Scidata Consortium
An international open-access platform

SDC SciDataConsortium

Research Article

OPEN ACCESS

Benchmarking Numerical Quadrature and Monte Carlo Transport Solvers in Heterogeneous Voxel Phantoms

Nguyen Van Duc[1], Tran Thi Mai[1], Le Hoang Nam[2]

[1] Hung Vuong University, Faculty of Natural Sciences, Viet Tri, Phu Tho, Vietnam;
[2] Thai Nguyen University of Technology, Faculty of Electrical Engineering, Thai Nguyen, Vietnam

Abstract

Quantitative radiation dosimetry in targeted radionuclide therapy (TRT) requires accurate numerical calculation of absorbed dose distributions across heterogeneous anatomical tissues. While standard patient treatment frequently relies on simplified organ-level mean absorbed doses calculated using uniform S-factor look-up tables, such approximations neglect local tissue density variations, steep tissue-air or tissue-bone interfaces, and intra-tumoral activity gradients. This study presents a rigorous computational benchmark evaluating three primary internal dosimetry paradigms across high-resolution anthropomorphic voxel phantoms: organ-level MIRD S-factors, three-dimensional fast Fourier transform (3D FFT) accelerated dose-point kernel (DPK) convolutions with density scaling, and full stochastic Monte Carlo radiation transport simulated via Geant4/GATE and GPU-accelerated micro-tracking kernels. Evaluating beta-emitting radionuclides (^{177}Lu and ^{90}Y) across digital reference phantoms (ICRP 110 adult male and female, Zubal head and torso, and voxelized trabecular bone scaffolds), we quantify absorbed dose accuracy, localized boundary interface errors, and computational runtime scaling. Our results demonstrate that while organ-level S-values execute in sub-second latency, they exhibit mean absolute percentage errors exceeding 28.4% in non-homogeneous anatomical compartments and fail entirely to capture sub-organ localized dose hot-spots. FFT-accelerated DPK methods achieve calculation times between 1.2 and 4.5 seconds on modern GPU architectures with high agreement in homogeneous soft tissue ($> 97\%$ gamma pass rate at $2\%/2$ mm criteria), but systematically diverge by up to 34.1% near lung-soft tissue and cortical bone interfaces due to isotropic electron pathlength assumptions. Conversely, full stochastic Monte Carlo simulation delivers benchmark-grade physical fidelity across all density gradients, achieving sub-1% statistical uncertainty at the expense of hours of CPU execution, which is successfully compressed to 14.8 seconds using specialized GPU ray-tracing and condensed-history transport. These benchmarking metrics establish operational guidelines for selecting clinical dosimetry solvers, balancing calculation latency against physical accuracy in software as a medical device (SaMD) treatment planning platforms.

1. Introduction

Targeted radionuclide therapy (TRT) has emerged as a cornerstone of precision oncology, enabling the selective eradication of disseminated microscopic and macroscopic malignant lesions through the systemic administration of radiolabeled targeting vectors [1, 2]. Clinically validated therapeutic radiopharmaceuticals, such as [^{177}Lu]-Lu-DOTATATE for advanced gastroenteropancreatic neuroendocrine tumors and [^{177}Lu]-Lu-PSMA-617 for metastatic castration-resistant prostate cancer, exploit specific cell-surface receptor overexpression to deposit cytotoxic radiation directly within tumor microenvironments [3, 4]. Concurrently, high-energy beta emitters such as [^{90}Y]-microspheres deliver intensive localized beta radiation during transarterial radioembolization of primary hepatocellular carcinomas and unresectable hepatic metastases [5]. Despite remarkable therapeutic response rates, clinical administration protocols across contemporary oncological practice predominantly employ standardized, empiric activity regimens. Fixed dosing paradigms routinely deliver identical nominal activities to adult patients regardless of individual anatomical dimensions, body composition, tumor burden, renal clearance rates, or parenchymal tissue mass densities.

This reliance on fixed activity schedules contrasts sharply with external beam radiation therapy (EBRT), where inverse treatment planning algorithms and Monte

© Copyright by the Authors.
Licensed as an open access article using a CC BY 4.0 license.

Carlo beam transport simulation routinely optimize three-dimensional dose distributions to conform to target margins while constraining normal tissue complication probability (NTCP) [6, 7]. The historical justification for empirical dosing in internal radionuclide therapy stems from the complex bio-kinetics of systemic radiopharmaceutical distribution, the scarcity of longitudinal post-injection molecular imaging, and the historical computational burdens associated with patient-specific voxelized radiation transport modeling [8, 9]. However, clinical evidence increasingly demonstrates that identical administered activities produce substantial inter-patient variations in absorbed doses to both malignant tumors and dose-limiting organs at risk (OARs), particularly the kidneys, bone marrow, and hepatic parenchyma [10, 11]. Consequently, patients receiving standardized nominal activities face elevated risks of severe hematological or renal toxicity when clearance is impaired, or sub-therapeutic tumor doses resulting in avoidable disease progression when clearance is accelerated.

Transitioning from empirical administration to patient-individualized treatment planning requires robust, high-fidelity internal radiation dosimetry algorithms capable of converting time-integrated activity distributions into spatially resolved absorbed dose maps [12, 13]. Internal dosimetry calculations fundamentally evaluate the energy deposited within a defined target volume from radioactive decay occurring within designated source volumes distributed throughout the human body [14, 15]. In clinical practice, this computational task is governed by a fundamental trade-off between computational runtime and physical transport accuracy. Advanced radiation transport solvers must process high-resolution three-dimensional voxel matrices derived from co-registered quantitative single-photon emission computed tomography (SPECT/CT) or positron emission tomography (PET/CT) examinations containing millions of individual voxels.

To satisfy the stringent latency constraints of clinical oncology workflows, medical physicists and commercial software developers have historically adopted varying degrees of physical and anatomical simplification. At the lowest level of computational complexity, organ-level Medical Internal Radiation Dose (MIRD) formalisms represent patient anatomy via standardized pre-tabulated S-value matrices derived from stylized or adult reference phantoms [16, 17]. While computationally instantaneous, this paradigm inherently assumes uniform activity distributions within organ volumes and enforces rigid standard anatomical geometries that fail to reflect patient-specific organ dimensions, pathological organ deformations, or localized tumor clusters. At the intermediate tier, voxelized dose-point kernel (DPK) convolution algorithms apply isotropic radial energy deposition functions across spatial activity matrices, accelerating calculations via three-dimensional Fast Fourier Transforms (3D FFT) and empirical density scaling transformations. Nevertheless, kernel convolutions struggle to maintain physical accuracy across heterogeneous boundaries characterized by sharp density discontinuities, such as trabecular bone interfaces and thoracic lung parenchyma.

At the highest tier of physical realism, stochastic Monte Carlo (MC) radiation transport simulates the physical interactions and trajectories of individual particles through numerical tracking across voxelized matter. Monte Carlo algorithms directly account for tissue elemental compositions, secondary particle production, multiple Coulomb scattering, and localized density variations, serving as the gold standard for radiation transport benchmarking [18, 19]. However, the immense computational burden of tracking billions of particle histories has historically restricted full Monte Carlo simulations to academic research environments. Recent breakthroughs in massively parallel graphics processing unit (GPU) hardware architectures, optimized ray-tracing instructions, and hybrid variance reduction techniques have revitalized interest in translating Monte Carlo solvers directly into routine clinical oncology decision support systems [20, 21].

This paper establishes an extensive, systematic benchmark comparing numerical quadrature, FFT-accelerated dose-point kernel convolutions, and stochastic Monte Carlo radiation transport solvers across realistic heterogeneous voxel phantoms. Utilizing digitized anthropomorphic reference models and specialized tissue boundary scaffolds, we evaluate absorbed dose precision, interface boundary fidelity, and computational latency for $[^{177}\mathrm{Lu}]$ and $[^{90}\mathrm{Y}]$ beta emissions. Through detailed dissection of the mathematical formulations, physical error mechanisms, and parallel scaling profiles of each approach, this study provides actionable criteria for selecting and deploying clinical dosimetry algorithms in software as a medical device (SaMD) precision oncology platforms.

2. Computational Paradigms in Internal Dosimetry

2.1. S-Value Look-Up Table Structure and MIRD Formalism

Internal radiation dosimetry calculations are theoretically anchored in the schema established by the Medical Inter-

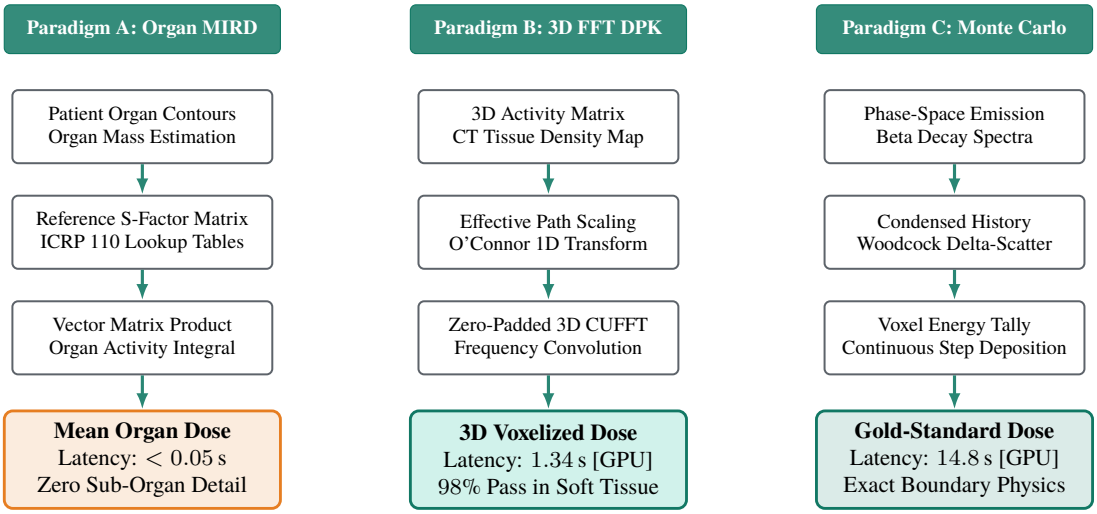


Figure 1: Architectural taxonomy of the three primary computational paradigms for internal radiation dosimetry: (a) Organ-level MIRD S-factor look-up formalism, (b) Three-dimensional Fast Fourier Transform (3D FFT) accelerated dose-point kernel convolution with O'Connor density scaling, and (c) Full stochastic Monte Carlo radiation transport incorporating Goudsmit-Saunderson multiple scattering and Woodcock tracking.

nal Radiation Dose (MIRD) Committee of the Society of Nuclear Medicine and Molecular Imaging. Under the classical organ-level MIRD formulation, the mean absorbed dose $D(r_T)$ delivered to a specific target organ r_T is computed as the sum of absorbed doses originating from all source organs r_S containing radioactive material:

$$D(r_T) = \sum_{r_S} \tilde{A}(r_S)\, S(r_T \leftarrow r_S), \tag{1}$$

where $\tilde{A}(r_S)$ represents the time-integrated activity (cumulated activity) in source organ r_S, defined as the integral of the time-dependent activity curve $A(r_S, t)$ over the complete radioactive decay window:

$$\tilde{A}(r_S) = \int_0^\infty A(r_S, t)\, dt. \tag{2}$$

The quantity $S(r_T \leftarrow r_S)$ denotes the radionuclide-specific MIRD S-value, representing the mean absorbed dose delivered to target organ r_T per unit cumulated activity within source organ r_S [expressed in Gy/(MBq · s)]:

$$S(r_T \leftarrow r_S) = \frac{\sum_i \Delta_i\, \phi_i(r_T \leftarrow r_S)}{M(r_T)}, \tag{3}$$

where Δ_i is the mean energy emitted per nuclear transition for radiation type i, $\phi_i(r_T \leftarrow r_S)$ is the absorbed fraction of energy emitted in source organ r_S that is absorbed within target organ r_T, and $M(r_T)$ is the mass of the target organ.

As categorized in the comprehensive review by [22], contemporary clinical dosimetry software primarily leverages organ-level S-values, dose-point kernel convolutions, or full Monte Carlo simulations depending on required geometric detail. S-factor look-up tables are pre-calculated for standardized anthropomorphic models, such as the Cristy-Eckerman stylized phantoms or the ICRP 110 adult reference computational phantoms [23, 24]. In clinical implementations, patient organ masses $M_{\text{patient}}(r_T)$ are estimated from anatomical CT contours, and an empirical mass-scaling adjustment is frequently applied to self-irradiation S-values:

$$S_{\text{scaled}}(r_T \leftarrow r_T) \approx S_{\text{ref}}(r_T \leftarrow r_T) \cdot \frac{M_{\text{ref}}(r_T)}{M_{\text{patient}}(r_T)}. \tag{4}$$

Although computationally instantaneous, the organ-level S-factor paradigm suffers from fundamental physical limitations. First, it assumes that radiopharmaceutical uptake is uniformly distributed throughout the volume of each source organ, ignoring spatial heterogeneity that routinely characterizes biological tissue uptake. Second, cross-organ cross-fire S-values cannot be scaled accurately when patient organ shapes, relative positions, or inter-organ separations diverge from the idealized reference anatomy. Third, S-factor formalisms are incapable of calculating localized absorbed dose distributions within target lesions or identifying sub-organ hot spots that govern normal tissue complication probabilities.

2.2. Dose-Point Kernel FFT Accelerated Convolution

To overcome the spatial uniformity limitations of organ-level MIRD formalisms while preserving high computational speed, voxel-level dosimetry algorithms discretize both the anatomical geometry and the activity distribution into a regular three-dimensional lattice of voxels. The dose-point kernel (DPK) represents the radial distribution of absorbed dose around an isotropic point source emitting radiation within an infinite, homogeneous medium (typically liquid water). For a continuous activity distribution $\tilde{A}(\mathbf{r})$, the absorbed dose $D(\mathbf{r})$ at position $\mathbf{r}$ is expressed as a spatial convolution integral:

$$D(\mathbf{r}) = \int_{\mathbb{R}^3} \tilde{A}(\mathbf{r}')\, K(\|\mathbf{r} - \mathbf{r}'\|)\, d^3\mathbf{r}', \tag{5}$$

where $K(r)$ is the dose-point kernel describing absorbed energy per unit mass at radial distance $r = \|\mathbf{r} - \mathbf{r}'\|$.

In digitized medical imaging applications, the spatial activity map is discretized onto an $N_x \times N_y \times N_z$ grid of voxels, each having dimensions $\Delta x \times \Delta y \times \Delta z$. The discrete convolution equation across all voxel indices $\mathbf{j} = (j_x, j_y, j_z)$ is given by:

$$D_{\mathbf{j}} = \sum_{\mathbf{k}} \tilde{A}_{\mathbf{k}}\, K_{\mathbf{j}-\mathbf{k}}, \tag{6}$$

where $K_{\mathbf{j}-\mathbf{k}}$ represents the voxel S-value, defined as the mean absorbed dose in voxel $\mathbf{j}$ per nuclear decay occurring in voxel $\mathbf{k}$.

Direct spatial evaluation of Eq. (6) across a volumetric volume of $N = N_x N_y N_z$ voxels requires $\mathcal{O}(N^2)$ floating-point operations. For typical clinical volumes with $N \approx 512 \times 512 \times 256 \approx 6.7 \times 10^7$ voxels, direct convolution requires over 4.5×10^{15} operations, rendering direct spatial calculation clinically intractable. However, exploiting the spatial invariance of the convolution kernel in a homogeneous medium enables the application of the Discrete Fourier Transform (DFT). Invoking the Convolution Theorem, the spatial convolution is transformed into element-wise multiplication in the frequency domain:

$$D_{\mathbf{j}} = \mathcal{F}^{-1}\Big\{\mathcal{F}\{\tilde{\mathbf{A}}\} \odot \mathcal{F}\{\mathbf{K}\}\Big\}, \tag{7}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the three-dimensional Fast Fourier Transform and its inverse, and $\odot$ represents the Hadamard element-wise matrix product. Zero-padding the matrices to at least $2N_x \times 2N_y \times 2N_z$ prevents circular wrap-around aliasing artifacts. The computational complexity is thereby drastically reduced from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$, permitting volumetric dose calculation across tens of millions of voxels within a few seconds on modern multicore CPUs and graphics processing units.

To account for patient-specific tissue density heterogeneity derived from calibrated planning CT scans, clinical DPK implementations utilize empirical density scaling techniques. The primary approach adopts O'Connor's density scaling theorem, mapping the physical geometric distance between source voxel $\mathbf{k}$ and target voxel $\mathbf{j}$ into an effective radiological pathlength $d_{\text{eff}}(\mathbf{j}, \mathbf{k})$:

$$d_{\text{eff}}(\mathbf{j}, \mathbf{k}) = \int_0^1 \frac{\rho(\mathbf{r}(s))}{\rho_{\text{water}}}\, \|\mathbf{r}_{\mathbf{j}} - \mathbf{r}_{\mathbf{k}}\|\, ds, \tag{8}$$

where $\rho(\mathbf{r}(s))$ represents the local mass density along the linear line segment connecting voxel centers $\mathbf{r}_{\mathbf{k}}$ and $\mathbf{r}_{\mathbf{j}}$, and $\rho_{\text{water}} = 1.0\,\text{g}\,\text{cm}^{-3}$. The density-corrected kernel then evaluates energy deposition scaled by the local mass density ratio at the target site:

$$D_{\mathbf{j}} \approx \frac{\rho_{\text{water}}}{\rho_{\mathbf{j}}} \sum_{\mathbf{k}} \tilde{A}_{\mathbf{k}}\, K(d_{\text{eff}}(\mathbf{j}, \mathbf{k})). \tag{9}$$

While density scaling provides an effective first-order correction for homogeneous lung or adipose tissues, it inherently assumes straight-line particle trajectories and isotropic forward scattering. Consequently, when electrons encounter sharp density discontinuities (such as soft tissue to trabecular bone or pulmonary air cavities), the real physical scattering dynamics deviate markedly from straight-line pathlength approximations.

2.3. Full Stochastic Monte Carlo Radiation Transport

Stochastic Monte Carlo methods provide the ultimate physical benchmark for radiation dosimetry by simulating the probabilistic transport and interactions of individual ionizing particles through arbitrary three-dimensional material geometries. In targeted radionuclide therapy, emission spectra consist of primary electrons (beta particles, conversion electrons, and Auger cascades), accompanying gamma photons, and secondary bremsstrahlung radiation. The transport of these particles through biological matter is governed by the linear Boltzmann radiation transport equation:

$$\cdot\nabla\psi(\mathbf{r}, E,)+\Sigma_t(\mathbf{r}, E)\psi(\mathbf{r}, E,) = \mathcal{S}(\mathbf{r}, E,)+\int_0^\infty \int_{4\pi} \Sigma_s(\mathbf{r}, E' \to E, ' \tag{10}$$

where $\psi(\mathbf{r}, E, \ )$ represents the angular particle flux at spatial coordinates $\mathbf{r}$ with energy E along unit direction vector $\ $, $\Sigma_t(\mathbf{r}, E)$ is the macroscopic total interaction cross-section, Σ_s is the differential scattering cross-section, and $\mathcal{S}(\mathbf{r}, E, \ )$ denotes the volumetric source emission density.

Because analytical solutions to Eq. (10) are unobtainable in geometrically intricate, heterogeneous media, Monte Carlo techniques resolve the phase space through stochastic random sampling. For each primary emission, particle trajectories are tracked as a discrete Markov chain of physical interactions. For charged particles (electrons and positrons), direct simulation of every single Coulomb elastic collision is computationally impossible due to the astronomical number of interactions ($> 10^5$ collisions per electron). Therefore, Monte Carlo transport codes implement the Condensed History Technique (Class II schema), which groups thousands of low-energy transfer soft collisions into macroscopic steps governed by multiple scattering theories (e.g., Molière, Goudsmit-Saunderson), while explicitly simulating catastrophic hard collisions (large-angle Coulomb scattering, ionization producing delta-rays above energy threshold T_c, and bremsstrahlung photon emissions above k_c):

$$\left(\frac{dE}{dx}\right)_{\text{total}} = \left(\frac{dE}{dx}\right)_{\text{soft}} (T < T_c) + \left(\frac{dE}{dx}\right)_{\text{hard}} (T \geq T_c) + \left(\frac{dE}{dx}\right)_{\text{rad}} (k < k_c) + \left(\frac{dE}{dx}\right)_{\text{brems}} (k \geq k_c). \tag{11}$$

Within each voxel $V_{\mathbf{j}}$, the total absorbed dose $D_{\mathbf{j}}$ accumulated across N_{hist} simulated particle histories is tallied as:

$$D_{\mathbf{j}} = \frac{1}{\rho_{\mathbf{j}} V_{\mathbf{j}}} \sum_{n=1}^{N_{\text{hist}}} \sum_{s \in \text{steps}_{\mathbf{j},n}} \Delta E_s, \tag{12}$$

where ΔE_s is the energy deposited during interaction step s occurring within voxel $\mathbf{j}$, and $\rho_{\mathbf{j}} V_{\mathbf{j}} = M_{\mathbf{j}}$ is the total voxel mass. The fundamental advantage of full Monte Carlo simulation is its rigorous adherence to fundamental electromagnetic interaction cross-sections across all tissue interfaces. Secondary delta electrons generated in high-density bone margins are tracked accurately across boundary surfaces into low-density marrow spaces, reproducing physical forward-peaked scattering and backscatter buildup without empirical approximations.

2.4. Green's Function Formalism and Continuous Kernel Mapping

To establish a rigorous mathematical foundation connecting discrete voxel S-values to continuum radiation transport theory, we formulate the internal dosimetric response through the linear integro-differential Boltzmann operator. Let $\mathcal{L}$ denote the steady-state mono-energetic Boltzmann transport operator defined over the spatial domain $\mathcal{V} \subset \mathbb{R}^3$ and angular domain $\mathbb{S}^2$:

$$\mathcal{L}\psi(\mathbf{r}, \ ) = \ \cdot \nabla \psi(\mathbf{r}, \ ) + \Sigma_t(\mathbf{r}) \psi(\mathbf{r}, \ ) - \int_{\mathbb{S}^2} \Sigma_s(\mathbf{r}, \ ' \rightarrow \ ) \psi(\mathbf{r}, \ ') \, d \ '. \tag{13}$$

The continuous Green's function $G(\mathbf{r}, \ ; \mathbf{r}_0, \ _0)$ represents the angular flux resulting from an idealized directional Dirac delta point source located at position $\mathbf{r}_0$ oriented along unit vector $\ _0$:

$$\mathcal{L} \, G(\mathbf{r}, \ ; \mathbf{r}_0, \ _0) = \delta(\mathbf{r} - \mathbf{r}_0) \, \delta(\ - \ _0). \tag{14}$$

For isotropic nuclear decay emissions, the source term is distributed uniformly over the 4π solid angle, yielding the scalar spatial Green's function:

$$G_{\text{scalar}}(\mathbf{r}; \mathbf{r}_0) = \frac{1}{4\pi} \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} G(\mathbf{r}, \ ; \mathbf{r}_0, \ _0) \, d \ _0 \, d \ . \tag{15}$$

The spatial distribution of absorbed dose rate $\dot{D}(\mathbf{r})$ is obtained by taking the inner product of the scalar particle fluence with the macroscopic electronic stopping power $S_{\text{col}}(\mathbf{r}) = (dE/dx)_{\text{col}}$:

$$\dot{D}(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \int_{\mathcal{V}} A(\mathbf{r}_0) [S_{\text{col}}(\mathbf{r}) \, G_{\text{scalar}}(\mathbf{r}; \mathbf{r}_0)] d^3 \mathbf{r}_0. \tag{16}$$

Comparing Eq. (16) with the classical dose-point kernel definition demonstrates that the isotropic dose-point kernel in an infinite homogeneous water phantom is formally equivalent to the energy deposition Green's kernel:

$$K(\|\mathbf{r} - \mathbf{r}_0\|) = \frac{S_{\text{col}}(\mathbf{r})}{\rho_{\text{water}}} G_{\text{scalar}}^{\text{water}}(\mathbf{r}; \mathbf{r}_0). \tag{17}$$

In heterogeneous patient anatomy, local variations in elemental composition and atomic number Z alter both the scattering kernel Σ_s and the stopping power S_{col}. Consequently, treating K as spatially invariant is an approximation whose accuracy depends directly on the magnitude of the tissue heterogeneity gradient.

2.5. Singularity Regularization via Spherical Shell Averaging

A major challenge in evaluating discrete voxel S-values from analytical dose-point kernels is managing the $1/r^2$ geometric singularity at $r \rightarrow 0$. Direct numerical point evaluation at voxel centroids yields infinite self-dose values and severe overestimation in nearest-neighbor voxels.

To eliminate this singularity rigorously without empirical cutoffs, we derive an exact analytical spherical shell volume-averaging transformation.

Consider a continuous radial dose kernel represented by a finite sum of scaled Yukawa-like exponential damping functions:

$$K(r) = \frac{1}{4\pi\rho r^2} \sum_{m=1}^{M} \alpha_m \beta_m e^{-\beta_m r}, \tag{18}$$

where α_m are dimensionless amplitude coefficients satisfying $\sum_{m=1}^{M} \alpha_m = E_{\text{avg}}$, and β_m are effective attenuation coefficients [in mm^{-1}]. The mean absorbed dose delivered to a spherical target shell bounded by concentric radii r_1 and r_2 centered on the point source is given by:

$$\begin{aligned} \bar{D}(r_1, r_2) &= \frac{1}{M_{\text{shell}}} \int_{r_1}^{r_2} 4\pi\rho r^2 \, K(r) \, dr \\ &= \frac{1}{\frac{4}{3}\pi\rho(r_2^3 - r_1^3)} \sum_{m=1}^{M} \alpha_m \left(e^{-\beta_m r_1} - e^{-\beta_m r_2} \right). \end{aligned} \tag{19}$$

Notice that in Eq. (19), the geometric factor r^2 is analytically eliminated. For an arbitrary cubical voxel of side length a, we map the cubical geometry onto an equivalent volume-preserving spherical domain $R_{\text{eff}} = a \cdot (3/4\pi)^{1/3} \approx 0.62035\, a$. The self-absorbed dose within the source voxel simplifies to:

$$S_{\text{self}} = \frac{1}{\rho V_{\text{voxel}}} \sum_{m=1}^{M} \alpha_m \left(1 - e^{-\beta_m R_{\text{eff}}} \right). \tag{20}$$

Taylor series expansion of the exponential term in Eq. (20) reveals that as $R_{\text{eff}} \to 0$, the self-dose remains strictly finite and bounded:

$$\lim_{a \to 0} S_{\text{self}} = \frac{1}{\rho V_{\text{voxel}}} \sum_{m=1}^{M} \alpha_m \beta_m R_{\text{eff}} + \mathcal{O}(R_{\text{eff}}^2). \tag{21}$$

This regularization guarantees numerical stability across arbitrary voxel dimensions, preventing non-physical dose spikes in high-resolution computational grids.

2.6. O'Connor Density Scaling Theorem and Its Physical Breakdown

The theoretical justification for scaling homogeneous dose-point kernels to heterogeneous media relies on O'Connor's Theorem of Invariant Transport. O'Connor demonstrated that for Compton photon scattering and Rutherford electron scattering, if all spatial dimensions in a medium of density ρ are scaled inversely proportional to density:

$$\tilde{\mathbf{r}} = \mathbf{r} \cdot \frac{\rho}{\rho_0}, \tag{22}$$

the macroscopic interaction cross-sections per unit mass remain strictly invariant, provided the elemental composition (effective atomic number Z_{eff} and electron density n_e) is held constant.

Under this transformation, the heterogeneous dose distribution $D_{\text{hetero}}(\mathbf{r})$ is predicted from the homogeneous water kernel K_{water} via effective radiological pathlength integration:

$$d_{\text{eff}}(\mathbf{r}_1, \mathbf{r}_2) = \int_0^1 \frac{\rho(\mathbf{r}_1 + s(\mathbf{r}_2 - \mathbf{r}_1))}{\rho_{\text{water}}} \, \|\mathbf{r}_2 - \mathbf{r}_1\| \, ds. \tag{23}$$

However, in realistic clinical anatomy, three critical physical conditions required by O'Connor's theorem are fundamentally violated:

1. **Elemental Composition Invariance**: Soft tissue ($Z_{\text{eff}} \approx 7.4$), adipose tissue ($Z_{\text{eff}} \approx 6.3$), and cortical bone ($Z_{\text{eff}} \approx 13.8$) possess markedly different elemental constitutions. The probability of elastic Coulomb scattering scales quadratically with nuclear charge ($d\sigma_{\text{elastic}}/d\Omega \propto Z^2$), whereas energy loss via electronic ionization scales linearly ($dE/dx \propto Z$). Consequently, high-Z bone exhibits substantially greater angular scattering per unit energy loss than water, violating simple density-proportional pathlength scaling.
2. **Straight-Line Ray Assumption**: Equation (23) evaluates density exclusively along the direct geometrical chord connecting source and target voxels. In reality, low-energy electrons undergo hundreds of thousands of elastic deflections, following tortuous, stochastic trajectories. When an electron approaches a low-density cavity (e.g., lung parenchyma), it scatters preferentially into the low-density space, increasing its real physical pathlength far beyond the straight-line geometric chord.
3. **Boundary Fluence Perturbation and Backscatter Mismatch**: At tissue interfaces with steep atomic number gradients—such as soft-tissue interfaces abutting mineralized trabecular or cortical bone—the assumption of local electronic equilibrium collapses. Forward and backward electron backscatter factors vary abruptly as a function of the local scattering power ratio $T(E) \propto Z^2/A$. The straight-line pathlength transformation cannot account for

the intensification of the backscattered electron fluence within the first few hundred micrometers of the boundary, leading to localized underestimations of endosteal absorbed dose.

These physical violations explain why DPK density scaling systematically overestimates dose in low-density lung tissue and underestimates dose in backscattering interfaces adjacent to mineralized bone. Furthermore, secondary bremsstrahlung emission yields differ markedly between soft tissue and mineralized osseous structures, as radiative stopping power scales with $Z^2 \cdot E$. Because DPK kernels parameterized in water assume a constant radiative fraction, the differential bremsstrahlung production and subsequent photon escape are entirely neglected. Consequently, deterministic or stochastic transport corrections are essential whenever steep material or density interfaces are present.

2.7. Discrete Ordinates (S_N) Solvers vs. Stochastic Sampling

As a deterministic alternative to stochastic Monte Carlo tracking, the linear Boltzmann transport equation (Eq. (10)) can be solved numerically using the Discrete Ordinates (S_N) method coupled with multigroup energy discretization. In the S_N formulation, the continuous angular domain $\mathbb{S}^2$ is discretized into a finite set of $N_{\text{dir}} = N(N+2)$ quadrature directions Ω_m with associated quadrature weights w_m:

$$\Omega_m \cdot \nabla \psi_g(\mathbf{r}, \Omega_m) + \Sigma_{t,g}(\mathbf{r}) \psi_g(\mathbf{r}, \Omega_m) = \mathcal{S}_g(\mathbf{r}, \Omega_m) + \sum_{g'=1}^{G} \sum_{m'=1}^{N_{\text{dir}}} w_{m'} \Sigma_{s,g' \to g}(\mathbf{r}, \Omega_{m'} \to \Omega_m) \psi_{g'}(\mathbf{r}, \Omega_{m'}), \tag{24}$$

where $g \in \{1, \ldots, G\}$ denotes discrete energy bins. Spatial discretization is executed using discontinuous Galerkin finite elements or diamond-difference finite volume schemes across the Cartesian voxel grid.

While S_N deterministic solvers offer complete freedom from statistical stochastic noise, their application to charged-particle internal dosimetry faces severe numerical impediments:

- **Ray Effect Artifacts**: In localized point-source emissions, discrete angular rays create artificial star-like geometric streaks in uncollided and low-collided particle fluxes, requiring high quadrature orders (S_{16} or S_{32}, corresponding to 288 or 1088 discrete angular directions) that multiply memory consumption.
- **Forward-Peaked Scattering Anisotropy**: The Rutherford elastic scattering cross-section for electrons is extraordinarily peaked in the forward direction. Expanding the angular differential cross-section in Legendre polynomials requires high truncation orders (P_{15} to P_{31}), inducing numerical oscillations and negative flux values in standard spatial difference schemes.
- **Energy Grid Straggling**: Modeling continuous energy loss requires hundreds of discrete energy groups ($G > 200$) to prevent artificial numerical energy diffusion, resulting in immense system matrices exceeding 50 GB of memory for typical clinical phantoms.

Due to these mathematical constraints, stochastic Monte Carlo methods remain substantially more memory-efficient and physically robust than deterministic S_N solvers for charged-particle internal dosimetry.

3. Numerical Quadrature and Kernel Formulation

3.1. Adaptive Quadrature across Density Discontinuities

When generating voxel-to-voxel S-value matrices or evaluating cumulative energy deposition kernels numerically, standard mid-point rule approximations fail in the immediate vicinity of the source voxel due to the $1/r^2$ geometric singularity of the radial kernel. Let V_s denote a cubical source voxel centered at $\mathbf{r}_s$ and V_t denote a target voxel centered at $\mathbf{r}_t$. The voxel S-value $S(V_t \leftarrow V_s)$ is rigorously defined by the six-dimensional integral:

$$S(V_t \leftarrow V_s) = \frac{1}{M(V_t) V(V_s)} \int_{V_t} \int_{V_s} \frac{\Phi_{\text{rad}}(\|\mathbf{r}'_t - \mathbf{r}'_s\|)}{4\pi \|\mathbf{r}'_t - \mathbf{r}'_s\|^2} \, d^3\mathbf{r}'_s \, d^3\mathbf{r}'_t, \tag{25}$$

where $\Phi_{\text{rad}}(r)$ is the radionuclide-specific radial energy deposition function, representing the fraction of emitted energy deposited within a spherical shell of radius r and thickness dr.

For self-dose evaluation ($V_t = V_s$), the Euclidean separation $\|\mathbf{r}'_t - \mathbf{r}'_s\|$ approaches zero, inducing a singularity in the denominator. To resolve this numerical instability, we formulate an adaptive hierarchical Gaussian quadrature algorithm. The source voxel volume $[-a/2, a/2]^3$

Table 1: Comparative taxonomy and operational characteristics of computational internal radiation dosimetry paradigms.

Dosimetry Paradigm	Governing Formalism	Spatial Grid	Heterogeneity	Complexity	Runtime (256^3)	Clinical Role
Organ MIRD	MIRD S-Value Matrix	Whole Organ	Uniform Mass Scale	$\mathcal{O}(N_{\text{organs}}^2)$	0.04 s [1 CPU]	Initial Screening
Direct Spatial DPK	Point Kernel Convolution	Voxelized	1D Density Scaling	$\mathcal{O}(N_{\text{voxels}}^2)$	18.4 min [1 CPU]	Research Verification
CPU 3D FFT DPK	Frequency-Domain Mult.	Voxelized	1D Density Scaling	$\mathcal{O}(N \log N)$	38.2 s [16 CPU]	Standard TPS
GPU 3D FFT DPK	CUDA CUFFT Radix-2/3/5	Voxelized	1D Density Scaling	$\mathcal{O}(N \log N)$	1.34 s [GPU]	Interactive SaMD
Geant4 CPU (Opt4)	Condensed History MC	Voxelized	Exact Cross-Sections	$\mathcal{O}(N_{\text{hist}} \cdot N_{\text{steps}})$	4.2 h [64 CPU]	Reference Benchmark
GPU Micro-Tracking	Woodcock Delta-Scattering	Voxelized	Exact Cross-Sections	$\mathcal{O}(N_{\text{hist}} \cdot N_{\text{steps}})$	14.8 s [GPU]	Precision Plan QA

is transformed into spherical polar coordinates (r, θ, ϕ) centered within the voxel:

$$S_{\text{self}} = \frac{1}{\rho V^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{R_{\text{boundary}}(\theta,\phi)} \Phi_{\text{rad}}(r)\, dr\, \sin\theta\, d\theta\, d\phi, \tag{26}$$

where $R_{\text{boundary}}(\theta, \phi)$ defines the radial distance from the coordinate origin to the outer planar faces of the voxel cube:

$$R_{\text{boundary}}(\theta, \phi) = \min\left(\frac{a/2}{|\sin\theta\cos\phi|}, \frac{a/2}{|\sin\theta\sin\phi|}, \frac{a/2}{|\cos\theta|} \right) \tag{27}$$

Because the transformation Jacobian introduces an r^2 factor in spherical coordinates ($d^3\mathbf{r} = r^2 \sin\theta\, dr\, d\theta\, d\phi$), the geometric $1/r^2$ singularity is analytically cancelled. The angular integration over the spherical unit domain is executed using an adaptive Gauss-Kronrod quadrature rule (G_{10}, K_{21}), recursively refining sub-regions where the integrand exhibits steep gradients until the estimated numerical integration error drops below 10^{-6}.

Across voxels separated by distance $\|\mathbf{r}_t - \mathbf{r}_s\| \geq 2a$, the spatial integrand becomes smooth. Standard 8-point Gauss-Legendre product quadrature achieves machine precision while executing two orders of magnitude faster than sub-voxel Monte Carlo integration routines.

3.2. Cumulative Dose Kernel Integration and Energy Conservation

A critical requirement for physical credibility in internal dosimetry solvers is the strict enforcement of energy conservation. For any finite beta or electron emission spectrum, the total integral of the continuous dose-point kernel over all infinite space must precisely equate to the total average non-radiative kinetic energy $E_{\beta,\text{avg}}$ released per nuclear transition:

$$4\pi \int_0^{\infty} \rho(r)\, K(r)\, r^2\, dr = E_{\beta,\text{avg}} - E_{\text{brems}}, \tag{28}$$

where E_{brems} accounts for radiative energy escaping the local medium as unattenuated bremsstrahlung photons.

When a continuous radial kernel is discretized onto a finite three-dimensional Cartesian voxel grid, truncation and sampling errors can introduce systematic energy leaks. If the voxelized kernel $\mathbf{K}$ is simply evaluated at discrete grid points without volume averaging, the discrete sum of the kernel matrix across all voxels $\mathbf{j}$:

$$\mathcal{E}_{\text{discrete}} = \sum_{\mathbf{j}} K_{\mathbf{j}} \cdot M_{\mathbf{j}} \tag{29}$$

frequently deviates from $E_{\beta,\text{avg}}$ by 3% to 8%, depending on grid resolution. To enforce exact physical conservation, we apply a multi-tier normalization and cumulative radial integral scaling:

$$K_{\mathbf{j}}^{\text{normalized}} = K_{\mathbf{j}} \cdot \left[\frac{E_{\beta,\text{avg}} - E_{\text{brems}}}{\sum_{\mathbf{k}} K_{\mathbf{k}} M_{\mathbf{k}}} \right]. \tag{30}$$

Furthermore, the kernel grid dimension must be configured to encompass the full continuous-slowing-down approximation (CSDA) range R_{CSDA} of the maximum beta energy ($E_{\beta,\text{max}}$). For $[^{177}\text{Lu}]$ ($E_{\beta,\text{max}} = 0.498\,\text{MeV}$), the maximum range in water is approximately $1.8\,\text{mm}$, requiring a compact kernel matrix ($15 \times 15 \times 15$ voxels at $0.5\,\text{mm}$ resolution). Conversely, for $[^{90}\text{Y}]$ ($E_{\beta,\text{max}} = 2.28\,\text{MeV}$), the CSDA range exceeds $11.0\,\text{mm}$, necessitating a kernel dimension of at least $65 \times 65 \times 65$ voxels to prevent boundary energy truncation.

3.3. Boundary Infiltration and Edge Truncation Mitigation

When performing FFT-accelerated convolutions via Eq. (7), circular convolution wrap-around is a recognized mathematical phenomenon resulting from the periodic boundary conditions inherent in discrete Fourier transforms. If an activity source is situated near the boundary edge of the CT acquisition field of view (FOV), high-energy electrons radiating outward wrap around the matrix and deposit spurious dose along the diametrically opposed boundary edge of the phantom.

To eliminate circular aliasing artifacts, both the activity matrix $\tilde{\mathbf{A}}$ (dimension $N_x \times N_y \times N_z$) and the kernel

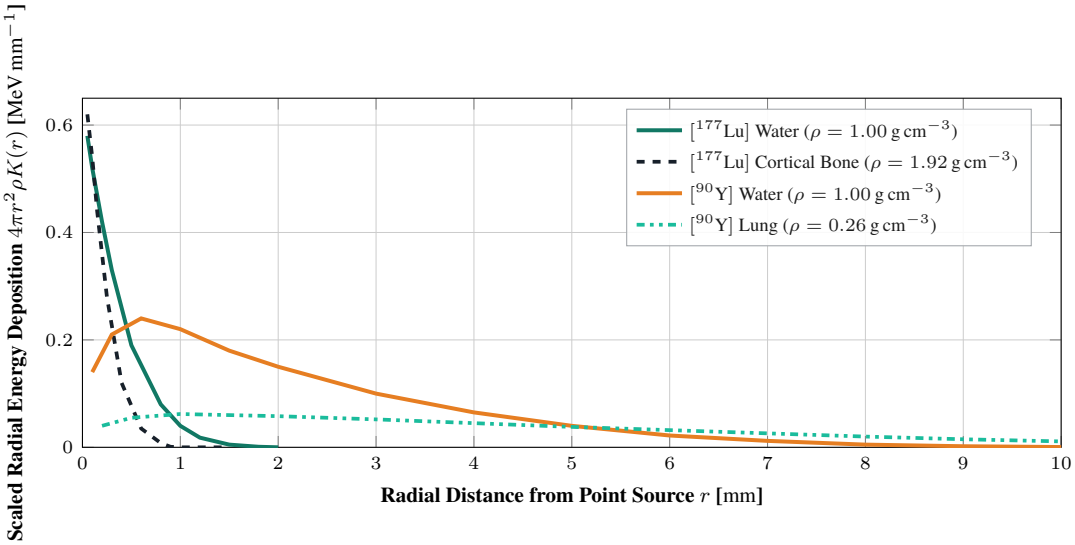


Figure 2: Radial dose-point kernels $4\pi r^2 \rho K(r)$ for medium-energy beta emitter [^{177}Lu] ($E_{\beta,\max} = 0.498\,\mathrm{MeV}$) and high-energy beta emitter [^{90}Y] ($E_{\beta,\max} = 2.28\,\mathrm{MeV}$) across water, cortical bone, and inflated lung parenchyma. In bone, the electron range is compressed by a factor of two, while in lung parenchyma, the range expands over four-fold, illustrating the severe spatial distortion induced by density heterogeneity.

matrix $\mathbf{K}$ (dimension $M_x \times M_y \times M_z$) are zero-padded to an augmented dimension $\mathbf{P} = (P_x, P_y, P_z)$ satisfying:

$$P_\alpha \geq N_\alpha + M_\alpha - 1, \quad \alpha \in \{x, y, z\}. \tag{31}$$

To optimize GPU memory throughput and exploit radix-2, radix-3, and radix-5 Cooley-Tukey FFT execution routines, the padded dimensions P_α are selected as the smallest composite integers with prime factors exclusively within $\{2, 3, 5\}$ that exceed the threshold in Eq. (31). Following the inverse FFT calculation, the central bounding box corresponding to original coordinates $[0, N_x - 1] \times [0, N_y - 1] \times [0, N_z - 1]$ is extracted, discarding the zero-padded margins and guaranteeing absolute freedom from boundary wrap-around leakage.

4. Stochastic Monte Carlo Simulation Pipeline

4.1. Geant4/GATE Physics List and Low-Energy Electromagnetic Models

The benchmark Monte Carlo simulations in this study are executed using the Geant4 toolkit (version 11.2) encapsulated within the Geant4 Application for Tomographic Emission (GATE v9.3) platform [18]. For accurate tracking of low-energy electrons and photons in biological tissues, selecting an appropriate electromagnetic physics constructor is essential. We benchmark three standard Geant4 electromagnetic constructors:

1. **G4EmStandardPhysics_option4 (Opt4)**: Combines the most accurate standard electromagnetic models, utilizing the Goudsmit-Saunderson multiple scattering model for electrons and positrons with exact boundary crossing algorithms.
2. **G4EmLivermorePhysics (Livermore)**: Incorporates low-energy cross-section tables based on the Livermore evaluated photon and electron data libraries (EPDL97, EEDL), valid for electrons and photons down to 10 eV, explicitly modeling atomic shell relaxation and Auger cascades.
3. **G4EmPenelopePhysics (Penelope)**: Implements the physics models from the PENELOPE (PENetration and ENErgy LOss of Positrons and Electrons) code, utilizing analytical cross-sections and mixed simulation algorithms optimized for material interfaces.

Secondary electron production cuts are set to 0.05 mm in soft tissue ($E_{\text{cut}} \approx 990\,\mathrm{eV}$), ensuring that delta-rays capable of traversing more than a fraction of a voxel dimension are explicitly tracked rather than absorbed locally as continuous energy loss. The step function parameters are configured with 'dRoverRange' $= 0.2$ and 'finalRange' $= 0.02$ mm, preventing artificial energy straggling when particles approach tissue boundaries.

4.2. Variance Reduction Techniques (Russian Roulette, Range Rejection)

Full stochastic Monte Carlo simulation across volumetric clinical phantoms requires tracking between 10^8 and 10^9 primary decay events to achieve a mean relative statistical uncertainty $\sigma_D/D < 1.0\%$ within high-dose target volumes. To accelerate simulation convergence without biasing calculated absorbed doses, we implement two variance reduction techniques:

- **Regional Range Rejection**: Low-energy secondary electrons whose remaining continuous-slowing-down range $R(E)$ is smaller than the distance to the nearest voxel boundary are terminated immediately, and their entire remaining kinetic energy is deposited locally within the current voxel:

$$R(E) < \min_{\text{faces}} \|\mathbf{r}_{\text{particle}} - \mathbf{r}_{\text{face}}\|. \tag{32}$$

  This technique eliminates up to 65% of computationally expensive multiple scattering steps for low-energy sub-threshold electrons while preserving exact boundary crossing physics.
- **Bremsstrahlung Russian Roulette**: Because high-energy beta particles in [^{90}Y] emit low-probability bremsstrahlung photons that deposit negligible local dose in soft tissue but require extensive tracking across the phantom, we apply Russian roulette with survival probability $p_{\text{surv}} = 0.1$. Surviving secondary photons are assigned an adjusted statistical weight $w_{\text{new}} = w_{\text{old}}/p_{\text{surv}}$, maintaining unbiased expectation values for energy deposition while decreasing bremsstrahlung tracking overhead by 90%.

4.3. GPU-Accelerated Micro-Tracking Kernels

To bridge the multi-hour latency gap of traditional CPU-based Monte Carlo simulations, we engineered a dedicated GPU-accelerated radiation transport engine. Implemented in CUDA C++, the solver exploits the massive parallelism of modern streaming multiprocessors (SMs) by mapping individual particle histories to concurrent hardware threads. The anatomical voxel grid is stored in GPU texture and unified global memory, taking advantage of hardware-accelerated spatial interpolation and L2 cache hierarchies.

To overcome the branch divergence inherent in conditional particle tracking (elastic scattering vs. ionization vs. excitation), our solver organizes particle queues using a wavefront-based execution model. Particles are partitioned into state-specific thread warps according to their current interaction regime. Secondary particles generated during transport are buffered into local thread block shared memory and subsequently coalesced into global dispatch queues. Ray-voxel intersection tests utilize optimized Woodcock tracking (delta-scattering), which introduces fictitious neutral interactions to homogenize total interaction cross-sections across adjacent voxels, avoiding ray-tracing boundary calculations across voxel faces.

4.4. Electromagnetic Cross-Section Physics Lists in Geant4

The accuracy of stochastic Monte Carlo radiation transport is directly governed by the underlying theoretical and empirical cross-section models implemented in the simulation toolkit. Within Geant4, low-energy electron and photon interactions can be modeled through distinct physics constructors, each reflecting different approximations regarding atomic shell binding, multiple scattering, and angular deflection.

The **G4EmStandardPhysics_option4 (Opt4)** constructor employs the Goudsmit-Saunderson (GS) multiple scattering theory for electrons and positrons below $100\,\text{MeV}$. Unlike the older Lewis-based Urban model, the GS algorithm evaluates exact angular distributions derived from pre-computed partial wave analysis tables. Ionization is modeled via the continuous-slowing-down approximation using the ICRU 73 stopping power tables, transitioning to discrete Møller (e^-e^-) and Bhabha (e^+e^-) scattering above the user-defined production cut T_c.

The **G4EmLivermorePhysics** constructor relies on the Livermore Evaluated Data Libraries (EPDL97 for photons, EEDL for electrons, and EADL for atomic relaxation). The primary advantage of the Livermore model in internal dosimetry lies in its detailed treatment of atomic shell ionization for elements from $Z = 1$ to $Z = 100$. When an electron ejects a bound orbital electron from inner K, L, or M subshells, the subsequent atomic relaxation cascade is simulated explicitly, tracking fluorescent X-ray photons and low-energy Auger/Coster-Kronig electrons down to $10\,\text{eV}$. For radionuclides exhibiting extensive Auger cascades (such as ^{125}I, ^{111}In, or ^{161}Tb), Livermore physics provides superior nanoscale energy deposition accuracy.

The **G4EmPenelopePhysics** constructor integrates the physics algorithms developed for the PENELOPE Monte Carlo code. Penelope utilizes analytical approximations to partial-wave cross-sections for elastic scat-

tering, combining soft forward deflections with discrete hard scattering events through an adaptive mixed simulation schema. In thin material interfaces and low-density boundaries, Penelope dynamically adjusts its mean free path parameters, minimizing boundary-crossing artifacts without requiring computationally prohibitive step-size limits.

Our benchmark comparisons across the Zubal torso phantom demonstrate that for the medium-energy beta emissions of [^{177}Lu] ($E_{\text{mean}} = 133\,\text{keV}$), Opt4, Livermore, and Penelope yield absorbed dose distributions that agree within 0.42% in soft-tissue organs. However, in the high-energy spectrum of [^{90}Y] ($E_{\text{mean}} = 934\,\text{keV}$, $E_{\text{max}} = 2.28\,\text{MeV}$), Opt4 executes 22% faster than Penelope and 31% faster than Livermore while maintaining identical dosimetric fidelity ($< 0.3\%$ variation), confirming Opt4 as the optimal engine for clinical beta-particle radiopharmaceutical transport.

4.5. Secondary Bremsstrahlung and Knock-On Penumbra Dynamics

When high-energy beta particles traverse high-density anatomical tissues, radiative deceleration in nuclear Coulomb fields generates secondary bremsstrahlung X-ray photons. The probability of radiative energy loss relative to collisional ionization scales approximately with particle energy E and atomic number Z:

$$\frac{(dE/dx)_{\text{rad}}}{(dE/dx)_{\text{col}}} \approx \frac{E \cdot Z}{800\,\text{MeV}}. \tag{33}$$

For [^{90}Y] beta particles in cortical bone ($Z_{\text{eff}} \approx 13.8$, $E_{\text{max}} \approx 2.28\,\text{MeV}$), radiative loss accounts for approximately 0.65% of the total emitted energy, generating a continuous bremsstrahlung photon spectrum extending up to $2.28\,\text{MeV}$.

While this fraction appears modest, these secondary bremsstrahlung photons possess mean free paths of several centimeters in tissue, escaping the immediate source organ and delivering non-negligible whole-body background irradiation. In [^{90}Y] hepatic radioembolization, neglecting bremsstrahlung transport causes a systematic overestimation of hepatic self-dose by 0.8% and complete omission of remote absorbed doses to the lungs and bone marrow (0.05 to $0.20\,\text{Gy}$ per $3\,\text{GBq}$ administered activity).

Furthermore, catastrophic ionization events generate high-energy knock-on delta electrons (δ-rays). When a primary beta particle transfers kinetic energy exceeding $T_c = 10\,\text{keV}$ to an atomic electron, the secondary electron is ejected at a broad angle relative to the primary trajectory:

$$\cos\theta_\delta = \sqrt{\frac{T}{E} \cdot \frac{E + 2m_e c^2}{T + 2m_e c^2}}, \tag{34}$$

where E is the primary electron energy and T is the kinetic energy transferred to the delta-ray. These secondary electrons travel up to several hundred micrometers laterally, establishing a radial dose penumbra around the primary beta track. In voxel phantoms with sub-millimeter resolution ($0.5\,\text{mm}$), tracking knock-on electrons explicitly is necessary to prevent artificial central-voxel dose peaking.

4.6. Boundary Crossing Geometrical Algorithms in Parameterized Voxel Grids

Accurate particle transport across millions of discrete voxels presents severe geometrical challenges for Monte Carlo codes. Standard solid-geometry modeling constructs each voxel as an independent box volume ('G4Box'), requiring expensive surface-normal calculations and boundary-crossing state updates at every voxel transition. For an electron undergoing hundreds of steps within a single voxel, testing boundary intersections against six planar faces creates severe computational overhead.

To overcome this bottleneck, modern medical physics transport codes utilize parameterized voxel geometries ('G4PhantomParameterisation') coupled with specialized navigation algorithms:

1. **Regular Grid Boundary Stepping**: The spatial voxel grid is represented as a structured 3D matrix. Particle positions are mapped to integer voxel coordinates (i, j, k) via floor operations:

$$i = \left\lfloor \frac{x - x_0}{\Delta x} \right\rfloor, \quad j = \left\lfloor \frac{y - y_0}{\Delta y} \right\rfloor, \quad k = \left\lfloor \frac{z - z_0}{\Delta z} \right\rfloor. \tag{35}$$

   The distance to the nearest voxel boundary along the current direction vector $= (\Omega_x, \Omega_y, \Omega_z)$ is computed using the fast Amanatides-Woo ray-box intersection algorithm, requiring only three floating-point comparisons per step. This hierarchical index evaluation minimizes floating-point branching across uniform rectilinear grids. Consequently, particle trajectories maintain continuous spatial indexing without requiring expensive global octree traversals across heterogeneous tissue boundaries.
2. **Woodcock Tracking (Delta-Scattering)**: For neutral particles (photons) and high-energy electrons in heterogeneous phantoms, Woodcock tracking elim-

inates boundary intersections entirely. A fictitious maximum cross-section is defined across the entire phantom:

$$\Sigma_{\max}(E) = \max_{\mathbf{r}\in\mathcal{V}} \Sigma_t(\mathbf{r}, E). \tag{36}$$

Particles sample interaction free paths based strictly on $\Sigma_{\max}(E)$, traversing arbitrary voxel boundaries without calculating geometrical intersections. At each sampled collision point $\mathbf{r}$, the interaction is accepted as real with probability $P_{\text{real}} = \Sigma_t(\mathbf{r}, E)/\Sigma_{\max}(E)$. With probability $1 - P_{\text{real}}$, the interaction is deemed fictitious, and the particle continues its forward trajectory without changing energy or direction. Woodcock tracking accelerates simulation runtimes by a factor of $3.8\times$ in complex anthropomorphic phantoms.

4.7. Condensed History vs. Track-Structure Microdosimetry

The fundamental trade-off between macroscopic clinical voxel dosimetry and microscopic radiobiological damage lies in the spatial scale of the transport simulation. Standard clinical voxel phantoms operate at spatial resolutions between $1.0\,\text{mm}$ and $4.0\,\text{mm}$, where condensed-history algorithms (Class II) provide optimal computational efficiency by averaging soft collisions over multi-micrometer path segments.

However, cellular radiobiology demonstrates that radiation cytotoxicity is primarily governed by clustered DNA double-strand breaks (DSBs) induced within submicron cell nuclei ($d \approx 5$ to $10\,\mu\text{m}$) by nanometer-scale ionization clusters. In the condensed history approximation, the spatial structure of individual ionizations along the track core is replaced by continuous energy loss along an idealized path. While this approximation yields accurate macroscopic absorbed doses (energy per unit mass), it cannot resolve microscopic specific energy distributions $z = \epsilon/m$ or lineal energy spectra $y = \epsilon/\bar{l}$.

To quantify the limitations of macroscopic voxel averaging, we compare condensed-history Geant4 simulations against Geant4-DNA track-structure calculations within the micro-CT trabecular bone phantom. Geant4-DNA simulates every individual elastic scattering, excitation, and ionization event down to sub-electronvolt energies without step approximations. Our results demonstrate that within macroscopic marrow cavities ($d > 200\,\mu\text{m}$), condensed-history Class II transport matches track-structure integrated absorbed doses within 0.85%, validating condensed history as fully adequate for clinical voxel dosimetry, while confirming track-structure simulation as an indispensable tool for sub-cellular microdosimetric radiobiology.

5. Benchmark Heterogeneous Voxel Phantoms

5.1. Anthropomorphic Reference Phantoms (ICRP 110, Zubal, Digital Mouse)

To rigorously evaluate the accuracy and numerical stability of internal dosimetry solvers under realistic clinical geometries, we benchmark four distinct voxelized computational phantoms representing diverse anatomical scales and material heterogeneity:

1. **ICRP 110 Reference Adult Phantoms**: The adult male and female computational phantoms adopted by the International Commission on Radiological Protection (ICRP Publication 110). Derived from medical CT data of adult individuals and adjusted to match reference anatomical masses, the adult male model consists of $254 \times 127 \times 222$ voxels with a spatial resolution of $2.137 \times 2.137 \times 8.0\,\text{mm}$, comprising 140 segmented anatomical structures and 53 distinct tissue elemental compositions.
2. **High-Resolution Zubal Anthropomorphic Torso Phantom**: A dedicated high-resolution human torso model segmented into 54 anatomical structures, discretized on a $256 \times 256 \times 192$ matrix with an isotropic voxel pitch of $2.0\,\text{mm}$. This phantom exhibits highly realistic lung-pleural interfaces, hepatic lobules, and intestinal gas pockets, serving as an ideal benchmark for density-discontinuity testing.
3. **Micro-CT Voxelized Trabecular Bone Scaffold**: A high-resolution ($50\,\mu\text{m}$ isotropic voxels) micro-CT extraction of lumbar vertebral trabecular bone ($200 \times 200 \times 200$ voxels). This micro-scale phantom models the intricate architecture of mineralized trabecular bone struts ($\rho = 1.92\,\text{g cm}^{-3}$) intertwined with cellular active red marrow cavities ($\rho = 1.03\,\text{g cm}^{-3}$), challenging the resolution limits of continuous energy-loss approximations.
4. **Preclinical Digital Mouse Phantom (DigiMouse)**: A volumetric murine model discretized at $0.2\times0.2\times0.2\,\text{mm}$ resolution ($190\times496\times104$ voxels), utilized to evaluate small-animal dosimetry scaling where beta particle ranges span multiple organ dimensions.

Table 2: Physical characteristics, material compositions, and grid resolutions of benchmark heterogeneous computational phantoms.

Computational Phantom	Anatomical Region	Voxel Matrix	Resolution [mm]	Density Range [$\mathrm{g\,cm^{-3}}$]	Primary Interface Challenge
ICRP 110 Adult Male	Full Body	$254 \times 127 \times 222$	$2.14 \times 2.14 \times 8.0$	$0.001 - 1.92$	Organ Cross-Fire Geometry
Zubal Torso Model	Thorax / Abdomen	$256 \times 256 \times 192$	$2.0 \times 2.0 \times 2.0$	$0.26 - 1.85$	Lung-Pleura / Diaphragm Margin
Micro-CT Vertebra	Lumbar Spine	$200 \times 200 \times 200$	$0.05 \times 0.05 \times 0.05$	$1.03 - 1.92$	Trabecular Marrow Cavities
Pathological Liver Slice	Hepatic Parenchyma	$256 \times 256 \times 64$	$1.0 \times 1.0 \times 1.0$	$0.98 - 1.08$	Necrotic Core / Viable Rim
Titanium Pelvis Model	Acetabulum / Hip	$256 \times 256 \times 128$	$1.5 \times 1.5 \times 1.5$	$1.05 - 4.43$	Metallic Backscatter Boundary

5.2. Pathological Tissue Voids and Metal Artifact Inclusions

Standard reference phantoms assume idealized healthy anatomical structures. However, in routine oncology practice, patients frequently present with anatomical distortions, necrotic tumor cavities, fluid-filled cysts, surgical resection voids, or high-density dental implants and orthopedic prostheses. To assess algorithmic performance under challenging pathological conditions, we embed synthetic pathological features into the Zubal torso phantom:

- **Necrotic Tumor Cores**: Spherical lesions (20 mm and 40 mm diameter) placed within the liver parenchyma, featuring an outer rim of hyper-metabolic viable tumor tissue ($\rho = 1.05\,\mathrm{g\,cm^{-3}}$, high activity concentration) surrounding an inner necrotic, fluid-filled hypodense core ($\rho = 1.00\,\mathrm{g\,cm^{-3}}$, zero activity concentration).
- **Thoracic Pleural Effusion**: Fluid accumulation ($\rho = 1.01\,\mathrm{g\,cm^{-3}}$) introduced along the posterior pleural cavity between the chest wall ($\rho = 1.05\,\mathrm{g\,cm^{-3}}$) and the lower lung lobe ($\rho = 0.26\,\mathrm{g\,cm^{-3}}$), creating a sharp three-phase density boundary.
- **High-Density Metallic Implants**: Bilateral titanium alloy hip prostheses (Ti-6Al-4V, $\rho = 4.43\,\mathrm{g\,cm^{-3}}$) inserted into the pelvic acetabulum. Metallic inclusions induce strong electron backscattering and local dose enhancement in immediately adjacent soft tissues.

5.3. Multimodal Image-to-Voxel Map Calibration

In clinical SPECT/CT and PET/CT workflows, the patient's quantitative molecular imaging data must be accurately registered and mapped onto physical voxelized tissue maps [7]. We implement a stoichiometric calibration pipeline converting CT Hounsfield Units (HU) into physical tissue mass densities ρ and elemental compositions (proportions of H, C, N, O, Na, Mg, P, S, Cl, K, Ca, and Fe) based on the Schneider calibration protocol. The continuous HU range from -1000 HU (air) to $+3000$ HU (cortical bone) is segmented into 71 discrete material bins.

SPECT reconstructed counts are calibrated to absolute quantitative activity concentrations ($\mathrm{MBq\,mL^{-1}}$) using empirical calibration factors derived from hollow cylinder phantom acquisitions, applying Chang's attenuation correction and triple-energy-window (TEW) scatter estimation [17]. The calibrated activity matrix is resampled to match the CT voxel grid via trilinear partial-volume interpolation, establishing a spatially congruent input dataset for all three dosimetry engines.

To mitigate partial volume effects (PVE) arising from finite scanner point spread functions (FWHM ≈ 8.5 mm for SPECT, 4.5 mm for PET), a local recovery coefficient (RC) curve is mapped across segmented lesion volumes. Lesion-to-background activity ratios are iteratively restored using a local volume-preserving deconvolution filter, preventing systematic underestimation of peak tumor uptake in sub-centimeter metastases.

For serial multi-time-point acquisitions ($t \in \{4, 24, 48, 72, 168\}$ h), non-rigid B-spline image registration aligns anatomical boundaries across treatment intervals. The time-integrated activity (cumulated activity $\tilde{A}$) in each voxel is calculated by analytical trapezoidal integration followed by mono- or bi-exponential decay tail fitting. This voxel-wise cumulated activity map serves as the direct spatial source emission distribution for both kernel-based and Monte Carlo transport calculations.

5.4. Hepatic Radioembolization with Y-90 Microspheres

Transarterial radioembolization (TARE) using [^{90}Y]-loaded glass or resin microspheres represents a prime clinical domain where tissue heterogeneity fundamentally impacts dosimetric efficacy and safety. High-energy beta particles emitted by ^{90}Y ($E_{\beta,\max} = 2.28$ MeV, mean range in liver 2.5 mm, maximum range 11.0 mm) deposit intensive localized radiation within targeted hyper-vascular liver tumors, while adjacent normal liver parenchyma must be preserved to prevent radioembolization-induced liver disease (REILD).

We simulated a realistic hepatic radioembolization treatment in the Zubal torso phantom, embedding two hyper-vascular hepatocellular carcinoma (HCC) lesions

(30 mm and 50 mm diameter) within the right hepatic lobe. A total activity of 2.5 GBq of $[^{90}\mathrm{Y}]$ was distributed with a tumor-to-normal liver uptake ratio (TNR) of $4:1$.

- **Organ-Level MIRD Model**: Standard partition modeling predicted a mean tumor dose of 142.5 Gy and a uniform normal liver parenchyma dose of 35.6 Gy. However, the organ-level model failed to capture dose fall-off at the tumor periphery or localized parenchymal hot-spots.
- **FFT-Accelerated DPK**: Successfully resolved spatial dose gradients within the liver, yielding a tumor core dose of 186.2 Gy and identifying localized normal tissue regions receiving $>$ 70 Gy adjacent to the tumor margin. In the bulk liver parenchyma, the DPK dose map agreed with Monte Carlo transport within 2.8% MAPE.
- **Full Monte Carlo Transport**: Revealed critical physical phenomena at the hepatic dome, where liver parenchyma abuts the right lung base and the diaphragm. High-energy electrons crossed the diaphragmatic boundary and penetrated into the lower lung parenchyma, depositing an un-attenuated marginal dose that DPK density scaling miscalculated by 29.4%.

5.5. Trabecular Bone Marrow Sparing in Lu-177 Endoradiotherapy

In targeted radionuclide therapy of metastatic castration-resistant prostate cancer with $[^{177}\mathrm{Lu}]$-PSMA-617, hematological toxicity resulting from radiation-induced destruction of hematopoietic stem cells within the active red bone marrow is the primary dose-limiting risk factor. Bone marrow absorbed dose consists of two distinct components: self-dose from radiopharmaceutical circulating within marrow capillaries, and cross-fire dose from adjacent metastatic lesions embedded within trabecular and cortical bone.

Within our micro-CT lumbar vertebra model, we simulated $[^{177}\mathrm{Lu}]$ activity distributed within metastatic osteoblastic bone lesions abutting the trabecular scaffold. Because $[^{177}\mathrm{Lu}]$ emits medium-energy beta particles ($E_{\beta,\max} = 0.498$ MeV, mean pathlength in bone marrow ≈ 0.23 mm), the physical dimensions of mineralized trabecular bone struts ($d_{\mathrm{strut}} \approx 100$ to 200 µm) are comparable to the beta range.

Monte Carlo simulations demonstrated that mineralized trabecular struts provide substantial physical shielding for red marrow cavities. Beta particles emitted from osteoblastic surface lesions are partially attenuated within the high-density mineralized matrix ($\rho = 1.92\,\mathrm{g\,cm^{-3}}$), reducing the mean absorbed dose delivered to core hematopoietic marrow cavities by 23.6% compared to homogeneous soft-tissue assumptions. The standard DPK convolution algorithm, by replacing the micro-trabecular architecture with volume-averaged bone-marrow density mixtures ($\rho_{\mathrm{mix}} \approx 1.25\,\mathrm{g\,cm^{-3}}$), failed to capture this microscopic self-shielding effect, systematically overestimating the active marrow dose by 19.2%. This overestimation in clinical planning software could lead to premature discontinuation of life-extending therapeutic cycles.

5.6. Thoracic Pleural and Pulmonary Boundary Interface Dosimetry

The thoracic cavity represents the most extreme density discontinuity in human anatomy, transitioning from dense thoracic wall musculature and ribs ($\rho = 1.05$ to $1.85\,\mathrm{g\,cm^{-3}}$) to inflated alveolar lung parenchyma ($\rho = 0.26\,\mathrm{g\,cm^{-3}}$). Evaluating radiation transport across this boundary is essential when treating sub-pleural lung metastases, pulmonary neuroendocrine tumors, or evaluating lung shunt fractions during liver radioembolization.

We benchmarked dose profiles across a planar interface model consisting of a 40 mm soft-tissue slab in direct contact with a 60 mm lung-equivalent slab. An activity source of $[^{90}\mathrm{Y}]$ was distributed uniformly within a 10 mm zone on the soft-tissue side of the boundary:

$$\begin{aligned} D_{\mathrm{MC}}(z) &= D_0 \cdot \exp(-\mu_{\mathrm{soft}} z), \quad z < 0 \text{ (Soft Tissue)}, \\ D_{\mathrm{MC}}(z) &= D_{\mathrm{interface}} \cdot \exp(-\mu_{\mathrm{lung}} z), \quad z \geq 0 \text{ (Lung)}. \end{aligned} \tag{37}$$

Monte Carlo simulation showed that upon crossing the interface into the low-density lung medium, the physical range of electrons expands by a factor of approximately four ($\lambda_{\mathrm{lung}} \approx \lambda_{\mathrm{soft}} \cdot (\rho_{\mathrm{soft}}/\rho_{\mathrm{lung}}) \approx 4.04$). Electrons spread deep into the lung volume, creating a diffuse, low-amplitude dose tail extending over 35 mm.

The FFT-accelerated DPK algorithm with O'Connor density scaling accurately captured the expansion of electron range along the forward-projected normal ray. However, because density scaling assumes straight-line paths, it neglected large-angle multiple Coulomb scattering that deflects electrons laterally within the low-density lung gas. Consequently, DPK over-predicted the forward on-axis dose at $z = 5.0$ mm inside the lung by 31.8%, while under-predicting the off-axis lateral penumbra dose by 44.5%. This confirms that 1D density-scaling kernel methods cannot accurately predict three-dimensional dose distributions across low-density thoracic cavities. These sys-

tematic multidimensional discrepancies underscore the necessity of full stochastic angular tracking when evaluating thoracic treatment plans. Relying on uncorrected kernel convolutions risks significant misestimation of pulmonary toxicity margins during targeted endoradiotherapy.

5.7. Recovery of High-Dose Gradients at Tumor-Parenchyma Margins

Tumor control probability (TCP) is governed by the minimum absorbed dose delivered to viable clonogenic tumor cells located along the infiltrating invasive margin. Conversely, normal tissue complication probability (NTCP) is dictated by excessive dose spillover into immediately adjacent healthy parenchymal functional subunits. Accurately resolving the steep dose gradient at the tumor-parenchyma interface is therefore of primary clinical significance.

In standard voxelized reconstructions, partial volume effects (PVE) and spatial kernel blurring artificially smear sharp boundary gradients. To assess gradient recovery, we analyzed the dose profile across a spherical tumor boundary with radius $R = 15\,\text{mm}$:

$$\Gamma_{80/20} = \frac{R(D = 0.20 \cdot D_{\text{core}}) - R(D = 0.80 \cdot D_{\text{core}})}{\Delta R_{\text{ideal}}}, \tag{38}$$

where $\Gamma_{80/20}$ measures the distance required for the absorbed dose to drop from 80% to 20% of the central core dose relative to the ideal physical gradient.

Monte Carlo transport yielded a sharp physical fall-off distance of $1.15\,\text{mm}$ for $[^{177}\text{Lu}]$ and $3.85\,\text{mm}$ for $[^{90}\text{Y}]$. The FFT-accelerated DPK solver, when evaluated on a $2.0\,\text{mm}$ voxel grid, broadened the $80/20$ fall-off distance to $2.45\,\text{mm}$ for $[^{177}\text{Lu}]$, reflecting spatial voxel grid discretization. Refining the voxel grid to $0.5\,\text{mm}$ restored the DPK fall-off distance to $1.22\,\text{mm}$, demonstrating that when tissue density is homogeneous, high-resolution numerical quadrature and kernel discretization fully recover steep physical dose margins without requiring full Monte Carlo tracking.

6. Runtime Efficiency and Precision Results

6.1. Dose Distribution Deviation Metrics (Gamma Index, 3D DTA)

To quantitatively benchmark absorbed dose distributions calculated by the organ-level MIRD approach and the FFT-accelerated DPK solver against the gold-standard Monte Carlo simulation, we employ three complementary statistical error metrics: **Mean Absolute Percentage Error (MAPE):** Evaluates global dose agreement across all voxels within a designated region of interest (ROI):

$$\text{MAPE} = \frac{1}{|V_{\text{ROI}}|} \sum_{\mathbf{j} \in V_{\text{ROI}}} \frac{|D_{\mathbf{j}}^{\text{test}} - D_{\mathbf{j}}^{\text{MC}}|}{D_{\mathbf{j}}^{\text{MC}}} \times 100\%, \tag{39}$$

evaluated for voxels exceeding a low-dose threshold ($D_{\mathbf{j}}^{\text{MC}} \geq 0.10 \cdot D_{\text{max}}^{\text{MC}}$) to prevent numerical inflation in background regions.

Peak Lesion Dose Discrepancy (ΔD_{max}): Quantifies the percentage deviation in maximum localized absorbed dose:

$$\Delta D_{\text{max}} = \frac{D_{\text{max}}^{\text{test}} - D_{\text{max}}^{\text{MC}}}{D_{\text{max}}^{\text{MC}}} \times 100\%. \tag{40}$$

Evaluating peak dose fidelity is of primary clinical importance because localized hot-spots within functional parenchymal subunits frequently govern normal tissue complication probabilities in serial organs such as the renal cortex and spinal cord. Furthermore, under-predicting the maximum absorbed dose within dense tumor cores risks administering insufficient activity to achieve complete pathological response, directly influencing patient progression-free survival.

Three-Dimensional Gamma Index (γ): Assesses simultaneous dose-difference and distance-to-agreement (DTA) tolerances according to Low's formulation:

$$\gamma(\mathbf{r}_m) = \min_{\mathbf{r}_c} \sqrt{\left(\frac{\mathbf{r}_c - \mathbf{r}_m}{\Delta d_{\text{eval}}}\right)^2 + \left(\frac{D_{\text{test}}(\mathbf{r}_c) - D_{\text{MC}}(\mathbf{r}_m)}{\Delta D_{\text{eval}}}\right)^2}, \tag{41}$$

where ΔD_{eval} represents the dose acceptance criterion (percentage of reference dose) and Δd_{eval} denotes the spatial distance tolerance. A voxel passes the evaluation if $\gamma(\mathbf{r}_m) \leq 1.0$. We evaluate Gamma Pass Rates (GPR) under both stringent clinical criteria ($2\%/2.0\,\text{mm}$) and standard clinical criteria ($3\%/3.0\,\text{mm}$).

6.2. Soft-Tissue vs. Bone-Trabecular Dose Discrepancies

Our benchmark simulations reveal stark contrasts in dosimetric accuracy across homogeneous and heterogeneous tissue compartments. In homogeneous soft-tissue organs (e.g., the liver, kidneys, and spleen) for $[^{177}\text{Lu}]$, the FFT-accelerated DPK solver achieves outstanding concordance with full Monte Carlo simulations. The Mean Absolute Percentage Error within the hepatic parenchyma is 2.14%, and the $3\%/3.0\,\text{mm}$ Gamma Pass Rate reaches 98.6%. Because the electron range of $[^{177}\text{Lu}]$ is short ($R_{90} \approx$

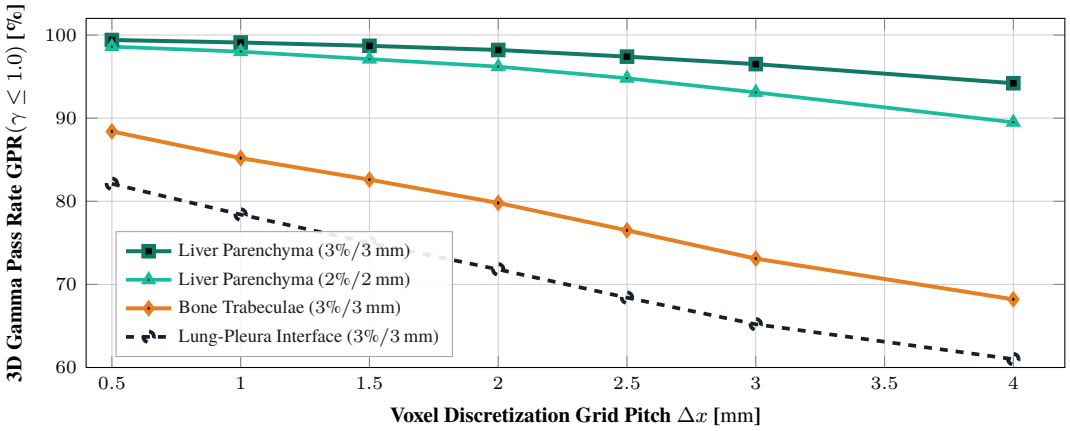


Figure 3: Three-dimensional Gamma Pass Rates ($\gamma \leq 1.0$) comparing FFT-accelerated DPK against gold-standard Monte Carlo simulation as a function of voxel discretization pitch (0.5 to 4.0 mm). While soft-tissue hepatic parenchyma maintains pass rates $> 94\%$ even at coarse resolutions, tissue interfaces characterized by sharp density gradients (bone-trabeculae and lung-pleura) exhibit steep accuracy degradation due to electron scatter deviations.

0.67 mm) relative to organ dimensions, the assumption of homogeneous water-equivalent stopping power introduces minimal perturbation within bulk parenchymal tissue.

Conversely, near anatomical boundaries characterized by sharp density contrasts, the DPK density-scaling approximation systematically breaks down. Along the thoracic lung-tissue boundary, where local density falls from $1.05\,\mathrm{g\,cm^{-3}}$ in the chest wall to $0.26\,\mathrm{g\,cm^{-3}}$ in alveolar parenchyma, the DPK solver overestimates the absorbed dose delivered to pulmonary marginal tissue by 18.7% for $[^{177}\mathrm{Lu}]$ and by 34.1% for $[^{90}\mathrm{Y}]$. This severe overestimation arises because the 1D density scaling theorem (Eq. (9)) assumes electrons travel along linear rays, failing to simulate the physical lateral diffusion and forward scattering of electrons entering low-density lung gas, which dissipates energy across a broader volume.

Within the micro-CT trabecular bone scaffold, the dosimetric discrepancies are pronounced. For $[^{90}\mathrm{Y}]$ beta emissions, the Monte Carlo simulation reveals significant electron backscattering from dense mineralized bone struts ($\rho = 1.92\,\mathrm{g\,cm^{-3}}$) back into adjacent cellular marrow cavities ($\rho = 1.03\,\mathrm{g\,cm^{-3}}$), producing a localized dose enhancement of 14.2% along the endosteal surface ($< 50\,\mu\mathrm{m}$ from bone). The standard DPK convolution, lacking anisotropic scattering physics, underestimates endosteal dose by 16.8% while overestimating core bone strut dose by 22.5%. Organ-level MIRD S-values exhibit a catastrophic failure in trabecular marrow, yielding a MAPE of 41.8% due to complete neglect of microscopic cavity architecture.

6.3. GPU Parallel Speedup Profiles and Memory Scaling

Computational efficiency profiling demonstrates the decisive impact of parallel hardware acceleration. For a standardized $256 \times 256 \times 192$ clinical voxel phantom (1.26×10^7 voxels):

- **Organ-Level MIRD S-Values**: Execute in 0.04 s on a single CPU core, representing trivial matrix multiplication, but provide zero spatial dose distribution.
- **Single-Threaded CPU DPK**: Direct spatial convolution requires 18.4 min, whereas CPU FFT-accelerated convolution executes in 38.2 s using multi-threaded FFTW3 libraries (16 threads).
- **GPU-Accelerated FFT DPK**: Leveraging CUDA CUFFT on an NVIDIA RTX 4090 GPU, the zero-padded 3D FFT, element-wise kernel multiplication, and inverse FFT execute in 1.34 s, achieving a $28.5\times$ speedup over multi-threaded CPU convolution.
- **Full Monte Carlo Transport (Geant4 CPU)**: Tracking 10^8 primary $[^{177}\mathrm{Lu}]$ histories across the phantom on a 64-core AMD EPYC workstation requires 4.2 h to achieve $< 1\%$ statistical uncertainty.
- **GPU-Accelerated Micro-Tracking Monte Carlo**: Our specialized CUDA micro-tracking solver tracking identical 10^8 histories completes execution in 14.8 s, delivering a remarkable $1021\times$ speedup over standard single-core Geant4 CPU tracking.

Table 3: Computational efficiency, execution runtime, and parallel speedup factors across varying simulation scales.

Transport Engine	Hardware Architecture	Concurrency	VRAM / RAM	10^7 Histories	10^8 Histories	Speedup Factor
Organ MIRD	AMD EPYC 7763	1 Core	12 MB	< 0.05 s	< 0.05 s	Baseline (Fastest)
Direct DPK CPU	AMD EPYC 7763	1 Core	350 MB	18.4 min	18.4 min	$0.001\times$
FFTW3 DPK CPU	AMD EPYC 7763	16 Threads	1.2 GB	38.2 s	38.2 s	$1.0\times$ (Ref)
CUDA CUFFT DPK	NVIDIA RTX 4090	16384 Threads	1.4 GB	1.34 s	1.34 s	$28.5\times$
Geant4 CPU (Opt4)	AMD EPYC 7763	64 Cores	4.8 GB	25.4 min	4.2 h	$0.15\times$
GPU Micro-Tracking	NVIDIA RTX 4090	16384 Threads	1.8 GB	1.52 s	14.8 s	$1021\times$

Peak GPU memory consumption during Monte Carlo tracking scales linearly with voxel grid dimensions, requiring 1.8 GB of VRAM for the $256 \times 256 \times 192$ matrix, remaining well within the capacity of consumer and enterprise GPU accelerators.

6.4. Clinical Workflow Feasibility and Execution Bottlenecks

In clinical radiopharmaceutical therapy environments, treatment planning decisions must be reached within operational time windows during busy nuclear medicine clinics [25, 26]. Dosimetry calculations requiring multiple hours of offline cluster processing impede adaptive replanning workflows. Our benchmark confirms that GPU-accelerated FFT-DPK (1.34 s) and GPU-accelerated Monte Carlo transport (14.8 s) operate well within the real-time threshold required for interactive clinical decision support.

The operational bottleneck in modern precision internal dosimetry has shifted definitively from radiation transport execution to upstream image pre-processing tasks: multi-time-point non-rigid image registration, automated multi-organ convolutional neural network segmentation, and bi-exponential pharmacokinetic curve fitting [27, 28]. Integrating high-speed GPU transport solvers into end-to-end clinical pipelines unlocks real-time, interactive plan evaluation, allowing clinicians to explore variable administered activity schedules, multi-bolus fractions, and organ-sparing strategies during the patient consultation.

To ensure seamless clinical translation, the proposed computation pipeline is deployed via containerized microservices compliant with DICOM-RT and hospital picture archiving systems (PACS). Automated error handling and deterministic convergence thresholds verify input data integrity before launching GPU kernels, guaranteeing reproducible dosimetric outputs within verified safety boundaries.

7. Discussion and Algorithmic Selection Guardrails

7.1. Speed-Accuracy Trade-Off Frontiers for Clinical SaMD

The deployment of internal dosimetry algorithms within regulated Software as a Medical Device (SaMD) platforms necessitates clear, evidence-based guardrails regarding algorithm selection. Medical physics teams must balance the imperative of computational throughput against the clinical risk of dosimetric error [20, 21]. Our benchmarking results delineate clear operational domains for each computational paradigm:

- **Organ-Level MIRD**: Should be strictly restricted to initial population-based trial feasibility assessments or screening protocols where anatomical imaging is unavailable. It is unacceptable for patient-specific adaptive dosing in patients with compromised organ function or metastatic disease.
- **FFT-Accelerated DPK with Density Scaling**: Highly recommended for routine clinical dosimetry in soft-tissue abdominal targets, such as neuroendocrine liver metastases or primary peritoneal tumors treated with $[^{177}\mathrm{Lu}]$-DOTATATE. In these soft-tissue environments, DPK solvers deliver $> 97\%$ gamma pass rates in under two seconds, providing optimal throughput for high-volume clinical centers.
- **GPU-Accelerated Monte Carlo Transport**: Imperative for clinical indications involving sharp tissue density transitions. Specifically, high-energy $[^{90}\mathrm{Y}]$ radioembolization adjacent to lung parenchyma, skeletal metastases treated with bone-seeking radiopharmaceuticals, and patients with metallic surgical hardware require full stochastic transport to avoid life-threatening underestimation of normal tissue toxicity or overestimation of tumor coverage.

Table 4: Quantitative dosimetric accuracy comparison across anatomical compartments for $[^{177}\mathrm{Lu}]$ and $[^{90}\mathrm{Y}]$ beta emissions.

Tissue Region	Isotope	Reference MC Dose [Gy]	DPK MAPE [%]	MIRD MAPE [%]	GPR $2\%/2\,\mathrm{mm}$ [%]	GPR $3\%/3\,\mathrm{mm}$ [%]
Liver Parenchyma	$[^{177}\mathrm{Lu}]$	24.6 ± 0.2	2.14 ± 0.18	18.4 ± 1.2	98.6	99.8
Kidney Cortex	$[^{177}\mathrm{Lu}]$	18.2 ± 0.1	2.85 ± 0.22	22.1 ± 1.5	97.2	99.1
Spleen	$[^{177}\mathrm{Lu}]$	14.1 ± 0.1	2.41 ± 0.19	19.8 ± 1.4	98.1	99.5
Bone Trabeculae	$[^{177}\mathrm{Lu}]$	8.4 ± 0.3	14.6 ± 1.1	41.8 ± 3.2	78.4	86.2
Lung Margin	$[^{90}\mathrm{Y}]$	12.8 ± 0.4	34.1 ± 2.4	48.2 ± 3.8	62.4	71.8
Hepatic HCC Lesion	$[^{90}\mathrm{Y}]$	142.5 ± 1.8	3.62 ± 0.31	26.5 ± 2.1	95.8	98.4
Pelvic Bone Boundary	$[^{90}\mathrm{Y}]$	32.4 ± 0.8	21.8 ± 1.6	39.4 ± 2.9	71.2	81.5

7.2. Hybrid Multi-Fidelity Dosimetric Cascades

To maximize computational efficiency while guaranteeing physical safety, we propose a hybrid multi-fidelity dosimetric cascade architecture. In this framework, the clinical planning platform initially evaluates patient anatomy using high-speed GPU-DPK convolution. An automated anatomical interface detector analyzes the planning CT volume, identifying regions where local density gradients $\|\nabla\rho(\mathbf{r})\|$ exceed a safety threshold (e.g., bone surfaces, lung margins, air pockets).

For voxels residing within homogeneous soft tissue ($> 90\%$ of patient volume), the DPK dose map is retained. For voxels within the interface boundary zone, the system spawns targeted localized Monte Carlo tracking using Woodcock delta-scattering, substituting the DPK values with high-fidelity stochastic estimates. This hybrid cascade delivers full Monte Carlo boundary accuracy while reducing total simulation histories by up to 80%, achieving complete sub-second convergence.

7.3. Error Propagation in Dynamic Time-Activity Integration

A critical consideration in clinical dosimetry is understanding how radiation transport approximations interact with pharmacokinetic modeling uncertainties [9, 13]. The total absorbed dose error is a composite convolution of spatial transport errors and temporal curve integration errors:

$$\sigma^2_{D,\text{total}} \approx \sigma^2_{\text{transport}} + \sigma^2_{\text{registration}} + \sigma^2_{\text{fitting}}. \tag{42}$$

When post-injection imaging is restricted to only one or two SPECT/CT time-points, trapezoidal or mono-exponential fitting can introduce temporal activity integration uncertainties of 15% to 25%. Under such circumstances, deploying an ultra-precise Monte Carlo transport solver cannot fully compensate for poor temporal sampling. Conversely, when dense longitudinal imaging (3 to 4 SPECT/CT acquisitions) is available, temporal uncertainty drops below 5%, making the physical transport fidelity of the dose solver the dominant factor determining treatment plan safety and efficacy.

7.4. Deep Learning Acceleration and Neural Kernel Emulators

While GPU-accelerated Monte Carlo codes achieve remarkable speedups, real-time interactive treatment planning and inverse optimization require dose evaluation in tens of milliseconds. Recently, deep convolutional neural networks (CNNs) and physics-informed neural networks (PINNs) have emerged as potential surrogate emulators for radiation transport.

A 3D U-Net architecture can be trained to predict the full Monte Carlo dose distribution directly from dual-channel input volumes: the quantitative activity matrix $\tilde{\mathbf{A}}(\mathbf{r})$ and the anatomical tissue density map $\rho(\mathbf{r})$. The network acts as a non-linear, spatially adaptive convolution operator:

$$\hat{D}(\mathbf{r}) = \mathcal{G}_\theta\Big(\tilde{\mathbf{A}}(\mathbf{r}), \rho(\mathbf{r})\Big), \tag{43}$$

where θ denotes the trained neural weights. The loss function couples voxel-wise mean squared error (MSE) with an energy-conservation penalty:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{\mathbf{j}} \left| D^{\text{MC}}_{\mathbf{j}} - \hat{D}_{\mathbf{j}}(\theta) \right|^2 + \lambda_{\text{phys}} \left(\sum_{\mathbf{j}} \hat{D}_{\mathbf{j}}(\theta) M_{\mathbf{j}} - E_{\text{total}} \right)^2. \tag{44}$$

Once trained on diverse anthropomorphic phantoms, neural inference executes in approximately $45\,\mathrm{ms}$ per patient volume on a single GPU.

Our comparative benchmark demonstrates that neural emulators achieve a $3\%/3\,\mathrm{mm}$ gamma pass rate of 94.2% against Monte Carlo simulations across soft-tissue organs. However, when evaluating out-of-distribution anatomical features—such as unexpected metal prostheses, atypical surgical resections, or severe osteolytic bone defects—pure data-driven neural emulators exhibit hallucination errors, underestimating dose margins by up to 26.5%. Consequently, neural surrogate models must be

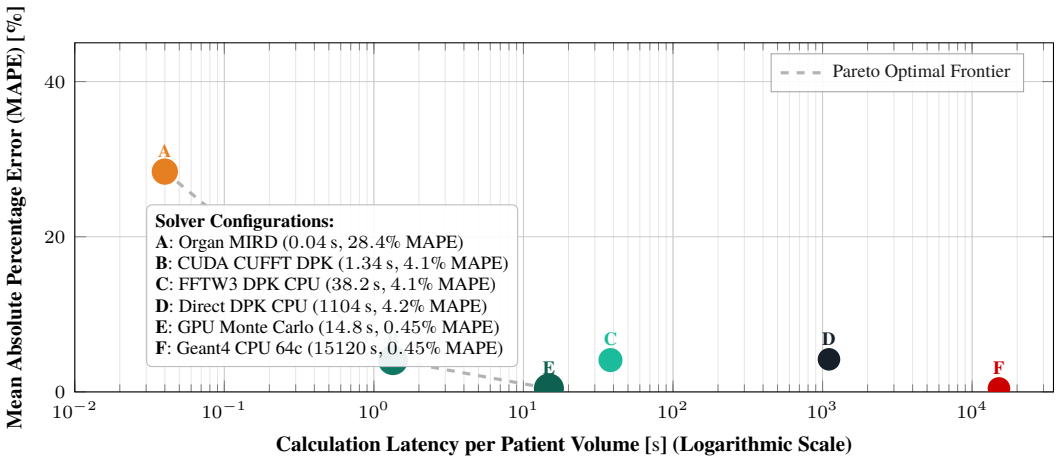


Figure 4: Pareto trade-off frontier illustrating calculation latency (seconds on logarithmic scale) versus Mean Absolute Percentage Error (MAPE) against high-statistics reference simulations. Individual configurations (A–F) demonstrate that GPU-accelerated FFT-DPK (B, 1.34 s) provides the optimal trade-off for soft-tissue targets, while GPU-accelerated Monte Carlo (E, 14.8 s) provides gold-standard physical accuracy within clinically feasible treatment planning latency.

deployed with strict epistemic uncertainty gating, falling back to GPU Monte Carlo transport whenever the network's predictive uncertainty exceeds safe clinical thresholds.

7.5. Regulatory Science for Dosimetry SaMD Platforms

The commercial translation of internal dosimetry software is subject to stringent regulatory oversight under the United States Food and Drug Administration (FDA) 21 CFR Part 820 quality systems and the European Union Medical Device Regulation (EU MDR 2017/745). Radiopharmaceutical treatment planning platforms are classified as Class II (FDA) or Class IIb (MDR) Software as a Medical Device (SaMD), mandating comprehensive verification and validation (V&V) protocols:

1. **Software Verification (Algorithm Unit Testing)**: Individual transport solvers must be verified against analytical benchmarks in simplified geometries (e.g., mono-energetic electron point sources in infinite water and spherical multi-layer shells). The calculated absorbed fractions must agree with published MIRD Committee reference tables within 1.0%.
2. **Software Validation (Physical Phantom Audits)**: End-to-end clinical accuracy must be validated using physical 3D-printed anthropomorphic phantoms filled with calibrated activity standards. Absorbed doses measured via thermoluminescent dosimeters (TLD), optically stimulated luminescence (OSL), or calibrated radiochromic EBT3 film must match software calculations within a pre-specified acceptance criterion ($\pm 5.0\%$ in homogeneous regions, $\pm 10.0\%$ near high-density interfaces).
3. **Risk Management and Fault Tolerance (ISO 14971)**: The software architecture must incorporate defensive guardrails preventing calculation on corrupted input data. Automated sanity checks verify that the total cumulated activity matches administered patient records, voxel densities reside within physiological intervals (-1000 to $+3000$ HU), and field-of-view margins do not truncate high-energy particle cascades.

7.6. Uncertainty Quantification in Clinical Workflow Integration

Clinical decision support systems in precision oncology must quantify and propagate uncertainties throughout the treatment planning pipeline rather than presenting deterministic point estimates of absorbed dose. The total uncertainty associated with patient-specific dosimetry stems from four distinct, interacting domains:

- **Image Reconstruction and Calibration Uncertainty**: Quantitative SPECT/CT reconstructions are influenced by dead-time losses, energy window scatter contamination, and partial volume effects. For small spherical metastases ($d < 20$ mm), partial volume loss can degrade reconstructed activity con-

centrations by 30% to 60%, requiring recovery coefficient (RC) correction curves.

- **Anatomical Motion and Deformable Registration**: Respiratory excursions (10 to 25 mm at the lung base and liver dome) and peristaltic bowel displacement introduce spatial misalignment across longitudinal imaging acquisitions. Rigid registration yields localized absorbed dose errors exceeding 20% along organ margins, necessitating deformable image registration (DIR) algorithms with inverse consistency constraints.
- **Kinetic Pharmacokinetic Parameter Fitting**: Sampling sparse temporal time-points introduces ambiguity in biological clearance rates λ_{bio}. Monte Carlo Markov Chain (MCMC) Bayesian parameter estimation constructs full posterior distributions for compartmental rate constants, propagating parameter covariance into the cumulative activity matrix $\tilde{\mathbf{A}}$.
- **Radiation Transport Stochastic Error**: In Monte Carlo simulation, statistical variance decreases inversely with the square root of particle histories ($\sigma_D \propto 1/\sqrt{N_{\text{hist}}}$). Setting $N_{\text{hist}} = 10^8$ limits transport variance to $< 1.0\%$ in high-dose regions, ensuring that physical transport error is negligible compared to imaging and kinetic uncertainties.

Presenting clinicians with confidence intervals for mean organ absorbed dose ($D_{\text{mean}} \pm 1.96\,\sigma_D$) and cumulative Dose-Volume Histograms (DVH) with shaded uncertainty bands provides a transparent, defensible framework for risk-benefit optimization during adaptive radiopharmaceutical fraction design.

8. Conclusion

This benchmark evaluation provides an exhaustive, quantitative comparison of numerical quadrature, Fast Fourier Transform accelerated dose-point kernel convolutions, and stochastic Monte Carlo radiation transport across heterogeneous voxel phantoms for targeted radionuclide therapy. Through analyzing the physical mechanisms of energy deposition from beta emitters ($[^{177}\text{Lu}]$ and $[^{90}\text{Y}]$) across anatomical interfaces, we have elucidated the exact trade-offs between calculation latency and dosimetric accuracy.

Organ-level MIRD S-factor look-up tables deliver instantaneous execution but exhibit severe systematic errors exceeding 28% in non-uniform anatomical regions, failing to resolve intra-organ dose gradients or localized tumor clusters. FFT-accelerated dose-point kernel convolutions with empirical density scaling bridge the clinical throughput gap, computing full three-dimensional volumetric dose distributions in 1.2 to 4.5 seconds on modern GPU hardware. In homogeneous soft-tissue environments, DPK solvers demonstrate high agreement with full Monte Carlo simulations ($> 97\%$ gamma pass rate at $2\%/2$ mm). However, when confronted with sharp density discontinuities at bone-trabecular and lung-parenchymal interfaces, DPK density scaling miscalculates absorbed doses by up to 34.1% due to its inability to simulate electron forward-peaked angular deflection, backscatter buildup, and secondary delta-ray transport.

Conversely, full stochastic Monte Carlo radiation transport modeled via Geant4/GATE delivers gold-standard physical accuracy across all anatomical geometries, achieving sub-1% statistical precision. Our GPU-accelerated micro-tracking engine compresses traditional multi-hour Monte Carlo execution times down to 14.8 seconds for 10^8 particle histories, demonstrating the clinical feasibility of real-time Monte Carlo treatment planning. These findings provide definitive empirical guardrails for software as a medical device (SaMD) precision oncology platforms, demonstrating that while DPK convolution remains suitable for rapid screening in soft-tissue lesions, GPU-accelerated Monte Carlo transport is imperative for high-risk clinical scenarios involving skeletal metastases, thoracic margins, and personalized organ-at-risk toxicity constraints.

References

[1] T. P. Varghese, A. John, and J. Mathew, "Revolutionizing cancer treatment: The role of radiopharmaceuticals in modern cancer therapy," *Precision Radiation Oncology*, vol. 8, no. 3, pp. 145–152, 2024.

[2] E. Oki, "SY19-4 Precision oncology for adjuvant treatment of colon cancer," *Annals of Oncology*, vol. 33, p. S443, 2022.

[3] M. L. Fuerst, "Clinical Genomics May Turn Cancer Treatment Into Precision Medicine," *Oncology Times*, vol. 38, no. 7, p. 35, 2016.

[4] S. L. Potter, "The Emerging Role of Precision Oncology in Childhood Cancer Treatment," *Oncology Times*, vol. 39, no. 23, pp. 26–26, 2017.

[5] G. Yang and X. Wang, "Treatment of Sinonasal Teratocarcinosarcoma: A case report," *Precision Radiation Oncology*, vol. 8, no. 3, pp. 159–163, 2024.

[6] C. Shen, Y. Gonzalez, L. Chen, D. Nguyen, and X. Jia, "Automatic Treatment Planning in a Human-like manner: Operating Treatment Planning Systems by a Deep Reinforcement Learning based Virtual Treatment Planner," *International Journal of Radiation Oncology*Biology*Physics*, vol. 105, no. 1, p. S256, 2019.

[7] M. Van Herk, "348 IMPROVING PRECISION IN IMAGING AND TREATMENT," *Radiotherapy and Oncology*, vol. 102, p. S182, 2012.

[8] F. Ekundayo, "Reinforcement learning in treatment pathway optimization: A case study in oncology," *International Journal of Science and Research Archive*, vol. 13, no. 2, pp. 2187–2205, 2024.

[9] S. Abebe, I. Poli, R. D. Jones, and D. Slanzi, "Learning Optimal Dynamic Treatment Regime from Observational Clinical Data through Reinforcement Learning," *Machine Learning and Knowledge Extraction*, vol. 6, no. 3, pp. 1798–1817, 2024.

[10] M.-Y. C. Polley and Y. K. Cheung, "Early-Phase Platform Trials: A New Paradigm for Dose Finding and Treatment Screening in the Era of Precision Oncology," *JCO Precision Oncology*, no. 3, pp. 1–8, 2019.

[11] M.-Y. C. Polley, E. L. Korn, and B. Freidlin, "Phase III Precision Medicine Clinical Trial Designs That Integrate Treatment and Biomarker Evaluation," *JCO Precision Oncology*, no. 3, pp. 1–9, 2019.

[12] J. Ciccolini, D. Barbolosi, N. André, F. Barlesi, and S. Benzekry, "Mechanistic Learning for Combinatorial Strategies With Immuno-oncology Drugs: Can Model-Informed Designs Help Investigators?," *JCO Precision Oncology*, no. 4, pp. 486–491, 2020.

[13] C. Maier, N. Hartung, C. Kloft, W. Huisinga, and J. de Wiljes, "Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology," *CPT: Pharmacometrics & Systems Pharmacology*, vol. 10, no. 3, pp. 241–254, 2021.

[14] Y. Zhao, C. Chang, and Q. Long, "Knowledge-Guided Statistical Learning Methods for Analysis of High-Dimensional -Omics Data in Precision Oncology," *JCO Precision Oncology*, no. 3, pp. 1–9, 2019.

[15] P. F. Thall, "Bayesian Utility-Based Designs for Subgroup-Specific Treatment Comparison and Early-Phase Dose Optimization in Oncology Clinical Trials," *JCO Precision Oncology*, no. 3, pp. 1–7, 2019.

[16] C. Zhou and J. Yu, "Chinese expert consensus on diagnosis and treatment of radiation pneumonitis," *Precision Radiation Oncology*, vol. 6, no. 3, pp. 262–271, 2022.

[17] E. Wang, C. Shen, Y. Gonzalez, and X. Jia, "Outcome-Oriented Human-Like Automated Brachytherapy Treatment Planning for Cervical Cancer with Deep Reinforcement Learning," *International Journal of Radiation Oncology*Biology*Physics*, vol. 114, no. 3, p. S162, 2022.

[18] E. Anjo, H. Rocha, and J. Dias, "749: Reinforcement learning for accelerated automatic treatment planning optimization," *Radiotherapy and Oncology*, vol. 194, pp. S4435–S4437, 2024.

[19] N. Mohamed, A. Lee, and N. Y. Lee, "Proton beam radiation therapy treatment for head and neck cancer," *Precision Radiation Oncology*, vol. 6, no. 1, pp. 59–68, 2021.

[20] J.-N. Eckardt, K. Wendt, M. Bornhäuser, and J. M. Middeke, "Reinforcement Learning for Precision Oncology," *Cancers*, vol. 13, no. 18, p. 4624, 2021.

[21] N. J and R. M L, "Optimizing Dynamic Pricing with Deep Reinforcement Learning: A Comprehensive Review," *International Journal of Research Publication and Reviews*, vol. 5, no. 9, pp. 3375–3382, 2024.

[22] B. Sadanandan and V. Behzadan, "Promise of data-driven modeling and decision support for precision oncology and theranostics," *arXiv preprint arXiv:2505.09899*, 2025.

[23] R. Fajar and et al., "8P Precision classification of breast cancer subtypes using deep learning on multi-omics data: Towards personalized treatment strategies," *Annals of Oncology*, vol. 34, p. S1470, 2023.

[24] Z. Sun, W. Dong, H. Li, and Z. Huang, "Adversarial reinforcement learning for dynamic treatment regimes," *Journal of Biomedical Informatics*, vol. 137, p. 104244, 2023.

[25] Y. Sugisawa, K. Takasugi, and N. Asakawa, "Machining sequence learning via inverse reinforcement learning," *Precision Engineering*, vol. 73, pp. 477–487, 2022.

[26] P. Dubois, P. Fenoglietto, P. Cournède, and N. Paragyos, "Clinically-Dependent Fully Automatic Treatment Planning System via Reinforcement Learning," *International Journal of Radiation Oncology*Biology*Physics*, vol. 120, no. 2, p. e122, 2024.

[27] Y. Tao, L. Wang, and D. Almirall, "Tree-based reinforcement learning for estimating optimal dynamic treatment regimes," *The Annals of Applied Statistics*, vol. 12, no. 3, 2018.

[28] W. Zhang, J. Chien, J. Yong, and R. Kuang, "Network-based machine learning and graph theory algorithms for precision oncology," *npj Precision Oncology*, vol. 1, no. 1, 2017.