$  denotes the one-dimensional spatial position, i.e. depth along the ray, and we define  $S^* = S + \frac{1}{2} \frac{dT}{dE}$  to obtain the standard diffusion form of straggling. We use a discontinuous Galerkin scheme to discretize the PDE in energy. The energy domain is divided into 128 equally-sized groups. The slowing down term is based on the Lax-Friedrichs scheme, whereas the straggling term utilizes the SIPG method. Polynomials up to second order are used as basis vectors (3 dof per energy element). After discretization in  $E$ , the system can be written as

$$\mathbf{M}_E \frac{\partial \psi_u(z)}{\partial z} + \mathbf{G} \psi_u(z) = 0, \quad (22)$$

where  $\mathbf{M}_E$  is the mass matrix for the energy basis and  $\mathbf{G}$  contains elements corresponding to absorption, slowing down and straggling. Here,  $\psi_u(z=0) = f(E)$  is the projection of a given energy distribution (Gaussian). We use a Crank-Nicolson scheme to solve this ODE system along the depth. The step size is determined such that it is smaller than the maximum step length (0.01 cm). The unscattered flux distribution is computed at the cell midpoints of the mesh. Based on the spatial and angular distribution of the beam source, a collection of start and end points for rays is generated. The solution is then summed over all ray positions within a beam source and the resulting space and energy dependent flux acts as a source term for the collided equation through inscattering.

### 3.3. Collided part: dynamical low-rank approximation

We now want to use DLRA to efficiently solve the costly collided part of the equation. Since DLRA is a method designed for time-dependent problems, we now transform from energy  $E$  to pseudo-time  $t(E) = E_{\max} - E$ , resulting in the following equation:

$$\partial_t \tilde{\Psi}_c(t, \mathbf{r}) = -\mathbf{\Omega} \cdot \nabla \frac{\tilde{\Psi}_c(t, \mathbf{r})}{S(t, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i \Gamma_i \frac{\tilde{\Psi}_c(t, \mathbf{r})}{S(t, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i \Gamma_{\text{in},i} \psi_u(t, \mathbf{r}, \mathbf{\Omega}). \quad (23)$$

Next, we will discretize this equation in angle and space to obtain a (pseudo)time dependent ordinary matrix differential equation.

#### 3.3.1. Angular discretization

We use a modal discretization of the angular variable. To this end, we define the real-valued spherical harmonics basis

$$m_\ell^k = \begin{cases} \frac{(-1)^k}{\sqrt{2}} \left( Y_\ell^k + (-1)^k Y_\ell^{-k} \right), & k > 0, \\ Y_\ell^0, & k = 0, \\ -\frac{(-1)^k}{\sqrt{2}} \left( Y_\ell^{-k} - (-1)^k Y_\ell^k \right), & k < 0, \end{cases}$$

which uses the spherical harmonics basis functions

$$Y_\ell^k(\mathbf{\Omega}) = \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-k)!}{(\ell+k)!}} e^{ik\varphi} P_\ell^k(\mu),$$

where  $P_\ell^k$  are the associated Legendre polynomials. Collecting all basis functions up to degree  $N$  in a vector

$$\mathbf{m} = (m_0^0, m_1^{-1}, m_1^0, m_1^1, \dots, m_N^N)^\top \in \mathbb{R}^m, \quad m = (N+1)^2$$

gives the modal approximation  $\psi_c(t, \mathbf{r}, \mathbf{\Omega}) \approx \mathbf{u}(t, \mathbf{r})^\top \mathbf{m}(\mathbf{\Omega})$ . Since we deal with monodirectional beam sources in this work, for a specific source with beam direction  $\mathbf{\Omega}_{\text{in}}$ , the uncollided flux  $\psi_u(t, \mathbf{r}, \mathbf{\Omega}) = \psi_u(t, \mathbf{r}) \delta(\mathbf{\Omega} - \mathbf{\Omega}_{\text{in}})$  at a specific point in space and time/energy is a scalar and the transformation from nodal to modal representation in angle can be done using the vector  $\mathbf{T}_M = \mathbf{m}(\mathbf{\Omega}_{\text{in}})$ . For a generalization to beams with a distribution in angle see [16]. Then, the  $P_N$  equations read

$$\begin{aligned} \partial_t \mathbf{u}(t, \mathbf{r}) = & -\mathbf{A} \cdot \nabla \frac{\mathbf{u}(t, \mathbf{r})}{S(t, \mathbf{r})} - \sum_{i \in \mathbb{A}} w_i(\mathbf{r}) \Sigma_{i,i}(t) \frac{\mathbf{u}(t, \mathbf{r})}{S(t, \mathbf{r})} \\ & + \sum_{i \in \mathbb{A}} w_i(\mathbf{r}) \mathbf{G}_i(t) \frac{\mathbf{u}(t, \mathbf{r})}{S(t, \mathbf{r})} + \sum_{i \in \mathbb{A}} w_i(\mathbf{r}) \mathbf{G}_i(t) \mathbf{T}_M \psi_u(t, \mathbf{r}), \end{aligned} \quad (24)$$

where  $\mathbf{A} \cdot \nabla := \mathbf{A}_1 \partial_x + \mathbf{A}_2 \partial_y + \mathbf{A}_3 \partial_z$  with  $\mathbf{A}_i := \int_{\mathbb{S}^2} \mathbf{m} \mathbf{m}^\top \mathbf{\Omega}_i d\mathbf{\Omega}$ .

The diagonal scattering matrix of material  $i$  is defined as  $\mathbf{G}_i(t) = (G_{i;pq})_{1 \leq p,q \leq m}$  where

$$G_{i;pp}(t, \mathbf{r}) = 2\pi \int_{[-1,1]} P_k(\mu_0) \Sigma_{s,i}(t, \mu_0) d\mu_0,$$

for Boltzmann and

$$G_{i;pp}(t) = -\frac{\xi_{1,i}}{2} k \cdot (k+1),$$

for Fokker-Planck, with  $p = l^2 + l + k$ ,  $0 \leq k \leq N$ ,  $|l| \leq k$ . The total cross section is then  $\Sigma_{t,i}(t) = G_{i;11}(t)$ . Note, that we further apply the extended transport correction to the scattering moments in the Boltzmann equation [45] and the correction described in [46,47] for the Fokker-Planck moments. Both methods aim at a "delta-function corrected expansion" [46], where the extended transport correction is chosen such that the truncated expansion matches moments 0 to  $N+1$  exactly [45].



### 3.3.2. Spatial discretization

In this work, we consider a full three-dimensional spatial domain, i.e., the spatial variable is  $\mathbf{r} = (x, y, z)^T \in \mathbb{R}^3$ . In order to sufficiently resolve the sharp gradients of the energy deposition near the Bragg peak, a higher-order method is necessary [13]. Here, we use a second order upwind scheme. For this, the spatial domain is split into  $n = n_x \cdot n_y \cdot n_z$  cells in the form of a structured rectangular grid with uniform one-dimensional grids  $x_1 \leq x_2 \leq \dots \leq x_{n_x+1}$ ,  $y_1 \leq y_2 \leq \dots \leq y_{n_y+1}$  and  $z_1 \leq z_2 \leq \dots \leq z_{n_z+1}$  with grid sizes  $\Delta_x, \Delta_y, \Delta_z$  respectively.

The cell of index  $\text{idx}(i, j, k) := (k-1) \cdot n_x \cdot n_y + (j-1) \cdot n_x + i$  is defined on  $I_{i,j,k} := [x_i, x_{i+1}] \times [y_j, y_{j+1}] \times [z_k, z_{k+1}]$  with  $i = 1, \dots, n_x$ ,  $j = 1, \dots, n_y$ ,  $k = 1, \dots, n_z$ . Assuming we want to solve the semi-discrete Eq. (24) from the previous section, we now collect the numerical solution in a matrix  $\mathbf{u}(t) \in \mathbb{R}^{n \times m}$  with entries  $u_{\text{idx}(i,j,k),p}(t)$ ,  $p = 1, \dots, m$ . We define the second order upwind difference stencils  $\mathbf{D}_{x,y,z}^{+,-} \in \mathbb{R}^{n \times n}$  as

$$\begin{aligned} \mathbf{D}_{x;\text{idx}(i,j,k),\text{idx}(i,j,k)}^{+,-} &= \frac{\pm 3}{2\Delta_x}, \quad \mathbf{D}_{x;\text{idx}(i,j,k),\text{idx}(i\pm 1,j,k)}^{+,-} = \frac{\mp 4}{2\Delta_x}, \quad \mathbf{D}_{x;\text{idx}(i,j,k),\text{idx}(i\pm 2,j,k)}^{+,-} = \frac{\pm 1}{2\Delta_x} \\ \mathbf{D}_{y;\text{idx}(i,j,k),\text{idx}(i,j,k)}^{+,-} &= \frac{\pm 3}{2\Delta_y}, \quad \mathbf{D}_{y;\text{idx}(i,j,k),\text{idx}(i,j\pm 1,k)}^{+,-} = \frac{\mp 4}{2\Delta_y}, \quad \mathbf{D}_{y;\text{idx}(i,j,k),\text{idx}(i,j\pm 2,k)}^{+,-} = \frac{\pm 1}{2\Delta_y} \\ \mathbf{D}_{z;\text{idx}(i,j,k),\text{idx}(i,j,k)}^{+,-} &= \frac{\pm 3}{2\Delta_z}, \quad \mathbf{D}_{z;\text{idx}(i,j,k),\text{idx}(i,j,k\pm 1)}^{+,-} = \frac{\mp 4}{2\Delta_z}, \quad \mathbf{D}_{z;\text{idx}(i,j,k),\text{idx}(i,j,k\pm 2)}^{+,-} = \frac{\pm 1}{2\Delta_z} \end{aligned}$$

These stencil matrices applied to the discretized solution directly correspond to the tensorized one-dimensional second order upwind stencil, i.e., for a function  $u(x)$  we have that

$$u'(x) = \frac{1}{2\Delta_x} (3u(x) - 4u(x - \Delta_x) + u(x - 2\Delta_x)) + O(\Delta x^2) \quad (\text{backward})$$

$$u'(x) = \frac{1}{2\Delta_x} (-3u(x) + 4u(x + \Delta_x) - u(x + 2\Delta_x)) + O(\Delta x^2) \quad (\text{forward}).$$

Further, we collect the eigenvectors of  $\mathbf{A}_{x,y,z} = \mathbf{V}_{x,y,z} \mathbf{\Lambda}_{x,y,z} \mathbf{V}_{x,y,z}^T$  in  $\mathbf{V}_{x,y,z}$  and split the eigenvalues collected in  $\mathbf{\Lambda}_{x,y,z} = \text{diag}(\lambda_1^{x,y,z}, \dots, \lambda_m^{x,y,z})$  into  $(\mathbf{\Lambda}_{x,y,z;p,p}^+)_{p=1,\dots,m} = \max(0, \lambda_p^{x,y,z})$  and  $(\mathbf{\Lambda}_{x,y,z;p,p}^-)_{p=1,\dots,m} = \min(0, \lambda_p^{x,y,z})$ . Then we obtain a large matrix differential equation of the form

$$\dot{\mathbf{u}}(t) = \mathbf{F}_S(\mathbf{u}(t)) + \mathbf{G}(t, \mathbf{u}(t)), \quad (25)$$

where

$$\begin{aligned} \mathbf{F}_S(\mathbf{u}(t)) &:= \left( \mathbf{D}_x^+ (\mathbf{V}_x \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_x^+ + \mathbf{D}_x^- (\mathbf{V}_x \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_x^- \right) \mathbf{V}_x^T \\ &\quad + \left( \mathbf{D}_y^+ (\mathbf{V}_y \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_y^+ + \mathbf{D}_y^- (\mathbf{V}_y \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_y^- \right) \mathbf{V}_y^T \\ &\quad + \left( \mathbf{D}_z^+ (\mathbf{V}_z \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_z^+ + \mathbf{D}_z^- (\mathbf{V}_z \mathbf{S}^{-1}(t) \mathbf{u}(t)) \mathbf{\Lambda}_z^- \right) \mathbf{V}_z^T \\ \mathbf{G}(t, \mathbf{u}(t)) &:= \sum_{i \in \mathbb{A}} \mathbf{w}_i \mathbf{S}^{-1}(t) \mathbf{u}(t) \mathbf{\Sigma}_{i,i} + \sum_{i \in \mathbb{A}} \mathbf{w}_i \mathbf{S}^{-1}(t) \mathbf{u}(t) \mathbf{G}_i \\ &\quad + \sum_{i \in \mathbb{A}} \mathbf{w}_i \psi_u(t) \mathbf{T}_M \mathbf{G}_i, \end{aligned}$$

here  $\mathbf{w}_i$  and  $\mathbf{S}(t) \in \mathbb{R}_+^{n \times n}$  are diagonal matrices containing the material composition weights and stopping power at each spatial grid point, respectively. Here,  $\psi_u$  is the uncollided solution from the ray tracer evaluated at the (same) spatial grid points. Since solving this at sufficiently high resolution is extremely costly we use the dynamical low-rank approximation to evolve the solution in (pseudo)-time.

### 3.3.3. Energy/time update using the dynamical low-rank approximation

For this, we further split Eq. (25) into its streaming and scattering parts using a first-order Lie splitting, i.e.,

$$\dot{\mathbf{u}}_I(t) = \mathbf{F}_S(\mathbf{u}_I(t)) \quad \mathbf{u}_I(t_0) = \mathbf{u}(t_0), \quad (26a)$$

$$\dot{\mathbf{u}}_{II}(t) = \mathbf{G}(t, \mathbf{u}_{II}(t)) \quad \mathbf{u}_{II}(t_0) = \mathbf{u}_I(t_1). \quad (26b)$$

For the streaming update (26a) we use the augmented BUG integrator from Section 2.2 to update  $\mathbf{u}_I$  in factorized form from  $\mathbf{U}^0, \mathbf{S}^0, \mathbf{V}^0$  to  $\mathbf{U}^{\frac{1}{2}}, \mathbf{S}^{\frac{1}{2}}, \mathbf{V}^{\frac{1}{2}}$ , by simply plugging in  $\mathbf{F}_S$  for the right hand side and using a fourth-order Runge-Kutta method for the time updates. Note, that we still do a full time step from  $t_0$  to  $t_1 = t_0 + \Delta t$  and the  $\frac{1}{2}$  in the exponent merely denotes that we have only completed the streaming part of the update, i.e.,  $\mathbf{U}^{\frac{1}{2}} \mathbf{S}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2},T} \approx \mathbf{u}_I(t_1)$ . The fourth-order Runge-Kutta method is primarily used to stabilize the stiff streaming operators in  $\mathbf{F}_S$ .

For the scattering update, we use the matrix projector splitting integrator for the collided flux and the augmented BUG integrator on the remainder according to [48]. This reduces the update to a K-, S- and L-step for inscattering from the uncollided flux and just the L-step for self-scattering of the collided flux. As discussed in [48], this reduction stems from the observation that the scattering operator will not change the solution's column space, leading to  $K$  and  $S$  equations that are identical except for their signs, therefore canceling each other out. Thus, to perform the scattering update (26b), we need to solve the following differential equations

1. *L*-step, collided:

$$\dot{\mathbf{L}}(t) = - \sum_{i \in \mathbb{A}} \mathbf{U}^{\frac{1}{2}, \top} \mathbf{w}_i \mathbf{S}^{-1}(t) \mathbf{U}^{\frac{1}{2}} \Sigma_{i,i} \mathbf{L}(t) + \sum_{i \in \mathbb{A}} \mathbf{U}^{\frac{1}{2}, \top} \mathbf{w}_i \mathbf{S}^{-1}(t) \mathbf{U}^{\frac{1}{2}} \mathbf{L}(t) \mathbf{G}_i.$$

$$\mathbf{L}(t_0) = \mathbf{S}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}, \top}.$$

Determine the updated basis using a QR decomposition of  $\mathbf{L}(t_1) = \tilde{\mathbf{S}}^{\frac{1}{2}} \tilde{\mathbf{V}}^{\frac{1}{2}}$ .

2. *K*-step, uncollided:

$$\dot{\mathbf{K}}(t) = \sum_{i \in \mathbb{A}} \mathbf{w}_i \mathbf{S}^{-1} \psi_u \mathbf{T}_M \mathbf{G}_i \mathbf{V}^{\frac{1}{2}},$$

$$\mathbf{K}(t_0) = \mathbf{U}^{\frac{1}{2}} \mathbf{S}^{\frac{1}{2}}$$

Determine the updated and augmented basis at  $t_1 = t_0 + \Delta t$  using a QR decomposition of  $[\mathbf{K}(t_1), \mathbf{U}^{\frac{1}{2}}] = \hat{\mathbf{U}} \mathbf{S}_K$ .

3. *L*-step, uncollided:

$$\dot{\mathbf{L}}(t) = \sum_{i \in \mathbb{A}} \mathbf{U}^{\frac{1}{2}, \top} \mathbf{w}_i \mathbf{S}^{-1} \psi_u \mathbf{T}_M \mathbf{G}_i,$$

$$\mathbf{L}(t_0) = \tilde{\mathbf{S}}^{\frac{1}{2}} \tilde{\mathbf{V}}^{\frac{1}{2}}.$$

Determine the updated and augmented basis at  $t_1 = t_0 + \Delta t$  using a QR decomposition of  $[\mathbf{L}(t_1), \tilde{\mathbf{V}}^{\frac{1}{2}}] = \hat{\mathbf{V}} \mathbf{S}_L$ .

4. *S*-step, uncollided:

$$\dot{\mathbf{S}}(t) = \sum_{i \in \mathbb{A}} \hat{\mathbf{U}}^\top \mathbf{w}_i \mathbf{S}^{-1} \psi_u \mathbf{T}_M \mathbf{G}_i \hat{\mathbf{V}},$$

$$\mathbf{S}(t_0) = (\hat{\mathbf{U}}^\top \mathbf{U}^{\frac{1}{2}}) \mathbf{S}^{\frac{1}{2}} (\tilde{\mathbf{V}}^{\frac{1}{2}, \top} \hat{\mathbf{V}})$$

and set  $\mathbf{S}^1 = \mathbf{S}(t_1)$ .

Similar to [16], we use an implicit Euler method for the first step to treat stiff scattering terms and an explicit Euler method for the time updates in the remaining steps.

**Remark 1.** Note that the complete numerical method as stated can be made locally conservative through minor adjustments [49], but not necessarily positivity preserving due to the use of the spherical harmonics method in angle. Conservation of particles and energy is however guaranteed for the uncollided flux equation through use of the ray-tracer.

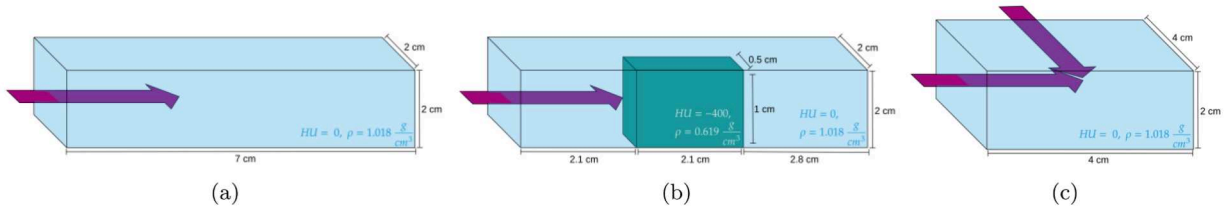
## 4. Numerical results

In the following, we present a number of numerical experiments in homogeneous and heterogeneous media to validate and compare our method against an otherwise equivalent full-rank method as well as a state-of-the-art Monte Carlo code (TOPAS MC) [35] with only electromagnetic interactions. The chosen experiments represent simplified scenarios from radiation therapy applications, ranging from irradiation of relatively homogeneous water-like tissue to strong heterogeneities (e.g. lung) and layered beams. Similar test cases are frequently used for the validation of new dose calculation algorithms [39,50–52]. The comparisons will be done based on visual as well as quantitative assessment of the results. Besides the deposited energy, we plot the relative differences normalized to percentages of the maximum reference dose and discuss relevant error metrics used in application. Specifically, we provide two metrics: 1. the *Gamma pass rate*, which quantifies percentage-wise agreement with a reference solution based on a tolerance in spatial and value distance that is evaluated for all cells with a dose above a specified threshold [53] 2. the range error of the integrated depth dose, where range is defined as the depth where the dose falls to 80% of its maximum behind the Bragg peak and the integrated depth dose is the deposited dose integrated over the two lateral dimensions in space. For the Gamma analysis we use a 2%/2 mm criterion and a threshold of 10% of the maximum dose.

We will first consider a single uni-directional beam in different media and later add a second beam with a different incoming angle to investigate the effect this has on the chosen ranks. All beams have a Gaussian spatial distribution in the two-dimensional plane perpendicular to the beam direction with 3 mm standard deviation and the initial beam energy follows a Gaussian distribution with standard deviation 1% of the mean beam energy.

Since the problem at full angular and spatial resolution is infeasible to solve at full rank, we will validate the low-rank method against the deterministic full-rank method on a coarser grid and then compare a higher resolution low-rank solution to the Monte Carlo reference. For the spatial resolution we always use a coarser gridsize of 1 mm and a finer one with 0.25 mm. Since initial numerical experiments indicate that more angular modes are required when using the Boltzmann collision operator, we choose  $N = 75$  for the coarser and  $N = 115$  for the finer resolution here. In case of the Fokker-Planck approximation, which uses a simplified collision operator,  $N = 19$  for the coarser and  $N = 75$  for the finer resolution have proven sufficient. The truncation tolerance used to adapt the rank is chosen as  $\vartheta = 0.01$  in all cases.





**Fig. 2.** Schematic representations of the domain size, composition and beam directions in the three numerical test cases: (a) homogeneous, water-like  $2 \times 2 \times 7$  cm domain with single beam in z-direction hitting middle of x-y-plane (b) heterogeneous  $2 \times 2 \times 7$  cm domain representing water with a lung insert with a single beam identical to (a) and (c) homogeneous, water-like  $2 \times 4 \times 4$  cm domain with two perpendicular beams pointing in z- and negative y-direction.

#### 4.1. Homogeneous case

First, we consider a 90 MeV proton beam in a homogeneous  $2 \times 2 \times 7$  cm box (see Fig. 2(a)). The composition of the water-like material is characterized as 0 Hounsfield Units (HU), where HU is the common unit used in CT scans to measure the radiodensity of different tissues [54]. These values are converted to material composition weights  $w_i$  and a density of  $1.018 \frac{g}{cm^3}$  using the conversion described in [55].

Fig. 3 shows an overview of two-dimensional slices and profiles of the deposited energy using the Boltzmann collision operator at different resolutions compared to the Monte Carlo reference and a full rank deterministic solver. It is apparent that DLRA can approximate the full-rank solution very well and at a much lower rank according to Fig. 4 (a). This is supported by a Gamma pass rate of 99.63%, a range error of 0.07 mm and low relative differences of maximum around -2.5% (see 3 (c)). It is also clear that the coarse grid resolution is not sufficient to fully capture the characteristics of the solution. When increasing both spatial and angular resolution we can however see that our model is capable of more closely approximating the Monte Carlo reference as well (see Fig. 3). Here we achieve an even higher pass rate of 99.99%, but somewhat higher relative differences of up to  $\pm 5\%$ <sup>1</sup>. The most significant deviation occurs directly behind the Bragg peak which can be seen especially in the depth profile and difference plots. Here, the solution takes on slightly negative values, which is most likely due to the nature of our angular discretization combined with the steep gradients in this area. Since the full and low-rank solutions match almost exactly, differences to the Monte Carlo reference should however be attributed to the choices of physics models and numerical discretizations.

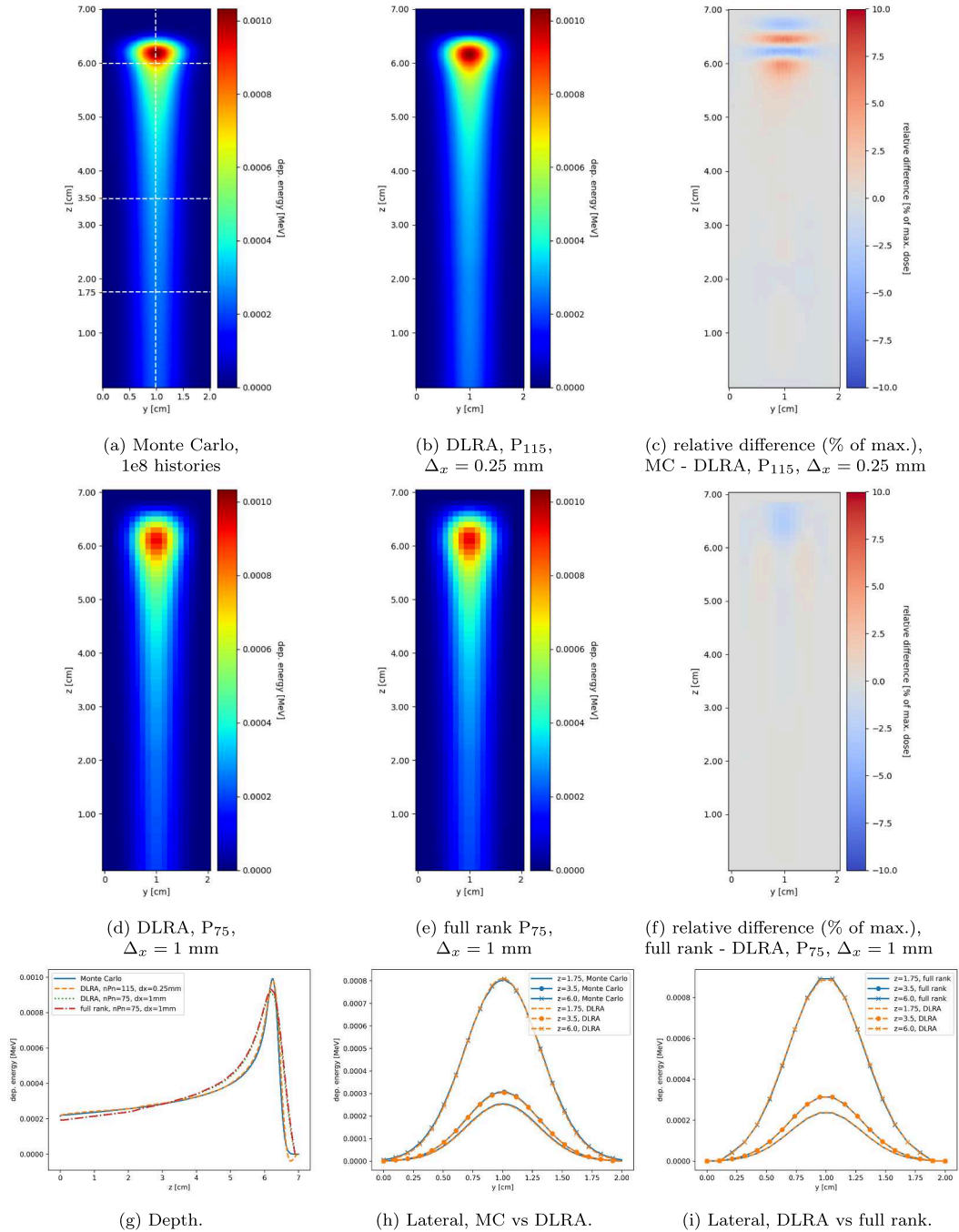
The Fokker-Planck approximation on the other hand shows smaller differences between the coarse and highly resolved solutions, but tends to underestimate the energy deposition in the region of the Bragg peak and overestimate it slightly in the build-up before the peak as can be seen in Fig. 5 (c) and (f). This might be attributed to the correction that we use to stabilize the solution [46], which has however proven to be necessary to avoid stronger artifacts near material boundaries during our studies. In Fig. 5 (g), the solution also exhibits similar negative areas as previously observed for the Boltzmann solver. Overall the agreement with the respective references is however still high, with Gamma pass rates of 99.65% compared to the full-rank solution and 99.58% compared to Monte Carlo. With -0.19 mm, the range error is higher than for the Boltzmann solver, but still below the resolution of the spatial grid. Fig. 6 further shows that the ranks for the Fokker-Planck approximation are consistently higher than for the Boltzmann equation. Interestingly, in contrast to the Boltzmann solver, the discretizations now affect the rank significantly and finer discretizations are not directly correlated with higher ranks. Here, the ranks are consistently higher for high energies when using a finer angular discretization, but are larger in the case of coarse angular discretizations for small energies. Due to the higher ranks and worse results compared to the Boltzmann solver for high resolutions, using the Fokker-Planck approximation (within the current framework) would only be beneficial for a cheap estimate at a low resolution for which the Boltzmann solver has not converged sufficiently.

#### 4.2. Heterogeneous case

Next, we consider a heterogeneous case mimicking a beam encountering strong heterogeneities along its path as might be the case for example in the head and neck region. For this, a lower density box with -400 HU, roughly corresponding to lung tissue, is placed inside the domain (see Fig. 2(b)). A proton beam with 80 MeV is directed at the box such that the center of the beam hits the interface between the materials. Again, the Hounsfield units are converted to material compositions and densities using the method in [55].

As in the homogeneous case, the full-rank solution is well-approximated by the dynamical low-rank approximation for both the Boltzmann and the Fokker-Planck solvers with Gamma pass rates of 99.82% and 99.71%, respectively. This can also be seen both in the 2D surface plots in Fig. 7 (d) and (e) (Boltzmann) and Fig. 8 (d) and (e) (Fokker-Planck) and the one-dimensional solution profiles in Figs. 7 and 8 (g) and (i). The Boltzmann solver also approximates the Monte Carlo reference well with a Gamma pass rate of 99.94%, although more deviations can be seen than in the homogeneous test case and there is a slightly higher range error of 0.29 mm. For example, the difference plot in Fig. 7(c) shows that the highest dose values in the Bragg peaks are underestimated and

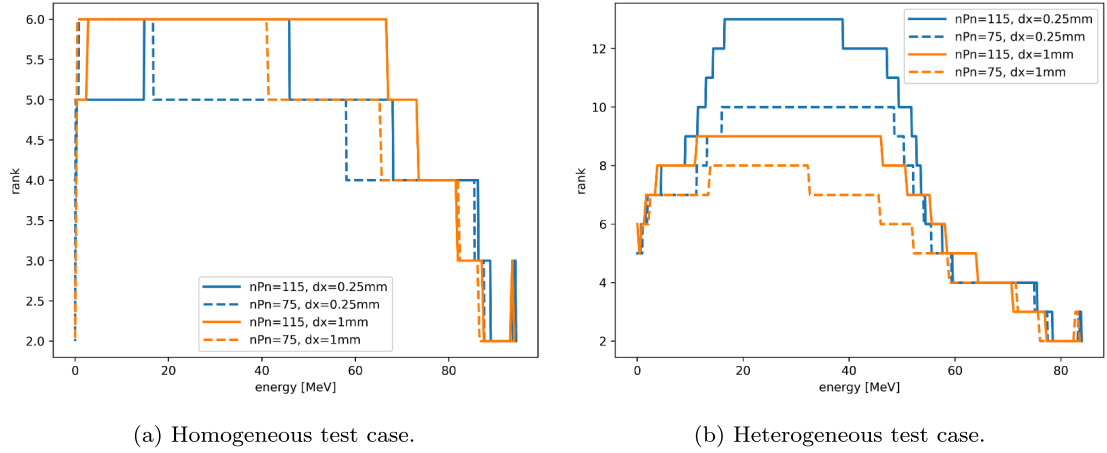
<sup>1</sup> Note that this shows that Gamma pass rates between the low and high resolution results can not be compared directly as the Gamma pass rate is defined based on continuous tolerance criteria but evaluated on a discretized grid.



**Fig. 3.** Surface plots (a)-(b) and (d)-(e), relative differences (c) and (f) and profiles (g)-(i) of deposited energy in the homogeneous test case with a single beam for the Monte Carlo reference compared to DLRA at high and low spatial and angular resolution and the full-rank deterministic method based on the Boltzmann equation. Lines in (a) indicate positions of profiles plotted in (g)-(i).

depth profiles in Fig. 7 (g) reveal differences in the local minimum behind the first peak. The ranks required for this more complex test case are expectedly higher than in the homogeneous case (see Fig. 4 (b)) and increase for increasingly fine spatial and angular resolutions. All chosen ranks are however still significantly smaller than the number of spatial cells or angular modes and lead to a large reduction of the system size (see also Section 5).

When comparing the Fokker-Planck approximation to the Monte Carlo reference in Figs. 8 (a) and (b) we observe similar results as in the homogeneous case as well. Again, we reach a high Gamma pass rate of 99.17%, but the solution tends to underestimate the highest energy depositions, there are small negative areas behind the Bragg peak and a higher range error than for Boltzmann with



**Fig. 4.** Ranks chosen by rank-adaptive integrator for different spatial and angular resolutions and the Boltzmann equation in (a) the homogeneous and (b) heterogeneous test case with a single beam.

0.39 mm. In this case, the sharp material boundary also tended to cause artifacts, which can still be slightly seen in Fig. 8 (b). While a stronger correction of the angular moments according to [46,47] reduces the artifacts, it also has a smoothing effect and dampens the height of the Bragg peak, creating a tradeoff between stability and preserving solution characteristics. Fig. 6 (b) further shows that ranks for different discretizations have a similar structure as in the homogeneous case, but are consistently higher, as already observed for the Boltzmann solver.

#### 4.3. Homogeneous with two beams

Finally, we consider a case with two 50 MeV beams entering the domain at different angles (see Fig. 2(c)). A treatment plan in proton therapy typically consists of a larger number of layered beamlets with different initial energies, positions and angles. Thus, this is a good test case to gauge the impact of a more complex beam source on rank and therefore computational costs and memory.

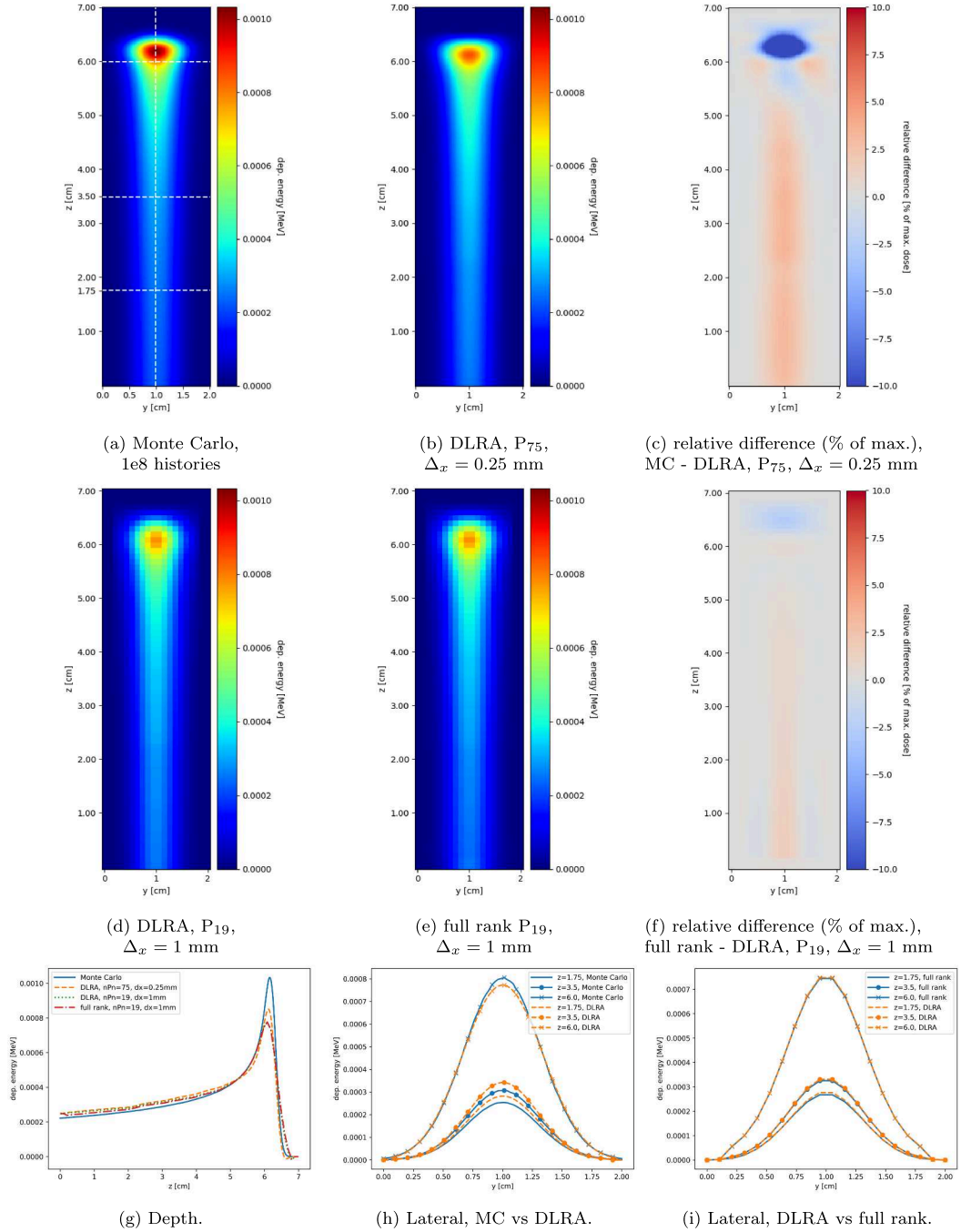
Figs. 9 and 10 show a comparison of Monte Carlo to DLRA as well as the ranks chosen using the same tolerance as in the previous test cases. The low-rank method manages to capture the basic dose characteristics well and achieves high Gamma pass rates of 99.86% (Boltzmann) and 99.90% (Fokker-Planck), but appears to be slightly smoothing in the sense of losing local "hot-spots" in the solution and corresponding sharp dose gradients. For the Boltzmann solver, the chosen ranks are on average roughly 2 points higher than in the homogeneous single beam test case for the finer angular resolution and even less when comparing the coarser resolutions (compare Figs. 4 (a) and 10 (a)). Interestingly, comparing Figs. 6 (b) and 10 (b), we see that the ranks are even lower in places than those of the single beam test case for the Fokker-Planck approximation. The higher complexity in angle is however reflected in the much quicker growth of ranks compared to the single beam case. Further, we have observed, from numerical experiments beyond the scope of this paper, that for a tolerance of  $\vartheta = 0.01$  the initial beam energy does not affect the rank for the Boltzmann solver, whereas for the Fokker-Planck solver ranks decrease for lower beam energies. Thus the lower ranks are likely at least partially due to the lower beam energy compared to Section 4.1.

For both approaches the ranks are mostly affected by the angular discretization, which is somewhat intuitive considering we effectively doubled the amount of relevant directions. Fig. 11 also shows that for both Boltzmann and Fokker-Planck, the first two modes of the angular basis indeed represent the directions of the two beams, thus capturing the forward movement of particles with some angular spread due to scattering.

Since in theory the rank quadratically impacts the computational costs (see Section 5), further studies are necessary to see when it is more efficient to treat beam sources separately or jointly. Our results however suggest that at the very least a relatively narrow spread in angle, which is typically used for beamlets originating from the same beam direction, can be incorporated without a significant increase in computational costs or memory. This would already reduce the amount of individual computations to only the 2–5 beam angles, compared to potentially thousands of pencil beams [56].

## 5. Memory, computational complexity and implementation

The dynamical low-rank approximation reduces the computational complexity from  $\mathcal{O}(n \cdot m)$  to  $\mathcal{O}(r^2 \cdot (n + m))$  and  $\mathcal{O}(r \cdot (n + m))$  with respect to runtime and memory, respectively [16]. Thus, for  $r \ll m, n$  we can expect significant cost reductions. In order to conduct our numerical experiments, the described method is implemented in Julia with GPU acceleration using CUDA.jl. The code to reproduce our numerical experiments is openly available on [GitHub](#). Since the full-rank method with a meaningful discretization

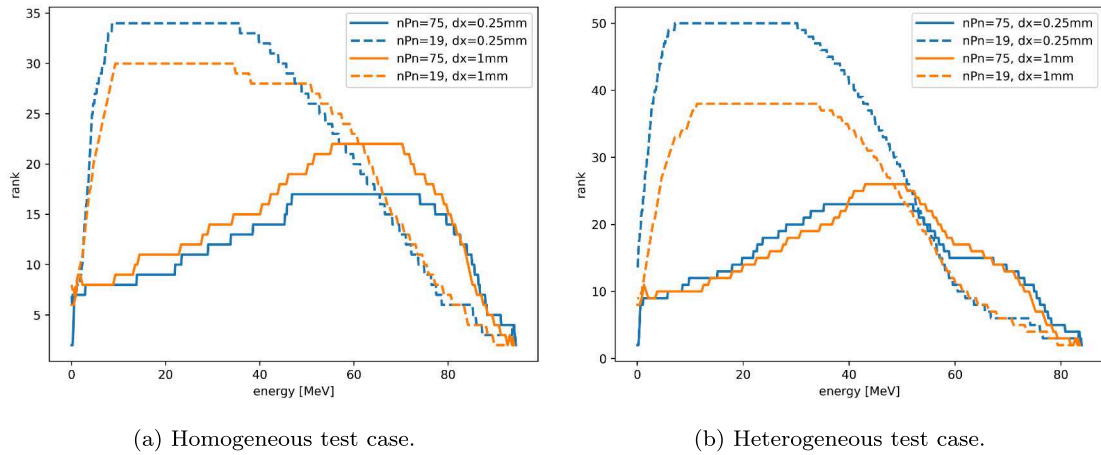


**Fig. 5.** Surface plots (a)-(b) and (d)-(e), relative differences (c) and (f) and profiles (g)-(i) of deposited energy in the homogeneous test case with a single beam for the Monte Carlo reference compared to DLRA at high and low spatial and angular resolution and the full-rank deterministic method based on the Fokker-Planck approximation. Lines in (a) indicate positions of profiles plotted in (g)-(i).

does not fit on the GPU, we further implement a full-rank CPU version using the same numerical methods and physical models. All experiments are run on a workstation PC<sup>2</sup> which is comparable to what can be expected in a clinical setting.

<sup>2</sup> CPU: Dual Intel Xeon Silver 4214R (24 cores, 48 threads) 2.9GHz, GPU: 3 NVIDIA A40, 46GB memory, 1.15–1.74GHz (only 1 GPU used here)





**Fig. 6.** Ranks chosen by rank-adaptive integrator for different spatial and angular resolutions and the Fokker-Planck approximation in (a) the homogeneous and (b) heterogeneous test case with a single beam.

**Table 1**

Theoretical and measured time and memory costs of DLRA vs. full rank.

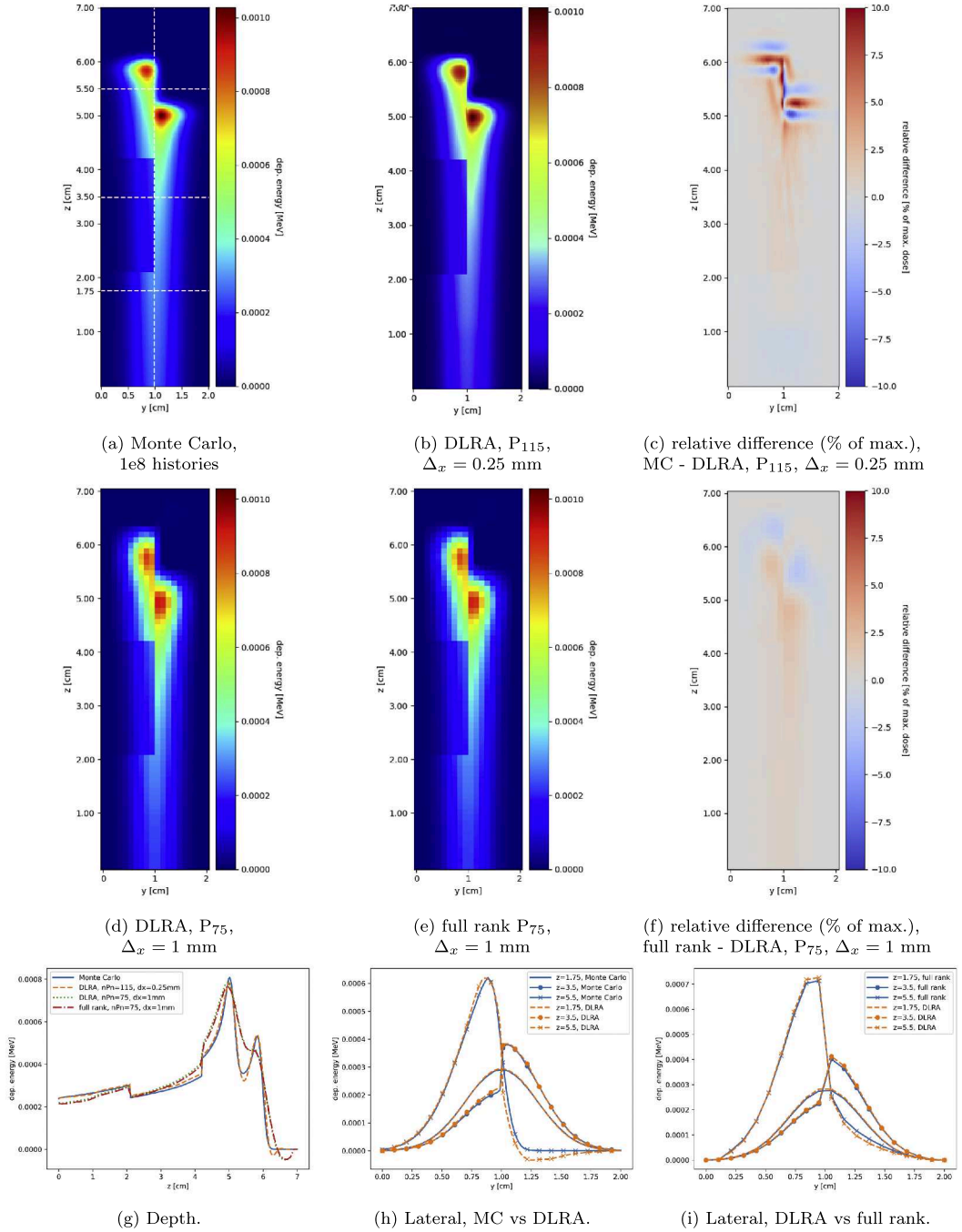
| <i>Boltzmann</i>     | full rank (CPU) | DLRA (GPU)      | % of full rank |
|----------------------|-----------------|-----------------|----------------|
| complexity           | 161 728 000     | 769 597.8992    | 0.4759%        |
| system size          | 161 728 000     | 161 226.3584    | 0.0997%        |
| runtime              | 12h             | 20.4s           | 0.04722%       |
| memory               | 41.4TiB         | 1.69GiB         | 0.0039%        |
| <i>Fokker-Planck</i> | full rank (CPU) | DLRA (GPU)      | % of full rank |
| complexity           | 11 200 000      | 11 348 642.8400 | 101.3272%      |
| system size          | 11 200 000      | 567 716.0000    | 5.0689%        |
| runtime              | 2684s           | 32.9s           | 1.2258%        |
| memory               | 2.84TiB         | 2.71GiB         | 0.0932%        |

Table 1 shows a comparison of the full-rank and DLRA Boltzmann and Fokker-Planck solvers with respect to their theoretical as well as measured costs<sup>3</sup> for the homogeneous test case with coarser spatial and angular resolution. We report the minimum execution time out of 20 runs, since the execution path is completely deterministic and thus variations are due to system overhead and background processes on the shared workstation. The theoretical values are computed using the formulas for computational complexity and memory from above and the average rank over time/energy steps is used for DLRA, which is 4.77 and 19.99 for Boltzmann and Fokker-Planck, respectively. We see that in case of the Boltzmann solver, DLRA reduces the costs to less than 0.5% in all cases and consistently achieves an even higher efficiency gain in practice, which demonstrates the combined effects of GPU implementation and complexity reduction. As predicted theoretically, the reductions are larger in terms of memory consumption.

Since for the Fokker-Planck approximation we both use lower orders of angular discretization and the average rank is higher than for Boltzmann, there are no theoretical complexity improvements in the case considered here. In practice, we can however see almost two orders of magnitude speed-up due to the GPU implementation. Thus, even for slightly higher rank and less high-dimensional problems, DLRA can be worthwhile by reducing the memory. Further, we see in Fig. 6 that the ranks for Fokker-Planck do not necessarily increase for finer spatial discretizations, so for more highly resolved problems we would also see larger theoretical gains with respect to runtime.

Note, that a fair comparison to the costs of the TOPAS MC reference computation is not as straightforward [57]. For  $10^8$  histories (primary particles) the computation time on the same workstation using up to 48 nodes is still in the ballpark of several hours. However, due to the different numerical discretization methods and included physics, the theoretical complexities are difficult to compare at a constant error. Further, the actual runtime can be almost arbitrarily reduced with more computational power due to the embarrassingly parallel nature of Monte Carlo methods. Thus, we give these values merely as a frame of reference.

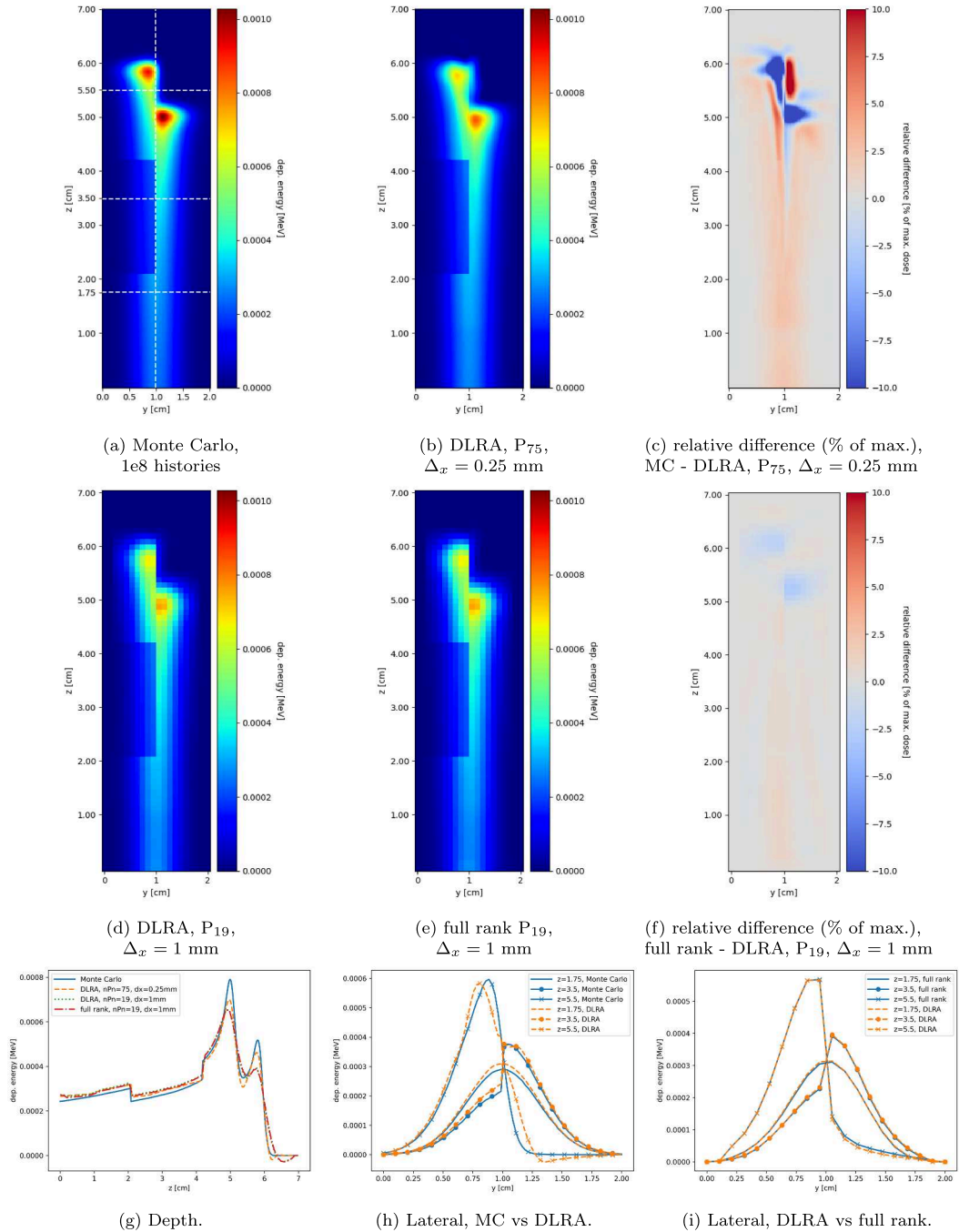
<sup>3</sup> The reported memory is total allocated memory according to TimerOutputs.jl. To avoid ambiguity we use GiB (Gibibyte) and TiB (Tebibyte) to denote the memory in the binary system, i.e., 1 GiB =  $1024^3$  bytes. Times are recorded using the same julia package with DLRA running on the GPU and the full rank method on the CPU with only the inherent linear algebra parallelization from julia.



**Fig. 7.** Surface plots (a)-(b) and (d)-(e), relative differences (c) and (f) and profiles (g)-(i) of deposited energy in the heterogeneous test case with a single beam for the Monte Carlo reference compared to DLRA at high and low spatial and angular resolution and the full-rank deterministic method based on the Boltzmann equation. Lines in (a) indicate positions of profiles plotted in (g)-(i).

## 6. Discussion

Our numerical results have demonstrated that the dynamical low-rank approximation can reduce the computational costs and memory associated with three-dimensional dose calculations in proton therapy by several orders of magnitude, both in theory and practice. In particular, we see that protons exhibit very low ranks even compared to other charged particles [16]. This can likely be attributed to the structure of their energy deposition curve, i.e., the dynamics are mostly dominated by advection, followed by close

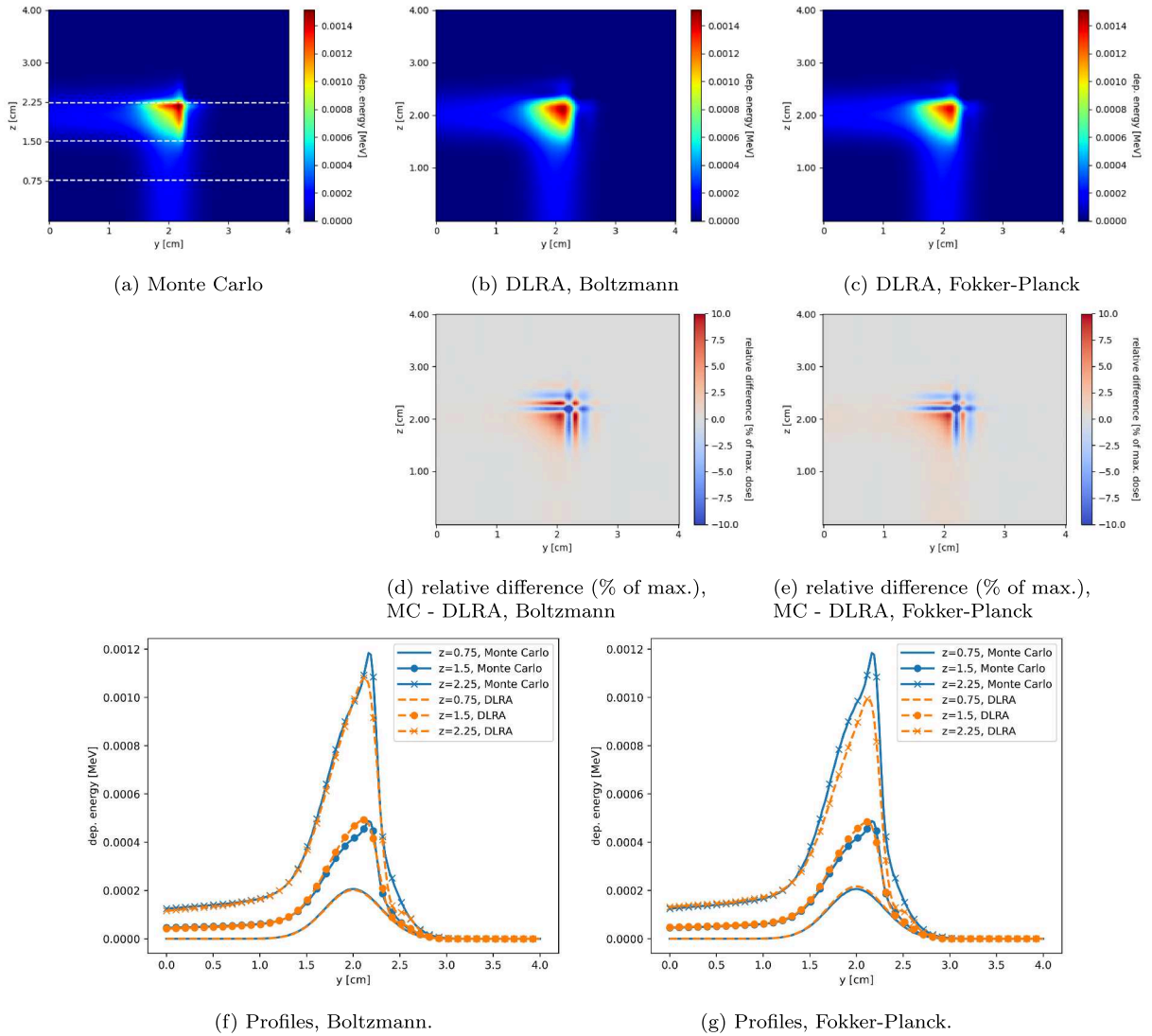


**Fig. 8.** Surface plots (a)-(b) and (d)-(e), relative differences (c) and (f) and profiles (g)-(i) of deposited energy in the heterogeneous test case with a single beam for the Monte Carlo reference compared to DLRA at high and low spatial and angular resolution and the full-rank deterministic method based on the Fokker-Planck approximation. Lines in (a) indicate positions of profiles plotted in (g)-(i).

to diffusive behavior for low energies. While ranks expectedly increase in heterogeneous materials, the fact that they do not increase significantly for several beam sources is promising for a use in treatment planning.

We investigated two variations of the transport model: the Boltzmann equation and a Fokker-Planck approximation. Overall, we observe slightly better results using the Boltzmann solver, both in terms of agreement with Monte Carlo, stability and ranks. Our comparative study between the Boltzmann and Fokker-Planck models reveals that, despite the Fokker-Planck model's simpler collision operator, the Boltzmann model provides more accurate results with lower computational ranks in our high-resolution framework. This suggests that the additional physical fidelity of the Boltzmann operator is crucial and is not a bottleneck when combined with DLRA.

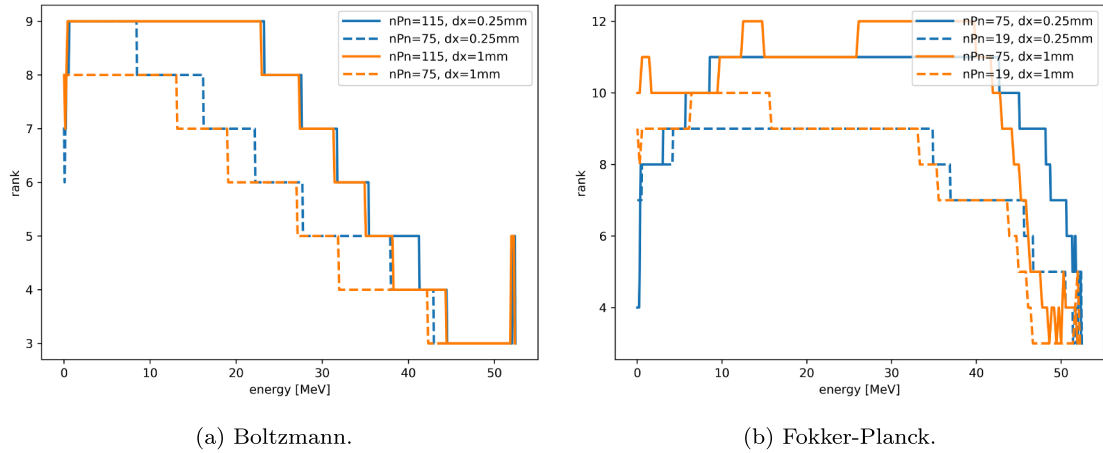




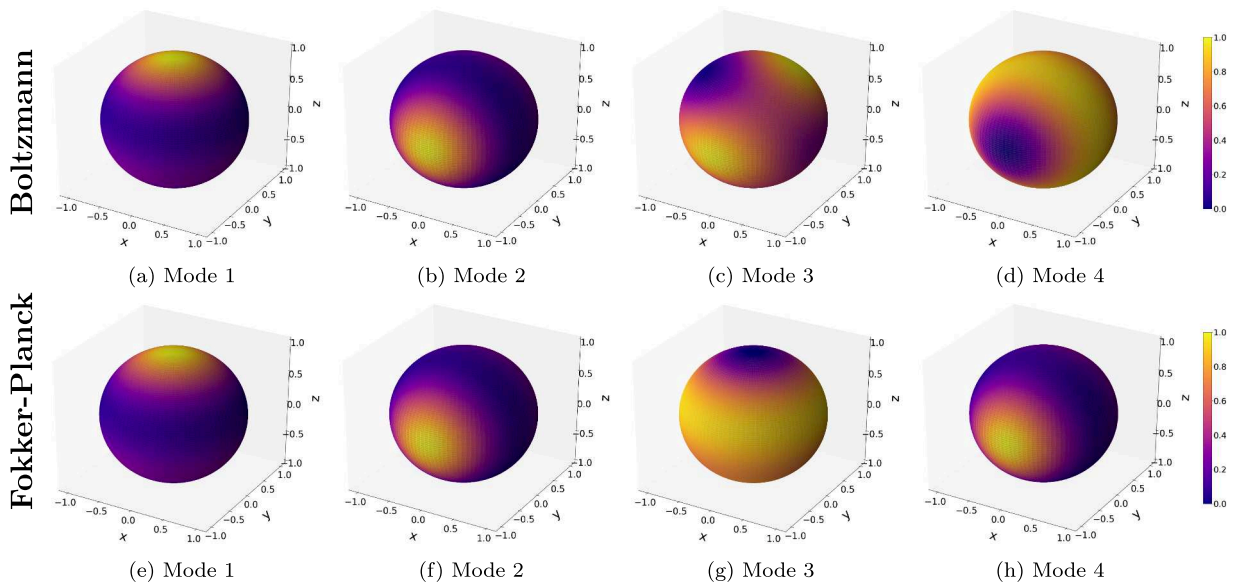
**Fig. 9.** Surface plots (a)-(c), differences (d)-(e) and profiles (f)-(g) of deposited energy for two perpendicular beams in a homogeneous material for the Monte Carlo reference compared to DLRA based on the Boltzmann equation and the Fokker-Planck approximation. Lines in (a) indicate positions of profiles plotted in (d)-(e).

For both approaches, the low-rank solution shows near perfect agreement with the equivalent full-rank solution. Comparisons to the Monte Carlo reference show a need for high spatial and angular resolution, as solution quality deteriorates with coarser discretization. Yet, being able to relatively cheaply compute the dose distribution on such fine grids speaks for the efficacy of the dynamical low-rank approximation. The choice of spatial and angular discretizations however needs to be reevaluated, also with view to “unphysical” behavior, i.e., negative regions in the solution. Further, the Fokker-Planck method heavily relies on the implicit scattering update and the correction described in [46,47] to avoid instabilities and artifacts at the material boundary, respectively. The correction however tends to overestimate the scattering moments [45] which could be a cause of the differences to Monte Carlo that we are seeing. Thus, future work could investigate whether the use of other, higher-order or locally refined spatial and angular discretizations can reduce the required expansion order and eliminate oscillations. Further, since protons undergo more physical interactions at low energies and thus energy loss is non-linear, a proportional, non-uniform energy/pseudo-time discretization could also be of interest to better capture rapid flux changes.

The currently used model for particle interactions still makes several simplifying assumptions. While the use of the continuous slowing down approximation and material composition model is common also in state-of-the-art Monte Carlo codes [58], incorporating nuclear interactions as well as absorption is necessary in order to achieve clinical precision [41]. Note, that we do not see the full impact of neglecting these effects in this work as we compare to a Monte Carlo method which is also limited to electromagnetic interactions. Fortunately, adding nuclear interactions or absorption does not majorly affect the numerical method described here and



**Fig. 10.** Ranks chosen by rank-adaptive integrator using (a) Boltzmann and (b) the Fokker-Planck approximation in the homogeneous test case with two beams.



**Fig. 11.** First four normalized dominant angular modes after one fifth of the time/energy steps.

is more a matter of accurate physical modeling or data (which can be challenging in itself). Based on the good agreement we achieve with the Monte Carlo reference and similar observations made in literature [15], neglecting straggling in the collided part of the equation appears to be a reasonable approximation. However, future work could also investigate strategies for handling a straggling term within the dynamical low-rank approximation of the collided equation.

## 7. Conclusion

In this work, we demonstrated that the computational costs and memory associated with deterministic solvers for a proton transport problem can be significantly reduced using the dynamical low-rank method. The chosen ranks are low, making a much higher spatial and angular resolution computationally feasible at little additional costs. We show very good agreement with a full-rank solver based on the same physical models and discretizations. Further, our results agree reasonably well with a TOPAS MC reference. Future work could explore the benefits of higher-order or refined discretizations, include more accurate physics models and extend the framework to full radiation therapy treatment plans.

## CRediT authorship contribution statement

**Pia Stammer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis; **Niklas Wahl:** Writing – review & editing, Software; **Jonas Kusch:** Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization; **Danny Lathouwers:** Writing – review & editing, Writing – original draft, Resources, Methodology.

## Data availability

A link to the repository containing code and data is included in the manuscript.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Pia Stammer reports financial support was provided by German National Academy of Sciences Leopoldina. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

Pia Stammer received funding from the German National Academy of Sciences Leopoldina for the project underlying this article, under grant number LPDS 2024-03.

## References

- [1] O.N. Vassiliev, T.A. Wareing, I.M. Davis, J. McGhee, D. Barnett, J.L. Horton, K. Gifford, G. Failla, U. Titt, F. Mourtada, Feasibility of a multigroup deterministic solution method for 3D radiotherapy dose calculations, *Int. J. Radiat. Oncol. Biol. Phys.* 72 (1) (2008) 220–227. <https://doi.org/10.1016/j.ijrobp.2008.04.057>
- [2] Y. Elcim, B. Dirican, O. Yavas, Dosimetric comparison of pencil beam and monte carlo algorithms in conformal lung radiotherapy, *J. Appl. Clin. Med. Phys.* 19 (5) (2018) 616–624. <https://doi.org/10.1002/acm2.12426>
- [3] I. Kawrakow, The effect of monte carlo statistical uncertainties on the evaluation of dose distributions in radiation treatment planning, *Phys. Med. Biol.* 49 (8) (2004) 1549–1556. <https://doi.org/10.1088/0031-9155/49/8/012>
- [4] M.W.K. Kan, P.K.N. Yu, L.H.T. Leung, et al., A review on the use of grid-based boltzmann equation solvers for dose calculation in external photon beam treatment planning, *Biomed. Res. Int.* 2013 (1) (2013) 692874. <https://doi.org/10.1155/2013/692874>
- [5] L. Eyges, Multiple scattering with energy loss, *Phys. Rev.* 74 (10) (1948) 1534–1535. <https://doi.org/10.1103/PhysRev.74.1534>
- [6] L. Hong, M. Goitein, M. Bucciolini, R. Comiskey, B. Gottschalk, S. Rosenthal, C. Serago, M. Urie, A pencil beam algorithm for proton dose calculations, *Phys. Med. Biol.* 41 (8) (1996) 1305. <https://iopscience.iop.org/article/10.1088/0031-9155/41/8/005/meta>
- [7] B. Schaffner, E. Pedroni, A. Lomax, Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation, *Phys. Med. Biol.* 44 (1) (1999) 27. <https://iopscience.iop.org/article/10.1088/0031-9155/44/1/004/meta>
- [8] M. Soukup, M. Fippel, M. Alber, A pencil beam algorithm for intensity modulated proton therapy derived from monte carlo simulations, *Phys. Med. Biol.* 50 (21) (2005) 5089. <https://iopscience.iop.org/article/10.1088/0031-9155/50/21/010/meta>
- [9] P. Yepes, A. Adair, D. Grosshans, D. Mirkovic, F. Poenisch, U. Titt, Q. Wang, R. Mohan, Comparison of monte carlo and analytical dose computations for intensity modulated proton therapy, *Phys. Med. Biol.* 63 (4) (2018) 045003. <https://iopscience.iop.org/article/10.1088/1361-6560/aaa845/meta>
- [10] P.A. Taylor, S.F. Kry, D.S. Followill, Pencil beam algorithms are unsuitable for proton dose calculations in lung, *Int. J. Radiat. Oncol.\*Biol.\*Phys.* 99 (3) (2017) 750–756. <https://doi.org/10.1016/j.ijrobp.2017.06.003>
- [11] J.E. Morel, T.A. Manteuffel, An angular multigrid acceleration technique for sn equations with highly forward-peaked scattering, *Nucl. Sci. Eng.* 107 (4) (1991) 330–342. <https://doi.org/10.13182/NSE91-A23795>
- [12] B. Turcksin, R.C. Jean, J.E. Morel, Angular multigrid preconditioner for krylov-based solution techniques applied to the sn equations with highly forward-peaked scattering, *Transp. Theory Stat. Phys.* 41 (1–2) (2012) 1–22. <https://doi.org/10.1080/00411450.2012.672944>
- [13] D. Lathouwers, Z. Perkó, An angular multigrid preconditioner for the radiation transport equation with fokker–planck scattering, *J. Comput. Appl. Math.* 350 (2019) 165–177. <https://doi.org/10.1016/j.cam.2018.10.010>
- [14] J. Kópházi, D. Lathouwers, A space–angle DGFEM approach for the boltzmann radiation transport equation with local angular refinement, *J. Comput. Phys.* 297 (2015) 637–668. <https://doi.org/10.1016/j.jcp.2015.05.031>
- [15] D. Lathouwers, A deterministic approach for proton transport, in: *Proceedings of M&C 2023 - the International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering*, American Nuclear Society – ANS, Niagara Falls, Ontario, Canada, 2023.
- [16] J. Kusch, P. Stammer, A robust collision source method for rank adaptive dynamical low-rank approximation in radiation therapy, *ESAIM: Math. Model. Numer. Anal.* 57 (2) (2023) 865–891. <https://doi.org/10.1051/m2an/2022090>
- [17] O. Koch, C. Lubich, Dynamical low-rank approximation, *SIAM J. Matrix Anal. Appl.* 29 (2) (2007) 434–454. <https://doi.org/10.1137/050639703>
- [18] E.W. Larsen, M.M. Miften, B.A. Fraass, I.A.D. Bruinvis, Electron dose calculations using the method of moments, *Med. Phys.* 24 (1) (1997) 111–125. <https://doi.org/10.1118/1.597920>
- [19] E. Kier, C. Lubich, H. Walach, Discretized dynamical low-rank approximation in the presence of small singular values, *SIAM J. Numer. Anal.* 54 (2) (2016) 1020–1038. <https://doi.org/10.1137/15M1026791>
- [20] C. Lubich, I.V. Oseledets, A projector–splitting integrator for dynamical low-rank approximation, *BIT Numer. Math.* 54 (1) (2014) 171–188. <https://doi.org/10.1007/s10543-013-0454-0>
- [21] G. Ceruti, C. Lubich, An unconventional robust integrator for dynamical low-rank approximation, *BIT Numer. Math.* 62 (1) (2022) 23–44. <https://doi.org/10.1007/s10543-021-00873-0>
- [22] G. Ceruti, J. Kusch, C. Lubich, A rank-adaptive robust integrator for dynamical low-rank approximation, *BIT Numer. Math.* 62 (4) (2022) 1149–1174. <https://doi.org/10.1007/s10543-021-00907-7>
- [23] G. Ceruti, J. Kusch, C. Lubich, A parallel rank-adaptive integrator for dynamical low-rank approximation, *SIAM J. Sci. Comput.* 46 (3) (2024) B205–B228. <https://doi.org/10.1137/23M1565103>
- [24] G. Ceruti, L. Einkemmer, J. Kusch, C. Lubich, A robust second-order low-rank BUG integrator based on the midpoint rule, *BIT Numer. Math.* 64 (3) (2024) 30. <https://doi.org/10.1007/s10543-024-01032-x>

- [25] J. Kusch, Second-order robust parallel integrators for dynamical low-rank approximation, *BIT Numer. Math.* 65 (3) (2025) 31. <https://doi.org/10.1007/s10543-025-01073-w>
- [26] L. Einkemmer, K. Kormann, J. Kusch, R.G. McClarren, J.-M. Qiu, A review of low-rank methods for time-dependent kinetic simulations, *J. Comput. Phys.* 538 (2025) 114191. <https://doi.org/10.1016/j.jcp.2025.114191>
- [27] L. Einkemmer, A. Moriggl, Semi-lagrangian 4d, 5d, and 6d kinetic plasma simulation on large-scale GPU-equipped supercomputers, *Int. J. High Perform. Comput. Appl.* 37 (2) (2023) 180–196.
- [28] J. Coughlin, J. Hu, U. Shumlak, Robust and conservative dynamical low-rank methods for the vlasov equation via a novel macro-micro decomposition, *J. Comput. Phys.* 509 (2024) 113055. <https://doi.org/10.1016/j.jcp.2024.113055>
- [29] C. Scalone, L. Einkemmer, J. Kusch, R.G. McClarren, et al., A multi-fidelity adaptive dynamical low-rank based optimization algorithm for fission criticality problems, *J. Sci. Comput.* 104 (1) (2025) 27. <https://doi.org/10.1007/s10915-025-02907-z>
- [30] J. Kusch, B. Whewell, R. McClarren, M. Frank, A low-rank power iteration scheme for neutron transport criticality problems, *J. Comput. Phys.* 470 (2022) 111587.
- [31] C.L. Leakeas, E.W. Larsen, Generalized fokker-planck approximations of particle transport with highly forward-peaked scattering, *Nucl. Sci. Eng.* 137 (3) (2001) 236–250. <https://doi.org/10.13182/NSE01-A2189>
- [32] T. Burlacu, D. Lathouwers, Z. Perkó, A deterministic adjoint-based semi-analytical algorithm for fast response change computations in proton therapy, *J. Comput. Theoret. Transport* 52 (1) (2023) 1–41. <https://doi.org/10.1080/23324309.2023.2166077>
- [33] B. Gottschalk, Radiotherapy Proton Interactions in Matter, 2018, <https://doi.org/10.48550/arXiv.1804.00022>
- [34] J.F. Janni, Proton range-energy tables, 1 kev–10 GeV, energy loss, range, path length, time-of-flight, straggling, multiple scattering, and nuclear interaction probability. part i. for 63 compounds, *At. Data Nucl. Data Tables* 27 (1982) 147. [https://doi.org/10.1016/0092-640X\(82\)90004-3](https://doi.org/10.1016/0092-640X(82)90004-3)
- [35] J. Perl, J. Shin, J. Schümann, B. Faddegon, H. Paganetti, TOPAS: An innovative proton monte carlo platform for research and clinical applications, *Med. Phys.* 39 (11) (2012) 6818–6837. <https://doi.org/10.1118/1.4758060>
- [36] E.J. Williams, W.L. N. Bragg, The passage of alpha- and beta- particles through matter and born's theory of collisions, *Proc. R. Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Character* 135 (826) (1997) 108–131. <https://doi.org/10.1098/rspa.1932.0023>
- [37] W.T. Scott, The theory of small-angle multiple scattering of fast charged particles, *Rev. Mod. Phys.* 35 (2) (1963) 231–313. <https://doi.org/10.1103/RevModPhys.35.231>
- [38] G. Molière, Theorie der streuung schneller geladener teilchen II mehrfach-und vielfachstreuung, *Zeitschrift für Naturforschung A* 3 (2) (1948) 78–97. <https://doi.org/10.1515/zna-1948-0203>
- [39] T. Burlacu, D. Lathouwers, Z. Perkó, Yet another dose algorithm (YODA) for independent computations of dose and dose changes due to anatomical changes, *Phys. Med. Biol.* 69 (16) (2024). <https://doi.org/10.1088/1361-6560/ad6373>
- [40] H.A. Bethe, Molière's theory of multiple scattering, *Phys. Rev.* 89 (6) (1953) 1256–1266. <https://doi.org/10.1103/PhysRev.89.1256>
- [41] W.D. Newhauser, R. Zhang, The physics of proton therapy, *Phys. Med. Biol.* 60 (8) (2015) R155. <https://doi.org/10.1088/0031-9155/60/8/R155>
- [42] M. Durante, H. Paganetti, Nuclear physics in particle therapy: a review, *Rep. Prog. Phys.* 79 (9) (2016) 096702. <https://doi.org/10.1088/0034-4885/79/9/096702>
- [43] P. Stammer, T. Burlacu, N. Wahl, J. Kusch, D. Lathouwers, A deterministic dynamical low-rank approach for charged particle transport, in: *Proceedings, ANS, Denver*, 2025, pp. 556–565. Preprint: [arXiv:2412.09484](https://arxiv.org/abs/2412.09484),
- [44] X. Zhang, X. Bai, M. Tang, A deterministic solver for the linear Boltzmann model of a single mono-directional proton beam, 2025, [arXiv:2504.00340\[math\]](https://arxiv.org/abs/2504.00340), <https://doi.org/10.48550/arXiv.2504.00340>
- [45] C.R. Drumm, F.c. Wesley, L. Leonard, J. Liscum-Powell, An analysis of the extended-transport correction with application to electron beam transport, *Nucl. Sci. Eng.* 155 (3) (2007) 355–366. <https://doi.org/10.13182/NSE07-A2668>
- [46] M. Landesman, J.E. Morel, Angular fokker-planck decomposition and representation techniques, *Nucl. Sci. Eng.* 103 (1) (1989) 1–11. <https://doi.org/10.13182/NSE89-A23655>
- [47] J.E. Morel, Fokker-planck calculations using standard discrete ordinates transport codes, *Nucl. Sci. Eng.* 79 (4) (1981) 340–356. <https://doi.org/10.13182/NSE79-340>
- [48] J. Kusch, L. Einkemmer, G. Ceruti, On the stability of robust dynamical low-rank approximations for hyperbolic problems, *SIAM J. Sci. Comput.* 45 (1) (2023) A1–A24. <https://doi.org/10.1137/21M1446289>
- [49] L. Einkemmer, J. Kusch, S. Schotthöfer, Construction of high-order conservative basis-update and Galerkin dynamical low-rank integrators, 2025, <https://doi.org/10.48550/arXiv.2311.06399>
- [50] O.N. Vassiliev, T.A. Wareing, J. McGhee, G. Failla, M.R. Salehpour, F. Mourtada, Validation of a new grid-based boltzmann equation solver for dose calculation in radiotherapy with photon beams, *Phys. Med. Biol.* 55 (3) (2010) 581–598. <https://doi.org/10.1088/0031-9155/55/3/002>
- [51] T. Pichard, G.W. Alldredge, S. Brull, B. Dubroca, M. Frank, An approximation of the \$m\_2\$ closure: application to radiotherapy dose simulation, *J. Sci. Comput.* 71 (1) (2017) 71–108. <https://doi.org/10.1007/s10915-016-0292-8>
- [52] T.-H. Nguyen-Bui, A. Larouze, R. Bourgeois, B. Dubroca, Multi-dimensional numerical modelling of heavy ion transport in matter using a fast and accurate method, 2021. <https://hal.science/hal-03243154>.
- [53] D.A. Low, W.B. Harms, S. Mutic, J.A. Purdy, A technique for the quantitative evaluation of dose distributions, *Med. Phys.* 25 (5) (1998) 656–661. <https://doi.org/10.1118/1.598248>
- [54] R. Bibb, D. Eggbeer, A. Paterson, 2 - Medical imaging, in: R. Bibb, D. Eggbeer, A. Paterson (Eds.), *Medical Modelling (Second Edition)*, Woodhead Publishing, Oxford, second edition edition, 2015, pp. 7–34. <https://doi.org/10.1016/B978-1-78242-300-3.00002-0>
- [55] W. Schneider, T. Bortfeld, W. Schlegel, Correlation between CT numbers and tissue parameters needed for monte carlo simulations of clinical dose distributions, *Phys. Med. Biol.* 45 (2) (2000) 459. <https://iopscience.iop.org/article/10.1088/0031-9155/45/2/314/meta>.
- [56] W. Cao, H. Rocha, R. Mohan, G. Lim, H.M. Goudarzi, B.C. Ferreira, J.M. Dias, Reflections on beam configuration optimization for intensity-modulated proton therapy, *Phys. Med. Biol.* 67 (13) (2022) 13TR01. <https://doi.org/10.1088/1361-6560/ac6fac>
- [57] C. Börgers, Complexity of monte carlo and deterministic dose-calculation methods, *Phys. Med. Biol.* 43 (3) (1998) 517. <https://doi.org/10.1088/0031-9155/43/3/004>
- [58] S. Huang, M. Kang, K. Souris, C. Ainsley, T.D. Solberg, J.E. McDonough, C.B. Simone, L. Lin, Validation and clinical implementation of an accurate monte carlo code for pencil beam scanning proton therapy, *J. Appl. Clin. Med. Phys.* 19 (5) (2018) 558–572. <https://doi.org/10.1002/acm2.12420>