Check for updates

Studies in Applied Mathematics

WILEY

STUDIES IN APPLIED MATHEMATICS

ORIGINAL ARTICLE OPEN ACCESS

A Unified Framework From Boltzmann Transport to Proton Treatment Planning

Andreas E. Kyprianou[1] | Aaron Pim[2] | Tristan Pryer[2]

[1]Department of Statistics, University of Warwick, Coventry, UK | [2]Department of Mathematical Sciences, University of Bath, Bath, UK

Correspondence: Tristan Pryer (tmp38@bath.ac.uk)

Received: 23 August 2025 | **Revised:** 9 June 2026 | **Accepted:** 30 July 2026

ABSTRACT

We develop a unified stochastic–deterministic framework for proton transport in radiotherapy. The deterministic formulation is based on a Boltzmann–Fokker–Planck (BFP) equation with continuous slowing-down, angular diffusion, and scattering terms, posed in a kinetic variational setting. The stochastic formulation describes proton trajectories by a track-length parameterized Markov process with energy loss, angular diffusion, and jump scattering. The main result is an operator-theoretic connection between these two descriptions. We identify the backward generator of the stochastic process with the adjoint BFP operator and show that the associated resolvent represents the Green's function of the deterministic transport equation. This gives a rigorous interpretation of deterministic fluence as an expected occupation density of stochastic particle paths. We then apply this duality to dose computation and treatment planning. Dose is represented equivalently as a deterministic fluence functional and as an expected accumulated stopping-power functional along stochastic tracks. Finally, we formulate a hybrid SDE–PDE optimization framework in which stochastic forward simulations are combined with deterministic adjoint equations for gradient computation. We prove existence and differentiability of the resulting objective functional and derive the corresponding first-order optimality system.

1 | Introduction

1.1 | Motivation

Proton therapy is an advanced form of radiotherapy that exploits the physical properties of charged particles to deliver highly localized radiation doses to tumors while minimizing collateral damage to healthy tissue [1].

The principal advantage of protons lies in their characteristic *Bragg peak*, a sharp energy deposition near the end of their range, which contrasts with the more diffuse dose profile of photons. This property enables precise dose shaping and better sparing of critical structures, making proton therapy particularly well suited for pediatric cases, head and neck cancers, and tumors adjacent to radiosensitive organs [2].

https://doi.org/10.1111/sapm.70285

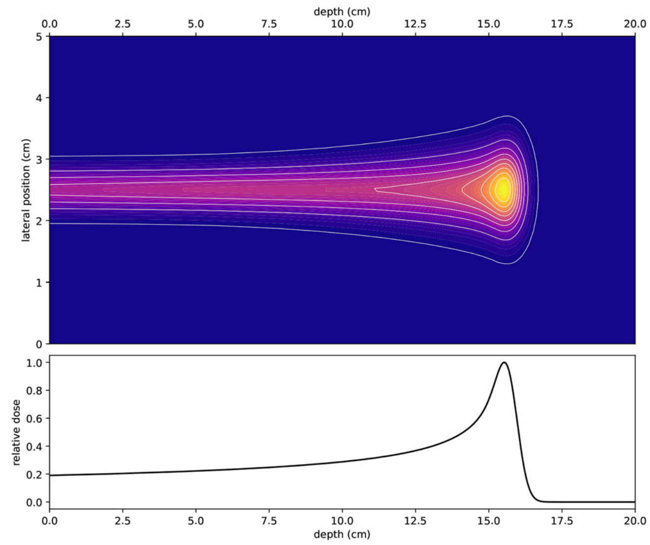

FIGURE 1 | Simulated dose profile of a proton beam illustrating the Bragg peak. The simulation was performed using MCsquare with $1.21 \cdot 10^7$ particles, a beam width of 2 mm, without nuclear interactions. The initial proton energy is 150 MeV with a 1% energy spread, and the dose is integrated along the plane orthogonal to the beam axis. Figure from [4].

Figure 1 shows a simulated proton dose profile illustrating the Bragg peak. The simulation, performed using MCsquare with $1.21 \cdot 10^7$ particles and a narrow beam width of 2 mm, captures the distinctive energy deposition behavior of 150 MeV protons in tissue-equivalent material [3].

Despite these advantages, the physics of proton transport is complex. Accurate modeling must account for energy loss due to ionization, multiple Coulomb scattering, and nonelastic nuclear reactions. Each of these effects introduces uncertainty and nonlocality into the transport process, making predictive modeling a formidable challenge. Two principal approaches have emerged in response: deterministic models based on integrodifferential transport equations, and stochastic models based on Monte Carlo simulation of individual particle tracks.

Deterministic approaches offer computational efficiency and analytic structure [4]. They are typically formulated using the Boltzmann or Boltzmann–Fokker–Planck (BFP) equation and serve as the foundation for many currently used clinical treatment planning systems [5]. However, they often rely on approximate closures or empirical fits to handle angular scattering and nuclear interactions [6]. In contrast, stochastic models offer high physical fidelity by directly simulating microscopic interactions [7, 8]. Monte Carlo methods are widely used for benchmarking and secondary dose calculations, but can easily become too expensive for routine optimization or real-time adaptation [9, 10].

This paper addresses the need for a rigorous mathematical connection between deterministic and stochastic frameworks for proton transport. This connection enables the development of hybrid computational strategies that combine the physical accuracy of Monte Carlo with the analytic structure and optimization capabilities of deterministic PDEs [11]. The resulting framework supports not only accurate dose calculation but also efficient and provably correct gradient-based treatment planning [12].

1.2 | Main Contributions

This work develops a unified mathematical theory of proton transport that links deterministic and stochastic formulations through functional analysis, operator theory, and stochastic calculus. Starting from a physical description of energy loss and scattering, we formulate the BFP equation in a variational framework. We define the transport operator in a kinetic graph space and establish monotonicity, maximality, and the existence of well-posed forward and adjoint problems. These results provide the foundation for the description of deterministic PDE-based transport.

In parallel, we introduce a stochastic model of proton motion defined via a system of stochastic differential equations (SDEs) indexed by *track length* rather than time. This model incorporates continuous energy loss, angular diffusion on the sphere, and discrete scattering events. We derive the infinitesimal generator of the stochastic process and show that it corresponds to the backward BFP operator.

Using semigroup theory and variational analysis, we prove that the deterministic operator is the adjoint of the stochastic generator and that the resolvent of the stochastic process coincides with the Green's function of the deterministic transport equation. This result provides a rigorous mathematical basis for identifying the *fluence* in the deterministic setting with the *expected occupation density* in the stochastic setting.

We then apply this framework to the problem of dose computation. We show that the classical expression

$$\text{dose} = \text{fluence} \times \text{mass stopping power}$$

emerges naturally and consistently from both deterministic and stochastic perspectives. This unification allows us to express dose as either a linear functional of the solution to the transport equation or as the expected accumulated energy loss of a stochastic particle.

Finally, we apply these ideas to formulate a hybrid optimization strategy for proton therapy treatment planning. The optimization problem is posed over the beam intensity function on the inflow boundary. The forward dose is computed using Monte Carlo simulations of the SDE model, while gradients are computed efficiently via deterministic adjoint PDEs. We prove existence of optimal controls under realistic constraints, derive the first-order optimality conditions in the form of a projected primal–dual system and propose a numerical algorithm based on alternating stochastic simulation and adjoint-based gradient updates.

1.3 | Relation to the Literature

The mathematical modeling of particle transport has been the subject of extensive work in both physics and applied mathematics. Deterministic approaches include discrete ordinates methods [13], Fokker–Planck approximations and moment-based closures, with clinical implementations often relying on simplifying assumptions to reduce computational cost [11, 14]. Stochastic methods, primarily Monte Carlo, provide detailed physics and are used extensively for verification and secondary dose estimates [7, 8]. However, the computational burden of Monte Carlo simulations poses challenges for real-time optimization settings [9, 10].

Several works have attempted to introduce adjoint methods into Monte Carlo frameworks, using techniques such as track reversal or score-function estimation [8, 15]. These approaches allow for sensitivity analysis but typically require specialized sampling techniques and are sensitive to variance. In contrast, our approach leverages the duality between stochastic generators and deterministic PDEs to obtain gradients via deterministic adjoint equations, while retaining the physical accuracy of Monte Carlo for forward dose computation. Moving forward, we believe this hybrid formulation will enable efficient and robust real-time treatment planning strategies [12, 16].

The mathematical connection between deterministic and stochastic transport is well known in kinetic theory but has not been fully developed in the context of proton therapy. Our work builds on ideas from operator theory, semigroup analysis, and stochastic calculus to establish this connection rigorously, with direct implications for computational radiotherapy [17–19].

The remainder of the paper is structured as follows. Section 2 introduces the physical processes governing proton transport, including energy loss and scattering. Section 3 formulates the deterministic transport equation in a variational framework and establishes the properties of the forward and adjoint transport operators. Section 4 develops the stochastic SDE model and derives its generator and associated Kolmogorov equations. Section 5 establishes the duality between the stochastic and deterministic operators and constructs the corresponding resolvents, thereby providing a unified theory of dose computation in both settings. Section 6 formulates the treatment planning problem as a PDE-constrained stochastic optimization problem and proposes a hybrid algorithm. Section 7 summarizes the results and discusses future directions.

2 | Physics of Proton Transport

To mathematically describe the evolution of protons, we work in the spatial-direction-energy phase space. Let $\boldsymbol{x} \in D \subset \mathbb{R}^3$ denote a position in a closed, bounded spatial domain, $\omega \in \mathbb{S}^2$ a unit vector describing the direction of motion, and $E \in \mathcal{E} := (E_{\min}, E_{\max}) \subset (0, \infty)$ an admissible energy. The phase space is then given by

$$\mathcal{C} = D \times \mathbb{S}^2 \times \mathcal{E}. \tag{1}$$

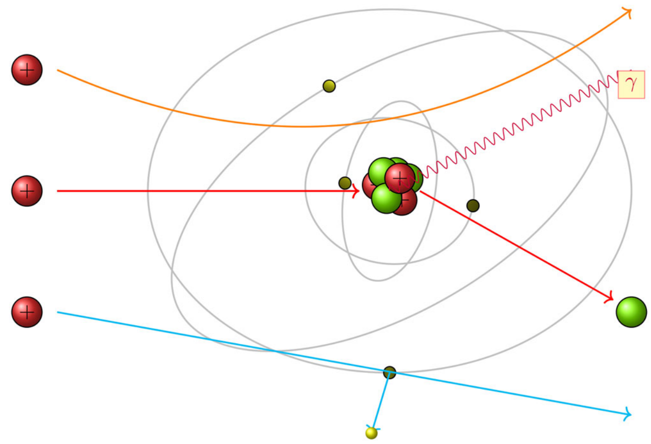

FIGURE 2 | The three main interactions of a proton with matter. A nonelastic proton–nucleus collision, an inelastic Coulomb interaction with atomic electrons, and elastic Coulomb scattering with the nucleus.

Elements of $\mathcal{C}$ will typically be written as $X = (\boldsymbol{x}, \omega, E)$; that is, capital letters denote points in $\mathcal{C}$, while $\boldsymbol{x} \in D$, $\omega \in \mathbb{S}^2$, and $E \in \mathcal{E}$. Variants of these dummy variables will be used throughout, but the notation will consistently distinguish spaces by font and capitalization. We also denote by $\boldsymbol{n}$ the outward-pointing unit normal on the spatial boundary ∂D.

Protons interact with the medium primarily through three mechanisms: inelastic collisions with atomic electrons, elastic Coulomb scattering with nuclei, and nonelastic nuclear reactions (see Figure 2). These interactions are characterized by cross sections, which describe the probability of an interaction occurring per unit length traveled. They form the basis for scattering kernels, stopping power functions, and reaction terms in the governing transport equations.

2.1 | Transport Processes in Proton Therapy

A proton traversing a medium undergoes multiple interactions that influence its trajectory, energy, and ultimately its dose deposition profile. In the absence of interactions, a proton moves in a straight line with constant kinetic energy E. This free-streaming motion is referred to as *transport*.

As a proton propagates through an atomic medium, it experiences collisional interactions with orbital electrons, transferring energy while largely maintaining its direction. This process, known as *inelastic electron–proton interaction*, is the primary mechanism of continuous energy loss and can lead to ionization or excitation of electrons. The frequency of such events is characterized by the cross section $\sigma_i(\boldsymbol{x}, E' \to E)$, where E' is the proton energy prior to the interaction and $E' - E$ is the energy lost.

Protons are also subject to small-angle deflections resulting from elastic Coulomb interactions with nuclei. This *elastic nucleus–proton interaction* preserves total kinetic energy but redistributes it between the proton and the nucleus, producing a net energy loss over many such events. The angular deflection is described by the cross section $\sigma_e(\boldsymbol{x}, \omega' \cdot \omega, E')$, where $\omega' \cdot \omega$ is the cosine of the scattering angle and E' is the precollision energy.

At higher energies, a proton may undergo a nonelastic nuclear reaction, producing secondary particles and incurring a discrete loss of energy. This *nonelastic proton–nucleus interaction* occurs with probability per unit path length given by the cross section $\sigma_{\mathrm{n}}(\boldsymbol{x}, \omega, E)$. Conditional on such an event, the proton transitions from an incoming state $(\boldsymbol{x}, \omega', E')$ to an outgoing state $(\boldsymbol{x}, \omega, E)$ with probability measure $\pi(\boldsymbol{x}, \omega', E'; d\omega, dE)$.

2.2 | Energy, Range, and Stopping Power

The range–energy relationship provides an estimate of how far protons of a given energy can penetrate into a homogeneous medium. Suppose a proton beam consists of particles with initial energy E_0. The corresponding range R_0 is often modeled by a power law

$$R_0 = \alpha E_0^p, \tag{2}$$

TABLE 1 | Range–energy relationship parameters for different media. The exponent p remains relatively constant across tissues, while α varies significantly with density and composition. The uncertainty for water reflects variability reported in [20–22].

Medium	p	α
Water	1.75 ± 0.02	0.00246 ± 0.00025
Muscle	1.75	0.0021
Bone	1.77	0.0011
Lung	1.74	0.0033

where $p \in [1, 2]$ and $\alpha > 0$ are empirical constants determined by the mass density and composition of the medium. For reference, representative parameter values are reported in Table 1; these values are not used in the analysis below, which only requires the structural assumptions stated explicitly in the operator estimates.

Let z denote the depth along the proton beam axis. At a given depth z, the remaining range of the proton beam is $R_0 - z$, and applying (2) to this residual range yields

$$E(z) = \alpha^{-\frac{1}{p}} (R_0 - z)^{\frac{1}{p}}. \tag{3}$$

The *stopping power*, defined as the energy loss per unit path length, follows by differentiation

$$S(z) := -\frac{dE(z)}{dz} = \frac{\alpha^{-\frac{1}{p}}}{p} (R_0 - z)^{\frac{1}{p} - 1}. \tag{4}$$

Inverting (3) to express z in terms of E and substituting into $S(z)$ yields the *Bragg–Kleeman rule*

$$S(E) = \frac{1}{\alpha p} E^{1-p}, \tag{5}$$

which shows how stopping power decreases with increasing energy, giving rise to the characteristic Bragg peak near the end of the range.

Remark 2.1 (Heterogeneous Tissue Effects). The range–energy relationship assumes a homogeneous medium, such as a water phantom, where α and p are constant. In heterogeneous tissues, the mass density and elemental composition vary with position, leading to spatial variations in α and discontinuities in the stopping power across tissue interfaces. In such cases, the range–energy relationship must be defined piecewise, with parameters adjusted for each region. This framework underlies practical models that incorporate CT-derived maps of tissue composition [23].

3 | Deterministic Formulation: The PDE Approach

3.1 | Functional Setup for Proton Transport

The mathematical analysis of the proton transport equation requires a precise functional analytic framework. We adopt standard Sobolev and Lebesgue space notations, following [24].

For a Hilbert space $\mathcal{X} = \mathcal{X}(\mathcal{C})$ over the phase space $\mathcal{C} = D \times \mathbb{S}^2 \times \mathcal{E}$, we write $\|\cdot\|_{\mathcal{X}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ for the norm and inner product, respectively. We denote the dual of $\mathcal{X}$ by $\mathcal{X}^*$, and write $\langle \cdot, \cdot \rangle_{\mathcal{X}(\mathcal{C})}$ for the duality pairing between $\mathcal{X}^*$ and $\mathcal{X}$ when the context is clear. Where needed for clarity, we write $\langle \cdot, \cdot \rangle_{\mathcal{X}, \mathcal{X}^*}$.

To set up the transport equation with boundary conditions, we first introduce function spaces adapted to the phase-space boundary. The spatial boundary is

$$\Gamma_x = \partial D \times \mathbb{S}^2 \times \mathcal{E}, \tag{6}$$

where $\boldsymbol{n}$ denotes the outward-pointing unit normal to ∂D. The energy boundary consists of

$$\Gamma_E^{\min} = D \times \mathbb{S}^2 \times \{E_{\min}\}, \qquad \Gamma_E^{\max} = D \times \mathbb{S}^2 \times \{E_{\max}\}. \tag{7}$$

The spatial inflow and outflow boundaries are

$$\Gamma_x^- := \{(\boldsymbol{x}, \omega, E) \in \Gamma_x : \omega \cdot \boldsymbol{n} < 0\}, \qquad \Gamma_x^+ := \{(\boldsymbol{x}, \omega, E) \in \Gamma_x : \omega \cdot \boldsymbol{n} \geq 0\}. \tag{8}$$

Since the energy drift is directed from high energy to low energy, $E = E_{\max}$ plays the role of an energy inflow face and $E = E_{\min}$ plays the role of an energy outflow face. We therefore write

$$\Gamma^- := \Gamma_x^- \cup \Gamma_E^{\max}, \qquad \Gamma^+ := \Gamma_x^+ \cup \Gamma_E^{\min}. \tag{9}$$

On the spatial part of the boundary we use the usual transport trace spaces

$$\mathrm{L}^2(\Gamma_x^\pm) := \left\{ \phi : \Gamma_x^\pm \to \mathbb{R} \,\middle|\, \int_{\Gamma_x^\pm} \phi^2 |\omega \cdot \boldsymbol{n}| d\gamma_x < \infty \right\}, \tag{10}$$

where $d\gamma_x$ denotes the product of surface measure on ∂D, surface measure on $\mathbb{S}^2$ and Lebesgue measure on $\mathcal{E}$. The corresponding inner product is

$$\langle \phi_1, \phi_2 \rangle_{\Gamma_x^\pm} := \int_{\Gamma_x^\pm} \phi_1 \phi_2 |\omega \cdot \boldsymbol{n}| d\gamma_x. \tag{11}$$

For notational simplicity, and since the main control is imposed on the physical inflow boundary, we continue to write $\mathrm{L}^2(\Gamma^-)$ and $\mathrm{L}^2(\Gamma^+)$ for the corresponding trace spaces, with the energy-boundary contribution weighted by the magnitude of the energy drift when it is explicitly used.

3.2 | The BFP Operator

The full BFP operator governing proton transport is denoted by $\vec{\mathrm{L}}$ and given by

$$\vec{\mathrm{L}}\, \phi := \vec{\mathrm{T}}\, \phi - \mu \Delta_\omega \phi, \tag{12}$$

where $\mu > 0$ is the angular diffusion coefficient and the transport operator $\vec{\mathrm{T}}$ is defined by

$$\vec{\mathrm{T}}\, \phi := \omega \cdot \nabla_{\boldsymbol{x}} \phi - \partial_E (S\phi) - \mathcal{S}\phi + \sigma_{\mathrm{n}} \phi. \tag{13}$$

Here, S is the stopping power from (5), and $\mathcal{S}$ is the scattering operator describing angular and energy redistribution due to collisions.

Each term in $\vec{\mathrm{L}}$ has a direct transport interpretation. The streaming term $\omega \cdot \nabla_{\boldsymbol{x}} \phi$ describes motion along the current direction of travel. The term $-\partial_E(S\phi)$ is the continuous slowing-down contribution: the stopping power S induces deterministic drift from high to low energy. The angular Fokker–Planck term $-\mu \Delta_\omega \phi$ models the cumulative effect of many small-angle elastic Coulomb deflections. Thus, μ is the effective angular diffusion coefficient appearing in the deterministic generator. The operator $\mathcal{S}$ is a scattering gain term, while $\sigma_{\mathrm{n}} \phi$ is the corresponding total interaction or attenuation term.

The operator used here should therefore be understood as a BFP approximation to the full linear Boltzmann transport equation: continuous energy loss is represented by the slowing-down drift, forward-peaked elastic scattering is represented by angular diffusion and nonelastic or large-angle events are retained through the integral scattering kernel.

We define $\mathcal{S}$ in terms of elastic and nonelastic components as

$$\mathcal{S}\phi(\boldsymbol{x}, \omega, E) = \int_{\mathbb{S}^2} \int_{\mathcal{E}} \phi(\boldsymbol{x}, \omega', E') \sigma_{\mathrm{n}}(\boldsymbol{x}, \omega', E') \pi(\boldsymbol{x}, \omega', E'; \omega, E) d\omega' dE', \tag{14}$$

where the total scattering cross section σ_{n} is decomposed into elastic and nonelastic parts

$$\sigma_{\mathrm{n}} = \sigma_{\mathrm{e}} + \sigma_{\mathrm{ne}}. \tag{15}$$

The kernel π defining postcollision transitions is then

$$\pi(\boldsymbol{x}, \omega', E'; \omega, E) = \frac{\sigma_{\mathrm{e}}(\boldsymbol{x}, \omega', E')}{\sigma_{\mathrm{n}}(\boldsymbol{x}, \omega', E')} \pi_{\mathrm{e}}(\boldsymbol{x}, \omega', E'; \omega) \delta_{E'}(E) + \frac{\sigma_{\mathrm{ne}}(\boldsymbol{x}, \omega', E')}{\sigma_{\mathrm{n}}(\boldsymbol{x}, \omega', E')} \pi_{\mathrm{ne}}(\boldsymbol{x}, \omega', E'; \omega, E), \tag{16}$$

where $\delta_{E'}$ denotes the Dirac density at energy E'. In this decomposition, $\sigma_{\mathrm{e}}(\boldsymbol{x}, \omega', E')$ and $\pi_{\mathrm{e}}(\boldsymbol{x}, \omega', E'; \omega)$ represent the elastic cross section and angular kernel from $(\boldsymbol{x}, \omega', E')$ to $(\boldsymbol{x}, \omega, E)$, while $\sigma_{\mathrm{ne}}(\boldsymbol{x}, \omega', E')$ and $\pi_{\mathrm{ne}}(\boldsymbol{x}, \omega', E'; \omega, E)$ model nonelastic scattering events with energy redistribution. The mixture kernel (16) defines a probability distribution over postscattering states (ω, E), conditioned on the incoming state $(\boldsymbol{x}, \omega', E')$.

In particular, we require that

$$\int_{\mathbb{S}^2} \int_{\mathcal{E}} \pi(\boldsymbol{x}, \omega', E'; \omega, E) d\omega dE = 1, \qquad \text{for all } (\boldsymbol{x}, \omega', E') \in \mathcal{C}, \tag{17}$$

so that π is a probability kernel in the outgoing variables. In addition, throughout the monotonicity argument below we assume the standard dissipativity condition

$$\langle (\sigma_{\mathrm{n}} - \mathcal{S}) v, v \rangle_{\mathrm{L}^2(\mathcal{C})} \geq 0 \qquad \text{for all admissible } v. \tag{18}$$

This condition holds, for example, under the usual contraction or detailed-balance hypotheses on the scattering kernel. It is stated separately because probability normalization alone ensures conservation of total scattering probability, whereas dissipativity in L^2 is the property required in the energy estimate.

3.3 | Functional Spaces and Operator Domains

We define the kinetic graph space

$$\vec{\mathscr{H}} := \left\{ \phi \in \mathrm{L}^2(\mathcal{C}) \,\middle|\, \nabla_\omega \phi \in \mathrm{L}^2(\mathcal{C})^3, \quad \vec{\mathrm{T}}\, \phi \in \mathrm{L}^2(\mathcal{C}) \right\}, \tag{19}$$

equipped with the graph norm

$$\|\phi\|^2_{\vec{\mathscr{H}}} := \|\phi\|^2_{\mathrm{L}^2(\mathcal{C})} + \left\| \sqrt{\mu} \nabla_\omega \phi \right\|^2_{\mathrm{L}^2(\mathcal{C})} + \left\| \vec{\mathrm{T}}\, \phi \right\|^2_{\mathrm{L}^2(\mathcal{C})}. \tag{20}$$

We define the subspace with homogeneous inflow boundary conditions as

$$\vec{\mathscr{H}}_0^- := \left\{ \phi \in \vec{\mathscr{H}} \,\middle|\, \phi = 0 \text{ on } \Gamma^- \right\} \tag{21}$$

corresponding to physical inflow constraints for the forward transport problem.

To define the range space of the BFP operator $\vec{\mathrm{L}}$, we introduce

$$\mathscr{V} := \mathrm{L}^2(D \times \mathcal{E}; H^{-1}(\mathbb{S}^2)), \tag{22}$$

which encodes L^2 regularity in position and energy, and dual Sobolev regularity in angle. This space is appropriate for the weak formulation of the degenerate elliptic transport equation.

Note that for its dual $\mathscr{V}^*$, we have continuous embeddings $\vec{\mathscr{H}}, \vec{\mathscr{H}}_0^- \subset \mathscr{V}^*$. We write $\langle \cdot, \cdot \rangle$ to denote the duality pairing between $\mathscr{V}$ and $\mathscr{V}^*$, omitting explicit mention of the spaces when context permits.

The natural boundary conditions for proton transport arise from specifying the incoming flux on the inflow boundary Γ^-. For solutions in $\overrightarrow{\mathscr{H}}$, we consider the boundary value problem

$$\overrightarrow{\mathrm{L}}\ \psi = 0 \quad \text{in } \mathcal{C}, \qquad \psi = g \quad \text{on } \Gamma^-, \tag{23}$$

where g is a prescribed inflow intensity. In the stochastic formulation, this corresponds to the *entry distribution* of particles, specifying the initial sampling law. At high energies, we assume decay of the solution as $E \to E_{\max}$, which reflects physical attenuation.

3.4 | The Backward Boltzmann–Fokker–Planck (bBFP) Operator

We define the functional analytic setting for the bBFP operator. The domain of the backward transport operator is

$$\overleftarrow{\mathscr{H}} := \left\{ \phi \in \mathrm{L}^2(\mathcal{C}) \,\middle|\, \overleftarrow{\mathrm{T}}\ \phi \in \mathrm{L}^2(\mathcal{C}),\ \nabla_\omega \phi \in \mathrm{L}^2(\mathcal{C})^3 \right\}, \tag{24}$$

with graph norm

$$\|\phi\|^2_{\overleftarrow{\mathscr{H}}} := \|\phi\|^2_{\mathrm{L}^2(\mathcal{C})} + \left\| \sqrt{\mu}\nabla_\omega \phi \right\|^2_{\mathrm{L}^2(\mathcal{C})} + \left\| \overleftarrow{\mathrm{T}}\ \phi \right\|^2_{\mathrm{L}^2(\mathcal{C})}. \tag{25}$$

We also define the subspace of functions vanishing on the outflow (i.e., terminal) boundary,

$$\overleftarrow{\mathscr{H}}_0^+ := \left\{ \phi \in \overleftarrow{\mathscr{H}} \,\middle|\, \phi = 0 \text{ on } \Gamma^+ \right\}. \tag{26}$$

This space enforces terminal conditions on the physical and energy outflow boundaries, corresponding to the vanishing of expected future contributions in the adjoint formulation.

The bBFP operator is defined formally as the dual of the forward operator $\overrightarrow{\mathrm{L}}$ with respect to the $\mathrm{L}^2(\mathcal{C})$ inner product

$$\left\langle \overrightarrow{\mathrm{L}}\ \phi, \psi \right\rangle = \left\langle \phi, \overleftarrow{\mathrm{L}}\ \psi \right\rangle \quad \text{for all } \phi \in \overrightarrow{\mathscr{H}}_0^-,\ \psi \in \overleftarrow{\mathscr{H}}_0^+. \tag{27}$$

Theorem 3.1 (Adjoint representation of the BFP operator). *It follows from the definition (27) that*

$$\overleftarrow{\mathrm{L}}\ \phi := \overleftarrow{\mathrm{T}}\ \phi - \mu \Delta_\omega \phi, \tag{28}$$

where the adjoint transport operator $\overleftarrow{\mathrm{T}}$ *is*

$$\overleftarrow{\mathrm{T}}\ \phi := -\omega \cdot \nabla_{\boldsymbol{x}} \phi + S\partial_E \phi - \mathcal{S}^* \phi + \sigma_{\mathrm{n}} \phi; \tag{29}$$

and the adjoint scattering operator $\mathcal{S}^*$ *is given by*

$$\mathcal{S}^* \psi(\boldsymbol{x}, \omega, E) := \sigma_{\mathrm{n}}(\boldsymbol{x}, \omega, E) \int_{\mathbb{S}^2} \int_{\mathcal{E}} \psi(\boldsymbol{x}, \omega', E') \pi(\boldsymbol{x}, \omega, E; \omega', E') d\omega' dE', \tag{30}$$

corresponding to a reversal of pre- and postcollision roles in the scattering process.

Proof. The duality relation (27) follows from formal integration by parts, under the assumption that boundary terms vanish due to the support properties of ϕ and ψ. Similar computations are carried out rigorously in, for example, [25]. □

3.5 | Admissible Inflow Boundary Conditions

The space of admissible inflow boundary data consists of functions in $L^2(\Gamma^-)$ that arise as traces of elements in $\vec{\mathscr{H}}$, the forward graph space. We define the trace space as

$$\mathrm{tr}(\Gamma^-) := \left\{ g \in L^2(\Gamma^-) \,\middle|\, \exists G \in \vec{\mathscr{H}} \ \text{ such that } \gamma^-(G) = g \right\}, \tag{31}$$

where γ^- denotes the trace operator on Γ^-. This space ensures compatibility between the boundary data and the variational formulation of the forward problem. In the stochastic formulation, $\mathrm{tr}(\Gamma^-)$ corresponds to the set of admissible entry distributions for particles entering the phase space.

3.6 | Reformulation as an Interior Problem

The natural variational space for the forward problem is $\vec{\mathscr{H}}_0^-$, comprising functions that vanish on the inflow boundary Γ^-. However, the boundary value problem

$$\vec{\mathrm{L}}\,\psi = 0 \quad \text{in } \mathcal{C}, \qquad \psi = g \quad \text{on } \Gamma^- \tag{32}$$

prescribes inhomogeneous boundary data g on Γ^-. To express this within the variational framework, we introduce a *lifting function* $\hat{g} \in \vec{\mathscr{H}}$ such that $\gamma^-(\hat{g}) = g$. Define the shifted unknown

$$u := \psi - \hat{g}, \tag{33}$$

so that $u \in \vec{\mathscr{H}}_0^-$ satisfies homogeneous inflow boundary conditions.
Substituting into the equation, we find that u satisfies the inhomogeneous problem

$$\vec{\mathrm{L}}\,u = -\vec{\mathrm{L}}\,\hat{g} := f \quad \text{in } \mathcal{C}, \qquad u = 0 \quad \text{on } \Gamma^-. \tag{34}$$

This allows us to recast the original boundary value problem as a variational equation: find $u \in \vec{\mathscr{H}}_0^-$ such that

$$\left\langle \vec{\mathrm{L}}\,u, \phi \right\rangle = \langle f, \phi \rangle = -\left\langle \vec{\mathrm{L}}\,\hat{g}, \phi \right\rangle, \qquad \forall \phi \in \vec{\mathscr{H}}_0^-. \tag{35}$$

This formulation is well defined provided that $\hat{g} \in \vec{\mathscr{H}}$ and $\vec{\mathrm{L}}\,\hat{g} \in \mathscr{V}^*$. The lifting $\hat{g}$ can be constructed by solving an auxiliary extension problem or interpolating g into a function in $\vec{\mathscr{H}}$ [cf. 24, Section 5.4].
In the probabilistic framework, this lifting corresponds to decomposing the solution ψ into two parts: a component $\hat{g}$ that encodes the entry distribution of particles on Γ^-, and a remainder u that captures the contribution of those trajectories that evolve within the domain after entering. Since the process is absorbed at the outflow boundary Γ^+ or at the minimal energy $E_{\min}$, the function u describes the internal evolution of particles conditioned on entering at Γ^-, the entrance boundary, and vanishes on the exit boundary Γ^+.

To establish well-posedness of (32) (and hence of the original boundary value problem), we show that the operator $\vec{\mathrm{L}}$ is both *monotone* and *maximal* [26, §7.1]. Under these conditions, existence and uniqueness of a solution $u \in \vec{\mathscr{H}}_0^-$ follow, and $\vec{\mathscr{H}}_0^-$ inherits a Hilbert space structure with inner product defined analogously to its graph norm [27, Section 6.3]. As we will see, in the stochastic setting, maximality of $\vec{\mathrm{L}}$ corresponds to well-posedness of the resolvent problem for $\overleftarrow{\mathrm{L}}$, ensuring that the backward process is defined throughout the entire phase space.

Lemma 3.2 (Monotonicity of the BFP Operator). *Assume that the angular diffusion coefficient $\mu \in L^\infty(D \times \mathcal{E}; \mathbb{R}^+)$, and that the stopping power $S \in W^{1,\infty}(D \times \mathcal{E})$ satisfies*

$$\lim_{E \downarrow E_{\min}} S(\boldsymbol{x}, E) \geq 0,\ \forall \boldsymbol{x} \in D, \qquad \partial_E S(\boldsymbol{x}, E) \leq 0,\ \forall (\boldsymbol{x}, E) \in D \times \mathcal{E}. \tag{36}$$

In addition, suppose that the nuclear scattering cross section satisfies $\sigma_{\mathrm{n}}(\boldsymbol{x}, \omega, E) \geq 0$, and that the scattering kernel π satisfies the normalization condition (17). Then, there exist constants $C_{\rightarrow}, C_{\leftarrow} > 0$ such that the BFP operator $\overrightarrow{\mathrm{L}}$ and its adjoint $\overleftarrow{\mathrm{L}}$ are monotone in the sense that

$$\begin{aligned} \left\langle \overrightarrow{\mathrm{L}}\, u, u \right\rangle &\geq C_{\rightarrow}\left(\|u\|_{\Gamma+}^2 + \|\sqrt{\mu}\nabla_\omega u\|_{\mathrm{L}^2(C)}^2 + \|u\|_{\mathrm{L}^2(C)}^2 + \lim_{E \downarrow E_{\min}} \|\sqrt{S}u\|_{\mathrm{L}^2(D\times\mathbb{S}^2)}^2 \right), \quad \forall u \in \overrightarrow{\mathscr{H}}_0^{-}, \\ \left\langle \overleftarrow{\mathrm{L}}\, \phi, \phi \right\rangle &\geq C_{\leftarrow}\left(\|\phi\|_{\Gamma-}^2 + \|\sqrt{\mu}\nabla_\omega \phi\|_{\mathrm{L}^2(C)}^2 + \|\phi\|_{\mathrm{L}^2(C)}^2 + \lim_{E \uparrow E_{\max}} \|\sqrt{S}\phi\|_{\mathrm{L}^2(D\times\mathbb{S}^2)}^2 \right), \quad \forall \phi \in \overleftarrow{\mathscr{H}}_0^{+}. \end{aligned} \tag{37}$$

Proof. We first consider the transport term. Integration by parts yields

$$\langle \omega \cdot \nabla_{\boldsymbol{x}} u, u \rangle = \frac{1}{2}\|u\|_{\Gamma+}^2 - \frac{1}{2}\|u\|_{\Gamma-}^2. \tag{38}$$

Since $u \in \overrightarrow{\mathscr{H}}_0^{-}$ vanishes on Γ^-, it follows that

$$\langle \omega \cdot \nabla_{\boldsymbol{x}} u, u \rangle = \frac{1}{2}\|u\|_{\Gamma+}^2 \geq 0. \tag{39}$$

Similarly, for the adjoint operator,

$$-\langle \omega \cdot \nabla_{\boldsymbol{x}} \phi, \phi \rangle = \frac{1}{2}\|\phi\|_{\Gamma-}^2 \geq 0, \quad \forall \phi \in \overleftarrow{\mathscr{H}}_0^{+}. \tag{40}$$

For the angular diffusion term

$$-\langle \mu\Delta_\omega v, v \rangle = \|\sqrt{\mu}\nabla_\omega v\|_{\mathrm{L}^2(C)}^2 \geq 0, \quad \forall v \in \overrightarrow{\mathscr{H}}_0^{-} \cup \overleftarrow{\mathscr{H}}_0^{+}. \tag{41}$$

We next consider the forward energy term

$$-\langle \partial_E(Su), u \rangle = -\langle S\partial_E u, u \rangle - \langle (\partial_E S)u, u \rangle \tag{42}$$

$$= \frac{1}{2}\lim_{E \downarrow E_{\min}} \|S^{1/2}u\|_{L^2(D\times\mathbb{S}^2)}^2 - \frac{1}{2}\langle (\partial_E S)u, u \rangle. \tag{43}$$

By the assumption $\partial_E S \leq 0$ and $S \geq 0$ near $E_{\min}$, both terms are nonnegative. For the backward energy term

$$\langle S\partial_E \phi, \phi \rangle = -\frac{1}{2}\langle \partial_E S \cdot \phi, \phi \rangle \geq 0. \tag{44}$$

Finally, for the scattering terms,

$$\langle (\sigma_{\mathrm{n}} - \mathcal{S})v, v \rangle_{\mathrm{L}^2(C)} \geq 0 \tag{45}$$

by the dissipativity assumption (18). The same argument applies to $\overleftarrow{\mathrm{L}}$ and $\mathcal{S}^*$ by adjointness. This completes the proof. □

We now discuss the validity of the assumptions made in Lemma 3.2. The assumption that the angular diffusion coefficient μ is nonnegative follows directly from the physical constraint that the elastic scattering cross-section σ_e is nonnegative. Since μ is derived from angular moments of σ_e, this condition is physically justified and ensures well-posed angular diffusion.

The assumption $S \in W^{1,\infty}(D \times \mathcal{E})$ reflects the use of regularized stopping power functions. For example, the Bragg–Kleeman model expresses stopping power as

$$S(\boldsymbol{x}, E) = \frac{1}{\alpha(\boldsymbol{x})p(\boldsymbol{x})}E^{1-p(\boldsymbol{x})}, \tag{46}$$

with $\alpha(\boldsymbol{x}) > 0$ and $p(\boldsymbol{x}) > 1$. This model satisfies the monotonicity and boundedness conditions provided the admissible energy domain satisfies $E_{\min} > 0$. In the unregularized case $E_{\min} = 0$, the stopping power diverges as $E \to 0$, violating the boundedness and integrability required for Lemma 3.2.

In practice, such singular models are not implemented directly due to their nonphysical behavior at low energies. Empirical stopping power laws (e.g., Bragg–Kleeman or Bethe–Bloch) are valid in the continuous slowing-down regime at high energies, but break down near rest energy. To address this, *screened stopping power models* are employed, where $S(\boldsymbol{x}, E)$ is regularized to remain bounded as $E \to 0$. These corrections incorporate low-energy screening and density effects. Further discussion can be found in [28].

We emphasize that these hypotheses are intended as regularized analytical assumptions rather than as a claim that all raw patient data are globally smooth. They are easy to satisfy in homogeneous media and in smoothed heterogeneous media, and they are also satisfied by Bragg–Kleeman or Bethe–Bloch-type stopping-power laws restricted to an interval $[E_{\min}, E_{\max}]$ with $E_{\min} > 0$ after the standard low-energy regularization used in numerical transport. In CT-derived geometries with sharp material interfaces, the coefficients are typically only piecewise smooth. The present theory then applies directly to mollified or voxelwise regularized coefficients, which are the objects actually passed to a stable deterministic or stochastic solver.

From the functional analytic perspective, the backward operator $\overleftarrow{\mathtt{L}} : \overleftarrow{\mathscr{H}}_0^{+} \to \mathscr{V}$ is closed, unbounded, and maximal monotone. It admits a well-defined resolvent operator. For any $\lambda > 0$, the resolvent

$$\overleftarrow{\mathtt{R}}_\lambda := (\lambda I + \overleftarrow{\mathtt{L}})^{-1} : \mathscr{V} \to \overleftarrow{\mathscr{H}}_0^{+} \tag{47}$$

defines the unique solution $u \in \overleftarrow{\mathscr{H}}_0^{+}$ to the variational problem

$$\lambda u + \overleftarrow{\mathtt{L}}\, u = v, \qquad \text{for } v \in \mathscr{V}. \tag{48}$$

In many classical settings, the case $\lambda = 0$ is excluded unless additional coercivity conditions are imposed to ensure invertibility of $\overleftarrow{\mathtt{L}}$. However, under the assumptions of Lemma 3.2, the operator $\overleftarrow{\mathtt{L}}$ contains dissipative terms, notably angular diffusion and energy drift, that induce strict energy decay and prevent nullspace degeneracy. As a result, the resolvent at zero

$$\overleftarrow{\mathtt{R}}_0 := \overleftarrow{\mathtt{L}}^{-1} \tag{49}$$

is well-defined and corresponds to the unique variational solution of the stationary problem $\overleftarrow{\mathtt{L}}\, u = v$ in $\overleftarrow{\mathscr{H}}_0^{+}$.

This property reflects the fundamentally dissipative, nonconservative nature of proton transport, where particles are eventually absorbed either spatially or energetically. In the stochastic formulation, this corresponds to finite lifetimes of sample paths, that is, almost sure termination of trajectories after a finite track length.

Lemma 3.3 (Maximality of the BFP Operator). *Let $S, \mu, \sigma_{\mathrm{n}}$, and π satisfy the assumptions of Lemma 3.2. Then, the operators $\overrightarrow{\mathtt{L}} : \overrightarrow{\mathscr{H}}_0^{-} \to \mathscr{V}$ and $\overleftarrow{\mathtt{L}} : \overleftarrow{\mathscr{H}}_0^{+} \to \mathscr{V}$ are maximal monotone. In particular,*

$$\operatorname{range}(I + \overrightarrow{\mathtt{L}}) = \operatorname{range}(I + \overleftarrow{\mathtt{L}}) = \mathscr{V}. \tag{50}$$

Proof. We begin by showing that $\operatorname{range}(I + \overrightarrow{\mathtt{L}})$ is closed in $\mathscr{V}$. Let $\{f_i\} \subset \operatorname{range}(I + \overrightarrow{\mathtt{L}})$ with $f_i \to f$ strongly in $\mathscr{V}$. By definition, for each i, there exists $u_i \in \overrightarrow{\mathscr{H}}_0^{-}$ such that

$$(I + \overrightarrow{\mathtt{L}}) u_i = f_i. \tag{51}$$

Taking the duality pairing with u_i, we obtain

$$\left\langle (I + \overrightarrow{\mathtt{L}}) u_i, u_i \right\rangle = \langle f_i, u_i \rangle. \tag{52}$$

By monotonicity from Lemma 3.2, the left-hand side is bounded below by the graph norm of u_i

$$\left\langle (I+\overrightarrow{\mathrm{L}})u_i, u_i \right\rangle \geq C\|u_i\|^2_{\overrightarrow{\mathscr{H}}} \tag{53}$$

for some constant $C > 0$. Since $\langle f_i, u_i\rangle \leq \|f_i\|_{\mathscr{V}} \|u_i\|_{\overrightarrow{\mathscr{H}}}$, we deduce that $\{u_i\}$ is bounded in $\overrightarrow{\mathscr{H}}_0^-$.

Hence, up to a subsequence, $u_i \rightharpoonup u$ weakly in $\overrightarrow{\mathscr{H}}_0^-$. Since $\overrightarrow{\mathrm{L}}$ is closed (as a sum of closed operators), the limit u satisfies $(I+\overrightarrow{\mathrm{L}})u = f$, showing that $f \in \mathrm{range}(I+\overrightarrow{\mathrm{L}})$. Thus, the range is closed in $\mathscr{V}$. The same argument applies to the adjoint operator $\overleftarrow{\mathrm{L}} : \overleftarrow{\mathscr{H}}_0^+ \to \mathscr{V}$, showing that $\mathrm{range}(I+\overleftarrow{\mathrm{L}})$ is closed as well.

To establish maximality, we invoke duality of monotone operators (see [26, Thm. 7.1]), which implies:

$$\left(\ker(I+\overrightarrow{\mathrm{L}})\right)^{\perp} = \mathrm{range}(I+\overleftarrow{\mathrm{L}}), \qquad \left(\ker(I+\overleftarrow{\mathrm{L}})\right)^{\perp} = \mathrm{range}(I+\overrightarrow{\mathrm{L}}). \tag{54}$$

It remains to show that $\ker(I+\overrightarrow{\mathrm{L}}) = \{0\}$ and similarly for $\overleftarrow{\mathrm{L}}$.

Suppose $u_0 \in \ker(I+\overrightarrow{\mathrm{L}})$, so that $(I+\overrightarrow{\mathrm{L}})u_0 = 0$. Then,

$$\left\langle (I+\overrightarrow{\mathrm{L}})u_0, u_0 \right\rangle = 0. \tag{55}$$

Using the decomposition from Lemma 3.2, we have

$$\begin{aligned} \left\langle (I+\overrightarrow{\mathrm{L}})u_0, u_0 \right\rangle &= \|u_0\|^2_{\mathrm{L}^2(C)} + \|\sqrt{\mu}\nabla_\omega u_0\|^2_{\mathrm{L}^2(C)} + \frac{1}{2}\|u_0\|^2_{\Gamma+} \\ &\quad + \frac{1}{2}\lim_{E\downarrow E_{\min}} \|\sqrt{S}u_0\|^2_{\mathrm{L}^2(D\times\mathbb{S}^2)} - \frac{1}{2}\langle(\partial_E S)u_0, u_0\rangle + \langle(\sigma_{\mathrm{n}} - \mathcal{S})u_0, u_0\rangle. \end{aligned} \tag{56}$$

Each term is nonnegative under the assumptions of Lemma 3.2, and therefore all must vanish. In particular,

$$\|u_0\|_{\mathrm{L}^2(C)} = \|\nabla_\omega u_0\|_{\mathrm{L}^2(C)} = \|u_0\|_{\Gamma+} = 0, \tag{57}$$

and so $u_0 = 0$ in $\overrightarrow{\mathscr{H}}_0^-$.

A similar argument applies to $\overleftarrow{\mathrm{L}}$ using its adjoint structure and monotonicity, yielding $\ker(I+\overleftarrow{\mathrm{L}}) = \{0\}$. Therefore, both operators are maximal monotone. $\square$

Theorem 3.4 (Existence and Uniqueness of Solutions). *Let the assumptions of Lemma 3.2 hold. Then, for each $f \in \mathscr{V}$, there exists a unique solution $u \in \overrightarrow{\mathscr{H}}_0^-$ satisfying*

$$\left\langle \overrightarrow{\mathrm{L}}\, u, v \right\rangle = \langle f, v\rangle, \qquad \forall v \in \overleftarrow{\mathscr{H}}_0^+. \tag{58}$$

Similarly, for each $f \in \mathscr{V}$, there exists a unique $z \in \overleftarrow{\mathscr{H}}_0^+$ such that

$$\left\langle \overleftarrow{\mathrm{L}}\, z, v \right\rangle = \langle f, v\rangle, \qquad \forall v \in \overrightarrow{\mathscr{H}}_0^-. \tag{59}$$

Proof. The result follows from the monotonicity of $\overrightarrow{\mathrm{L}}$ and $\overleftarrow{\mathrm{L}}$ established in Lemma 3.2, and the maximality proven in Lemma 3.3. By the standard theory of maximal monotone operators in Hilbert spaces (see, e.g., [27, §6.3]), the variational problems admit unique solutions $u \in \overrightarrow{\mathscr{H}}_0^-$ and $z \in \overleftarrow{\mathscr{H}}_0^+$. $\square$

4 | Stochastic Formulation: The SDE Approach

We now reformulate the proton transport problem from a probabilistic perspective. Whereas the deterministic formulation models the evolution of particle density via an integro-differential operator, the stochastic formulation describes a Markovian trajectory in the phase space C. This viewpoint models the physical path along which energy is deposited by a single initial proton, accounting for its complete stochastic life cycle, including continuous energy loss (Coulomb interaction), elastic scattering and the possible generation of secondary or tertiary protons at nonelastic nuclear interactions. These secondary particles further transport residual energy (cf. [29]).

This trajectory-based representation naturally expresses dose as an accumulated quantity along individual tracks and aligns with Monte Carlo simulation techniques widely used for dose computation and verification in clinical proton therapy. Mathematically, the stochastic framework yields a dual characterization of the deterministic model, the infinitesimal generator of the process corresponds to the adjoint transport operator and the associated resolvent defines a probabilistic representation of fluence.

This section introduces the stochastic process and its governing SDEs, and establishes its correspondence with the deterministic theory developed earlier.

Unlike conventional SDEs, which are indexed by time, we follow the convention of the nuclear physics literature and parameterize the process by *track length*.

We define the stochastic process

$$Y = (Y_\ell,\ \ell \geq 0) = ((\boldsymbol{x}_\ell, \omega_\ell, E_\ell),\ \ell \geq 0), \tag{60}$$

where $\boldsymbol{x}_\ell \in D$, $\omega_\ell \in \mathbb{S}^2$, and $E_\ell \in \mathcal{E}$ denote the proton's position, direction, and energy at track length ℓ. The process Y describes the trajectory of a single proton evolving in the phase space $C = D \times \mathbb{S}^2 \times \mathcal{E}$.

The evolution continues until a random stopping length Λ defined by

$$\Lambda := \inf\{\ell > 0 : E_\ell = E_{\min} \text{ or } \boldsymbol{x}_\ell \notin D\}, \tag{61}$$

that is, until the proton either exhausts its energy or leaves the spatial domain. The stopping time Λ depends on the initial condition $(\boldsymbol{x}_0, \omega_0, E_0)$, though this dependence is suppressed in notation.

To ensure the process is defined for all $\ell \geq 0$, we extend the state space by introducing a cemetery state $\dagger$, and define

$$Y_\ell := \begin{cases} (\boldsymbol{x}_\ell, \omega_\ell, E_\ell), & \ell < \Lambda, \\ \dagger, & \ell \geq \Lambda. \end{cases} \tag{62}$$

The state $\dagger$ is absorbing, once entered, the process remains there for all subsequent track lengths. To maintain consistency, we adopt the convention that any test function $f : C \cup \{\dagger\} \to \mathbb{R}$ satisfies $f(\dagger) = 0$. This ensures that $\dagger$ contributes nothing to expectations, integrals, or occupation measures, allowing us to work on the extended domain.

4.1 | Stochastic Model for Proton Transport

We now provide the differential form of the stochastic evolution of the proton configuration Y_ℓ, defined on the event $\{\ell < \Lambda\}$. The dynamics are governed by a system of SDEs on C, incorporating both continuous and jump-driven components.

Energy loss is driven by a continuous stopping power term and discrete jumps due to nuclear scattering. Spatial motion proceeds along the current direction ω_ℓ, while angular variation arises from both Brownian diffusion and jump-induced deflections. Angular diffusion is modeled by Brownian motion $B_\ell \in \mathbb{R}^3$, projected onto the tangent space of $\mathbb{S}^2$ via the cross-product $\omega_{\ell-} \wedge dB_\ell$ and correction term $-\omega_{\ell-} d\ell$, ensuring that $\omega_\ell \in \mathbb{S}^2$ is maintained.

Discrete scattering events are described by a random jump measure $N(Y_{\ell-}; d\ell, d\omega', dE')$, with predictable compensator

$$\tilde{N}(Y; d\ell, d\omega', dE') = \sigma_{\mathrm{n}}(Y)\pi(Y; \omega', E')d\omega' dE' d\ell, \qquad \forall Y \in C,\ \omega' \in \mathbb{S}^2,\ E' \in \mathcal{E}, \tag{63}$$

where the scattering components σ_{n}, σ_{e}, σ_{ne}, and the kernel π were introduced in (15).

The resulting system of SDEs governing $Y_\ell = (\boldsymbol{x}_\ell, \omega_\ell, E_\ell)$ is given by

$$
\begin{aligned}
dE_\ell &= -S(Y_{\ell-})d\ell - \int_{\mathcal{E}}\int_{\mathbb{S}^2}(E_{\ell-} - E')N(Y_{\ell-}; d\ell, d\omega', dE'),\\
d\boldsymbol{x}_\ell &= \omega_{\ell-}d\ell, \qquad\qquad (64)\\
d\omega_\ell &= -2\mu(Y_{\ell-})\omega_{\ell-}d\ell + \sqrt{2\mu(Y_{\ell-})}\omega_{\ell-}\wedge dB_\ell + \int_{\mathcal{E}}\int_{\mathbb{S}^2}(\omega' - \omega_{\ell-})N(Y_{\ell-}; d\ell, d\omega', dE').
\end{aligned}
$$

See [30]. With this convention the angular diffusion coefficient in the infinitesimal generator agrees with the coefficient μ in the BFP operator.

Theorem 4.1 (Variational Representation of the Infinitesimal Generator). *Let $Y_\ell = (\boldsymbol{x}_\ell, \omega_\ell, E_\ell)$ be the stochastic process defined by the SDE (64), extended with a cemetery state † after the stopping time Λ, as described above. Let $\overleftarrow{\mathrm{L}} : \overleftarrow{\mathscr{H}} \to \mathscr{V}$ denote the bBFP operator defined in (28). Then, for any $v \in \overleftarrow{\mathscr{H}}_0^{+}$, the function*

$$
u(\ell, X) := \mathbb{E}_X[v(Y_\ell)\mathbf{1}_{\ell<\Lambda}], \qquad \forall X = (\boldsymbol{x}, \omega, E) \in C, \tag{65}
$$

satisfies the Kolmogorov backward equation

$$
\partial_\ell u + \overleftarrow{\mathrm{L}}\, u = 0 \quad \text{in } \mathscr{V}, \qquad u(0) = v. \tag{66}
$$

Moreover, $u \in C([0,\infty); \overleftarrow{\mathscr{H}}) \cap C^1((0,\infty); \mathscr{V})$, and the operator $\overleftarrow{\mathrm{L}}$ generates a strongly continuous contraction semigroup on $\overleftarrow{\mathscr{H}}_0^{+}$.

Proof. Define the operator family

$$
T_\ell v(X) := \mathbb{E}_X[v(Y_\ell)\mathbf{1}_{\ell<\Lambda}], \qquad \forall v \in \overleftarrow{\mathscr{H}}_0^{+},\ X \in C. \tag{67}
$$

The indicator ensures that the process is stopped upon exiting the domain or reaching minimal energy and the semigroup $(T_\ell)_{\ell\geq 0}$ satisfies:

$$
T_0 = I, \quad T_{\ell_1+\ell_2} = T_{\ell_1}T_{\ell_2}, \quad \lim_{\ell\downarrow 0}\|T_\ell v - v\|_{\overleftarrow{\mathscr{H}}} = 0, \quad \forall v \in \overleftarrow{\mathscr{H}}_0^{+}. \tag{68}
$$

Hence, (T_ℓ) is a C_0-semigroup on $\overleftarrow{\mathscr{H}}_0^{+}$, and by the Hille–Yosida theorem [26, §4.4], it admits a densely defined infinitesimal generator $-\overleftarrow{\mathrm{L}}$ on $\overleftarrow{\mathscr{H}}_0^{+}$, given by

$$
-\overleftarrow{\mathrm{L}}\, v := \lim_{\ell\downarrow 0}\frac{T_\ell v - v}{\ell}, \quad \text{in } \mathscr{V}. \tag{69}
$$

By Lemma 3.3, $\overleftarrow{\mathrm{L}}$ agrees with the variational backward BFP operator, and is closed, monotone, and maximal. Setting $u(\ell) := T_\ell v$, we have

$$
u \in C([0,\infty); \overleftarrow{\mathscr{H}}) \cap C^1((0,\infty); \mathscr{V}), \tag{70}
$$

and u satisfies the evolution equation

$$
\partial_\ell u(\ell) + \overleftarrow{\mathrm{L}}\, u(\ell) = 0, \qquad u(0) = v. \tag{71}
$$

In variational form, for all $\phi \in \overrightarrow{\mathscr{H}}_0^{-}$ and almost every $\ell > 0$,

$$
\langle \partial_\ell u(\ell), \phi\rangle + \left\langle \overleftarrow{\mathrm{L}}\, u(\ell), \phi\right\rangle = 0. \tag{72}
$$

This completes the proof. □

Theorem 4.1 provides an $\mathrm{L}^2(\mathcal{C})$ formulation of standard results in stochastic analysis [31, Chapter 8], which are more commonly stated in an $\mathrm{L}^\infty(\mathcal{C})$ setting. In such formulations, the infinitesimal generator of a diffusion process governs the evolution of expectations of test functions via the Kolmogorov backward equation. A derivation of these ideas in the context of kinetic equations and transport theory is given in [32].

In this spirit, when a transition density $p(\ell, X, Z)$ on $[0,\infty)\times\mathcal{C}\times\mathcal{C}$ exists, we may write the semigroup as

$$T_\ell v(X) = \int_{\mathcal{C}} p(\ell, X, Z) v(Z) dZ, \qquad \forall X \in \mathcal{C}. \tag{73}$$

By testing Equation (72) against Dirac sequences, we obtain the weak form of the Kolmogorov backward equation for the transition density

$$\partial_\ell p + \overleftarrow{\mathrm{L}}_X\, p = 0, \quad \text{in } \mathscr{V}, \tag{74}$$

where the subscript X indicates that $\overleftarrow{\mathrm{L}}$ acts on the first spatial argument of $p(\ell, X, Z)$.

In the distributional sense, this formulation enforces the initial condition

$$p(0, X, Z) = \delta_X(Z), \tag{75}$$

as well as the absorbing boundary condition

$$p(\ell, X, Z) = 0, \quad \text{for } X \in \Gamma^+,\ \ell > 0. \tag{76}$$

Theorem 4.2 (Duality of the Semigroups). *Suppose that the transition probability density $p(\ell, X, Z)$ of the process Y_ℓ exists with respect to the base measure on C. Then, for fixed $X \in C$, the map $Z \mapsto p(\ell, X, Z)$ satisfies the Kolmogorov forward (Fokker–Planck) equation*

$$\partial_\ell p + \overrightarrow{\mathrm{L}}_Z\, p = 0, \quad \textit{in } \mathscr{V}, \tag{77}$$

with initial condition

$$p(0, X, Z) = \delta_X(Z), \tag{78}$$

and absorbing boundary condition

$$p(\ell, X, Z) = 0, \quad \textit{for } Z \in \Gamma^+,\ \ell > 0. \tag{79}$$

Proof. Since both $\overleftarrow{\mathrm{L}}$ and $\overrightarrow{\mathrm{L}}$ are densely defined, closed, monotone, and maximal on $\overleftarrow{\mathscr{H}}_0^{+}$ and $\overrightarrow{\mathscr{H}}_0^{-}$, respectively, they generate strongly continuous semigroups $\exp(-s\overleftarrow{\mathrm{L}})$ and $\exp(-s\overrightarrow{\mathrm{L}})$; see [33, 34]. These semigroups satisfy the duality identity

$$\left\langle \psi, \exp(-s\overleftarrow{\mathrm{L}})\phi \right\rangle = \left\langle \exp(-s\overrightarrow{\mathrm{L}})\psi, \phi \right\rangle, \qquad \forall \phi \in \overleftarrow{\mathscr{H}}_0^{+},\ \psi \in \overrightarrow{\mathscr{H}}_0^{-}. \tag{80}$$

This follows by expanding the exponential as a power series in s, applying (27), and noting that the boundary conditions are preserved under the domain definitions of $\overrightarrow{\mathrm{L}}$ and $\overleftarrow{\mathrm{L}}$.

Assuming a transition density $p(\ell, X, Z)$ exists, for $\psi \in \overrightarrow{\mathscr{H}}_0^{-}$, by differentiating $\ell \mapsto \langle \psi, \exp(-\ell\overleftarrow{\mathrm{L}})\phi\rangle$, we note that, on the one hand,

$$\langle \psi, \exp(-\ell\overleftarrow{\mathrm{L}})\phi\rangle = \int_C \int_C \psi(X) p(\ell, X, Z) \phi(Z) dX dZ. \tag{81}$$

On the other hand, applying (80) and differentiating again with respect to ℓ tells us that

$$\partial_\ell p(\ell, X, Z) + \overrightarrow{\mathrm{L}}_Z\, p(\ell, X, Z) = 0 \tag{82}$$

in the distributional sense. The initial condition follows from $T_0 = I$, and the boundary condition $p(\ell, X, Z) = 0$ for $Z \in \Gamma^+$ and $\ell > 0$ reflects the absorption of the forward process at the outflow boundary.

The value $p(\ell, X, X)$ may be strictly positive for $\ell > 0$ depending on the recurrence properties of the process and does not vanish in general. □

The adjoint relationship between $\overleftarrow{\mathtt{L}}$ and $\overrightarrow{\mathtt{L}}$ provides a rigorous explanation for why the deterministic BFP equation governs proton fluence in phase space. The backward operator $\overleftarrow{\mathtt{L}}$ generates the evolution of expectations of path-dependent functionals, while its adjoint $\overrightarrow{\mathtt{L}}$ governs the forward evolution of proton densities and fluence.

This distinction underpins the roles of the two formulations in proton radiotherapy. The backward equation is used to compute expected observables, such as dose or track-weighted quantities, by propagating initial values backward along the stochastic trajectories. In contrast, the forward equation propagates phase-space distributions and is naturally aligned with PDE-based treatment planning.

This duality forms the mathematical foundation for hybrid computational approaches that combine Monte Carlo simulation with deterministic transport solvers. It allows us to translate between pathwise and density-based representations, a capability we will exploit in the optimization framework developed in the following sections.

5 | Deterministic and Stochastic Resolvents

In the deterministic framework, the operators $\overrightarrow{\mathtt{L}} : \overrightarrow{\mathscr{H}}_0^- \to \mathscr{V}$ and $\overleftarrow{\mathtt{L}} : \overleftarrow{\mathscr{H}}_0^+ \to \mathscr{V}$ are unbounded, closed, and maximal monotone. Their resolvent operators are defined for $\lambda \geq 0$ by

$$\overrightarrow{\mathtt{R}}_\lambda := (\lambda I + \overrightarrow{\mathtt{L}})^{-1}, \qquad \overleftarrow{\mathtt{R}}_\lambda := (\lambda I + \overleftarrow{\mathtt{L}})^{-1}, \tag{83}$$

and correspond to variational solutions of the stationary equations

$$\lambda u + \overleftarrow{\mathtt{L}}\, u = v, \qquad u \in \overleftarrow{\mathscr{H}}_0^+, \tag{84}$$

$$\lambda \phi + \overrightarrow{\mathtt{L}}\, \phi = v, \qquad \phi \in \overrightarrow{\mathscr{H}}_0^- . \tag{85}$$

These equations are well-posed even when $\lambda = 0$, due to the strict monotonicity of $\overleftarrow{\mathtt{L}}$ and $\overrightarrow{\mathtt{L}}$, as established in Lemma 3.2, and the maximality proven in Lemma 3.3. This ensures that $\overrightarrow{\mathtt{R}}_\lambda$ and $\overleftarrow{\mathtt{R}}_\lambda$ are everywhere-defined bounded linear operators $\mathscr{V} \to \overrightarrow{\mathscr{H}}_0^-$ and $\mathscr{V} \to \overleftarrow{\mathscr{H}}_0^+$, respectively.

Equation (84) represents the expected cumulative contribution of a source term v along stochastic proton tracks, and admits a probabilistic interpretation as discussed below. Equation (85) describes the propagation of fluence to a point $X \in \mathcal{C}$ from a specified inflow distribution. For homogeneous boundary conditions, the forward problem is defined in $\overrightarrow{\mathscr{H}}_0^-$, which corresponds to zero incoming fluence on Γ^-. But, in general, inhomogeneous boundary data $\mathrm{g} \in \mathrm{tr}(\Gamma^-)$ must be lifted into $\overrightarrow{\mathscr{H}}$ and subtracted out, as explained in Section 3.1.

5.0.0.1 | Stochastic Interpretation of the Backward Resolvent. In the stochastic formulation, the resolvent $\overleftarrow{\mathtt{R}}_\lambda$ of the generator $\overleftarrow{\mathtt{L}}$ has a natural interpretation as the Laplace transform of the semigroup generated by the SDE (64). For $v \in \mathscr{V}$ and $X \in \mathcal{C}$,

$$\overleftarrow{\mathtt{R}}_\lambda\, [v](X) = \mathbb{E}_X\left[\int_0^\infty \mathrm{e}^{-\lambda \ell} v(Y_\ell) \mathbf{1}_{\ell < \Lambda} d\ell\right], \tag{86}$$

where Λ is the stopping time at which the particle exits the spatial domain or reaches minimal energy. This expression represents the λ-discounted mean occupation time of the process Y_ℓ weighted by the test function v.

This interpretation provides a direct connection between the deterministic resolvent equation and the sampling structure of time-discounted Monte Carlo estimates. In particular, the representation of $\overleftarrow{\mathtt{R}}_\lambda\, v$ as a Laplace transform of the expected occupation integral aligns closely with classical results in stochastic analysis, where resolvent operators are understood as expectations of occupation measures over the process lifetime.

For any bounded measurable function v vanishing on the cemetery state †, the backward resolvent has a probabilistic representation

$$\overleftarrow{\mathrm{R}}_{\lambda}[v](X) = \mathbb{E}_X\left[\int_0^\infty \mathrm{e}^{-\lambda\ell} v(Y_\ell)\mathbf{1}_{\ell<\Lambda} d\ell\right], \qquad X \in C, \tag{87}$$

where Λ is the stopping time at which the process exits the domain or reaches the minimum energy $E_{\min}$. The integrand vanishes for $\ell \geq \Lambda$ due to the cemetery convention $v(\dagger) = 0$, and the representation automatically enforces the boundary condition $\overleftarrow{\mathrm{R}}_{\lambda}[v] = 0$ on Γ^+.

Since the transport process is nonconservative, the integral remains finite even when $\lambda = 0$. The unregularized resolvent

$$\overleftarrow{\mathrm{R}}[v](X) = \mathbb{E}_X\left[\int_0^\infty v(Y_\ell)\mathbf{1}_{\ell<\Lambda} d\ell\right] \tag{88}$$

is therefore well-defined and describes the expected total occupation time in configuration space. This coincides with the variational solution to $\overleftarrow{\mathrm{L}}\, u = v$ with $u|_{\Gamma^+} = 0$, justified analytically by the maximal monotonicity of $\overleftarrow{\mathrm{L}}$.

Theorem 5.1 (Existence of the Resolvent Density via Hypoellipticity). *Suppose that the stochastic process $Y_\ell = (\boldsymbol{x}_\ell, \omega_\ell, E_\ell)$ is Malliavin nondegenerate in the sense that its Malliavin covariance matrix is almost surely invertible. In addition, assume that the transition probabilities of Y_ℓ are absolutely continuous with respect to the Lebesgue measure on C. Then, for $\lambda \geq 0$, the resolvent measure*

$$\overleftarrow{\mathrm{R}}_{\lambda}(X; d\boldsymbol{x}' d\omega' dE'), \qquad X = (\boldsymbol{x}, \omega, E) \in C, \tag{89}$$

admits a density $\mathrm{r}_\lambda(X, (\boldsymbol{x}', \omega', E'))$ such that

$$\overleftarrow{\mathrm{R}}_{\lambda}(X; d\boldsymbol{x}' d\omega' dE') = \mathrm{r}_\lambda(X, (\boldsymbol{x}', \omega', E'))d\boldsymbol{x}' d\omega' dE'. \tag{90}$$

As a consequence, the resolvent operator admits the kernel representation

$$\overleftarrow{\mathrm{R}}_{\lambda}[v](X) = \int_C \mathrm{r}_\lambda(X, Z)v(Z)dZ, \qquad X \in C \tag{91}$$

for any $v \in \mathrm{L}^1(C)$. In particular, the formula holds for $v \in \mathrm{L}^2(C)$, which is the natural setting in the variational framework.

Proof. By definition, the resolvent measure satisfies

$$\overleftarrow{\mathrm{R}}_{\lambda}(X; A) = \mathbb{E}_X\left[\int_0^\infty \mathbf{1}_{Y_\ell \in A}\mathbf{1}_{\ell<\Lambda}\mathrm{e}^{-\lambda\ell} d\ell\right], \qquad X \in C \tag{92}$$

for any measurable set $A \subset C$. This expresses the expected occupation time of the process Y_ℓ in the set A, discounted by $\mathrm{e}^{-\lambda\ell}$ and stopped at the exit time Λ.

Under the Malliavin nondegeneracy assumption, the process Y_ℓ admits a smooth transition density $p(\ell, X, Z)$ for each $\ell > 0$, satisfying

$$\mathbb{P}_X(Y_\ell \in A,\ \ell < \Lambda) = \int_A p(\ell, X, Z)dZ \tag{93}$$

for all measurable $A \subset C$. The existence and regularity of p follow from Hörmander's theorem, owing to the angular diffusion term in the generator $\overleftarrow{\mathrm{L}}$, which ensures hypoellipticity [35, 36].

Substituting this into the resolvent expression, we obtain

$$\overleftarrow{\mathrm{R}}_{\lambda}(X; dZ) = \int_0^\infty \mathrm{e}^{-\lambda\ell} p(\ell, X, Z)d\ell\, dZ. \tag{94}$$

Hence, by Fubini's theorem, we define the density

$$\mathtt{r}_\lambda(X,Z) := \int_0^\infty \mathrm{e}^{-\lambda\ell} p(\ell,X,Z) d\ell. \tag{95}$$

For any $v \in \mathrm{L}^1(\mathcal{C})$, we obtain

$$\overleftarrow{\mathtt{R}}_\lambda [v](X) = \int_\mathcal{C} \mathtt{r}_\lambda(X,Z) v(Z) dZ. \tag{96}$$

This proves the existence of a jointly measurable integral kernel $\mathtt{r}_\lambda$ and the corresponding kernel representation of the resolvent operator. If higher regularity is required, one may take $\phi \in \mathrm{H}^k(\mathcal{C})$. $\square$

To make a formal connection between the resolvent kernel $\mathtt{r}_\lambda$ and the elliptic problem (48), we begin with the backward Kolmogorov equation (74) for the transition density $p(\ell,X,Z)$ and its exponentially weighted form. Define $q_\lambda(\ell,X,Z) := \mathrm{e}^{-\lambda\ell} p(\ell,X,Z)$. Then

$$\partial_\ell q_\lambda(\ell,X,Z) = -\lambda \mathrm{e}^{-\lambda\ell} p(\ell,X,Z) + \mathrm{e}^{-\lambda\ell}\partial_\ell p(\ell,X,Z) = -\lambda q_\lambda(\ell,X,Z) - \overleftarrow{\mathtt{L}}_X\, q_\lambda(\ell,X,Z). \tag{97}$$

Hence, q satisfies the evolution equation

$$\partial_\ell q_\lambda(\ell,X,Z) + (\lambda I + \overleftarrow{\mathtt{L}}_X) q_\lambda(\ell,X,Z) = 0 \tag{98}$$

in the distributional sense. Integrating over $\ell \in (0,\infty)$ and using

$$\mathtt{r}_\lambda(X,Z) = \int_0^\infty q_\lambda(\ell,X,Z) d\ell, \tag{99}$$

we obtain

$$\int_0^\infty \partial_\ell q_\lambda(\ell,X,Z) d\ell + (\lambda I + \overleftarrow{\mathtt{L}}_X)\mathtt{r}_\lambda(X,Z) = \lim_{\ell\to\infty} q_\lambda(\ell,X,Z) - q_\lambda(0,X,Z) + (\lambda I + \overleftarrow{\mathtt{L}}_X)\mathtt{r}_\lambda(X,Z). \tag{100}$$

Since the process Y_ℓ is dissipative and $p(\ell,X,Z) \to 0$ as $\ell \to \infty$, we conclude

$$(\lambda I + \overleftarrow{\mathtt{L}}_X)\mathtt{r}_\lambda(X,Z) = \delta_X(Z). \tag{101}$$

This shows that the resolvent kernel $\mathtt{r}_\lambda(X,Z)$ is the Green's function for the elliptic operator $(\lambda I + \overleftarrow{\mathtt{L}}_X)$, subject to the natural terminal condition $\mathtt{r}_\lambda(X,Z) = 0$ for $X \in \Gamma^+$.

The reader will also note that, providing $S > 0$, the solution to the SDE (74) always immediately moves away from its initial state due to depleting energy and kinetic motion. This means that we additionally have in this setting that $r_\lambda(X,Z) = 0$ for $Z \in \Gamma^-$.

A similar argument replacing the role of $q_\lambda(\ell,X,Z)$ in (97) by $\int_\mathcal{C} q_\lambda(\ell,X,Z) v(X) dZ$ shows that

$$(\lambda I + \overleftarrow{\mathtt{L}}_X)\, \overleftarrow{\mathtt{R}}_\lambda [v](X) = v(X) \tag{102}$$

with $\overleftarrow{\mathtt{R}}_\lambda [v](X) = 0$ for $X \in \Gamma^+$. This also justifies the preemptive choice of notation $\overleftarrow{\mathtt{R}}_\lambda := (\lambda I + \overleftarrow{\mathtt{L}})^{-1}$.

While the parameter $\lambda > 0$ traditionally plays the role of an exponential killing rate, appearing as the Laplace dual variable in resolvent representations of the semigroup, it is natural in this setting to consider the case $\lambda = 0$. Due to the nonconservative nature of the underlying stochastic process (which terminates at finite track length with probability one), the resolvent remains well-defined even when $\lambda = 0$. This case corresponds to unweighted occupation time and captures the true lifetime-integrated response of the transport system without artificial discounting.

Next, define for $\lambda \geq 0$,

$$\overrightarrow{\mathtt{R}}_\lambda [\phi](Z) = \int_\mathcal{C} \phi(X)\mathtt{r}_\lambda(X,Z) dX \tag{103}$$

for any $\phi \in \mathrm{L}^1(C)$. An analogous argument applied to the forward equation (77) yields

$$(\lambda I + \overrightarrow{\mathrm{L}}_Z)\mathrm{r}_\lambda(X, Z) = \delta_X(Z) \tag{104}$$

subject to the condition $\mathrm{r}_\lambda(X, Z) = 0$ for $Z \in \Gamma^+$, as inherited from (79). In both (104) and (101), the distributional delta on the right-hand side encodes the identity action of the resolvent on test functions. The adjoint structure ensures consistency between the forward and backward Green's functions. As with (102), we can similarly show that

$$(\lambda I + \overleftarrow{\mathrm{L}}_Z)\, \overrightarrow{\mathrm{R}}_\lambda\, [\phi](Z) = \phi(Z) \tag{105}$$

with $\overrightarrow{\mathrm{R}}_\lambda\, [\phi](Z) = 0$ for $Z \in \Gamma^-$. Again, we are led to understand our preemptive choice of notation $\overrightarrow{\mathrm{R}}_\lambda := (\lambda I + \overrightarrow{\mathrm{L}})^{-1}$. If we enforce $v \in \overleftarrow{\mathscr{H}}_0^{+} \cap \mathscr{V}$ and $\phi \in \overrightarrow{\mathscr{H}}_0^{-} \cap \mathscr{V}$, then we see the link between (84) and (102), as well as (85) and (105). Moreover,

$$\langle \phi, \overleftarrow{\mathrm{R}}_\lambda\, [v] \rangle = \langle \phi, (\lambda I + \overleftarrow{\mathrm{L}})^{-1} v \rangle = \langle (\lambda I + \overrightarrow{\mathrm{L}})^{-1} \phi, v \rangle = \langle \overrightarrow{\mathrm{R}}_\lambda\, [\phi], v \rangle, \tag{106}$$

where we have used the duality of $\overleftarrow{\mathrm{L}}$ and $\overrightarrow{\mathrm{L}}$, showing the natural duality of the operators $\overleftarrow{\mathrm{R}}_\lambda$ and $\overrightarrow{\mathrm{R}}_\lambda$, for $\lambda \geq 0$.

Remark 5.2 (Resolvent and kernel extension to Γ^-). We note that, in either forward or backward form, $r_\lambda(X, Z)$ should be thought of as the (λ-discounted if $\lambda > 0$) density of fluence incoming at Z generated by an infinitesimal outgoing source at X, informally a "flux capacitor" accumulating expected flux along tracks. Importantly, we also note that many of the above identities can be extended by allowing $X \in \Gamma^-$. Indeed $r_\lambda(X, Y)$ is perfectly well defined for $X \in \Gamma^-$, thanks to the directional nature of the SDE (64). That is, we can initiate the SDE from the configuration $X \in \Gamma^-$ and it will immediately enter C and spend a positive amount of time before exiting on Γ^+; as such the semigroup (65) and hence the associated resolvent (88) are well defined. Extending to $X \mapsto r_\lambda(X, Y)$ to the boundary Γ^- also allows us to extend the notion of $\overrightarrow{\mathrm{R}}_\lambda\, [\phi]$ in (103) to accommodate for functions ϕ which are supported on the entrance boundary Γ^-.

The extension to the inflow boundary can be explained as consistent with the formulation of

$$\lambda\psi + \overrightarrow{\mathrm{L}}\, \psi = \phi, \qquad \phi \in \mathscr{V}. \tag{107}$$

This is done by writing

$$\psi = u + \hat{\phi} \qquad \text{such that} \qquad \lambda u + \overrightarrow{\mathrm{L}}\, u = -(\lambda + \overrightarrow{\mathrm{L}})\hat{g} \quad \text{in } C \tag{108}$$

such that $\hat{\phi} = \phi$ on Γ^- and $u \in \overrightarrow{\mathscr{H}}_0^{-}$, in particular, $u = 0$ on Γ^-.

5.0.0.2 | Dose Computation. In both the stochastic and deterministic settings, the dose represents the total energy deposited per unit mass in the absorbing medium. The fundamental relation governing dose is the classical fluence-dose formula

$$\text{dose} = \text{fluence} \times \text{mass stopping power}. \tag{109}$$

In our formulation, this identity emerges naturally from both the stochastic and PDE-based perspectives, via a common dose operator acting on a notion of fluence.

5.1 | Dose in the Stochastic Framework

We are interested in viewing dose from a purely spatial perspective. If $S(\boldsymbol{x}, E)$ is the stopping power and $\rho(\boldsymbol{x})$ is the mass density, then the expected contribution to dose from a single proton starting at $X \in \Gamma^-$, as observed through a test function v on C, is

$$\overleftarrow{\mathrm{R}}\, [vS/\rho](X) = \int_C \mathrm{r}_0(X, (\boldsymbol{x}, \omega, E)) v(\boldsymbol{x}, \omega, E) \frac{S(\boldsymbol{x}, E)}{\rho(\boldsymbol{x})} d\boldsymbol{x} d\omega dE, \tag{110}$$

where we have taken advantage of Remark 5.2. If there is inflow boundary data g on Γ^-, then the dose profile is given by

$$\langle g, \overleftarrow{\mathtt{R}}\,[vS/\rho]\rangle = \int_C \mathtt{r}_0^{(g)}(\boldsymbol{x},\omega,E)v(\boldsymbol{x},\omega,E)\frac{S(\boldsymbol{x},E)}{\rho(\boldsymbol{x})}d\boldsymbol{x}d\omega dE, \tag{111}$$

where

$$\mathtt{r}_0^{(g)}(\boldsymbol{x},\omega,E) = \int_{\Gamma^-} \mathtt{g}(X)\mathtt{r}_0(X,(\boldsymbol{x},\omega,E))dX \tag{112}$$

so that the pointwise dose density is

$$\mathtt{r}_0^{(g)}(\boldsymbol{x},\omega,E)\frac{S(\boldsymbol{x},E)}{\rho(\boldsymbol{x})}, \tag{113}$$

thus conforming to the paradigm that

$$\text{dose} = \text{fluence} \times \text{mass stopping power}. \tag{114}$$

Finally, the spatial dose $\boldsymbol{x} \mapsto \mathcal{D}(\boldsymbol{x})$ received by the medium at position $\boldsymbol{x}$ is given by marginalizing over angle and energy

$$\mathcal{D}(\boldsymbol{x}) = \int_{\mathcal{E}} \frac{S(\boldsymbol{x},E)}{\rho(\boldsymbol{x})}\int_{\mathbb{S}^2} \mathtt{r}_0^{(g)}(\boldsymbol{x},\omega,E)d\omega dE. \tag{115}$$

5.2 | Dose in the Deterministic Framework

In the deterministic setting, the fluence ψ satisfies the forward BFP equation

$$\overrightarrow{\mathtt{L}}\,\psi = 0 \quad \text{in } C, \qquad \gamma^-\psi = g \quad \text{on } \Gamma^-, \tag{116}$$

where g is the prescribed inflow trace. Equivalently, by using the lifting construction described above, this boundary-driven problem may be rewritten as an interior problem for $u = \psi - \widehat{g}$ with homogeneous inflow condition.
The dose is then obtained by integrating the fluence against the energy loss per unit mass

$$\mathcal{D}(\boldsymbol{x}) = \int_{\mathcal{E}} \frac{S(\boldsymbol{x},E)}{\rho(\boldsymbol{x})}\int_{\mathbb{S}^2} \psi(\boldsymbol{x},\omega,E)d\omega dE. \tag{117}$$

In differential form, the dose density is

$$\psi(\boldsymbol{x},\omega,E)\frac{S(\boldsymbol{x},E)}{\rho(\boldsymbol{x})}, \tag{118}$$

reaffirming the fundamental relation

$$\text{dose} = \text{fluence} \times \text{mass stopping power}. \tag{119}$$

The two calculations for spatial dose are equivalent because, as per (105) with $\lambda = 0$, the solution to (116) is identified as

$$\psi(\boldsymbol{x},\omega,E) = \overrightarrow{\mathtt{R}}_\lambda\,[\mathtt{g}](\boldsymbol{x},\omega,E) = \int_{\Gamma^-} \mathtt{g}(X)\mathtt{r}_0(X,(\boldsymbol{x},\omega,E))dX = \mathtt{r}_0^{(g)}(\boldsymbol{x},\omega,E), \tag{120}$$

where we have again taken advantage of Remark 5.2, and hence (118) agrees with (113).

6 | Hybrid SDE–PDE Optimization for Treatment Planning

In this section, we propose a proof-of-concept framework for treatment planning in proton therapy that leverages the dual stochastic-deterministic formulation developed in this work. Our goal is to show how the SDE model of individual proton transport and the PDE-based variational formulation can be combined to guide algorithm design and theoretical analysis under uncertainty.

The stochastic formulation captures the physical randomness of elastic scattering, energy loss, and nonelastic interactions through an SDE for the full proton life cycle, simulated via Monte Carlo methods. The deterministic framework, by contrast, provides a variational structure suited to adjoint-based sensitivity analysis and optimization, via the forward BFP equation and its resolvent.

Our hybrid strategy exploits this duality, dose is computed using high-fidelity stochastic simulations for fixed configurations, while gradients are evaluated efficiently using deterministic adjoint PDEs. This enables scalable control under uncertainty while preserving physical accuracy.

Uncertainty in treatment planning arises not only from particle-scale randomness but also from anatomical and physiological variability, modeled by a parameter θ. For each realization θ, the stochastic and deterministic models agree; however, the optimization must be robust across the distribution of θ. This induces a nested structure, the inner expectation captures particle variability for fixed θ, while the outer expectation averages dose over scenario uncertainty.

This structure motivates a class of hybrid algorithms combining forward Monte Carlo solvers with deterministic adjoint methods. Our formulation aligns with recent efforts to incorporate uncertainty into radiotherapy planning [37], while remaining compatible with classical constrained optimization approaches such as [38].

6.1 | Problem Formulation

We aim to determine an optimal proton beam intensity distribution $g : \Gamma^- \to \mathbb{R}_+$, prescribing the number of protons injected at each position $\boldsymbol{x}$, direction ω, and energy E on the inflow boundary of phase space. The goal is to select g such that the resulting dose approximates a prescribed clinical target while respecting physical and biological constraints.

The delivered fluence ψ_g^θ depends not only on g, but also on a random parameter $\theta \in \Theta$, which models uncertainty in anatomical and physical characteristics. These include patient-specific variations in tissue density $\rho^\theta(\boldsymbol{x})$, stopping power $S^\theta(\boldsymbol{x}, E)$, and domain geometry.

For each scenario θ, the forward transport problem yields a fluence ψ_g^θ, from which dose is computed via the scenario-dependent operator

$$\mathcal{D}_\theta \left[\psi_g^\theta\right](\boldsymbol{x}) = \int_{\mathcal{E}} \frac{S^\theta(\boldsymbol{x}, E)}{\rho^\theta(\boldsymbol{x})} \int_{\mathbb{S}^2} \psi_g^\theta(\boldsymbol{x}, \omega, E)\, d\omega dE. \tag{121}$$

As θ is random, we consider the expected dose

$$\mathbf{E}_\theta\left[\mathcal{D}_\theta \left[\psi_g^\theta\right](\boldsymbol{x})\right] = \int_\Theta \mathcal{D}_\theta \left[\psi_g^\theta\right](\boldsymbol{x}) \mathrm{d}\mathbb{P}(\theta). \tag{122}$$

In general, the dose operator does not commute with expectation:

$$\mathbf{E}_\theta\left[\mathcal{D}_\theta \left[\psi_g^\theta\right]\right] \neq \mathcal{D}\left[\mathbf{E}_\theta \left[\psi_g^\theta\right]\right]. \tag{123}$$

This noncommutativity arises because both ψ_g^θ and the operator $\mathcal{D}_\theta$ depend on θ.

Consequently, the expected dose must be approximated using Monte Carlo sampling. Drawing scenarios $\{\theta^{(i)}\}_{i=1}^N$, solving the transport equation for each $\theta^{(i)}$, computing the corresponding dose and averaging yields an empirical estimate of $\mathbf{E}_\theta[\mathcal{D}_\theta[\psi_g^\theta]]$. This structure underpins robust treatment planning across uncertain physical and anatomical parameters [16].

We impose box constraints on the control to reflect physical bounds on beam delivery

$$0 \le g(\boldsymbol{x}, \omega, E) \le g_{\max}(\boldsymbol{x}, \omega, E), \qquad \text{a.e. on } \Gamma^-, \tag{124}$$

where $g_{\max} \in L^\infty(\Gamma^-)$ encodes machine-specific limits. The admissible control set is

$$\mathcal{U}_{\mathrm{ad}} := \left\{ g \in L^2(\Gamma^-) \,\middle|\, 0 \le g \le g_{\max} \text{ a.e. on } \Gamma^- \right\}, \tag{125}$$

a closed, convex subset of $L^2(\Gamma^-)$.

In the deterministic setting, the dose depends only on g, and we define the cost functional

$$\mathcal{J}(g;\theta) := \int_D w(\boldsymbol{x})\Big(\mathcal{D}_\theta\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{\text{prescribed}}(\boldsymbol{x})\Big)^2 d\boldsymbol{x} + \frac{\alpha}{2}\|g\|^2_{\mathrm{L}^2(\Gamma^-)}, \tag{126}$$

where $d_{\text{prescribed}}$ is the target dose, $w(\boldsymbol{x})$ encodes clinical priorities (e.g., tumor vs. healthy tissue) and $\alpha > 0$ is a regularization parameter penalizing large intensities.

When uncertainty is explicitly modeled, both ψ_g^θ and $\mathcal{D}_\theta$ depend on the random parameter $\theta \in \Theta$. The expected cost becomes

$$\mathbf{E}_\theta[\mathcal{J}(g;\theta)] := \mathbf{E}_\theta\left[\int_D w(\boldsymbol{x})\Big(\mathcal{D}_\theta\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{\text{prescribed}}(\boldsymbol{x})\Big)^2 d\boldsymbol{x}\right] + \frac{\alpha}{2}\|g\|^2_{\mathrm{L}^2(\Gamma^-)}. \tag{127}$$

This formulation corresponds to risk-neutral optimization, the control g is deterministic, but the objective penalizes the *expected* mismatch between delivered and prescribed dose. Alternative strategies, such as penalizing variance or minimizing worst-case deviation, lie within the broader class of robust optimization. See [16] for a comparison of such models.

Lemma 6.1 (Interchange of Gâteaux Derivative and Expectation). *Let Θ be a probability space and suppose that for each $\theta \in \Theta$, the map $g \mapsto \mathcal{J}(g;\theta)$ is Gâteaux differentiable on a Banach space $\mathcal{X}$ with derivative $\mathcal{J}'(g;\theta) \in \mathcal{X}^*$. Assume that*

1. *For each $g \in \mathcal{X}$ and $h \in \mathcal{X}$, the map $\theta \mapsto \langle \mathcal{J}'(g;\theta), h\rangle$ is measurable;*
2. *There exists an integrable function $M \in L^1(\Theta)$ such that for all g in a neighborhood of g_0 and all unit vectors $h \in \mathcal{X}$,*

$$\left|\langle \mathcal{J}'(g;\theta), h\rangle\right| \le M(\theta) \qquad \textit{for a.e. } \theta \in \Theta. \tag{128}$$

$$\left\langle \mathcal{J}'(g), h\right\rangle := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathbf{E}_\theta[\mathcal{J}(g+\varepsilon h;\theta)] = \mathbf{E}_\theta\left[\left\langle \mathcal{J}'(g;\theta), h\right\rangle\right]. \tag{129}$$

Then, the expected functional $\mathbf{E}_\theta[\mathcal{J}(g;\theta)]$ is Gâteaux differentiable, and its derivative satisfies

The box constraints $0 \le g \le g_{\max}$ reflect physical limitations on beam intensity delivery imposed by the accelerator hardware. These constraints define a closed and convex admissible set $\mathcal{U}_{\text{ad}} \subset \mathrm{L}^2(\Gamma^-)$, which is weakly sequentially closed. This structural property ensures the existence of minimizers via the direct method in the Calculus of Variations.

In the optimality conditions, the box constraints give rise to a variational inequality that characterizes admissible descent directions. Equivalently, the gradient step for g must be projected onto $\mathcal{U}_{\text{ad}}$, leading to a projected gradient or semismooth update rule. This formulation naturally accommodates practical control limits within the optimization algorithm.

Theorem 6.2 (Existence of an Optimal Control). *Let $\alpha > 0$, and suppose the stopping power $S \in W^{1,\infty}(D \times \mathcal{E})$, the density $\rho \in L^\infty(D;(0,\infty))$, and the spatial weighting function $w \in L^\infty(D)$. Assume*

1. *For each $g \in \mathcal{U}_{\text{ad}}$, the stochastic forward problem admits a unique fluence $\psi_g^\theta \in \vec{\mathcal{H}}$ for almost every $\theta \in \Theta$;*
2. *The mapping $g \mapsto \psi_g^\theta$ is linear and continuous from $\mathrm{L}^2(\Gamma^-)$ into $\vec{\mathcal{H}}$, uniformly in θ;*
3. *The dose operator $\mathcal{D} : \vec{\mathcal{H}} \to L^2(D)$ is linear and bounded;*
4. *For all $g \in \mathcal{U}_{\text{ad}}$, the integrand*

$$\theta \mapsto \left\|\mathcal{D}\left[\psi_g^\theta\right] - d_{\text{prescribed}}\right\|^2_{L^2(D;w)} \tag{130}$$

is integrable over Θ.

$$\mathbf{E}_\theta[\mathcal{J}(g)] := \mathbf{E}_\theta\left[\left\|\mathcal{D}\left[\psi_g^\theta\right] - d_{\text{prescribed}}\right\|^2_{L^2(D;w)}\right] + \frac{\alpha}{2}\|g\|^2_{\mathrm{L}^2(\Gamma^-)} \tag{131}$$

Then, the expected cost functionalis weakly lower semicontinuous on $\mathrm{L}^2(\Gamma^-)$ and admits at least one minimizer $g^\star \in \mathcal{U}_{\text{ad}}$.

Proof. Let $\{g_n\} \subset \mathcal{U}_{\text{ad}}$ be a minimizing sequence

$$\lim_{n\to\infty} \mathbf{E}_\theta[\mathcal{J}(g_n)] = \inf_{g\in\mathcal{U}_{\text{ad}}} \mathbf{E}_\theta[\mathcal{J}(g)]. \tag{132}$$

Since $\mathcal{U}_{\text{ad}}$ is convex, closed, and bounded in $\mathrm{L}^2(\Gamma^-)$, it is weakly sequentially compact. By the Banach–Alaoglu theorem, there exists a subsequence (still denoted g_n) and a weak limit $g^\star \in \mathcal{U}_{\text{ad}}$ such that

$$g_n \rightharpoonup g^\star \quad \text{in } \mathrm{L}^2(\Gamma^-). \tag{133}$$

By the linearity and continuity of $g \mapsto \psi_g^\theta$ and of $\mathcal{D}$, the composition $g \mapsto \mathcal{D}[\psi_g^\theta]$ is linear and continuous from $\mathrm{L}^2(\Gamma^-)$ to $L^2(D)$ for each θ. Therefore, the map

$$g \mapsto \left\| \mathcal{D}\left[\psi_g^\theta\right] - d_{\text{prescribed}} \right\|_{L^2(D;w)}^2 \tag{134}$$

is convex and continuous, hence weakly lower semicontinuous.

Now, the function

$$f_n(\theta) := \left\| \mathcal{D}\left[\psi_{g_n}^\theta\right] - d_{\text{prescribed}} \right\|_{L^2(D;w)}^2 \tag{135}$$

satisfies $f_n(\theta) \to f(\theta)$ pointwise a.e. in θ, where $f(\theta) := \left\| \mathcal{D}[\psi_{g^\star}^\theta] - d_{\text{prescribed}} \right\|^2$. Since each $f_n(\theta) \geq 0$ and $\mathbf{E}_\theta[f_n]$ is bounded, Fatou's lemma yields

$$\mathbf{E}_\theta[f] \leq \liminf_{n\to\infty} \mathbf{E}_\theta[f_n]. \tag{136}$$

Thus, the expected fidelity term is weakly lower semicontinuous.

The regularization term $\frac{\alpha}{2}\|g\|_{\mathrm{L}^2}^2$ is convex and strongly continuous, hence also weakly lower semicontinuous. Summing both contributions gives

$$\mathbf{E}_\theta[\mathcal{J}(g^\star)] \leq \liminf_{n\to\infty} \mathbf{E}_\theta[\mathcal{J}(g_n)], \tag{137}$$

and $g^\star \in \mathcal{U}_{\text{ad}}$ is a minimizer. □

Remark 6.3 (Practical meaning of the hypotheses). The assumptions in the preceding theorem are standard sufficient conditions for the direct method. They are readily satisfied for regularized transport coefficients on bounded energy intervals and for smoothed heterogeneous media. The assumptions that S and ρ are bounded and that ρ is bounded away from zero are consistent with the finite range of physical tissue densities used in treatment planning after CT-to-material conversion. The continuity of the control-to-fluence map follows from the well-posedness of the forward transport problem established earlier, provided the constants in the transport estimates are uniform over the uncertainty parameter θ.

For sharply heterogeneous patient geometries, these conditions should be interpreted as applying to the regularized coefficient fields used by the numerical solver. The theorem is not intended to assert that discontinuous raw CT data satisfy global $W^{1,\infty}$ assumptions without preprocessing; rather, it gives a well-posed optimization result for the regularized models on which stable deterministic and stochastic transport computations are based.

6.2 | First-Order Optimality Conditions

We now derive the first-order necessary conditions for optimality in the risk-neutral formulation. These conditions are expressed in terms of the directional derivative of the cost functional and are evaluated using the adjoint method.

For each fixed uncertainty realization $\theta \in \Theta$, the forward fluence ψ_g^θ solves the transport equation

$$\vec{\mathrm{L}}^\theta \psi_g^\theta = 0 \quad \text{in } C, \qquad \psi_g^\theta = g \quad \text{on } \Gamma^-, \tag{138}$$

and the adjoint variable z_g^θ solves the backward transport equation

$$\overleftarrow{\mathrm{L}}^{\theta} z_g^\theta = -\frac{qS^\theta}{\rho^\theta}\left(\mathcal{D}\left[\psi_g^\theta\right] - d_{\text{prescribed}}\right) \quad \text{in } \mathcal{C}, \qquad z_g^\theta = 0 \quad \text{on } \Gamma^+. \tag{139}$$

These forward-adjoint pairs allow us to compute the Gâteaux derivative of the cost functional with respect to the control. For each θ, the directional derivative of $\mathcal{J}(\mathrm{g}, \theta)$ in the direction $h \in \mathrm{L}^2(\Gamma^-)$ is given by

$$\delta\mathcal{J}(\mathrm{g}, \theta)[h] = \left\langle z_g^\theta, h\right\rangle_{\Gamma^-} + \alpha\langle \mathrm{g}, h\rangle_{\mathrm{L}^2(\Gamma^-)}. \tag{140}$$

Averaging over the distribution of θ yields the Gâteaux derivative of the expected objective

$$\nabla_{\mathrm{g}}\mathbf{E}_\theta\left[\mathcal{J}(\mathrm{g})\right] = \mathbf{E}_\theta\left[z_g^\theta\right] + \alpha \mathrm{g}. \tag{141}$$

This characterization provides the first-order optimality condition for the unconstrained problem. A control $\mathrm{g}^\star$ satisfies

$$\nabla_{\mathrm{g}}\mathbf{E}_\theta[\mathcal{J}(\mathrm{g}^\star)] = 0 \tag{142}$$

if and only if it is a stationary point. In the presence of control constraints, this gradient enters a variational inequality formulation, as developed in subsequent sections.

Theorem 6.4 (Gâteaux Differentiability of the Expected Cost Functional). *Let $\alpha > 0$, and suppose that for each $\theta \in \Theta$, the control-to-fluence map $\mathrm{g} \mapsto \psi_g^\theta \in \vec{\mathcal{H}}$ is linear and continuous and that the dose operator $\mathcal{D} : \vec{\mathcal{H}} \to L^2(D)$ is linear and bounded. Assume further that the integrand*

$$\theta \mapsto \left\|\mathcal{D}\left[\psi_g^\theta\right] - d_{prescribed}\right\|^2_{L^2(D;w)} \tag{143}$$

is integrable with respect to $\mathbb{P}(\theta)$ for each $\mathrm{g} \in \mathcal{U}_{\mathrm{ad}}$. Then the expected cost functional

$$\mathcal{J}(\mathrm{g}) := \mathbf{E}_\theta\left[\int_D w(\boldsymbol{x})\left(\mathcal{D}\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{prescribed}(\boldsymbol{x})\right)^2 d\boldsymbol{x}\right] + \frac{\alpha}{2}\|\mathrm{g}\|^2_{\mathrm{L}^2(\Gamma^-)} \tag{144}$$

is Gâteaux differentiable on $\mathcal{U}_{\mathrm{ad}} \subset \mathrm{L}^2(\Gamma^-)$. For any direction $h \in \mathrm{L}^2(\Gamma^-)$, the directional derivative is

$$\delta\mathcal{J}(\mathrm{g}; h) = \mathbf{E}_\theta\left[2\int_D w(\boldsymbol{x})\left(\mathcal{D}\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{prescribed}(\boldsymbol{x})\right)\cdot \mathcal{D}\left[\psi_h^\theta\right](\boldsymbol{x}) d\boldsymbol{x}\right] + \alpha\langle \mathrm{g}, h\rangle_{\mathrm{L}^2(\Gamma^-)}. \tag{145}$$

Proof. Fix $\mathrm{g}, h \in \mathrm{L}^2(\Gamma^-)$. For each $\theta \in \Theta$, the mapping $\mathrm{g} \mapsto \mathcal{D}[\psi_g^\theta]$ is linear and continuous by assumption. Therefore, the inner fidelity term

$$\mathcal{J}(\mathrm{g}, \theta) := \int_D w(\boldsymbol{x})\left(\mathcal{D}\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{\text{prescribed}}(\boldsymbol{x})\right)^2 d\boldsymbol{x} \tag{146}$$

is a quadratic functional of g and hence Gâteaux differentiable with pointwise directional derivative

$$\delta\mathcal{J}(\mathrm{g}, \theta; h) = 2\int_D w(\boldsymbol{x})\left(\mathcal{D}\left[\psi_g^\theta\right](\boldsymbol{x}) - d_{\text{prescribed}}(\boldsymbol{x})\right)\cdot \mathcal{D}[\psi_h^\theta](\boldsymbol{x}) d\boldsymbol{x}. \tag{147}$$

Since $\mathrm{g} \mapsto \mathcal{D}[\psi_g^\theta]$ is continuous and bounded uniformly over θ on bounded sets, and since

$$|\delta\mathcal{J}(\mathrm{g}, \theta; h)| \leq C(\theta)\|h\|_{\mathrm{L}^2(\Gamma^-)} \tag{148}$$

for an integrable function $C(\theta)$ (e.g., a bound on $\|\mathcal{D}[\psi_h^\theta]\|$), the conditions of Lemma 6.1 apply. We may therefore interchange expectation and differentiation

$$\delta\mathcal{J}(\mathrm{g}; h) = \mathbf{E}_\theta[\delta\mathcal{J}(\mathrm{g}, \theta; h)] + \alpha\langle \mathrm{g}, h\rangle_{\mathrm{L}^2(\Gamma^-)}. \tag{149}$$

This proves the Gâteaux differentiability of $\mathcal{J}$ and provides the desired representation. □

Corollary 6.5 (Gradient Representation via the Adjoint Problem). *Let $\psi_g^\theta \in \vec{\mathscr{H}}$ denote the fluence associated to control g under uncertainty realization $\theta \in \Theta$, defined by the forward problem*

$$\vec{\mathrm{L}}^{\theta}\, \psi_g^\theta = 0 \quad in\ C, \qquad \psi_g^\theta = g \quad on\ \Gamma^-. \tag{150}$$

Let $z_g^\theta \in \overleftarrow{\mathscr{H}}_0^{+}$ solve the adjoint transport equation

$$\overleftarrow{\mathrm{L}}^{\theta}\, z_g^\theta = -\frac{qS^\theta}{\rho^\theta}\left(\mathcal{D}\left[\psi_g^\theta\right] - d_{\text{prescribed}}\right) \quad in\ C, \qquad z_g^\theta = 0 \quad on\ \Gamma^+. \tag{151}$$

Then, the Gâteaux derivative of $\mathcal{J}$ satisfies

$$\delta\mathcal{J}(g;h) = \mathbf{E}_\theta\left[\left\langle z_g^\theta, h\right\rangle_{\Gamma^-}\right] + \alpha\langle g, h\rangle_{\mathrm{L}^2(\Gamma^-)}. \tag{152}$$

In particular, the $\mathrm{L}^2(\Gamma^-)$-gradient is given by

$$\nabla_g\mathcal{J}(g) = \mathbf{E}_\theta\left[z_g^\theta\right] + \alpha g. \tag{153}$$

6.3 | Stochastic Gradient Estimators

In the hybrid formulation, the fluence ψ_g^θ is computed via stochastic forward simulations depending on a random parameter $\theta \in \Theta$, while the adjoint variable z_g^θ is obtained by solving a deterministic PDE conditioned on θ. This results in stochastic variation in both the forward dose and the corresponding adjoint source term.

To compute the expected gradient $\nabla_g\mathbf{E}_\theta[\mathcal{J}(g)]$, we use a Monte Carlo estimator. Let $\{\theta_i\}_{i=1}^N$ be independent samples from the uncertainty distribution. For each θ_i:

1. Solve the forward problem to obtain the fluence $\psi_g^{\theta_i}$;
2. Evaluate the dose $\mathcal{D}[\psi_g^{\theta_i}]$;
3. Solve the adjoint equation

$$\overleftarrow{\mathrm{L}}^{\theta_i}\, z_g^{\theta_i} = -\frac{qS^{\theta_i}}{\rho^{\theta_i}}\left(\mathcal{D}\left[\psi_g^{\theta_i}\right] - d_{\text{prescribed}}\right), \qquad z_g^{\theta_i} = 0 \text{ on } \Gamma^+; \tag{154}$$

4. Compute the adjoint trace $z_g^{\theta_i}|_{\Gamma^-}$.

The empirical gradient estimator is then defined as

$$\nabla_g\mathcal{J}_N(g) := \frac{1}{N}\sum_{i=1}^N z_g^{\theta_i} + \alpha g. \tag{155}$$

As $N \to \infty$, the estimator converges strongly in $\mathrm{L}^2(\Gamma^-)$ to the true gradient $\mathbf{E}_\theta[z_g^\theta] + \alpha g$, under mild integrability conditions on $\theta \mapsto z_g^\theta$.

The variance of the Monte Carlo estimator $\nabla_g\mathcal{J}_N(g)$ can be substantial if the dose $\mathcal{D}[\psi_g^\theta]$ is highly sensitive to θ. Standard techniques such as importance sampling may be employed to reduce variance [39]. More advanced strategies, such as multifidelity corrections or Gaussian process surrogates, are effective when forward solves are expensive and adjoint source terms vary smoothly with respect to θ.

6.4 | Stochastic Variational Inequalities for Constrained Optimization

The admissible set $\mathcal{U}_{\text{ad}} \subset \mathrm{L}^2(\Gamma^-)$, defined by box constraints on the control g, is closed and convex. In the presence of stochastic uncertainty, the first-order optimality condition for minimizing the expected cost functional subject to $g \in \mathcal{U}_{\text{ad}}$ is approximated using a finite-sample Monte Carlo estimator of the gradient.

Let $\nabla_g \mathcal{J}_N(g) := z_g^N + \alpha g$ denote the empirical gradient estimator, where z_g^N is the sample average of adjoint traces computed from N realizations of the uncertainty parameter θ. The stochastic variational inequality associated with this setting is

$$\left\langle z_g^N + \alpha g, \tilde{g} - g \right\rangle_{\Gamma^-} \geq 0 \qquad \forall\, \tilde{g} \in \mathcal{U}_{\text{ad}}. \tag{156}$$

This condition characterises stationarity with respect to the empirical gradient estimate and ensures that no admissible variation decreases the sampled cost functional.

The variational inequality (156) leads to a projected stochastic gradient update of the form

$$g_{k+1} = \mathcal{P}_{\mathcal{U}_{\text{ad}}}\left(g_k - \tau_k \nabla_g \mathcal{J}_N(g_k)\right) = \mathcal{P}_{\mathcal{U}_{\text{ad}}}\left(g_k - \tau_k \left(z_{g_k}^N + \alpha g_k\right)\right), \tag{157}$$

where $\tau_k > 0$ is the step size and $\mathcal{P}_{\mathcal{U}_{\text{ad}}}$ is the orthogonal projection onto the admissible set in $\mathrm{L}^2(\Gamma^-)$. Under appropriate assumptions on step sizes and control of the gradient variance, this iteration converges almost surely to a solution of the true variational inequality as $N \to \infty$.

The projection onto the admissible set is computed pointwise almost everywhere as

$$\mathcal{P}_{\mathcal{U}_{\text{ad}}}[g](\boldsymbol{x}, \omega, E) = \min\{g_{\max}(\boldsymbol{x}, \omega, E), \max\{0, g(\boldsymbol{x}, \omega, E)\}\}. \tag{158}$$

This ensures that the control remains feasible throughout the iteration and enforces the physical constraints imposed by the treatment delivery system.

6.5 | Stochastic Realizations and Empirical Primal–Dual Systems

In practice, the expected fluence and dose quantities in the hybrid optimization framework cannot be computed exactly. Instead, they are approximated using a finite number of stochastic realisations. Let $\{\theta^{(i)}\}_{i=1}^N$ be independent samples from the uncertainty distribution $\mathbb{P}_\theta$, and let $\{\psi_g^{(i)}\}_{i=1}^N$ be the corresponding forward fluence solutions, that is, $\psi_g^{(i)} := \psi_g^{\theta^{(i)}}$. We define the empirical average of these realizations as

$$\bar{\psi}_g^N := \frac{1}{N} \sum_{i=1}^N \psi_g^{(i)}. \tag{159}$$

If the stopping power and density are deterministic, the corresponding empirical dose may be written as

$$\mathcal{D}\left[\bar{\psi}_g^N\right](\boldsymbol{x}) := \int_{\mathcal{E}} \frac{S(\boldsymbol{x}, E)}{\rho(\boldsymbol{x})} \int_{\mathbb{S}^2} \bar{\psi}_g^N(\boldsymbol{x}, \omega, E)\, d\omega dE. \tag{160}$$

If S^θ or ρ^θ depends on the uncertainty realization, the sample-average dose must instead be computed as

$$\mathcal{D}_N[g](\boldsymbol{x}) := \frac{1}{N} \sum_{i=1}^N \int_{\mathcal{E}} \frac{S^{\theta^{(i)}}(\boldsymbol{x}, E)}{\rho^{\theta^{(i)}}(\boldsymbol{x})} \int_{\mathbb{S}^2} \psi_g^{\theta^{(i)}}(\boldsymbol{x}, \omega, E)\, d\omega dE. \tag{161}$$

This is consistent with the noncommutativity of expectation and the dose operator discussed above.

We define the empirical adjoint z_g^N as the solution to the adjoint PDE with source determined by the empirical dose discrepancy:

$$\overleftarrow{\mathrm{L}}\, z_g^N = -\frac{qS}{\rho}\left(\mathcal{D}\left[\bar{\psi}_g^N\right] - d_{\text{prescribed}}\right), \qquad z_g^N = 0 \quad \text{on } \Gamma^+. \tag{162}$$

Together, these yield the empirical primal–dual system:

$$\overrightarrow{\mathrm{L}}\,\bar{\psi}_g^N = 0, \qquad \text{in } \mathcal{C}, \qquad \bar{\psi}_g^N = g, \qquad \text{on } \Gamma^-,$$
$$\overleftarrow{\mathrm{L}}\,z_g^N + \frac{qS}{\rho}\left(\mathcal{D}\left[\bar{\psi}_g^N\right] - d_{\text{prescribed}}\right) = 0, \qquad \text{in } \mathcal{C}, \qquad z_g^N = 0, \qquad \text{on } \Gamma^+, \tag{163}$$
$$g^\star - \Pi_{[0,g_{\max}]}\left(-\frac{1}{\alpha}z_g^N\right) = 0, \qquad \text{on } \Gamma^-.$$

Remark 6.6 (Stochastic Variational Inequality from the Empirical System). The projection formula from Remark 6.4 applies to the empirical adjoint z_g^N as well. Each sample-based approximation satisfies the variational inequality

$$\left\langle z_g^N + \alpha g^\star, \tilde{g} - g^\star \right\rangle_{\Gamma^-} \geq 0 \qquad \forall\, \tilde{g} \in \mathcal{U}_{\text{ad}}. \tag{164}$$

This characterises a sample-wise approximation of the true stochastic variational inequality and supports the use of empirical gradients in projected optimization.

The empirical adjoint z_g^N serves as a Monte Carlo gradient estimator for the expected fidelity functional. As $N \to \infty$, it converges (in probability) to the true adjoint $z_g := \mathbf{E}_\theta[z_g^\theta]$, and the empirical system recovers the deterministic optimality conditions. In practice, small N yields efficient approximate updates and variance reduction techniques can be used to stabilize and accelerate convergence.

6.6 | Algorithmic Optimization Strategy

We now present a numerical implementation of the hybrid SDE–PDE framework for robust treatment planning. This algorithm combines high-fidelity stochastic forward simulations with deterministic adjoint-based gradient computation to solve the risk-neutral optimization problem under physical uncertainty.

At each iteration, the current control g^k is used to generate N realizations of the stochastic fluence ψ_g^θ. These are averaged to yield an empirical estimate $\bar{\psi}_{g^k}^N$ of the expected fluence. The corresponding dose discrepancy is used to construct the source term in the adjoint BFP equation, which is solved deterministically to produce the empirical gradient direction z^k. A projected gradient update is then applied to ensure feasibility with respect to box constraints (Algorithm 1).

The projected-gradient form above uses τ_k for the algorithmic step size and reserves α for the Tikhonov regularization parameter. In the limiting case where the adjoint state is frozen and the box-constrained quadratic subproblem is solved pointwise, this update reduces to the projection formula $g = \Pi_{[0,g_{\max}]}(-z/\alpha)$.

The accuracy of the gradient estimate depends on the number of Monte Carlo samples N used to compute $\bar{\psi}_{g^k}^N$. Early iterations may tolerate high variance and hence use small N, while later stages may require more accurate gradients. Adaptive sampling schemes, in which N increases with k or is chosen based on variance estimates, can improve efficiency and convergence robustness.

Remark 6.7 (Computational role of the hybrid formulation). The proposed hybrid SDE–PDE method should be understood as an adjoint-gradient strategy rather than as a replacement for Monte Carlo dose simulation. Let m denote the number of independent control variables, for example, beam weights or boundary degrees of freedom, and let N denote the number of stochastic samples used to estimate the forward dose. A finite-difference Monte Carlo gradient requires repeated forward estimates under control perturbations and therefore has a cost that grows linearly with m. In contrast, the hybrid formulation uses stochastic simulation to estimate the forward dose or fluence for the current control, but obtains the gradient through deterministic adjoint traces on Γ^-. Thus, the gradient construction is essentially independent of m, apart from the cost of evaluating the adjoint trace against the chosen control basis.

This places the method between direct Monte Carlo optimization, which retains stochastic forward simulation but lacks an efficient adjoint gradient, and fully deterministic adjoint optimization, which computes both the forward and adjoint states from a deterministic transport solver.

The method does not remove the cost of stochastic forward sampling. Its computational advantage is that it avoids recomputing Monte Carlo dose estimates for each control perturbation. If the uncertainty enters only through stochastic

ALGORITHM 1 | Hybrid SDE–PDE Optimization Algorithm

Input. Initial control $g^0 \in \mathcal{U}_{\mathrm{ad}}$, regularization parameter $\alpha > 0$, step sizes $\{\tau_k\}_{k\geq 0}$, tolerance $\varepsilon > 0$, sample size N
Output. Approximate minimizer $g^\star$
while $\|z^k + \alpha g^k\|_{\mathrm{L}^2(\Gamma^-)} > \varepsilon$ **do**
  (i) Forward Solve. Simulate N samples $\{\psi_{g^k}^{\theta_i}\}_{i=1}^N$ using the SDE model with inflow g^k
    Compute the empirical average $\bar{\psi}_{g^k}^N := \frac{1}{N}\sum_{i=1}^N \psi_{g^k}^{\theta_i}$
    Evaluate dose $\mathcal{D}[\bar{\psi}_{g^k}^N]$
  (ii) Adjoint Solve. Solve

$$\overleftarrow{\mathrm{L}}\, z^k = -\frac{qS}{\rho}\left(\mathcal{D}[\bar{\psi}_{g^k}^N] - d_{\text{prescribed}}\right), \qquad z^k = 0 \text{ on } \Gamma^+$$

  (iii) Control Update.

$$g^{k+1} = \Pi_{[0,g_{\max}]}\big(g^k - \tau_k(z^k + \alpha g^k)\big)$$

  (iv) Convergence Check. If $\|z^k + \alpha g^k\|_{\mathrm{L}^2(\Gamma^-)} \leq \varepsilon$, **break**
end while
Return: $g^\star = g^k$

particle histories for fixed transport coefficients, one deterministic adjoint solve may be used after the forward estimate. If the coefficients depend on scenario parameters θ, one may either solve scenario-conditioned adjoint problems and average their traces, or use the reduced empirical adjoint system described below. In both cases, the adjoint formulation avoids finite-difference gradient scaling in the number of controls.

6.7 | Numerical Demonstration in One Dimension

To illustrate the hybrid optimization strategy, we present a simple numerical experiment in a one-dimensional water phantom. The spatial domain is an 8.5 cm homogeneous slab of water, discretized uniformly in depth and energy. A bank of 20 pencil beams is constructed with varying initial energies, each modeled as a Gaussian energy distribution. Dose and fluence profiles are computed using a Monte Carlo simulation of a stopping-power-driven SDE.

The goal is to match a prescribed dose profile delivering uniform dose to a target region between 3 and 6 cm depth. A weighted least-squares cost functional is minimized over nonnegative linear combinations of the pencil beams, corresponding to intensity-modulated inputs. The resulting weights define the optimal beam configuration, from which dose and fluence are computed.

Figure 3 shows the optimized fluence in depth-energy coordinates, along with the relative input spectrum and resulting depth-dose curve. The modulation of Bragg peaks across the beam bank produces a spread-out Bragg peak (SOBP) that conforms to the target region.

7 | Conclusion

This work develops a unified stochastic-deterministic framework for modeling proton transport in radiotherapy. Starting from a stochastic process describing individual proton tracks, we established its connection to the BFP equation via variational theory and dual semigroup structures. The resulting resolvent formulation links expected occupation measures with deterministic PDE solutions, providing a rigorous basis for interpreting dose as both a stochastic functional and a variational quantity.

The adjoint relationship between forward and backward transport enables a dual characterization of fluence and dose, with implications for uncertainty quantification, control, and optimization. As a proof of concept, we applied this

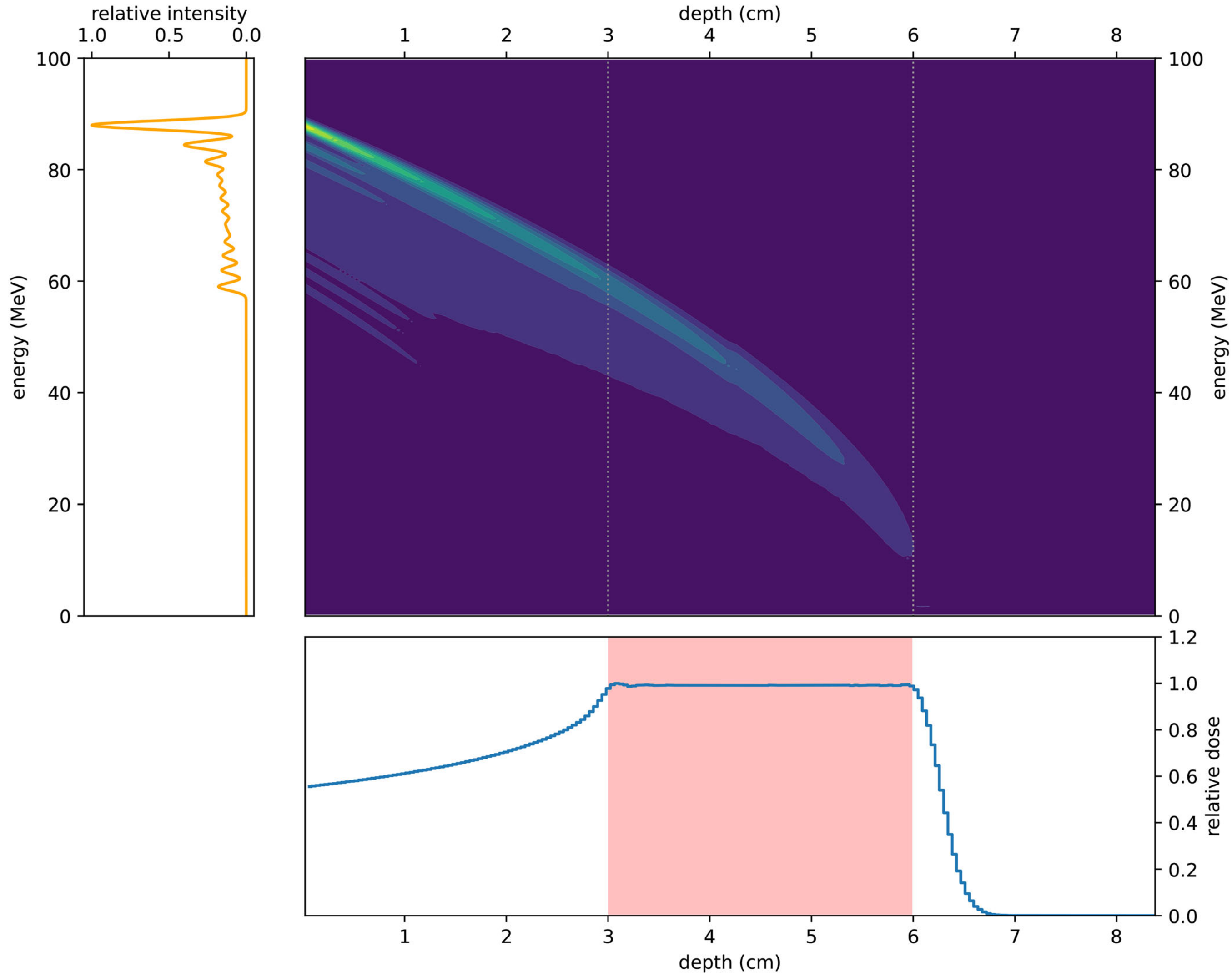

FIGURE 3 | Numerical result of the 1D dose optimization experiment. **Top left**: Relative energy spectrum of the optimized beam. **Bottom**: Resulting depth-dose curve. **Top right**: Fluence in depth-energy coordinates. The target region between 3 and 6 cm is shown in red.

framework to a robust treatment planning problem, using stochastic forward models and deterministic adjoint-based gradient computation. A simple numerical demonstration illustrated how the theory underpins hybrid algorithm design.

Beyond optimization, the stochastic-deterministic duality supports theoretical analysis, variance reduction, and model reduction strategies. It provides a coherent foundation for incorporating uncertainty into transport-based methods, bridging probabilistic simulation, and PDE-based inference in a mathematically consistent way.

Acknowledgments

All authors are grateful for support from the EPSRC program grant EP/W026899/2 and TP for support from Leverhulme RPG-2021-238.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no data sets were generated or analyzed during the current study.

References

1. W. D. Newhauser and R. Zhang, “The Physics of Proton Therapy,” *Physics in Medicine & Biology* 60, no. 8 (2015): R155–R209.

2. H. Thomas and B. Timmermann, “Paediatric Proton Therapy,” *British Journal of Radiology* 93, no. 1107 (2020): 20190601.

3. K. Souris, J. A. Lee, and E. Sterpin, “Fast Multipurpose Monte Carlo Simulation for Proton Therapy Using Multi- and Many-Core CPU Architectures,” *Medical Physics* 43, no. 4 (2016): 1700–1712.

4. B. S. Ashby, V. Chronholm, D. K. Hajnel, et al., “Efficient Proton Transport Modelling for Proton Beam Therapy and Biological Quantification,” *Journal of Mathematical Biology* 90, no. 5 (2025): 1–33.

5. L. Lin, S. Huang, M. Kang, et al., “A Benchmarking Method to Evaluate the Accuracy of a Commercial Proton Monte Carlo Pencil Beam Scanning Treatment Planning System,” *Journal of Applied Clinical Medical Physics* 18, no. 2 (2017): 44–49.

6. G. C. Pomraning, “The Fokker-Planck Operator as an Asymptotic Limit,” *Mathematical Models and Methods in Applied Sciences* 2, no. 01 (1992): 21–36.

7. A. Haghighat, *Monte Carlo Methods for Particle Transport* (CRC Press, 2020).

8. I. Lux, *Monte Carlo Particle Transport Methods* (CRC Press, 2018).

9. A. Fredriksson, “A Characterization of Robust Radiation Therapy Treatment Planning Methods—From Expected Value to Worst Case Optimization,” *Medical Physics* 39, no. 8 (2012): 5169–5181.

10. J. Unkelbach and H. Paganetti, “Robust Proton Treatment Planning: Physical and Biological Optimization,” *Seminars in Radiation Oncology* 28, no. 2 (2018): 88–96.

11. H.-G. Roos, M. Stynes, and L. Tobiska, *Robust Numerical Methods for Singularly Perturbed Differential Equations: Convection-Diffusion-Reaction and Flow Problems* (Springer, 2008).

12. H. Paganetti, *Proton Therapy Physics* (CRC Press, 2025).

13. A. Calloo, M. Evans, H. Lockyer, F. Madiot, T. Pryer, and L. Zanetti, “Cycle-Free Polytopal Mesh Sweeping for Boltzmann Transport,” *Numerical Algorithms* 101 (2026): 2157–2180.

14. J. C. Wagner, D. E. Peplow, S. W. Mosher, and T. M. Evans, “Review of Hybrid (Deterministic/Monte Carlo) Radiation Transport Methods, Codes, and Applications at Oak Ridge National Laboratory,” *Progress in Nuclear Science and Technology* 2, no. 104 (2011): 808–814.

15. F. C. DiFilippo, M. Goldstein, B. A. Worley, and J. C. Ryman, “Adjoint Monte Carlo Methods for Radiotherapy Treatment Planning,” *Transactions of the American Nuclear Society* 74 (1996): 14–16.

16. F. Georgiou and T. Pryer, “Scotty: A Robust Optimisation Framework for Proton Therapy Treatment Planning,” Preprint, 2025.

17. J. Louis Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, vol. 170 (Springer, 1971).

18. B. Oksendal, *Stochastic Differential Equations: An Introduction With Applications* (Springer Science & Business Media, 2013).

19. G. A. Pavliotis, “Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations,” in *Texts in Applied Mathematics*, vol. 60 (Springer, 2014).

20. S. N. Boon, *Dosimetry and Quality Control of Scanning Proton Beams* (PhD thesis, University of Groningen, 1998).

21. T. Bortfeld, “An Analytical Approximation of the Bragg Curve for Therapeutic Proton Beams,” *Medical Physics* 24, no. 12 (1997): 2024–2033.

22. H. E. S. Pettersen, M. Chaar, I. Meric, O. H. Odland, J. R. Sølie, and D. Röhrich, “Accuracy of Parameterized Proton Range Models; A Comparison,” *Radiation Physics and Chemistry* 144 (2018): 295–297.

23. A. M. G. Cox, L. Hattam, A. E. Kyprianou, and T. Pryer, “A Bayesian Inverse Approach to Proton Therapy Dose Delivery Verification,” *Proceedings of the Royal Society A* 480 (2024): 20230836.

24. L. C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, vol. 19 (American Mathematical Society, 1998), xviii+662, ISBN: 0-8218-0772-2.

25. R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology, Evolution Problems. II*, vol. 6, eds. C. Bardos, M. Cessenat, A. Kavenoky, et al. (Springer-Verlag, Berlin, 1993), Translated from the French by Alan Craig.

26. H. Brézis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, vol. 2.3 (Springer, 2011).

27. J.-L. Guermond Alexandre Em, *Theory and Practice of Finite Elements*, Applied Mathematical Sciences, vol. 159 (Springer, 2004).

28. V. Chronholm and T. Pryer, “Geometry, Energy and Sensitivity in Stochastic Proton Dynamics,” Preprint, 2025.

29. J. D. Smith, R. Lehoucq, and B. Franke, “A Stochastic Calculus Approach to Boltzmann Transport,” *Nuclear Science and Engineering* 199 (2025): 220–234.

30. A. Crossley, K. Habermann, E. Horton, J. Koskela, A. E. Kyprianou, and S. Osman, “Jump Stochastic Differential Equations for the Characterisation of the Bragg Peak in Proton Beam Radiotherapy,” *Proceedings of the Royal Society A* 481 (2025): 20240687.

31. B. Øksendal, *Stochastic Differential Equations: An Introduction With Applications*, 6th ed. (Springer, 2003), ISBN 978-3-540-04758-9.

32. G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in Applied Mathematics, vol. 60 (Springer, 2014), ISBN: 978-1-4939-1322-6.

33. K.-J. Engel, R. Nagel, and S. Brendle, *One-Parameter Semigroups for Linear Evolution Equations*, vol. 194 (Springer, 2000).

34. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, vol. 44 (Springer Science & Business Media, 2012).

35. J.-M. Bismut, "Martingales, the Malliavin Calculus and Hypoellipticity Under General Hörmander's Conditions," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 56, no. 4 (1981): 469–505.

36. L. Hörmander, "Hypoelliptic Second Order Differential Equations," *Acta Mathematica* 119 (1967): 147–171.

37. M. Bangert, P. Hennig, and U. Oelfke, "Analytical Probabilistic Modeling for Radiation Therapy Treatment Planning," *Physics in Medicine & Biology* 58, no. 16 (2013): 5401–5419.

38. T. Bortfeld and W. Schlegel, "Optimization of Beam Orientations in Radiation Therapy: Some Theoretical Considerations," *Physics in Medicine & Biology* 38, no. 2 (1993): 291–304.

39. J. E. Hoogenboom, "Zero-Variance Monte Carlo Schemes Revisited," *Nuclear Science and Engineering* 160, no. 1 (2008): 1–22.