

Hellenic Mediterranean University

Polytechnic School

Department of Electrical & Computer Engineering

In affiliated with

Foundation for Research and Technology - Hellas (FORTH)

Institute of Computer Science
Computational Biomedical Laboratory (CBML)

Diploma Thesis

Machine Learning Techniques for Accelerating Monte Carlo Simulations in Radiotherapy and Optimizing Treatment Planning

Author:

Konstantinos Salonikiou

Supervisor:

Manolis Tsiknakis — Katerina Nikiforaki — Kostas Marias

Heraklion, Crete, Greece

March 17, 2026

"Στα πιο σκοτεινά σου μονοπάτια θα είμαι εκεί,
να φωτίζω κάθε σπιθαμί απο το σκοτάδι,
να φωτίζω με αγάπη κάθε τι όπως εσύ μου έδειξες."

Αφιερωμένο στην μαμά μου.

Abstract

This thesis presents a comprehensive framework for accelerated radiotherapy dose prediction and inverse treatment plan optimization, establishing a novel paradigm that bridge deterministic radiation transport physics with modern deep learning architectures In contrast to other methods in the literature that correlate masks (PTV, OAR contours) with dose distributions by abstraction of anatomical segmentation without taking into account the actual beam physics, the proposed framework is rooted in the fundamental physical quantities governing radiation transport, allowing it to generalize to virtually any treatment geometry and planning objective.

We developed two complementary sub-components: (1) An open-source implementation of a Linear Boltzmann Transport Equation (LBTE) solver based on the discrete ordinate method with multigroup energy discretization that provide the computational means to extract physics-informed features directly from DICOM data, specifically the RTPLAN file; (2) A physics-informed neural network architecture that learns to predict dose distributions from the LBTE-derived features with clinically accurate results and orders-of-magnitude faster computation than the traditional solution algorithms. The theoretical basis spans from Quantum Electrodynamics (Klein-Nishina cross-sections for Compton scattering) to Classical Charged Particle Physics (Bethe-Bloch stopping power for electron transport), to numerical transport methods, with all cross-sections, energy deposition mechanisms, and spatial discretization schemes obtained from first principles.

The main novelty is the creation of novel input features - primarily the uncollided fluence Ψ_{unc} that include the total beam geometry, MLC aperture configurations, gantry angles, and fluence modulation patterns as computed via analytical ray-tracing similar to commercial solvers like Acuros XB. The physics-based representation of this feature set is what differentiates our approach from previous works: While previous methods would incorrectly predict dose deposition within the PTV regardless of the actual beam configuration, our framework will accurately predict the correct dose distribution for any arbitrary treatment plan, including unconventional geometries where dose is deposited outside the traditional target volumes. This capability has opened up unprecedented possibilities for inverse planning optimization using deep learning, since the network can perceive the physical characteristics of each beam and field shape.

The neural network component utilizes a carefully-designed hybrid architecture combining a 3D U-Net for local feature extraction (e.g., tissue interfaces, density gradients, penumbral regions) with a Fourier Neural Operator (FNO) for capturing global spectral properties (e.g., beam geometry, scattering kernels, long-range energy transport). Spatial attention fusion adaptively combines local and global information on a voxel-by-voxel basis, while residual learning predicts corrections to the TERMA baseline, effectively learning the scattered dose and secondary electron transport contributions that comprise the computationally expensive iterative component of transport solvers. The use of the softplus activation function allows for physics-consistent non-negative outputs while the use of gradient-based loss functions improves accuracy at charged-particle equilibrium break-downs near tissue interfaces.

Validation of the forward dose-prediction model on 52 VMAT and 26 IMRT lung-cancer patients demonstrates that the framework meets clinical quality requirements on both delivery modalities. For VMAT, the model achieves a mean gamma pass rate of 98.17% (5%/3 mm criteria, 10% dose threshold) with 92.3% of patients exceeding the 95% clinical acceptability threshold and a mean absolute error of 1.71 ± 0.17 Gy. For IMRT — the first evaluation of this framework on static-gantry delivery — the mean gamma pass rate is 89.59% with a mean absolute error of 0.84 ± 0.18 Gy, demonstrating cross-modality generalisation: the same physics-based features enable accurate dose prediction for both continuous-arc and static-gantry deliveries without any modality-specific fine-tuning. The physics-informed nature of the approach enables the maximum dose to be deposited in the PTV while minimizing the dose to organs-at-risk, ultimately leading to improved therapeutic outcomes and reduced side effects associated with treatment. The large number of parameters (109 million) is sufficient to allow for the framework to achieve inference times suitable for real-time adaptive radiotherapy applications.

The trained dose engine has been used as a "frozen" or "fixed" and "differentiable" physics module in an end-to-end inverse treatment planning pipeline (Chapter 12), and

beyond just predicting the dose to be delivered, the inverse planning network uses a new "fluence-first" architecture to predict the fluence maps and MU for each field to be delivered based solely on the anatomy of the patient, completely eliminating the need for iterative optimization using a conventional TPS. A novel structure-angle sheaf graph block is included in the inverse planning network to model the inter-beam dose interactions between fields as learned transport maps on a bipartite graph that includes signed attention. In addition, positive (support) and negative (risk) message channels allow the architecture to simultaneously maximize target coverage and minimize OAR exposure by directionally diffusing sheaves. A bilevel curriculum training method is employed to train the network. In Phase 1 of this process, the network is anchored to supervised fluence targets. In Phase 2, the network is then trained on differentiable clinical physics including the loss functions associated with the constraints defined by the dose volume histograms (DVH-constraint hinge losses) and the Adjoint Clinical Transport Prior (ACTP) that maps voxel level clinical objectives to beamlet level fluence targets via the adjoint of the sparse dose kernel previously computed. Convex fluence projection with TV (total variation) and angular smoothness regularization is then applied to the fluence predictions made by the network to ensure that they can be delivered by the MLC.

The key connection between forward prediction as outlined in Chapter 11 and inverse planning as outlined in Chapter 12 is the sparse dose transport kernel. Both the uncollided fluence and the TERMA can be computed using the same ray tracing operators derived from the Boltzmann equation used for feature extraction as described in Chapter 11. Thus the frozen dose prediction network in Chapter 11 will produce a 3D dose map directly from the predicted fluence maps in a single forward pass. The gradients from the clinical loss function will then travel backwards through the frozen dose network and the sparse kernels to reach the fluence predictor; thus completing the optimization loop without having to invoke a traditional treatment planning system (TPS).

A validation study was completed on a subset of 50 VMAT lung cancer patients (AIRTP, 60 Gy/30 fr) demonstrates clinically competitive results with subsecond inference ($\sim$323 ms per plan): gamma pass rate of $85.20 \pm 7.96\%$ (5%/3 mm, 10% threshold), OAR sparing that matches or exceeds clinical reference plans (Lung V20 -0.5 pp, Cord D0.03cc $+12.0$ pp pass-rate improvement), and a characteristic coverage–protection trade-off where PTV V95%Rx pass rate is 14 pp lower than reference clinical plans but 4 out of 7 OAR objectives were achieved equal or higher rates. Most notably, the most clinically challenging case (a patient with bilateral pulmonary metastasis) had the lowest gamma score but produced a physically valid plan covering all of the target volumes appropriately protecting the organs at risk, which serves as direct evidence that the model learned the "goal" of radiotherapy as opposed to simply learning the patterns in the ground truth plan.

Keywords: Radiotherapy, Dose prediction, Inverse planning, Boltzmann transport equation, Physics-informed neural networks, Fourier Neural Operator, U-Net, Deep learning, Gamma index, VMAT, IMRT, AXB, MC, Lung cancer, Real-time optimization

Contents

Nomenclature **12**

Chapter 1 Introduction **14**
1.1 Motivation and Clinical Context . . . 14
1.2 Problem Statement . . . 15
1.3 Thesis Objectives and Contributions . . . 16

Chapter 2 Clinical Background: Intensity-Modulated Radiotherapy and Volumetric Modulated Arc Therapy **18**
2.1 Introduction . . . 18
2.2 Fundamental Geometric and Dosimetric Concepts . . . 18
2.2.1 Source-to-Axis Distance and Isocentric Geometry . . . 18
2.2.2 Field Size Definition and Beam Collimation . . . 19
2.2.3 Isodose Curves: Definition and Clinical Utility . . . 19
2.2.4 Monitor Units and Dose Calibration . . . 20
2.2.5 Photon Beam Energy Selection . . . 21
2.3 Multileaf Collimator: Design, Constraints, and Dynamic Motion . . . 22
2.3.1 Physical Design and Geometric Properties . . . 22
2.3.2 Mechanical Constraints and Leaf Motion Dynamics . . . 23
2.3.3 Control Points and Delivery Synchronization . . . 24
2.4 Intensity-Modulated Radiotherapy: Technical Implementation . . . 24
2.4.1 Beam Arrangement and Optimization Strategy . . . 24
2.4.2 Monitor Units and Delivery Efficiency . . . 26
2.4.3 Dosimetric Characteristics for Lung Cancer . . . 26
2.5 Volumetric Modulated Arc Therapy: Principles and Implementation 27
2.5.1 Historical Development and Technical Innovation . . . 27
2.5.2 Arc Geometry and Control Point Sampling . . . 28
2.5.3 Delivery Efficiency and Monitor Unit Utilization . . . 30
2.5.4 Dosimetric Comparison: VMAT versus IMRT for Lung Cancer . . . 30
2.6 Clinical Implementation Considerations for Lung Cancer Radiotherapy 31
2.6.1 Technique Selection Guidelines . . . 31
2.6.2 Motion Management Integration . . . 31
2.6.3 Quality Assurance and Delivery Verification . . . 32
2.7 Summary . . . 32

Chapter 3 Literature Review: Photon Dose Calculation Algorithms **34**
3.1 Introduction . . . 34
3.2 Classification Framework for Dose Calculation Algorithms . . . 34
3.2.1 Type A Algorithms: Correction-Factor Methods . . . 35
3.2.2 Type B Algorithms: Kernel-Based Convolution and Superposition Methods . . . 35
3.2.3 Type C Algorithms: Rigorous Radiation Transport Solvers . 38
3.3 Comparative Algorithm Performance and Clinical Selection Criteria . 41
3.3.1 Quantitative Accuracy Benchmarks . . . 41
3.3.2 Clinical Selection Guidelines by Treatment Site . . . 41
3.4 Computational Efficiency and Clinical Workflow Integration . . . 42
3.5 Current Limitations and Future Directions . . . 43
3.6 Summary . . . 44
3.7 Research Gap and Thesis Contribution . . . 44

3.7.1 Lack of Open-Source Deterministic Transport Implementation 44
3.7.2 Computational Cost Prohibits Inverse Planning Integration 44
3.7.3 No Integration with Physics-Informed Machine Learning . . 44

Chapter 4 Machine Learning for Dose Prediction 46
4.1 Introduction . . . 46
4.2 Foundations of Dose Prediction and Accelerated Dose Calculation . . 47
4.3 Deep Learning Architectures for Dose Prediction . . . 48
4.3.1 U-Net and Convolutional Neural Networks . . . 48
4.4 Multi-Task Learning: Simultaneous Dose and Fluence Prediction . . 49
4.5 Transformer-Based Architectures . . . 49
4.6 Generative and Physics-Informed Models . . . 50
4.6.1 Generative Adversarial Networks . . . 50
4.7 Diffusion Models . . . 51
4.8 Physics-Informed Neural Networks . . . 52
4.9 Two-Stage Physics-Guided Planning . . . 54
4.10 Uncertainty Quantification and Evidential Learning . . . 55
4.11 Comparative Performance and Clinical Impact . . . 56
4.12 Challenges, Limitations, and Future Directions . . . 57
4.12.1 Data Scarcity and Heterogeneity . . . 57
4.12.2 Non-Uniqueness of Optimal Solutions . . . 58
4.13 Interpretability and Clinical Trust . . . 58
4.14 Deliverability and Physical Constraints . . . 58
4.15 Summary . . . 59

Chapter 5 Lung Cancer Radiotherapy: Characteristics and Challenges 61
5.1 Clinical Applications and Treatment Modalities . . . 61
5.2 Unique Dosimetric Challenges in Lung Radiotherapy . . . 61
5.3 Comparative Evaluation of Dose Calculation Algorithms . . . 62
5.3.1 Pencil Beam and Correction-Based Algorithms . . . 62
5.3.2 Accuracy Limitations in Lung . . . 62
5.4 Anisotropic Analytical Algorithm (AAA) . . . 63
5.4.1 Performance in Lung SBRT . . . 63
5.5 Acuros XB (AXB): Grid-Based Boltzmann Transport Solver . . . 64
5.5.1 Validation Against Monte Carlo and Measurements . . . 64
5.5.2 Computational Efficiency . . . 65
5.6 Monte Carlo Simulation: The Gold Standard . . . 65
5.7 Summary of Algorithm Performance in Lung . . . 65
5.8 The Uniqueness and Significance of Acuros XB Implementation . . . 66
5.8.1 Scarcity of LBTE-Based Implementations . . . 66

Chapter 6 Theoretical Foundation: Radiation Transport Physics 67
6.1 Historical Context and Clinical Importance . . . 67
6.2 Radiotherapy Treatment Planning Overview . . . 67
6.3 Types of Ionizing Radiation . . . 67
6.4 Energy Ranges in Medical Physics . . . 68
6.5 Megavoltage Photon Beam Production in Medical Linear Accelerators 68
6.5.1 Electron Acceleration and Waveguide Transport . . . 68
6.6 Bremsstrahlung Target and Beam Modulation . . . 68
6.7 Photon Interactions in Tissue . . . 69
6.8 Radiobiological Basis of Radiotherapy . . . 69
6.8.1 Mechanisms of DNA Damage Induction . . . 69

6.9 DNA Double-Strand Breaks and Cellular Fate . . . 70
6.10 Interaction Cross Section Units . . . 72
6.11 Photon Interactions with Matter . . . 72
6.11.1 Overview of Interaction Processes . . . 72
6.12 Photoelectric Effect . . . 72
6.13 Total Photon Attenuation . . . 74
6.14 Electron Interactions with Matter . . . 74
6.15 Dose Quantities and Units . . . 76
6.15.1 Absorbed Dose . . . 76
6.16 Kerma . . . 76

Chapter 7 The Boltzmann Transport Equation 77
7.1 General Form . . . 77
7.2 Coupled Photon-Electron System . . . 77

Chapter 8 The Acuros XB Algorithm 78
8.1 Algorithm Overview . . . 78
8.2 Discrete Ordinates (S_N) Method . . . 78
8.3 Energy Groups (Multigroup) . . . 78
8.4 Spatial Discretization . . . 80
8.5 Legendre Expansion of Scattering . . . 80
8.6 Mathematical Framework: Spherical Harmonics and Angular Discretization . . . 80
8.6.1 Coordinate System and Notation . . . 80
8.7 Angular Fluence Expansion . . . 80
8.8 Discrete Ordinates Quadrature . . . 82
8.9 Scattering Kernel Theory . . . 82
8.9.1 General Scattering Operator . . . 82
8.10 Legendre Moments of Scattering . . . 82
8.11 Cross-Section Generation Pipeline . . . 83
8.11.1 Multigroup Energy Discretization . . . 83
8.11.2 Cross-Section Physics Models . . . 83
8.11.3 Material Library: 44 Materials . . . 102
8.11.4 Cross-Section Visualizations . . . 103
8.11.5 Solver Consumption: From YAML to Voxel Fields . . . 111
8.11.6 Per-Material Data Structure . . . 112
8.12 Energy Deposition and Dose Calculation . . . 113
8.13 Energy Deposition Mechanisms . . . 113
8.14 Restricted vs. Unrestricted Stopping Power . . . 113
8.15 Energy Deposition Cross Section . . . 113
8.16 Dose Calculation Formula . . . 113
8.17 Numerical Methods for Transport . . . 114
8.18 Spatial Discretization . . . 114
8.19 Transport Sweep Algorithm . . . 114
8.20 Diamond Difference Scheme . . . 114
8.21 Iterative Solution Strategy . . . 116
8.22 Required Cross Sections . . . 117
8.23 Data Sources . . . 117
8.24 Multigroup Condensation . . . 117
8.25 Key Parameters of the Acuros XB Algorithm . . . 117
8.26 Transport Cutoff Energies . . . 117
8.27 Energy Group Structure . . . 118

8.28 Angular Quadrature Order (Discrete Ordinates) . . . 118
8.29 Scattering Expansion (Legendre Order) . . . 119
8.30 Continuous Slowing Down (CSD) and Stopping Power . . . 119
8.30.1 Dose Calculation Formula . . . 119
8.31 Spatial Discretization and Numerical Stability . . . 120
8.31.1 Linear Discontinuous Galerkin Method . . . 120
8.31.2 Positivity Fix and Step Fallback . . . 120
8.32 Summary of Key Parameters . . . 121

Chapter 9 Implementation Details 122
9.1 Scope and Goal . . . 122
9.2 Notation . . . 122
9.3 Acuros XB Equations from Reference Documentation . . . 122
9.4 Solver Equations (Implemented Form) . . . 123
9.5 Discretization and Numerical Methods . . . 124
9.5.1 Angular Quadrature . . . 125
9.5.2 Spatial Transport Sweep . . . 125
9.5.3 Scattering Source Evaluation . . . 125
9.6 Mathematical Derivation: From LBTE to Iterative Solver . . . 125
9.6.1 Stage 1: Operator Formulation of the Continuous LBTE . . . 125
9.6.2 Stage 2: Angular and Energy Discretization . . . 126
9.6.3 Stage 3: Source Decomposition (Uncollided/Collided Split) . . . 127
9.6.4 Stage 4: Fixed-Point Iteration Construction . . . 128
9.6.5 Stage 5: Electron Transport with CSD Energy Coupling . . . 128
9.6.6 Summary: Complete Mapping from A19–A36 to Iteration Scheme . . . 129
9.7 Algorithmic Workflow . . . 130
9.7.1 Photon Transport Iteration . . . 130
9.7.2 Electron Transport with Energy Marching . . . 130
9.7.3 Solution Pipeline . . . 130
9.7.4 Complete Coupled Photon-Electron Dose Calculation Algorithm . . . 131
9.8 Transport Along Characteristic Lines and Quadrature Properties . . . 135
9.8.1 Characteristic Line Formulation . . . 135
9.8.2 Angular Quadrature Weight Properties . . . 135
9.8.3 Multigroup Cross-Section Averaging . . . 136
9.9 Detailed Scattering Operator Formulation . . . 136
9.9.1 Precomputed Weighted Legendre Matrices . . . 136
9.9.2 Full Photon-Photon Scattering Operator . . . 137
9.9.3 Photon-to-Electron Coupling Operator . . . 137
9.9.4 Electron-Electron Scattering Operator . . . 137
9.9.5 Coupled Electron Scattering During Source Iteration . . . 137
9.10 Ray Tracing Geometry and First-Collision Sources . . . 138
9.10.1 BEV Sampling and Ray Geometry . . . 138
9.10.2 Siddon Raymarching Algorithm . . . 138
9.10.3 Beam Cell Fluence Prefactor . . . 138
9.10.4 Angular Bin Mapping . . . 139
9.10.5 First-Collision Sources . . . 139
9.11 Parallel Sweep Algorithm and Domain Decomposition . . . 139
9.11.1 Octant Decomposition for Parallel Processing . . . 139
9.11.2 Wavefront Parallelism . . . 139

9.11.3 Spatial Domain Decomposition . . . 140
9.11.4 Load Balancing Metric . . . 140
9.11.5 GPU Thread Mapping . . . 140
9.12 Boundary Conditions, Convergence Criteria, and Error Propagation . 141
9.12.1 Vacuum Boundary Conditions . . . 141
9.12.2 Convergence of Source Iteration . . . 141
9.12.3 Error Propagation in Spatial Discretization . . . 142
9.13 Preprocessing and Source Construction Algorithms . . . 142
9.13.1 CT Data to Transport Coefficients . . . 142
9.13.2 RTPLAN to Uncollided Photon Flux . . . 143
9.14 Physical Quantities and Units Reference . . . 143
9.15 Implementation Insights and Numerical Validation . . . 143
9.15.1 Memory-Efficient Scattering via Material Indirection . . . 143
9.15.2 Half-Segment Attenuation Correction . . . 144
9.15.3 Energy Conservation Constraint . . . 144
9.15.4 Photon Downscatter Constraint . . . 144
9.15.5 Mean Free Path Validation . . . 144
9.16 Computational Optimization . . . 144
9.17 Beam Model and Source Sampling . . . 145
9.18 IEC-61217 Coordinate System and Mathematical Framework . . . 145
9.18.1 Coordinate System Definitions . . . 145
9.18.2 Gantry Rotation Kinematics . . . 145
9.18.3 Collimator and Couch Rotations . . . 146
9.18.4 Beam Basis Vector Derivation . . . 146
9.18.5 Explicit Rotation Matrices . . . 147
9.19 Material Tables and Element Coverage . . . 147

Chapter 10 Clinical Dataset and Validation Result 149
10.1 Lung Cancer Clinical Dataset . . . 149
10.2 Dataset Generation and Characteristics . . . 149
10.3 Beam Configuration and Delivery Parameters . . . 152
10.4 Organs at Risk and Helper Structures . . . 159
10.5 Plan Quality Evaluation Metrics . . . 159
10.6 Dosimetric Characteristics . . . 160
10.6.1 Organ-at-Risk Radiosensitivity and Molecular Mechanisms . 160
10.7 Density Heterogeneity and Transport Considerations . . . 162

Chapter 11 Physics-Informed Dose Prediction via Hybrid 3D U-Net and Fourier Neural Operator Architecture 164
11.1 Physical Foundation: From Uncollided Fluence to Dose Prediction . 164
11.2 Novelty of the Proposed Approach . . . 165
11.3 Interpretation: The Network Learns Scattered Dose . . . 166
11.4 Architecture Design Philosophy . . . 166
11.5 Input Feature Description . . . 167
11.6 Network Architecture: Dual-Branch Hybrid Design . . . 168
11.6.1 Positional Encoding . . . 170
11.6.2 Enhanced U-Net Branch (Local Features) . . . 170
11.6.3 Enhanced FNO Branch (Global Features) . . . 170
11.6.4 Spatial Attention Fusion . . . 171
11.6.5 Training Configuration . . . 172
11.6.6 Mathematical Formulation Summary . . . 172
11.7 Training Dataset and Generalization . . . 172

11.8 Training Convergence . . . 173
11.9 Validation Results and Gamma Analysis . . . 173
11.9.1 Best-Case Clinical Examples: VMAT and IMRT . . . 176

Chapter 12 Fluence-First Inverse Planning via Rotary Transformers and Sheaf-Structured Graph Diffusion 179
12.1 Introduction . . . 179
12.2 Related Work . . . 179
12.3 Problem Formulation . . . 180
12.4 System Architecture . . . 180
12.5 Input Representation and Encoding . . . 180
12.5.1 Multi-Modal Inputs . . . 180
12.5.2 K-Statistics Per Angle (12-dim) . . . 183
12.5.3 Structure-Angle Edge Features (16-dim) . . . 183
12.5.4 BEV Convolutional Encoder . . . 183
12.5.5 Token Composition . . . 183
12.6 Canonical Representation on $S^1 \times (u, v)$ . . . 183
12.6.1 Angular Fourier Basis . . . 184
12.6.2 Spatial DCT Basis . . . 184
12.6.3 Encode and Decode Maps . . . 184
12.7 Rotary Transformer Backbone . . . 185
12.7.1 Cross-Attention with Spatial Tokens . . . 185
12.7.2 Rotary Self-Attention Layers . . . 185
12.7.3 Rotary Position Embedding (RoPE) . . . 185
12.8 Structure-Angle Sheaf Graph Block . . . 185
12.8.1 Motivation . . . 185
12.8.2 Graph Topology . . . 186
12.8.3 Signed Sheaf Attention . . . 186
12.8.4 Sheaf Transport Maps and Diffusion . . . 187
12.8.5 Sheaf Laplacian: Formal Definition . . . 188
12.8.6 Sheaf Diffusion Equation and Oversmoothing Control . . . 189
12.9 Fluence-First Decoder and Convex Projection . . . 189
12.9.1 FiLM-Conditioned Spatial Decoder . . . 189
12.9.2 Residual Branch with Bounded Gain . . . 189
12.9.3 MU Simplex Head . . . 190
12.9.4 Convex Fluence Projection . . . 190
12.10 Differentiable Physics Module . . . 191
12.10.1 Sparse Dose Transport . . . 191
12.10.2 Derivation of Transport Kernels from Boltzmann Transport 191
12.10.3 Frozen Dose Network . . . 192
12.10.4 Bilevel Inner Refinement . . . 192
12.11 Loss Functions and Two-Phase Training . . . 193
12.11.1 Phase 1: Anchor Losses (Supervised) . . . 193
12.11.2 Phase 2: Anchor + Physics Losses . . . 194
12.11.3 Dynamic Gradient Balancing . . . 196
12.12 Experiments . . . 196
12.12.1 Dataset and Implementation . . . 196
12.12.2 Training Dynamics . . . 196
12.12.3 Bilevel Loss Landscape . . . 197
12.12.4 Validation Metrics on Trimmed Cohort . . . 197
12.12.5 Target Coverage and Plan Quality . . . 197

12.12.6 OAR Endpoint Metrics . . . 201
12.12.7 Global Similarity and Gamma . . . 201
12.12.8 Clinical Objective Pass Rates . . . 201
12.12.9 Clinical Case Examples . . . 201
12.12.10 Comparison to Recent Literature . . . 201
12.12.11 Worst-Case Clinical Scenario: Bilateral Multi-Metastatic Cancer . . . 204
12.12.12 Fast Lung Inverse Planning Landscape (Related Work) . . . 207
12.12.13 Cross-Study Key Metrics (Including Ours) . . . 207
12.12.14 Multi-site Inverse-Planning Comparison . . . 209

Chapter 13 Conclusion **211**
13.1 Physics-Informed Feature Extraction and the Novelty of Ψ_{unc} . . . 211
13.2 Forward Dose Prediction: Architecture and Validation . . . 211
13.3 Inverse Planning: Learning Radiotherapy Goals, Not Copying Plans . 212
13.4 Limitations and Future Directions . . . 213
13.5 Closing Remarks . . . 213

Appendices **215**

A Gamma Index Analysis **215**
1.1 Theoretical Foundation of Gamma Analysis . . . 215
1.1.1 The Dose Distribution Comparison Problem . . . 215
1.2 Mathematical Definition of the Gamma Index . . . 215
1.3 Geometric Interpretation . . . 215
1.4 Pass/Fail Criterion . . . 215
1.5 Gamma Pass Rate Calculation . . . 216
1.6 Computational Implementation of Gamma Index . . . 216
1.7 Naive Algorithm ($O(N^2)$ complexity) . . . 216
1.8 Search Radius Optimization . . . 216
1.9 Fast Gamma Algorithm (Wendling et al.) . . . 217
1.10 Implementation in Our Solver . . . 217
1.11 Global vs Local Gamma . . . 218
1.12 Global Gamma (Recommended) . . . 218
1.13 Local Gamma . . . 218
1.14 Dose Threshold . . . 218

B Source Model and Beam Fluence Calculation **219**
2.1 Beam Energy Spectrum . . . 219
2.2 Angular Fluence Distribution . . . 219
2.3 MLC Aperture Modeling . . . 219
2.4 Multi-Leaf Collimator Geometry . . . 219
2.5 MLC Aperture Function . . . 220
2.6 VMAT Fluence Integration . . . 220
2.7 Control Point Interpolation . . . 220
2.8 Cumulative Meterset Weight . . . 220
2.9 Source Building Algorithm . . . 221
2.10 Angular Quadrature Assignment . . . 221
2.11 Siddon Ray Tracing for Voxel Coverage . . . 221

C Materials Library Data Structure **223**
3.1 Energy Group Structure 223
3.1.1 Photon Energy Groups 223
3.2 Electron Energy Groups 223
3.3 Legendre Expansion Orders 223
3.4 Material Entry Format 224

Nomenclature and List of Symbols

Symbol	Description
Greek Letters	
Ψ	Angular fluence [particles cm^{-2} s^{-1} sr^{-1}]
Ψ_{unc}	Uncollided angular fluence
Ψ_{coll}	Collided (scattered) angular fluence
Φ	Scalar fluence (integrated over angle) [particles cm^{-2} s^{-1}]
σ_t	Total macroscopic cross-section [cm^{-1}]
σ_s	Scattering cross-section [cm^{-1}]
σ_a	Absorption cross-section [cm^{-1}]
$\sigma_{\gamma\gamma}$	Photon-photon (Compton) scattering cross-section
$\sigma_{\gamma e}$	Photon-to-electron transfer cross-section
σ_{ee}	Electron-electron (Møller) scattering cross-section
μ	Direction cosine, $\mu = \cos\theta$
θ	Polar angle
ϕ	Azimuthal angle
ρ	Mass density [g cm^{-3}]
τ	Optical thickness, $\tau = \sigma_t \cdot s$
α	Reduced photon energy, $\alpha = E_\gamma / m_e c^2$
β	Particle velocity relative to speed of light, $\beta = v/c$
γ	Lorentz factor, $\gamma = 1/\sqrt{1-\beta^2}$
Latin Letters	
D	Absorbed dose [Gy]
E	Particle energy [MeV]
T	Kinetic energy [MeV]
S_R	Restricted stopping power [MeV cm^{-1}]
S_{coll}	Collisional stopping power
S_{rad}	Radiative stopping power
$\mathbf{r}$	Position vector
$\mathbf{\Omega}$	Unit direction vector
G_γ	Number of photon energy groups
G_e	Number of electron energy groups
L_γ	Legendre expansion order for photon scattering
L_e	Legendre expansion order for electron scattering
M	Number of discrete ordinates directions, $M = 2N^2$
N	Discrete ordinates order (S_N)
P_ℓ	Legendre polynomial of degree ℓ
Y_ℓ^m	Spherical harmonic function
w_m	Quadrature weight for direction m
q	Volumetric source term [particles cm^{-3} s^{-1}]
I	Mean excitation energy [eV]
Z	Atomic number
A	Atomic mass number
Subscripts and Superscripts	
g, g'	Energy group indices
m	Discrete ordinate (direction) index

(continued on next page)

(continued from previous page)

Symbol	Description
ℓ	Legendre moment index
i, j, k	Spatial voxel indices
in, out	Incoming and outgoing cell faces
unc, coll	Uncollided and collided components
Operators	
∇	Gradient operator
$\mathbf{\Omega} \cdot \nabla$	Streaming operator
T	Transport operator
S	Scattering operator
Acuros XB Solver Symbols	
g, k	Photon (g) and electron (k) energy group indices
$\sigma_{s,\ell}$	Legendre moment of scattering cross section
σ^e_{ED}	Energy deposition coefficient for dose
q^γ, q^e	External photon/electron sources
$q^{\gamma\gamma}, q^{\gamma e}, q^{ee}$	Scattering/production source terms
ΔE_k	Electron energy group width
$A(u, v)$	BEV aperture transmission
dA	BEV cell area
SAD	Source-to-axis distance
$K^{(\ell)}$	Legendre folding matrix for moment ℓ
T^γ	Discrete photon sweep operator
Constants	
c	Speed of light, 2.998×10^{10} cm/s
m_e	Electron rest mass, 0.511 MeV/c^2
r_e	Classical electron radius, 2.818×10^{-13} cm
N_A	Avogadro's number, 6.022×10^{23} mol^{-1}
Acronyms	
LBTE	Linear Boltzmann Transport Equation
S_N	Discrete Ordinates method (order N)
CSD	Continuous Slowing Down
STL	Square Tchebychev-Legendre (quadrature)
DD	Diamond Difference (spatial scheme)
CSDA	Continuous Slowing Down Approximation
HU	Hounsfield Units (CT number)
TPS	Treatment Planning System
MLC	Multi-Leaf Collimator
VMAT	Volumetric Modulated Arc Therapy
DVH	Dose-Volume Histogram
PTV	Planning Target Volume
OAR	Organ at Risk

Chapter 1 Introduction

1.1 Motivation and Clinical Context

Note on Scope. The registered title of this diploma thesis is *Machine Learning Techniques for Accelerating Monte Carlo Simulations in Radiotherapy and Optimizing Treatment Planning*. During the course of the project, publicly available clinical dose data computed with the Acuros XB deterministic solver (an LBTE-based algorithm equivalent in accuracy to Monte Carlo) became accessible. We determined that building the entire framework on this data was substantially more advantageous: it yields faster inference, higher clinical accuracy, and direct compatibility with LBTE physics, producing a more impactful and immediately deployable contribution than a Monte Carlo acceleration approach alone. A formal title amendment was intended — the scientific work carried out corresponds to the title *Physics-Informed Neural Acceleration of Boltzmann Transport Equation Solvers for Radiotherapy Dose Calculation and Inverse Planning* — however, the next departmental council session at which title changes could be approved was delayed, and submission could not wait as it was required for the award of the degree. The registered title was therefore left unchanged for bureaucratic reasons. It is worth noting that the deviation is minimal in scope: the only algorithmic difference is the substitution of a stochastic Monte Carlo solver with a deterministic LBTE solver (Acuros XB), both of which solve the same underlying radiation transport problem and achieve equivalent dosimetric accuracy.

External beam radiotherapy plays a critical role in the treatment of lung cancer, with over 60% of patients receiving radiation as part of their curative or palliative therapy, and the advancement of new technologies, including but not limited to IMRT and VMAT, has significantly increased dose conformity and enabled higher doses to be delivered to tumors while avoiding adjacent OAR's [1, 2]. However, the degree to which the clinical success of these advanced technologies relies upon accurate dose calculation algorithms is critical to their success. Errors of 5–10% in areas of electronic disequilibrium are common and can result in geographic misses or unacceptable toxicity to OAR's when interfacing tumor and lung tissue[3, 4].Radiotherapy seeks to provide the maximum dose possible to the PTV while providing the least amount of exposure to OAR's; therefore, maximizing therapeutic efficacy, minimizing toxic side effects of treatment, and improving overall patient outcome is directly related to the fundamental objective of radiotherapy.

In addition to the computational challenges of accurately calculating dose distributions in radiotherapy, there are also computational challenges of performing multiple dose calculations. Thousands of dose calculations are required during the optimization process of inverse treatment planning, adaptive radiotherapy may require daily image guided replanning, and emerging applications of real-time MRI guided radiotherapy will require dose calculations to be performed within seconds. For lung cancer specifically, an additional challenge is posed by respiratory motion: the tumor and surrounding anatomy moves continuously throughout the delivery of the treatment, thus requiring real-time plan optimization that can account for anatomical changes within the treatment fraction. Current clinical workflows do not allow for real-time adaptation, and represent a significant limitation in the treatment of thoracic malignancies where motion management is of utmost importance.

These limitations have driven the development of accelerated dose calculation methods that maintain the same level of clinical accuracy as conventional methods, but significantly reduce the time required to perform the calculation. This thesis presents a comprehensive framework for predicting dose distributions and optimizing inverse treatment plans that represents a fundamentally new approach to solving these problems compared to all previous approaches described in the literature.

1.2 Problem Statement

The majority of current dose calculation techniques are limited by a major trade-off between computational speed and accurate representation of dose distribution. Monte Carlo (MC) simulations are recognized as the "gold standard" for dose calculation with dose accuracy within 2%; however, the MC method is computationally intensive, taking 10–360 minutes to perform each dose calculation based on desired levels of statistical precision and [5, 6]. and therefore makes it difficult to utilize thousands of dose calculations during inverse planning optimization loops for treatment plans[7, 5]. On the other hand, fast analytical Type A algorithms, including Pencil Beam, allow for rapid calculation of dose (i.e., $< 5 seconds$); however, these algorithms demonstrate large systematic dose errors in low-density lung tissue ranging from 10–46%, which is clearly not acceptable for thoracic radiotherapy[8, 9]. Intermediate Type B algorithms, including Collapsed Cone Convolution (CCC), (10-30 seconds) and Anisotropic Analytical Algorithm (AAA), provide significantly better accuracy than fast Type A algorithms; however, both CCC and AAA systematically overestimate lung doses by 1–3% [10, 11].

Recently developed commercial deterministic transport solvers, primarily Acuros XB (Varian Medical Systems), appear to be a middle ground to this computational dilemma and are capable of producing dose accuracy comparable to MC methods (0.6–2% difference in dose) while allowing for clinically practical computation times of approximately 40-95 seconds[3, 12]. however, Acuros XB is a proprietary closed-source system that limits the ability of researchers to reproduce results and to innovate new radiation dose calculation methods. Additionally, even the fastest deterministic solvers are still too slow for use in real-time adaptive applications that require dose calculations to be performed within seconds.

Deep learning has been identified as a potential alternative to solve this problem. Neural networks can learn complex mappings from input features to dose distributions that could result in second or less inference time. However, nearly all prior works on deep learning for dose prediction suffer from a significant limitation: virtually all prior work utilizes anatomical segmentation masks (binary volume defining PTV, OARs and body contour) and/or geometric features such as signed distance transforms. Therefore, the methods proposed in these papers abstractly correlate anatomical structure to dose distribution without utilizing any of the actual beam physics or treatment delivery parameters. As a result, these methods also have several important weaknesses. First, they are unable to generalize across different beam configurations a network trained on seven-field IMRT will not predict dose for VMAT plans unless completely retrained. Second, and most importantly, since anatomy-based methods inherently imply that the dose should be deposited to the PTV, they are unable to predict the correct dose distributions for unconventional treatment plans in which the radiation is intentionally delivered outside traditional target volumes. For example, consider a hypothetical treatment plan design in which the plan is designed to deliver a dose to a region outside of the PTV. The current methods would incorrectly predict a dose concentration within the PTV because they have learned a spurious relationship between the target contours and dose deposition, instead of learning the actual physics of radiation delivery.

These fundamental limitations make the current deep learning approaches inappropriate for inverse planning optimization problems in which the goal is to find the optimal beam configuration(s), rather than simply to predict the dose from the predetermined anatomical relationships. Since the models currently being used are unable to perceive actual beam geometries, MLC aperture shapes, or fluence modulation patterns, the models are unable to serve as differentiable dose engines in optimization loops. These limitations motivate the creation of physics-informed neural network architectures that include domain knowledge from radiation transport physics and that directly encode the physical characteristics of treatment delivery.

1.3 Thesis Objectives and Contributions

This thesis develops a multi-disciplinary approach which combines deterministic transport physics with deep neural networks to accelerate the process of predicting radiation therapy doses and optimizing inverse treatment plans. The proposed multi-disciplinary approach is based on true physics models and not on abstract anatomical correlations. Therefore, it can be generalized to an arbitrary treatment geometry. Also, this multi-disciplinary approach will enable new research to be conducted toward developing real time adaptive treatment planning. The main goals and achievements of the dissertation are:

1. **Open-source LBTE Solver and Physics-Based Feature Extraction:** Development of an open and customizable code for a deterministic Linear Boltzmann Transport Equation (LBTE) solver to calculate radiation doses during radiotherapy treatments. The LBTE solver uses the discrete ordinates (S_N) method and multigroup energy discretization. Additionally, the solver provides the computational means to extract novel physics-inspired features from DICOM data. The uncollided fluence Ψ_{unc} is calculated using analytical ray tracing from the RTPLAN file. The RTPLAN file encodes all of the physical information of the beam geometry, MLC aperture configuration, gantry angles, and fluence modulation patterns, just like commercially available solvers, such as Acuros XB.

2. **Novel Physics-Informed Feature Engineering:** For the first time in the literature, this study presents an extensive collection of physically-based input parameters for neural networks to predict doses based upon physics principles, including: uncollided photon fluence; TERMA (Total Energy Released per unit MAss); electron density; beam direction vector maps; and distance-to-surface maps. In contrast to the use of generic segmentation masks by all previous methods for dose prediction, our new features are designed to physically represent the actual delivery process of radiation beams and fields for the network to understand how each beam was actually delivered. Therefore, the physically-based approach will allow the model to correctly predict dose distributions regardless of the treatment geometry, including non-conventional treatment geometries in which dose may be delivered outside of the "traditional" target volume a feature that has never been demonstrated or modeled by any other method in the literature.

3. **Hybrid Neural Architecture with Attention-Based Fusion:** Development of a well-designed hybrid framework using 3D U-Net to extract the local features from images (e.g., tissue interfaces, density gradients, the penumbra region in which charged particle equilibrium is violated) and a Fourier Neural Operator to learn the global spectra (e.g., beam geometry, scattering kernel, long range energy transport), as well as an attention-based spatial fusion mechanism, so the hybrid model can adaptively combine local and global information at each voxel location; also, the architecture captures how the local spatial spread of energy dictates the dose deposition patterns in the vicinity of tissue interfaces.

4. **Residual Learning for Scattered Dose Prediction:** Implementation of a physics-informed residual-learning approach, in which the network estimates corrections to the dose predictions generated by the TERMA baseline. In this way, the network learns to predict both the scattered dose and the secondary-electron transport components that are computationally intensive parts of the iterative process associated with transport solvers. Use of the softplus activation function for physically meaningful output of non-negative dose values and use of gradient-based loss functions help to improve the accuracy of the dose calculations at interfaces, where clinical accuracy is of the highest importance.

5. **End-to-End Inverse Planning via Fluence-First Architecture:** The Dose-Prediction Network for contribution four has been pre-trained and has been integrated into a new

Inverse Treatment Planning System (ITPS) where it acts as a differential Physics Module. This ITPS produces Deliverable Fluence Maps and Monitor Unit Weights directly from Patient Anatomy and eliminates the need for the traditional Iterative TPS Optimisation Loop. The Inverse-Planning Network employs three Architectural Innovations. (I) A Rotary Transformer backbone that utilizes its RoPe's natural encoding of the Periodic Gantry Geometry of VMAT arcs to allow the Self-Attention Mechanism to reason about Angular Distance Between Beams. (II) A Structure-Angle Sheaf Graph Block, where Inter-Beam Dose Interactions are modeled as Learned Signed Transport Maps on a Bipartite Graph. Positive (Support), and Negative (Risk) Message Channels Simultaneously Reinforce Target Coverage and Suppress Organ-at-Risk Exposure Through Directional Sheaf Diffusion. (III) A Bilevel Curriculum Training Strategy that First Anchors the Network on Supervised Fluence Targets (Phase 1, 20 Epochs) and Subsequently Introduces Differentiable Clinical Physics Including DVH Constraint Hinge Losses and an Adjoint Clinical Transport Prior (ACTP) That Maps Voxel-Level Clinical Objectives to Beamlet-Level Fluence Targets Via the Adjoint of the Precomputed Sparse Dose Kernel (Phase 2, 40 Epochs). Convex fluence projection with total-variation and angular smoothness regularisation ensures MLC-deliverable output, while the sparse Boltzmann-derived ray-tracing operators developed for feature extraction in Contribution 1 compute uncollided fluence and TERMA from the predicted fluence maps, closing the physics loop. Validation on 50 VMAT lung-cancer patients demonstrates clinically competitive results at subsecond inference ($\sim$323 ms per plan): gamma pass rate of $85.20 \pm 7.96\%$ (5%/3 mm, 10% threshold), OAR sparing that matches or exceeds clinical reference plans (Lung V20 -0.5 pp, Cord D0.03cc $+12.0$ pp pass-rate improvement), and a characteristic coverage–protection trade-off where PTV V95%Rx is 14 pp below clinical plans while four of seven OAR constraints are met at equal or higher rates. Crucially, the most demanding case — a patient with bilateral pulmonary metastases — produced the lowest gamma score yet generated a physically coherent plan covering all target volumes with appropriate organ sparing, providing direct evidence that the model internalised the *goal* of radiotherapy rather than memorising ground-truth plan patterns.

6. **Clinical Validation for Lung Cancer VMAT and IMRT:** Clinical validation of the forward dose-prediction engine was conducted on 52 VMAT and 26 IMRT lung cancer patients using the 5%/3 mm gamma criterion with a 10% dose threshold. For VMAT, the model achieves a mean gamma pass rate of 98.17% (median 99.16%), with 92.3% of patients exceeding the 95% clinical threshold and a mean absolute error of 1.71 ± 0.17 Gy. For IMRT — the first evaluation of this framework on static-gantry delivery — the mean gamma pass rate is 89.59% (median 91.69%) with MAE of 0.84 ± 0.18 Gy. The lower IMRT gamma reflects the sharper fluence discontinuities inherent to step-and-shoot delivery and the predominantly VMAT training distribution, yet over half of IMRT patients exceed the 90% threshold without any modality-specific fine-tuning, demonstrating the cross-modality generalisability conferred by physics-based input features. The significance of this combined validation for the management of lung cancer is primarily due to the need for rapid replanning as a result of respiratory motion and the ultimate goal of providing optimal PTV coverage while limiting OAR exposure to improve patient outcome.

Chapter 2 Clinical Background: Intensity-Modulated Radiotherapy and Volumetric Modulated Arc Therapy

2.1 Introduction

Radiotherapy as an external beam technique, has evolved significantly from conventional 2D to conformally shaped 3 dimensional techniques which have improved the therapeutic index of oncology by increasing the radiation dose delivered to tumors while minimizing the dose to surrounding normal tissues. Intensity-Modulated Radiotherapy (IMRT), and Volumetric Modulated Arc Therapy (VMAT) are now the standard of care for most thoracic cancers that are complicated by the presence of multiple normal structures within the irradiation field. These techniques use advanced algorithms and dynamic movement of the multi leaf collimator (MLC) to create a precise distribution of dose to the target(s)[1, 1, 2]. In this section we will examine both the technical aspects of these treatments, including the delivery parameters and dosimetries of each treatment type, and how they can be used clinically to treat patients with lung cancer. We also will provide information on the machine level parameters which affect treatment delivery.

2.2 Fundamental Geometric and Dosimetric Concepts

2.2.1 Source-to-Axis Distance and Isocentric Geometry

Modern Linear Accelerators (LINACS) are isocentric. That means that the radiation source on the end of a rotating gantry remains a fixed distance away from a mechanical isocenter — the spot in space around which the gantry, the collimator, and the treatment couch all move[13, 14]. The "source-to-axis" distance, (also known as source-isocenter distance) is set at100 cm for the overwhelming majority of LINACS used today in the clinic[15, 13]. This particular distance is the result of a series of compromises made by engineers. These compromises include: Adequate clearance for full360° gantry rotationto avoid collision with patient anatomy and other gantry head components; Enough room to allow for the placement of beam shaping devices (such as flattening filters, jaws, and multi-leaf collimators) in the beam path; Clinically acceptable dose delivery rates at the patient's surface (i.e typically 400 cGy/min to 600 cGy/min at 100 cm for megavoltage photon beams); Ergonomic accessibility of the isocenter height to facilitate patient positioning by radiation therapists[13, 14].

Other than the above specified configuration there are alternative SAD arrangements available for specialty applications. Historical ^{60}Co teletherapy units were designed with an 80 cm SAD. The rationale behind this design was to increase the dose rate, while compensating for the relatively low specific activity of ^{60}Co sources, (average photon energy = 1.25 MeV) In addition, it reduced the physical dimensions of treatment vaults. However, it did so at the expense of reduced clearance for larger patients and increased penumbras due to the larger physical source sizes (1 cm to 2 cm diameter)[13]. Gamma Knife radiosurgery units utilize 40 cm focal distance to maximize dose fall-off gradient for intracranial stereotactic treatments[14]. Gamma Knife radiosurgical units use a 40 cm focal distance to optimize the steepness of the dose fall-off gradient for stereotactic treatments of intracranial tumors. For radiotherapy of thoracic and abdominal malignancies however, the 100 cm SAD standard provides the best balance between geometric accuracy and clinical practicability.

The isocenter is placed within the patient anatomy at the geometrical center of the Planning Target Volume (PTV) or at a clinically strategic position which enables optimal beam arrangementduring isocentric treatment planning [13]. The Source-to-Skin Distance (SSD), which is defined as the distance from the radiation source to the patient's skin surface along the central beam axis, varies with patient thickness and isocenter depth.

$$\mathrm{SSD} = \mathrm{SAD} - d_{\mathrm{iso}}, \tag{2.1}$$

where d_{iso} is the depth of the isocenter below the skin surface[15]. For a typical thoracic treatment with SAD = 100 cm and isocenter positioned at 10 cm depth, the SSD equals 90 cm.

Isocentric Geometry and SAD/SSD Relationship

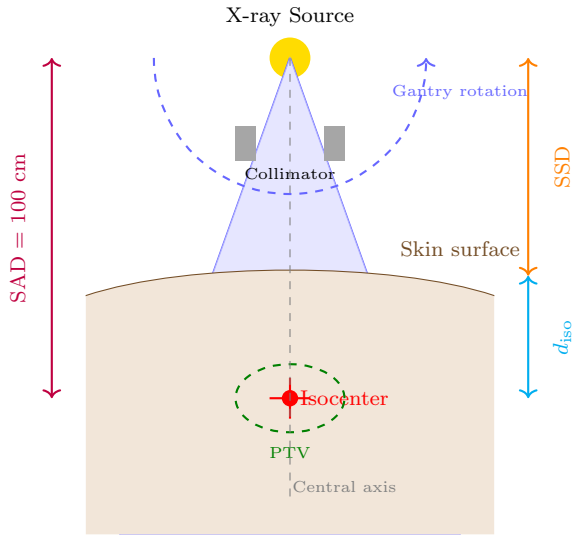


Figure 2.1: Isocentric geometry showing the relationship between source-to-axis distance (SAD), source-to-surface distance (SSD), and isocenter depth (d_{iso}). The radiation source maintains a fixed SAD of 100 cm to the isocenter while the gantry rotates around the patient.

2.2.2 Field Size Definition and Beam Collimation

Radiation field dimensions are defined in the isocenter plane (for SAD technique) or on the patient's surface (for SSD technique). These radiation field dimensions are defined by the lateral distance between the 50 % isodose lines[13]. Physical apertures are formed by two orthogonal pairs of collimator jaws constructed from a high-atomic number tungsten alloy. The transmission of the primary beam through these collimators is generally less than 1 %[14]. By convention, the movement of the X-jaws (or lateral jaws), which are perpendicular to the gantry rotation plane, and the Y-jaws (or longitudinal jaws), which are parallel to the gantry rotation plane, determine the final shape of the physical aperture[13].

For a given rectangle field of dimension X x Y that has been defined at the isocenter, the equivalent square field size used for dose calculations can be determined using the empirical equation[15, 13]:

$$r_{\mathrm{eq}} = \frac{4 \cdot A}{P} = \frac{4 \cdot (X \cdot Y)}{2(X + Y)}, \tag{2.2}$$

where A is the field area and P is the field perimeter. This 4 A/P formula provides accurate approximation for most clinical field shapes but should be verified for highly elongated geometries (e.g., 5 cm × 40 cm mantle fields)[15].

2.2.3 Isodose Curves: Definition and Clinical Utility

The use of Isodose Curves is an essential tool in creating 2-Dimensional and 3- Dimensional illustrations of Dose Distribution, which is defined as all of the areas where absorbed radiation has been delivered to the body in the exact same amount (expressed as a percent of a

Reference Dose). The Reference Dose is usually the highest dose delivered to the area at the Isocenter or Prescription Point[13, 15]. To produce these 2-Dimensional and 3-Dimensional illustrations, the Treatment Planning System will take data from the Dose Matrices produced by interpolating between the Dose Values calculated on each Voxel Grid and display the Isodose Curves as Color-Coded Overlays on the Images used for Treatment Planning. By displaying the Isodose Curves as Color-Coded Overlays on the Images used for Treatment Planning, it allows the Clinician to visually assess the Coverage of the Planning Target Volume (PTV) (by determining if the Prescription Isodose Curve intersects with the PTV Boundaries), the Sparing of Critical Organs at Risk (by determining which Isodose Percentages intersect with Structures of Concern), and the Low-Dose Bath Areas (Areas of low-dose irradiation) of Normal Tissue Surrounding the PTV [13]. Due to the Clinical Utility of using Isodose Visualizations in evaluating Treatment Plans for Intensity Modulated Radiotherapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT), clinicians can rapidly identify Underdosed Regions within Targets, Overdosed Hotspots Exceeding the International Commission on Radiation Units and Measurements (ICRU) Dose Limits, and Unintended Low-Dose Regions that could contribute to Late Toxicity or Secondary Malignancy Risk[13].

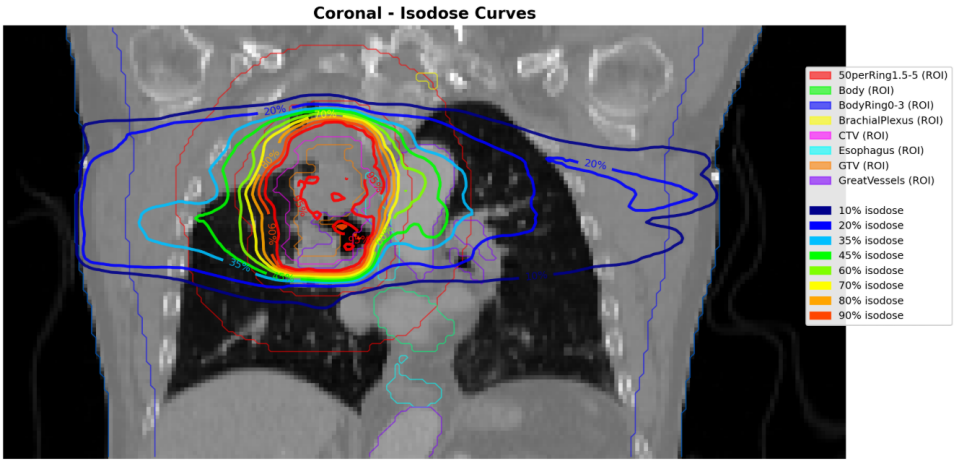


Figure 2.2: Schematic representation of isodose curves overlaid on an axial CT slice. The 90% isodose (red) represents the prescription dose and should encompass the planning target volume (CTV-PTV). Lower isodose percentages extend outward, demonstrating dose fall-off. The 45% isodose (green) and 20% isodose (blue) show the low-dose bath affecting surrounding tissues including the lungs.

2.2.4 Monitor Units and Dose Calibration

The basis for delivering radiation in external beam radiotherapy is the Monitor Unit (MU). The MU represents the amount of accumulated charge measured by a Transmission Ionization Chamber positioned in the radiation beam within the Treatment Head of the Linear Accelerator (LINAC)[15, 13]. According to an internationally recognized standard based on protocols from the American Association of Medical Physicists (AAPM), International Atomic Energy Agency (IAEA), and International Commission on Radiation Units and Measurements (ICRU), clinical LINACs have been standardized to deliver 100 MU with a dose of 100 MU deposits 1.00 Gy (equivalent to 100 cGy) at the Reference Calibration Point under Normalization Conditions[15, 16].

Photon Beam Normalization Conditions are specified as follows:Field Size: 10 cm × 10 cm at the isocenter; SAD = 100 cm; depth of maximum dose $d_{\max}$ along the beam central axis in water-equivalent phantom (approximately 1.5 cm for 6 MV X-rays, approximately 3.5 cm for 18 MV); No Beam Modifiers (Wedges, Compensators, Blocks)[15, 16]. Under these conditions,

the Dose Per Monitor Unit $D_0 = 1.00\,\text{cGy/MU}$, which provides a common reference point for calculating doses among institutions and Treatment Planning Systems.

The Total Number of Monitor Units needed to deliver the prescribed dose D_{rx} to a point at depth d for a field of Radius is determined using the formula[15]:

$$\text{MU} = \frac{D_{\text{rx}}}{D_0 \cdot \text{TPR}(d,r) \cdot S_c(r_c) \cdot S_p(r) \cdot \text{OAR}(d,x) \cdot \text{WF}}, \tag{2.3}$$

where TPR(d,r) represents the tissue-phantom ratio that accounts for the attenuation of the dose with respect to the depth and scattering of the radiation; $S_c(r_c)$ represents the collimator scatter factor (also referred to as Sc or phantom scatter factor); $S_p(r)$ is the phantom scatter factor; OAR refers to the off-axis ratio; WF represents the wedge transmission factor (which will be unity for open fields);[15]. The total MU in treatments using intensity modulated radiation therapy with dynamic motion of the multi leaf collimator is the sum of the individual MUs for each control point or segment.

2.2.5 Photon Beam Energy Selection

Clinical linear accelerators create megavoltage X-rays through bremsstrahlung interaction of accelerated electrons with a high atomic number material (tungsten or gold)[14]. Electron kinetic energy at a maximum is usually the way that nominal beam energies are defined for clinical applications and typically denoted in megavolt (MV). Clinically used beam energies range from 4 MV to 25 MV[13]. The two most commonly used photon energies for thoracic radiotherapy are 6 MV and 10 MV, with 15 MV or 18 MV The two most commonly used photon energies for thoracic radiotherapy are[17, 13].

There are many dosimetric factors that the selection of a photon beam energy will affect. High-energy beams (¿ 10 MV) have the following characteristics[17, 13, 14]. (1) they penetrate further into a phantom than lower energy beams because their attenuation coefficients are smaller; this characteristic is advantageous for treating deep-seated tumors but could be disadvantageous for superficial tumors where dose build-up regions extend many centimeters into the underlying tissues; (2) they cause less damage to the skin, and therefore provide better skin-sparing effects, which results in the depth of maximum dose being shifted from 1.5 cm at 6 MV to 3.5 cm at 18 MV;(3) they produce smaller widths for lateral penumbras due to smaller scattering angles resulting from Compton scatter, which allows for better conformity to small fields during IMRT and VMAT treatments; and (4) they produce an increased amount of neutron contamination as a result of the photonuclear reactions that occur within the treatment head components (γ, n) which have thresholds around 10 MeV), which raises concern about the risk of secondary malignancies particularly in pediatric patients and long-term survivors[17, 14].

In general, when performing lung cancer IMRT and VMAT treatments, 6 MV photon energy is preferred over higher energies for several reasons[2, 17]. First, since the majority of the thoracic anatomical structures have relatively moderate thicknesses (15 cm to 25 cm anterior-posterior separation) there is no need to utilize the enhanced penetration characteristics of higher-energy beams. Second, 6 MV beams deposit less total energy (integral dose) in the patient compared to higher-energy beams, which may decrease the risk of late toxicities. Third, the transmission and leakage radiation produced by multileaf collimators increases with beam energy — measurements have shown that MLC transmission is approximately 1.8 % at 6 MV versus 2.5 % at 15 MV and 3.5 % at 25 MV— resulting in higher out-of-field doses and potentially higher risks of secondary cancers for higher-energy IMRT[17]. Fourth, the shallower buildup region of the 6 MV (1.5 cm) also provides better deposition of the dose in the anterior chest wall targets compared to the buildup regions of the 15 −−18 MV beams where the build-up region extends to 3 −−3.5 cm, which may result in underdosing of superficial disease[17].

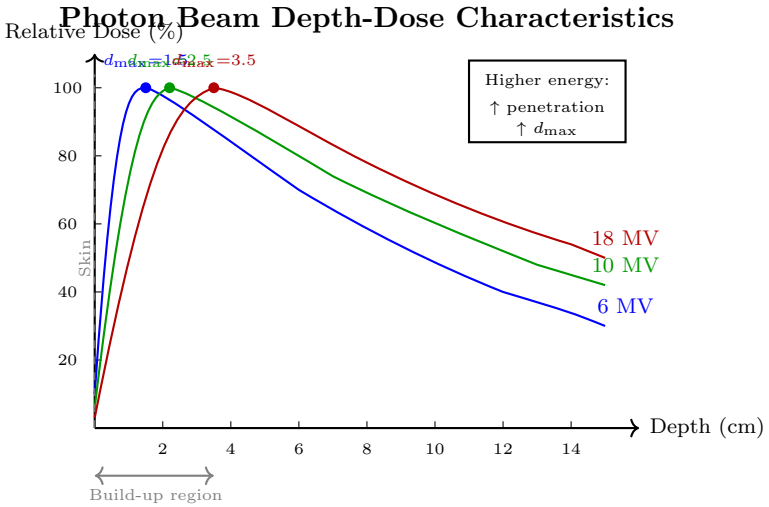


Figure 2.3: Percentage depth-dose curves for clinical photon beam energies. Higher energy beams exhibit increased penetration and deeper depth of maximum dose ($d_{\max}$). For lung cancer radiotherapy, 6 MV is generally preferred due to adequate penetration and shorter build-up region.

2.3 Multileaf Collimator: Design, Constraints, and Dynamic Motion

2.3.1 Physical Design and Geometric Properties

The multileaf collimator (MLC) represents the key enabler for intensity modulated radiotherapy. It consists of two sets of leaves (carriages) of individual tungsten alloy leaves which are motorized on opposite sides of the radiation beam and therefore can be moved to define the aperture of the radiation beam with a very high degree of spatial detail[13, 18]. There are a number of commercially available MLC systems with different designs for their leaves. Two types have been most widely used clinically; one type has a single focused leaf design (e.g., Varian Medical Systems Millennium MLC, HD120 MLC), and the other has a double focused leaf design (e.g., Elekta Agility MLC, Siemens OPTI-FOCUS MLC). The side angles of the leaves in the single focused design are directed towards the source of the radiation, which minimizes interleaf transmission. In contrast, both leaf ends and sides of the double focused design are directed towards the radiation source[18, 19].

The width of the leaves is the critical factor determining the resolution of the intensity modulation and the ability to achieve close conformality of the radiation beam to small volume targets[20, 21]. As such, modern commercial MLC systems utilize variable leaf widths depending upon the anatomy being treated[18, 13]. The Varian Millennium 120-leaf MLC system utilizes three separate leaf widths; central 40 leaf pairs are 5 mm width (projected to 100 cm SAD),and provide the highest resolution modulation over a 20 cm × 20 cm central area for treatments of the head and neck, and central thoracic areas; the remaining 20 leaf pairs are 10 mm wide and allow the maximum field size to reach 40 cm × 40 cm[20]. The Varian HD120 MLC utilizes uniform 2.5 mm wide leaves throughout all 64 leaf pairs, and allows for higher conformality in stereotactic applications[18].The Elekta Agility system provides 160 leaves, each 5 mm width, providing 40 cm × 40 cm maximum field[13].

Studies examining dosimetry indicate that decreasing the leaf width from 10 mm to 5 mm or 3 mm significantly improves conformity index (CI) and homogeneity index (HI) for small target volumes (PTV $< 100\,\text{cm}^3$), with relative improvements in CI of 8 % to 15 %[20, 21]. However, these benefits do not extend to larger target volumes (PTV $> 500\,\text{cm}^3$); in these cases, plans developed using leaf widths of 5 mm and 10 mmwill yield virtually identical plan quality metrics[20]. A major drawback to utilizing narrower leaf widths is the increase in total

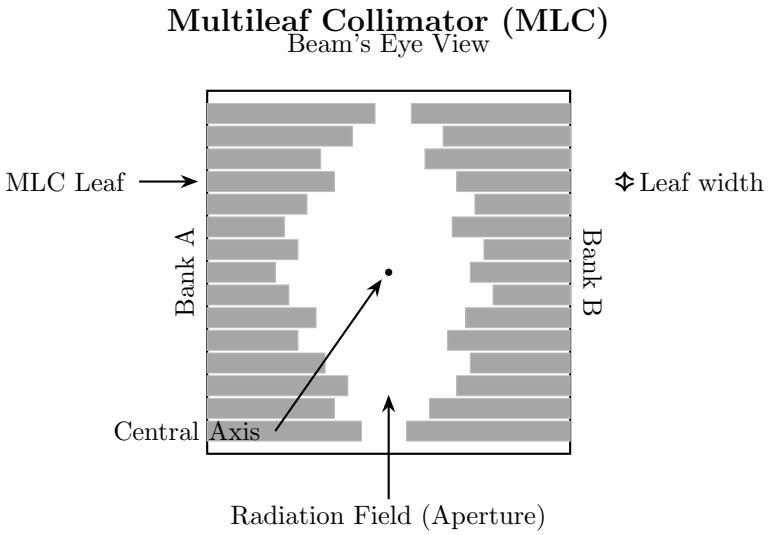


Figure 2.4: Beam's eye view of a multileaf collimator (MLC) showing opposing leaf banks (A and B) shaping the radiation aperture to conform to the planning target volume (PTV). Each leaf moves independently to create irregular field shapes. CAX = central axis.

monitor units; plans created with leaves that are3 mm wide will require 20 % to 40 % more MU than those created with 10 mm wide leaves, primarily as a result of the finer discretization of the beamlets and slower maximum leaf speeds (1.5 cm s^{-1} for 3 mm MLC versus 2.5 cm s^{-1} for 10 mm MLC). This results in longer treatment delivery times and an increased integral dose[20].

2.3.2 Mechanical Constraints and Leaf Motion Dynamics

For optimal leaf motion during dynamic IMRT delivery, MLC leaf motion must respect a variety of mechanical and dosimetric constraints as determined during both the treatment plan optimization phase and during delivery of the treatment plan[18, 22, 23]. Maximum leaf velocity is an MLC dependent variable as well. Typical maximum leaf velocities include: Varian Millennium 2.5 cm s^{-1}; Varian HD120 3.0 cm s^{-1}; Elekta Agility 3.5 cm s^{-1}; Siemens 160-leaf MLC 4.0 cm s^{-1}[18, 24]. As such these velocity limits create minimum transit times when moving large distances using a single leaf. An example of this includes a leaf traveling the entire 40 cm wide MLC field at 2.5 cm s^{-1} requires 16 stransit time, creating an absolute minimum time that any given segment of the sliding window IMRT delivery can take to complete.

The general specification for leaf positioning accuracy dictates that each leaf must achieve the commanded position within ±1 mm regardless of leaf speed[25, 26].Achieving this level of accuracy is increasingly difficult at higher leaf speeds due to mechanical inertia and delays associated with the servo control loop used to control leaf movement [27]. , thus treatment planning systems restrict maximum leaf speed during optimization and will often limit leaves to 80% to 90% of the hardware maximum to ensure the best possible leaf positioning accuracy[1, 18].

A third significant constraint involves the minimum separation required between opposing leaf pairs. To minimize radiation leakage into the gap between the two opposing leaves, a minimum separation (typically 0.2 cm to 1.0 cm isimposed[22, 28]. his constraint makes it even more complicated to optimize highly modulated intensity patterns where very small beam widths are desired. When the opposing leaves do not meet the minimum separation requirement, one leaf pair must be completely retracted while the adjacent leaf pair provides compensation, increasing the effects of under-dosage caused by the tongue-and-groove effect[22].

This also defines how wide a single carriage position MLC field can be treated dynamically

with a single carriage position[28].n the case of the Varian Millennium MLC, the maximum leaf over-travel is 16 cm past the center of the field, therefore, providing a maximum of about 14 cmfield width for single carriage position dynamic IMRT delivery (allowing for 5 mm offset from the jaw edge to avoid transmission of energy between leaf ends)[28]. Wider targets require either multi-carriage delivery (where the MLC carriage position is shifted during beam-on), or dual isocenter techniques with carefully feathered field junctions [28].

2.3.3 Control Points and Delivery Synchronization

A control point represents an instantaneous temporal snapshot that describes the configuration of all delivery parameters for radiation therapy including gantry angle, collimator angle, couch angle, MLC leaf positions (both left and right for all leaf pairs), jaw positions (X1, X2, Y1, Y2), dose rate, and total monitor units delivered[29, 30, 1].Control points are temporally ordered to define the treatment plan and intermediate machine states are interpolated using linear interpolation between explicitly defined control points[30, 1].

In Dynamic Multileaf Collimator (DMLC) IMRT, also referred to as sliding window, control points are produced on a millisecond basis(approximately 50 milliseconds to 100 milliseconds), resulting in 200 to 400 controlpoints per beam for a standard delivery that requires 100 MU to 300 MU[29, 30]. In addition, the LINAC's control system utilizes continuous real-time feedback loops to verify that actual leaf positions, gantry angle and total monitor units delivered agree with prescription values within specified tolerances (typically ±1 mm for leaf position, ±0.5° for gantry angle, ±2 % for MU)[13]. If any of the parameters exceed the tolerance limits, the beam will be immediately interrupted and an error condition will be identified by the physicist for review.

Step-and-Shoot (Static Multileaf Collimator or SMLC) IMRT deliveries use control points to describe discrete segments with static leaf position[30, 20]. Beam-off occurs during transitions between segments; leaves move with the beam-off and the beam is activated again after verification that all leaves have reached their target positions within the tolerance limits[25]. Standard SMLC plans can utilize from 5 to 30 segments per beam where the number of segments provides a balance between delivery efficiency and intensity modulation of the beam (i.e., larger number of segments produce better intensity modulation but result in longer delivery times because of beam-off intervals and position verification overhead (approximately 2 s to 5 s per segment)[29, 30].

The total MU for a multi-segment field is apportioned among control points based upon the optimized intensity pattern[29, 30]. The MU assigned to control point n is calculated as:

$$\mathrm{MU}_n = \mathrm{MU}_{\mathrm{total}} \times \frac{I_n}{\sum_{i=1}^{N} I_i}, \tag{2.4}$$

where I_n is the optimized intensity (dose contribution) for control point n, and N is the total number of control points[29]. Independent verification of segment MU via analytical calculation or independent Monte Carlo dose calculation is recommended as part of patient-specific quality assurance[25].

2.4 Intensity-Modulated Radiotherapy: Technical Implementation

2.4.1 Beam Arrangement and Optimization Strategy

Intensity modulated radiotherapy uses multiple static gantry positions either coplanar or non-coplanar to deliver highly conformal dose distributions through inverse optimization of beam intensity patterns[31, 32]. for thoracic maligmancies, standard beam arrangements usually consist of 5 to 9 angles selected manually or equidistantly to maximize dose deposition in the

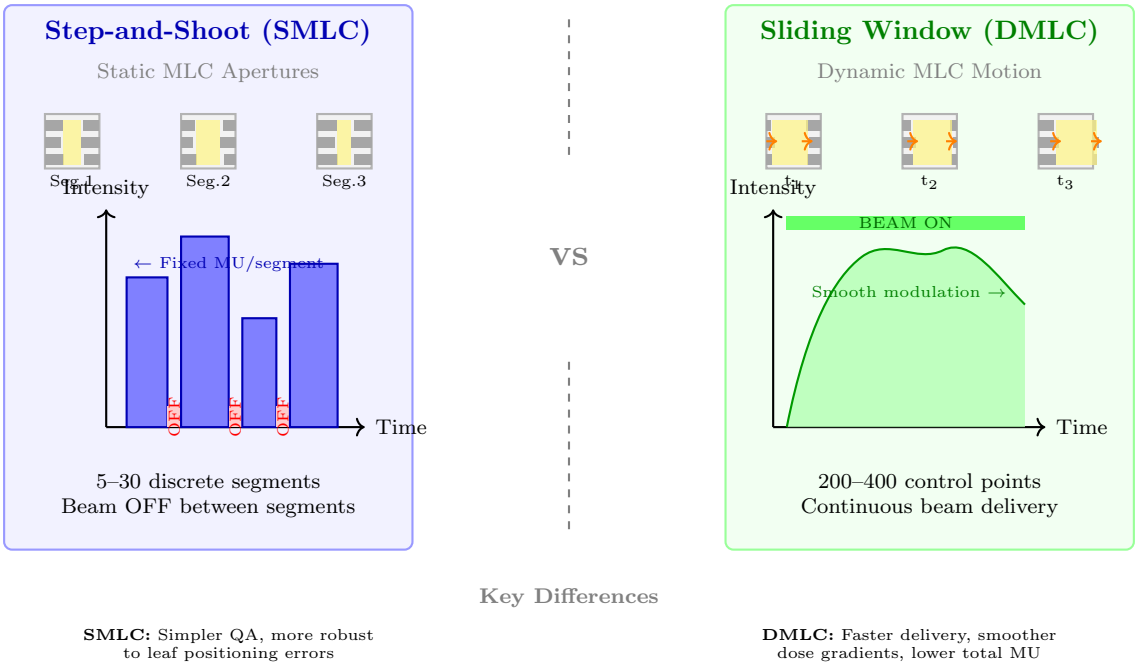


Figure 2.5: Comparison of step-and-shoot (SMLC) and sliding window (DMLC) IMRT delivery modes. **Left:** SMLC delivers discrete segments with fixed MLC positions and beam-off periods during leaf transitions. **Right:** DMLC maintains continuous beam delivery with dynamically moving leaves, producing smoother intensity profiles and faster treatment times. The top row shows MLC aperture snapshots, while intensity-time profiles illustrate the fundamental difference in modulation approach.

planning target volume while minimizing beam entrance and exit doses through critical organs at risk[2, 28].

Standard lung cancer IMRT beam configurations include[2, 28, 8] (1) seven-field arrangement with gantry angles at0°, 51°, 102°, 154°, 206°, 257°, 309°, which provide near uniform angular sampling while avoiding coplanar geometry that concentrates dose in anterior posterior or lateral directions; (2) nine-field arrangement at40°intervals (0°, 40°, 80°, . . . , 320°),provides fine angular sampling at cost of increased total monitor units and delivery time; (3) oblique-only arrangements that avoid anterior (340°–20°) and posterior (160°–200°) beams to reduce heart, esophagus, and spinal cord exposure, specifically used for centrally located tumors adjacent to mediastinal structures.

Collimator angle — the rotation of the mlc assembly about the beam central axis — is generally set to 0°(leaves perpendicular to superior-inferior patient axis) or optimized to align leaf motion with the longest dimension of the target projection, thereby minimizing tongue-and-groove effects and leaf travel distance[31, 13].Advanced optimization algorithms may select collimator angles individually for each beam based on target geometry and organ at-risk proximity, with reported improvements in organ sparing of 5 % to 15 % for selected head-and-neck cases[31].

Couch rotation (also termed table angle) remains fixed at 0°(couch longitudinal axis aligned with gantry rotation plane) for the majority of imrt treatments, maintaining coplanar beam geometry[13]. non-coplanar beam arrangements incorporating couch rotations (90° or 270°) have been studied for selected cases where posterior oblique beams offer dosimetric advantage—for example apical lung tumors where conventional coplanar beams must pass through extensive areas of contralateral lung and mediastinum[33]. However, non-coplanar imrt results in substantially increased treatment planning complexity, possible couch-gantry collision requirement for manual verification, and longer patient setup times, limiting its routine clinical

application[33].

The iso center is positioned at the geometric centroid of the planning target volume or at a strategically selected location within the high-dose region to minimize off-axis distances for all beams, therefore reducing dosimetric perturbations from off-axis beam softening and required field sizes[13, 15]. For lung SBRT employing tight PTV margins (3 mm to 5 mm), for lung SBRT with tight PTV margins (3mm – 5mm), accurate placement of the isocenter is critical and requires IGRT utilizing onboard cone-beam CT or orthogonal kV imaging immediately before each fraction[2].

2.4.2 Monitor Units and Delivery Efficiency

A number of factors contribute to the high levels of monitor units seen in IMRT compared to 3D-CRT: First, IMRT is inherently inefficient – large areas of the beam are blocked by moving leafs to produce the desired modulated intensity distribution; second, because a larger area is being blocked than needed to create the desired modulated intensity distribution, more radiation is required at the entrance surface of the patient to overcome the blocked radiation [31, 1]. The typical range of total monitor units used for each fraction in IMRT treatment plans is generally in the order of 300 to 800 monitor units per fraction for conventional fractionation (1.8 Gy to 2.0 Gy per fraction), and corresponds to an average ratio of 1.5 to 4.0 for monitor units to prescribed dose[31].The use of ablatively high doses (10 Gy to 20 Gy per fraction), forlung SBRT resultsin total monitor units per fraction exceeding 2000 to 5000 MU[2].

Increased monitor units result in a number of negative consequences[31, 1]: (1) , including: (1) an increase in the integrated dose delivered to the entire body (i.e., the total energy absorbed by the patient) which may be associated with an increased risk of late toxicity and secondary malignancy; (2) a longer time that the beam is "on," resulting in an increased sensitivity to intra-fraction motion, which is especially problematic when treating lung tumors that have significant respiratory excursion; (3) an increased rate of wear on the components of the LINAC (e.g., magnetron, electron gun, motorized leafs) and thus an increase in the need for maintenance of these devices and the cost of this maintenance; (4) a decrease in patient throughput as a result of the longer treatment delivery times.

The length of time required for treatment delivery of IMRT is typically measured as the time elapsed from the start of delivery of the first beam until the end of delivery of the last beam (exclusive of time spent setting up the patient and acquiring images). The typical range for treatment delivery time is between10 min to 18 min for thoracic treatments employing 7 to 9 beams and dynamic leaf motion[2, 1]. These delivery times consist of: (1) beam-on time for the actual delivery of the dose (cumulative time from (6 min to 10 min cumulative across all beams);(2) time required to move the gantry from one position to another (8 −−15 s per rotation, approximately 2 min to 3 min total); beam start-up time (magnetron charging, dose rate stabilization, approximately 5 s to 10 s per beam, 1 min to 1.5 min total); (3) time required to turn on the beam (about 1 minute to 1.5 minutes total); (4) verification and recording overhead (time required to perform flatness and symmetry checks, communicate with the RAVS, etc. or about 2 s to 5 s per beam)[1, 29].

2.4.3 Dosimetric Characteristics for Lung Cancer

Studies comparing plan optimization using IMRT vs. 3D-CRT for the treatment of lung cancer have shown that IMRT consistently improves several key dosimetric parameters[2, 34]in comparison to 3D-CRT. Specifically, for PTV sizes ranging from 100 cm^3 to 500 cm^3, the use of IMRT will result in a range of benefits including: (1) an improvement in conformity index (defined as the ratio of the volume of the PTV receiving the prescribed dose divided by the PTV volume), of approximately 15 % to 30 %,with average conformity index CI values of approximately 1.05 to 1.15 for IMRT compared to 1.3 to 1.5 for 3D-CRT[2]; (2) an improvement in homogeneity

index (HI, defined as the ratio of the maximum dose received by the PTV to the prescribed dose), of approximately 5 % to 10 %,with average HI values of 1.05 to 1.10compared to 1.10 to 1.20 for 3D-CRT[2]; (3) a reduction in mean lung dose of approximately 8 % to 15 %, especially advantageous for patients with bilateral disease or those who require large fields through the mediastinum[34].

However, because of the larger number of beam directions used with IMRT and the beam modulation process required to generate the optimal dose distribution, the low-dose spread of IMRT is greater than that seen with 3D-CRT. The manifestation of this phenomenon can be observed in elevated V5 (volume of tissue receiving at least ≥5 Gy) for total lung, contralateral lung and heart[2, 34]. Radiation pneumonitis is one of the most common late complications associated with thoracic irradiation and the theoretical concern regarding the increased risk of pneumonitis with IMRT is based on epidemiological studies demonstrating that lung V5 is a strong predictor of Grade ≥2 radiation induced pneumonitis[35]. Therefore, in order to minimize the adverse effects associated with the low-dose spread inherent in IMRT plans for the treatment of lung cancer, the design of IMRT beam arrangements should be constrained to optimize low-dose metrics (V5, V10, mean lung dose)[2].

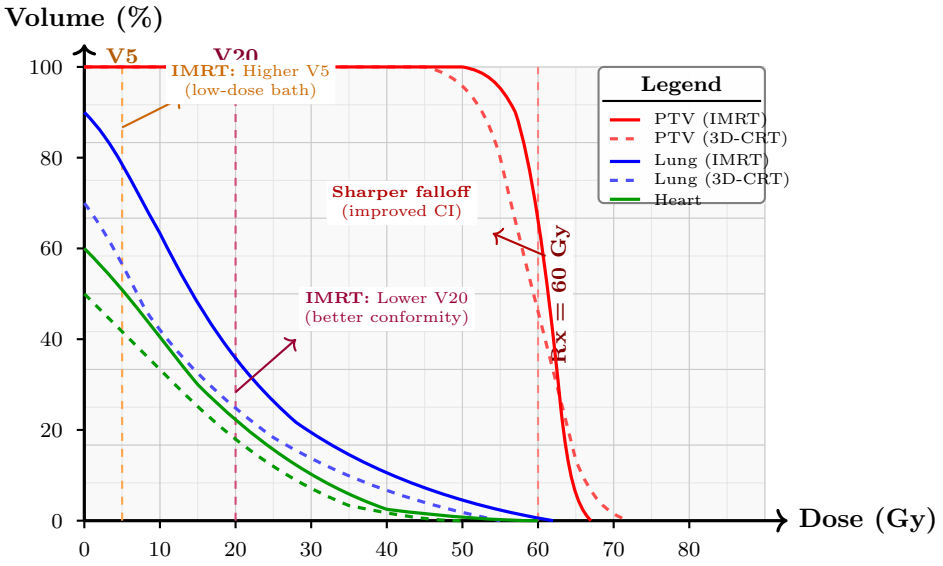


Figure 2.6: Dose-volume histogram (DVH) comparison between IMRT and 3D-CRT for lung cancer radiotherapy. Key observations: (1) IMRT achieves superior PTV conformity with sharper dose falloff at prescription dose; (2) V20 for organs at risk is reduced with IMRT, indicating better high-dose sparing; (3) however, IMRT exhibits elevated V5 due to low-dose spread from multiple beam angles. Clinical constraint lines for V5 and V20 are shown, as these metrics correlate with radiation pneumonitis risk.

2.5 Volumetric Modulated Arc Therapy: Principles and Implementation

2.5.1 Historical Development and Technical Innovation

Beginning with the definition of volumetric modulated arc therapy (VMAT), it is evident that VMAT has created a new paradigm in the delivery of intensity-modulated radiotherapy (IMRT). In 1995, Yu had first developed an idea called "intensity-modulated arc therapy" (IMAT)[36, 1]. A major technological advancement made possible by Otto in 2008 was the introduction of practical optimization methods allowing for simultaneous changes in three independent parameters over the course of continuous gantry rotation: (1) the gantry's rotational speed (between from $0.1\,^\circ\,s^{-1}$ to $6.0\,^\circ\,s^{-1}$, dependent upon the LINAC model used); (2) MLC aperture shape via

dynamic leaf motion; and (3) the dose rate (which can range from 0 to 600 MU/min for conventional LINACs, up to 0 to 2400 MU/min for flattening-filter-free beams)[1, 1].

As such, previous IMAT implementations had been limited to using a fixed dose rate and a constant gantry speed and had relied exclusively on MLC motion to achieve intensity modulation[36]. These constraints have greatly restricted the potential optimization options because the gantry rotation speed must be sufficient enough to allow for the maximum amount of time required for any leaves to move at any given angle - typically resulting in gantry speeds less than $1\,^{\circ}\,s^{-1}$ and treatments lasting longer than 10 min to 15 min per arc[36]. The ability of VMAT to vary the gantry speed enables the gantry to decelerate (momentarily stop) at critical angles where complex MLC motion is necessary while increasing the gantry speed through those areas where less critical angles exist; ultimately decreasing the treatment time to approximately 1.5 min to 3 min per arc while producing dose distributions that are equal to or better than those produced by static field IMRT[1, 1].

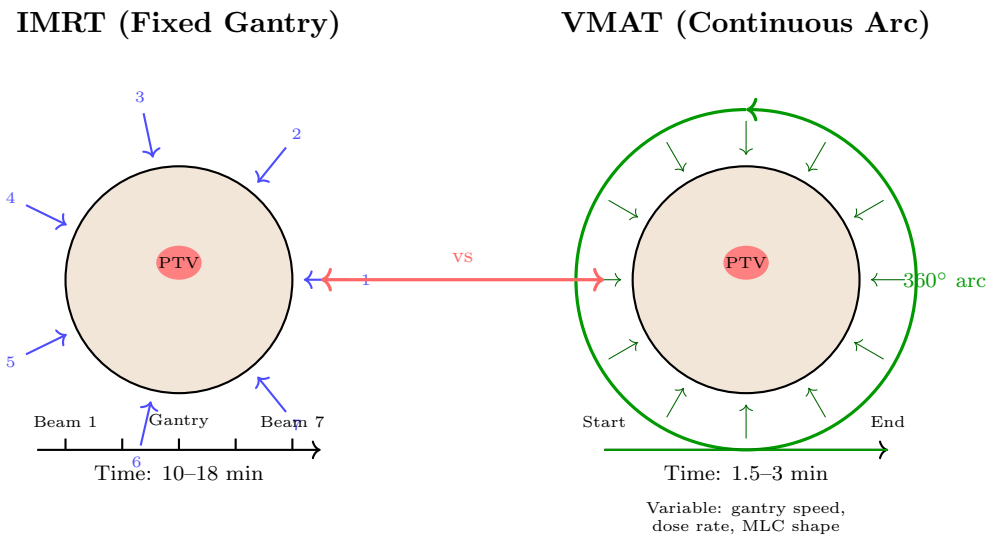


Figure 2.7: Comparison of IMRT and VMAT delivery modes. IMRT employs multiple static gantry positions (typically 7–9 beams) with beam-off during gantry rotation, requiring 10–18 minutes. VMAT delivers dose during continuous gantry rotation with simultaneous variation of gantry speed, dose rate, and MLC aperture, reducing delivery time to 1.5–3 minutes per arc.

2.5.2 Arc Geometry and Control Point Sampling

Optimization of VMAT treatment delivery models the treatment delivery process as a sequence of static control points that have been sampled over time at regular angular intervals throughout the gantry arc trajectory[1, 1]. A major factor in determining both the quality of optimization and the computational cost associated with optimization, is the density of the angular sampling of the control points. In particular, sparse sampling of the control points (e.g., 10° intervals, 36 control points per 360° arc) allows for the rapid initial optimization of VMAT plans, however it restricts the ability of the optimizer to create highly modulated intensity patterns [1]. Conversely, dense sampling of the control points (e.g., 2° provide for high angular resolution and allow for the creation of highly modulated intensity patterns; however, they increase the dimensionality of the optimization problem such that optimization times become prohibitively long and convergence may be difficult[1].

The progressive sampling strategy developed by Otto to address these competing demands for optimization quality and computational efficiency was based on an iterative approach[1, 1]. Specifically, initial optimization using coarse angular sampling of the control points (10°–20°

intervals)was used to rapidly identify promising beam aperture shapes and intensity patterns. Following the identification of a promising initial plan, the density of the control points were progressively increased using multiple optimization cycles (5°, 3°, 2° intervals),and the final optimized plan employed 2° control point spacing which corresponded to approximately 180 control points per 360° arc or approximately 0.33 s temporal resolution for a gantry rotating at $6\,^{\circ}\,s^{-1}$[1].

Thoracic VMAT arc trajectories are characterized into three categories[37, 2]: (1) full arc (FA-VMAT) – the gantry rotates 360° (180° clockwise to 180° counterclockwise, or 0° to 360° in single direction); (2) partial arc (PA-VMAT) – the gantry rotation is restricted to avoid beams entering contralateral lung or critical anterior /posterior structures. The typical range for PA-VMAT is 180°–180° (avoiding anterior 330°–30° and posterior 150°–210° sectors); (3) double partial arc (2PA-VMAT) – two partial arcs having complementary angular ranges (e.g., Arc 1: 180°–30°, Arc 2: 30°–180°). The double partial arc VMAT trajectory offers higher angular sampling density in areas of interest (i.e. critical regions) while maintaining beam avoidance sectors[2, 37].

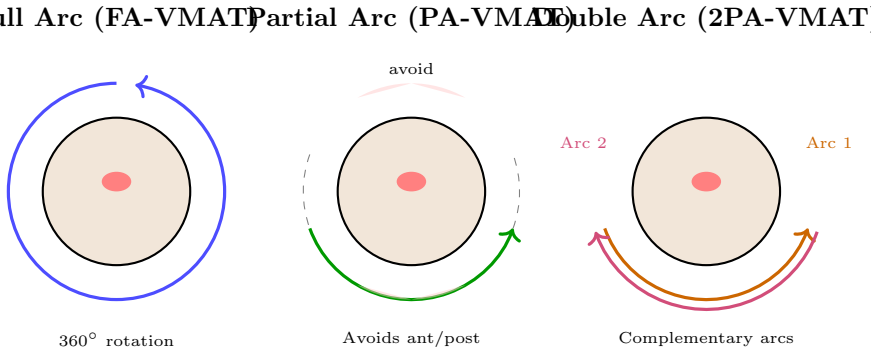


Figure 2.8: VMAT arc configurations for lung cancer radiotherapy. Full arc (FA-VMAT): continuous 360° rotation. Partial arc (PA-VMAT): restricted arc avoiding anterior/posterior beam entrance through critical structures. Double partial arc (2PA-VMAT): two complementary arcs providing enhanced angular sampling while maintaining beam avoidance sectors.

For most VMAT treatments the collimator angle is set to either 30° or 330° for full arcs, and is held constant throughout the arc to prevent MLC interleaf transmission patterns from aligning with PTV boundaries[1, 37]. There are some modern VMAT implementations where the collimator angle dynamically rotates during gantry rotation, and there have been reports of improved conformality for concave target geometries ranging from 3 % to 8 %[1].

Couch angle remains fixed at 0° ffor conventional coplanar VMAT[1]. However, in order to better manage collisions between moving components within the radiation therapy system (i.e. gantry, collimator, and couch), there has recently been the development of non-coplanar VMAT (referred to as 4πVMAT). Non-coplanar VMAT utilizes continuous couch rotation that is synchronized with gantry motion, thereby allowing the radiation therapy system to navigate collision-free trajectories through the joint gantry-couch configuration space[33]. Non-coplanar VMAT also requires sophisticated collision detection algorithms that account for actual machine geometry and patient surface meshes generated from planning CT scans. Additionally, graph search optimization (i.e. Dijkstra's algorithm) is used to determine the optimal gantry-couch trajectories[33]. Dosimetric studies show that non-coplanar VMAT can reduce mean organ-at-risk doses by 8 % to 15 % relative to coplanar VMAT for certain lung cases, particularly apical tumors where non-coplanar arcs can provide superior sparing of contralateral lung and brachial plexus[33]. However, the clinical implementation of non-coplanar VMAT is currently limited due to extended optimization times (2 h to 4 h versus 20 min to 40 min for coplanar VMAT) and required collision verification.

2.5.3 Delivery Efficiency and Monitor Unit Utilization

The most significant clinical advantages of using VMAT compared to IMRT are that VMAT will significantly reduce the time it takes to provide radiation therapy treatment[1, 1, 2]. Studies have been conducted that compare VMAT and IMRT treatments in terms of time to treat a patient for lung cancer. These studies have reported that the average time for delivering a VMAT treatment ranges from approximately 1.5 min to 3.5 minminutes for single-arc plans and 3 min to 6 min Additionally, these studies have reported that this represents an approximate 50–70% reduction in treatment time as compared to IMRT (10 min to 15 min for seven-to-nine-field plans)[2, 1]. This increased efficiency of VMAT over IMRT results from removing the time required for the gantry to rotate between individual static fields, as well as combining all beam on time together in one continuous arc.

VMAT also results in a moderate decrease in total monitor units delivered by the linear accelerator when compared to IMRT. Monitor unit reductions associated with VMAT typically range from 10 % to 30 % percent less than those observed with IMRT[2, 1].For example, if a lung SBRT fraction prescribed 12 Gy, were treated with single-arc VMAT, double-arc VMAT, and nine-field IMRT, the resulting MU values would be: single-arc VMAT 1800 to 2500 MU, double-arc VMAT 2200 to 3200 MU, nine-field IMRT 2800 to 4000 MU[2]. The reduction in MU delivered by VMAT occurs because VMAT continuously delivers dose while rotating the gantry of the linear accelerator. In contrast, IMRT requires the gantry to transition and verify segments between each beam position, thus introducing "beam off" overheads which can contribute to reduced MU delivered. The benefit of reduced MU delivered by VMAT is largely due to its ability to more efficiently utilize the beam on time.

Reduced delivery times associated with VMAT treatments result in numerous clinical benefits[1, 2]: (1) a reduction in motion uncertainty related to intrafraction motion (especially important for lung tumors that exhibit respiratory motion excursions ranging from 5 mm to 20 mm; (2) improved patient comfort and compliance, which reduces voluntary motion artifacts; (3) increased patient through put, allowing additional patients to receive treatment within fixed LINAC operating hours; and (4) reduced anesthesia time for pediatric patients or adults requiring sedation.

2.5.4 Dosimetric Comparison: VMAT versus IMRT for Lung Cancer

The results from various institutional comparisons between VMAT and IMRT dosimetry for treatment of lung cancer depend upon the location of the tumor and the arrangement of the arcs[2, 38, 1]. Peripheral Lung Tumors: Peripheral lung tumors were defined as planning target volumes that did not extend into either the proximal bronchial tree or the mediastinum within a distance of 2 cm In these cases, the authors reported the following dosimetric results for VMAT and IMRT:[2, 38] Conformity Index: The conformity indexes (CI) for single-arc VMAT and IMRT were found to be equivalent in statistical analysis (1.08 to 1.12), However, the conformity index was significantly improved (1.05 to 1.09, $p < 0.05$); Homogeneity Index: The homogeneity index (HI) for double partial-arc VMAT was greater than for both IMRT (1.04 to 1.07) relative to IMRT (1.08 to 1.12, $p < 0.01$) and single-arc VMAT (1.06 to 1.10); Target Coverage: Double-arc VMAT resulted in a slightly better PTV D_{95} (dose to 95 % of PTV volume) relative to IMRT (mean difference 1.2 % of prescription dose, $p < 0.05$)[2].

Organs at Risk: The authors identified some trade-offs when comparing VMAT and IMRT dosimetry for OARs:[2, 38]: Total Lung V5: All VMAT arrangements demonstrated an elevated total lung V5 relative to IMRT (partial-arc VMAT 8 %, single-arc 15 %, double-arc 12 %), This increase can be attributed to the fact that VMAT's 360° (or near-360°) angular sampling deposits low entrance doses around the circumference of the patient.Total Lung V20 and V30: Double partial-arc VMAT demonstrated lower total lung V20 and V30 than IMRT (V20: -2.3 %, $p < 0.05$; V30: -1.8 %, $p < 0.05$) Thus, double partial-arc VMAT results in better sparing of high dose regions, while still producing a larger low-dose "bath.[2] "Mean Lung Dose:

The mean lung dose was equivalent between IMRT and double-arc VMAT(difference <0.5 Gy, $p > 0.1$) however, it was significantly elevated for single-arc VMAT (1.2 Gy, $p < 0.01$)[38].

Spinal Cord Maximum Dose: The spinal cord maximum dose was reduced for VMAT compared to IMRT (VMAT −2.8 Gy, $p < 0.01$), The authors attribute this to VMAT's superior ability to dynamically move the radiation beam away from small series organs during arc motion[2]. Heart Mean Dose and V30: Heart mean dose and V30 were elevated for single-arc VMAT compared to IMRT (heart mean dose 1.5 Gy, V30 3.2 %, $p < 0.05$) The reason for this is the fact that single-arc VMAT produces an anterior beam entrance through the heart. However, the authors point out that the use of partial arcs that do not enter the heart at the most anterior angles (330°–30°)can mitigate this disadvantage[2, 38].

Central Lung Tumors: Central lung tumors refer to those where the planning target volume includes mediastinal lymph nodes. For central lung tumors, double partial-arc VMAT has been shown to provide dosimetric superiority over IMRT based on several metrics: CI improved by 8 % ($p < 0.01$), HI improved by 4 % ($p < 0.05$), esophagus mean dose reduced by 3.2 Gy ($p < 0.01$), spinal cord maximum dose reduced by 4.5 Gy ($p < 0.001$), and total lung V20 decreased by 2.1 % ($p < 0.05$)[2]. The authors attribute these advantages to the increased angular sampling and modulation flexibility provided by the two complementary arcs used in double partial-arc VMAT, which are especially beneficial for concave target geometries surrounding critical structures.

2.6 Clinical Implementation Considerations for Lung Cancer Radiotherapy

2.6.1 Technique Selection Guidelines

Methodology for Choosing Between VMAT & IMRT for Lung Cancer Treatment Technique selection for VMAT or IMRT for the treatment of lung cancer should rely on both tumor-specific & patient-specific factors [2, 13] These include: (1) Tumor Location - Peripheral tumors >2 cm from mediastinum will allow for better lung sparing using either Partial-Arc VMAT or IMRT w/ an Oblique Beam Arrangement. Double-Partial Arc VMAT allows for better dose reduction to Organ at Risk than IMRT for Central Tumors that are adjacent to Esophagus, Heart, or Proximal Airways [2]; 2) Amplitude of Respiration-Induced Tumor Motion - Tumors that exhibit excursions >10 mm are best treated using VMAT's shorter delivery time (2 min to 3 min) to reduce artifacts caused by intrafractional motion, compared to IMRT which has longer delivery times (10 min to 15 min)allowing for greater motion-induced blurring of Dose[1]; (3) Patient Tolerance - Elderly or Dyspneic patients who have difficulty maintaining the supine position for prolonged periods prefer VMAT to IMRT; (4) Concerns about Toxicity at Low Doses - Patients with pre-existing Interstitial Lung Disease or Reduced Pulmonary Reserve ($FEV_1 < 50$ % predicted) may prefer IMRT to VMAT because IMRT can provide a Lower Lung V5, thereby potentially reducing the risk of Pneumonitis[35].

2.6.2 Motion Management Integration

Respiratory-induced motion of the tumor represents a significant challenge for lung radiation therapy, with tumor motion amplitudes ranging from 3 mm to 30 mm (median 8 mm to 12 mm) n the superior-inferior direction[39]. Several methods of managing respiratory-induced motion are employed clinically[39, 2]: (1) Internal Target Volume (ITV) Approach - A planning CT is obtained during Free Breathing or as Four-Dimensional CT (4DCT) composite, encompassing all positions of the tumor within one respiratory cycle; PTV expansion accommodates the entire motion envelope, resulting in increased irradiation of normal tissues surrounding the PTV, but ensuring target coverage; (2) Respiratory Gating - The beam is delivered only during certain phases of the respiratory cycle (typically during the End-Exhalation phase) when the position of the tumor is most consistent and predictable, resulting in PTV margins of 3 mm to 5 mm but

doubling to quadrupling treatment delivery times due to Duty Cycle Losses; (3) Tumor Tracking - Real-time location of the tumor position is achieved through implanted fiducial markers and stereoscopic X-ray imaging, along with Dynamic MLC or Robotic Couch Adjustments to maintain beam-target alignment, resulting in minimal PTV margins 2 to 4 due to duty cycle losses; (3) Tumor tracking: Real-time localization of tumor position via implanted fiducial markers and stereoscopic X-ray imaging, with dynamic MLC or robotic couch adjustments to maintain beam-target alignment, enabling minimal PTV margins (2 mm to 3 mm) , but requiring specialized equipment (CyberKnife, Vero, or Varian TrueBeam with gating capabilities)[39].

ntegrating motion management into VMAT creates new challenges[39, 2]. Respiration gating with VMAT results in interruptions in gantry motion and dose delivery to occur during beam-holding phases (segments of respiratory cycle when the tumor position is outside the gating window), followed by resumed synchronized gantry-MLC-dose rate motion once the gating window is reopened. The interruptions and resumes involved in this cycling process complicate delivery verification and increase the risk of delivery of incorrect doses due to cumulative gantry positioning inaccuracy[39]. Therefore, the majority of Institutions utilize ITV based VMAT for moving targets, while accepting moderate additional PTV volumes (5 mm to 10 mm toensure that the treatment remains robust despite motion uncertainty, while maintaining the delivery efficiency benefits of VMAT[2].

2.6.3 Quality Assurance and Delivery Verification

Quality assurance (QA) for intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) requires that specific to each patient, dose distribution as planned be verified against dose distribution delivered to the patient. Dosimetric verification is a means to compare the planned dose distribution with the dose distribution delivered[13, 40]. Standard QA protocols include[40]: (1) Dose distribution verification: absolute dose measurements in a phantom at either isocenter or high-dose/low gradient regions, and an acceptable agreement between the planned and measured doses as ±3 % (2) Dose distribution verification: two-dimensional or three-dimensional dose measurements in a phantom using detector arrays (i.e., MapCHECK, ArcCHECK, Delta4), or radiochromic films; and analyzed via gamma index criteria (commonly 3%/3 mm global normalization, 90 % passing threshold); (3) Positional accuracy of multileaf collimator (MLC): log file analysis of commanded versus actual leaf positions during delivery, with a tolerance commonly of ±1 mm RMS error[40, 13].

Due to the continuously changing gantry angle, dose rate, and MLC position during delivery of VMAT, the required QA will be more complex than for static-field IMRT[40].To measure the dose delivered from all gantry angles during rotational phantom delivery, cylindrical detector arrays (i.e., ArcCHECK, Octavius 4D) can be used to reconstruct three-dimensional dose distributions for comparison to planning system calculations[40]. Plans created for VMAT lung cancer treatment will have gamma analysis pass rates greater than 95 % using 3%/3 mm criteria, whereas similar plans calculated with AAA algorithms will have pass rates of 88 % to 92 %— indicating that the Type C dose calculation algorithms are superior to the Type A dose calculation algorithms in modeling the heterogeneity of the lung [3].

2.7 Summary

Intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) are both sophisticated examples of current radiation delivery technologies that allow for increased flexibility and control over dose delivery in 3-Dimensional space. Both modalities offer advantages in the delivery of lung cancer treatments; specifically VMAT has been shown to reduce delivery times by 50 to 70%, and improve conformity relative to IMRT; however VMAT also increases the amount of low-dose spread. Due to these differences in delivery efficiencies and conformity, the decision to use either IMRT or VMAT should be made on a case-by-case basis,

based on factors such as tumor location, respiratory motion amplitude, and patient tolerance. Regardless of which modality is chosen, both modalities will require thorough and rigorous commissioning and quality assurance testing prior to delivery to the patient to provide safe and accurate dose delivery in the complex anatomy of the thorax.

Chapter 3 Literature Review: Photon Dose Calculation Algorithms

3.1 Introduction

The accurate prediction of absorbed dose distribution in three-dimensional space is a basic necessity for today's radiotherapy treatment planning systems using external beam treatments. To achieve an average dose accuracy within 5 %, as recommended by ICRU, requires the implementation of dose calculation algorithms able to predict dose distribution accuracies within 3 %, taking into account all uncertainties in patient positioning, machine calibration, and beam delivery[41]. In particular, these requirements become extremely difficult in those areas of anatomy characterized by large variations in tissue heterogeneity, such as the thoracic area, where there is a variation of three orders of magnitude in the physical density among structures close to each other, i.e., air-filled cavities ($0.001\,\mathrm{g\,cm^{-3}}$), lung parenchyma ($0.035\,\mathrm{g\,cm^{-3}}$ to $0.3\,\mathrm{g\,cm^{-3}}$), soft tissue ($0.9\,\mathrm{g\,cm^{-3}}$ to $1.1\,\mathrm{g\,cm^{-3}}$), and cortical bone ($1.4\,\mathrm{g\,cm^{-3}}$ to $1.8\,\mathrm{g\,cm^{-3}}$)[3, 42].

The linear Boltzmann equation governs the physical phenomena governing the behavior of photons during energy transfer and absorption in the presence of spatially varying densities of material. The Boltzmann equation represents the balance between the particle current, collision loss, scattering gain, and external source terms[4, 43]. Due to the complex nature of the Boltzmann equation, exact analytical solutions exist only for very simplified geometrical configurations, thus requiring the development of approximate methods that strike a balance between computational feasibility and dosimetric accuracy[44, 45]. This review presents a comprehensive evaluation of photon dose calculation algorithms based on their hierarchical treatment of radiation transport physics, with special attention paid to mathematical formulations, computational implementations, accuracy benchmarks, and clinical applicability.

3.2 Classification Framework for Dose Calculation Algorithms

External-beam dose calculation algorithms are commonly grouped into three hierarchical categories labeled Type A, Type B, and Type C, reflecting the level of detail provided for modeling secondary electron transport physics in heterogeneous media[46, 47].This hierarchical categorization framework was developed through a consensus-building process involving members of the radiation oncology physics community and illustrates progressively greater levels of physical rigor and computational complexity[48, 47].

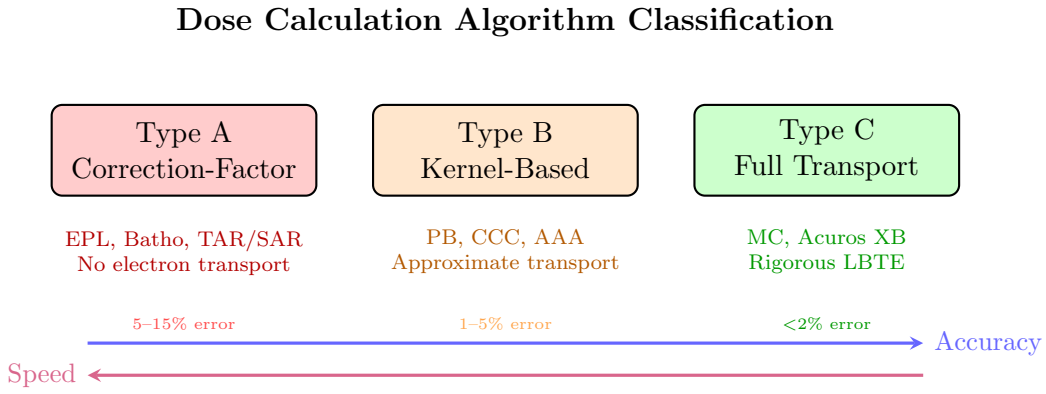


Figure 3.9: Hierarchical classification of photon dose calculation algorithms based on treatment of electron transport physics.

3.2.1 Type A Algorithms: Correction-Factor Methods

Type A algorithms represent the first generation of heterogeneity correction methods developed primarily in the late 1970s and early 1980s due to the then-severe limitations imposed by available computer hardware on the size of models that could be reasonably implemented[47, 49]. Type A algorithms do not explicitly consider the modifications in secondary electron transport produced by tissue heterogeneities; rather, they utilize empirical correction factors that were obtained using simple scaling laws applied to dose distributions calculated using homogeneous, water equivalent media[48, 47]. The fundamental assumption of Type A algorithms is that the effects of heterogeneity may be approximated by applying one-dimensional density corrections along the beam central axes, while minimizing consideration of lateral scatter perturbations[49].

Some of the most commonly used Type A methods include the Equivalent Path Length (EPL) algorithm, which modifies the dose along the beam central axis by converting the paths of the heterogeneous tissue to water-equivalent depth through density weighted summation; the Effective Attenuation Coefficient (EAC) algorithm, which accounts for beam quality dependent attenuation coefficients; the Batho Power Law correction, which was originally developed for calculating doses to lungs and employs density raised to power correction factors; and the Ratio of Tissue Air Ratios (TAR) algorithm, which uses pre-calculated values of TAR to modify dose distributions to account for heterogeneity effects[47, 49].

Due to the fact that Type A algorithms operate on a single dimension, their computational efficiency is excellent, allowing for complete three dimensional dose distributions to be generated in approximately5 s[41, 49]. Unfortunately, the superior computational efficiency of Type A algorithms comes at the expense of significant inaccuracies in the representation of the physics of dose distribution in heterogeneous media. Specifically, Type A algorithms are incapable of accurately modeling lateral photon scatter perturbations caused by differences in scattering coefficients among different tissue densities[8, 50]. Additionally, Type A algorithms neglect the significant changes in secondary electron range associated with traversals across boundaries separating materials of differing density-a feature of considerable importance in low-density lung tissue, where secondary electron ranges can exceed 2 cm for megavoltage photon beams[8, 3].

Quantitative validation studies have clearly shown that Type A algorithms cannot provide adequate predictions of dose distributions in heterogeneous anatomies. For example, in lung tissue, systematic underestimations of dose can approach $5--15\,\%$ for photon energies (4 MV to 18 MV),and increase monotonically with increasing beam energy due to increased electron disequilibrium at higher energies [8, 41]. For SBRT techniques employing small field sizes and steep dose gradients, the degree of error can exceed 20 % in low-density regions[46]. Therefore, the systematic inaccuracies inherent in Type A algorithms render them unsuitable for modern radiotherapy applications, and they have largely been abandoned in contemporary clinical practice[46, 49].

3.2.2 Type B Algorithms: Kernel-Based Convolution and Superposition Methods

Type B algorithms, also termed model-based dose calculation methods, represent a substantial advancement in computational dosimetry by incorporating physics-based models of radiation transport and energy deposition[44, 47]. These methods decompose the dose calculation problem into two conceptually distinct components: the computation of primary photon interaction sites throughout the patient volume, and the subsequent modeling of energy deposition from those interaction sites via precalculated dose deposition kernels[44, 51].

Mathematical Formulation: TERMA and Kernel Convolution The foundational mathematical framework for Type B algorithms is the convolution integral[44, 51]:

$$D(\mathbf{r}) = \int_V \mathrm{TERMA}(\mathbf{s}) \cdot K(\mathbf{r}-\mathbf{s}, \rho_{\mathrm{eff}})\, dV(\mathbf{s}), \tag{3.5}$$

where $D(\mathbf{r})$ is the absorbed dose at position $\mathbf{r}$, TERMA represents the Total Energy Released per unit MAss, K is the dose deposition kernel encoding the spatial distribution of energy transport from interaction sites, and ρ_{eff} represents effective density corrections applied to account for tissue heterogeneity[44].

The TERMA distribution quantifies the total radiant energy transferred from the primary photon field to the medium per unit mass at each spatial location[44, 44]. For a polyenergetic photon beam with energy spectrum $\phi(E)$, the TERMA at position $\mathbf{r}$ is computed via ray-tracing integration[44, 52]:

$$\text{TERMA}(\mathbf{r}, E) = \phi_0(E) \cdot \frac{\mu(\mathbf{r}, E)}{\rho(\mathbf{r})} \cdot \exp\left[- \int_{\gamma(\mathbf{s}_0 \to \mathbf{r})} \mu(\mathbf{s}, E)\, ds \right], \tag{3.6}$$

where $\phi_0(E)$ is the incident energy fluence at energy E at the phantom surface, $\mu(\mathbf{r}, E)/\rho(\mathbf{r})$ is the mass attenuation coefficient at position $\mathbf{r}$, and the exponential term accounts for attenuation of primary photons along the ray path γ from source $\mathbf{s}_0$ to $\mathbf{r}$[44, 52]. The implementation of TERMA calculation employs Siddon's ray-tracing algorithm, which efficiently determines the intersection lengths of rays with the discrete voxel grid of the patient CT volume[53].

The dose deposition kernel $K(\mathbf{r}-\mathbf{s})$ encodes the three-dimensional spatial distribution of energy deposited at location $\mathbf{r}$ resulting from a primary photon interaction at $\mathbf{s}$[44, 54]. Kernels are generated via Monte Carlo simulation in homogeneous water phantoms, tracking the transport and energy deposition of all secondary particles (Compton-scattered photons, photoelectrons, Compton recoil electrons, pair-production electrons and positrons) produced by monoenergetic point sources[55, 54]. For clinical photon beams with continuous energy spectra, polyenergetic kernels are constructed via weighted superposition of monoenergetic kernels:

$$K_{\text{poly}}(\mathbf{r}) = \sum_{j=1}^{N_E} w_j(E_j) \cdot K_{\text{mono}}(\mathbf{r}; E_j), \tag{3.7}$$

where w_j represents the spectral weight at energy bin E_j, derived from phase-space measurements or Monte Carlo simulation of the treatment head[56].

Pencil Beam Algorithm The pencil beam convolution algorithm simplifies the three-dimensional kernel to a product of one-dimensional depth-dose deposition along the beam central axis and lateral spread functions perpendicular to the beam axis[57, 58]. The dose at position (x, y, z) is computed as:

$$D(x, y, z) = \iint_A \Phi(x', y') \times K_{\text{PB}}(x - x', y - y', z; \rho_{\text{eff}})\, dx' dy', \tag{3.8}$$

where $\Phi(x', y')$ is the energy fluence distribution at the phantom surface (incorporating effects of flattening filter, jaws, and multileaf collimators), and K_{PB} is the pencil beam kernel derived from narrow-beam measurements or Monte Carlo simulations[57].

Heterogeneity correction in pencil beam algorithms employs equivalent path length scaling along the beam axis[57, 58]. The radiological depth d_{rad} is computed by density-weighted integration:

$$d_{\text{rad}}(z) = \int_0^z \frac{\rho(z')}{\rho_{\text{water}}}\, dz', \tag{3.9}$$

and dose is referenced from the kernel at radiological depth rather than geometric depth[57]. This one-dimensional correction scheme adequately models density effects on beam attenuation and primary dose deposition but fails to account for three-dimensional electron transport perturbations and lateral scatter asymmetry in heterogeneous media[8, 10].

The systematic overestimation of dose by pencil beam algorithms in low-density lung tissue is well-documented in the literature. Woo and Cunningham demonstrated dose errors ranging from 4.9 % at 4 MV to 14 % at 18 MV for lung treatments, with error magnitude increasing with beam energy due to extended secondary electron ranges[8]. For proton therapy, the failure of pencil beam methods is even more catastrophic. Taylor et al. conducted a multi-institutional phantom study demonstrating dose overestimation up to 31 % within internal gross tumor volumes and 46 % within planning target volumes for lung treatments, leading to their explicit recommendation that "pencil beam algorithms for proton dose calculation in lung are clinically unacceptable"[9].

Collapsed Cone Convolution Algorithm The collapsed cone convolution (CCC) method addresses the limitations of pencil beam algorithms by discretizing the full three-dimensional dose deposition kernel into a finite number of conical directions emanating radially from each interaction point[44, 57]. The dose is computed as:

$$D(\mathbf{r}) = \sum_{i=1}^{M} \int_{\text{cone}_i} \text{TERMA}(\mathbf{s}) \cdot K_i(\mathbf{r} - \mathbf{s}, \rho_{\text{path}})\, dV(\mathbf{s}), \tag{3.10}$$

where M is the number of discrete cone directions (typically 84 to 200), K_i is the kernel function for cone i, and ρ_{path} represents the average mass density computed along each cone ray via ray-tracing through the patient CT volume[44, 59].

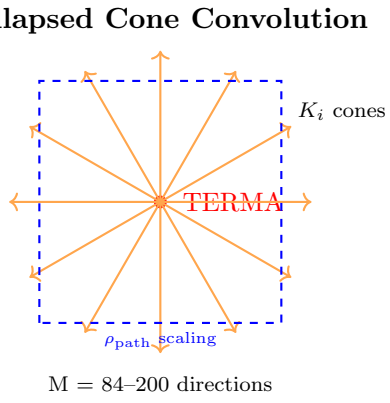


Figure 3.10: Collapsed cone convolution: energy transport from interaction point via discrete cone directions with density scaling.

The collapsed cone discretization reduces the computational complexity of full three-dimensional convolution from $O(N^7)$ to $O(N^4 M)$, where N is the number of voxels along each spatial dimension[44]. This reduction enables clinically practical calculation times of 10 s to 30 s for intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) plans while maintaining substantially improved accuracy relative to pencil beam methods[10, 48].

Inhomogeneity correction in CCC algorithms employs radiological path length scaling within each cone. The effective radiological distance d_{eff} along cone i from interaction point $\mathbf{s}$ to dose deposition point $\mathbf{r}$ is computed as:

$$d_{\text{eff}}(\mathbf{s} \to \mathbf{r}) = \int_{\gamma(\mathbf{s}\to\mathbf{r})} \frac{\rho(\mathbf{s}')}{\rho_{\text{water}}}\, ds', \tag{3.11}$$

and the kernel is referenced at $K_i(d_{\text{eff}})$ rather than geometric distance $|\mathbf{r} - \mathbf{s}|$[44, 10]. Advanced implementations incorporate kernel tilting to account for beam divergence and oblique incidence

angles[10]:

$$K_{\text{tilted}}(\mathbf{r}, \hat{\Omega}_{\text{beam}}) = K_{\text{cone}}(\mathbf{R}(\theta) \cdot (\mathbf{r} - \mathbf{s})), \tag{3.12}$$

where $\mathbf{R}(\theta)$ is a rotation matrix aligning the kernel central axis with the local beam direction, thereby reducing dose errors from 8 % to 2 % in off-axis heterogeneous regions[10].

Validation studies demonstrate that CCC algorithms achieve $2--5\,\%$ accuracy for most clinical scenarios, representing substantial improvement over pencil beam methods[48, 10]. However, approximations inherent to cone discretization and one-dimensional density scaling along cone rays produce residual errors in extremely low-density regions ($\rho < 0.2\,\text{g}\,\text{cm}^{-3}$), where electron disequilibrium effects extend beyond the geometric span of individual cones[46].

Anisotropic Analytical Algorithm Anisotropic analytical models Anisotropic Analytical Algorithms (AAAs), such as the Anisotropic Analytical Algorithm (AAA) implemented in the Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA), use pencil-beam algorithms with spatially varying scatter kernels obtained from Monte Carlo simulations[60, 61]. The AAA decomposes the total dose into three physically different components:

$$D_{\text{total}}(\mathbf{r}) = D_{\text{primary}}(\mathbf{r}) + D_{\text{scatter}}(\mathbf{r}) + D_{\text{electron}}(\mathbf{r}), \tag{3.13}$$

where D_{primary} represents energy deposition by primary photons (including their locally generated secondary electrons), D_{scatter} accounts for dose from photons that have undergone one or more Compton scattering interactions, and D_{electron} captures contamination electrons originating from beam-modifying devices (flattening filter, jaws, multileaf collimators)[60, 61].

The scatter component uses anisotropic kernels that depend upon the scattering angle and are sampled in sixteen discrete lateral directions to simulate the forward-peaked nature of Compton scattering at mega-voltage energies[60].Heterogeneity correction scales the primary dose deposition function radiologically along the beamlet propagation direction and uses electron-density-based scaling for the photon scatter kernels in the lateral direction[60, 61]. Calculating the dose using this method takes about 5 s to 10 s for IMRT/VMAT plans, which is relatively computation fast among the clinically deployed methods[11, 48].

Although the AAA is more sophisticated than many other simple Type B methods, it has been documented to have a significant systematic dose overestimation in low density lung tissues. A comparison was made between the AAA and the Acuros XB code for lung SBRT plans. The AAA produced dose values to the PTV that were 1.33 % to 3.34 % greater than those of the Acuros XB code depending on the energy of the beams (6 MV versus 10 MV flattening-filter-free beams) and calculation grid resolution[11]. Gamma analysis using 3%/3 mm criteria criteria gave pass rates of 86 % for AAA compared to 100 % for the Acuros XB code when compared to Monte Carlo simulations[3].When the phantom contained very low density lung material($0.035\,\text{g}\,\text{cm}^{-3}$), AAA gamma agreement decreased to 11 %, whereas the gamma agreement of the Acuros XB code remained at 86 % agreement[3].

The reason that the AAA consistently overestimates dose in lung is due to a lack of modeling of lateral electron transport perturbations and charged particle disequilibrium at tissue/lung interfaces[3, 42]. The build-up curves of the AAA after placing low density inserts exhibit spatial differences, beginning at the interface rather than at least several millimeters downstream of the interface as would be expected based on rigorous transport calculations[3]. This leads to errors in the calculated dose predictions for lung SBRT treatment plans that may result in a misrepresentation of adequate target coverage and therefore geographic tumor miss and potentially compromised local control unless addressed through dose prescription adjustments specific to the algorithm[46].

3.2.3 Type C Algorithms: Rigorous Radiation Transport Solvers

Type C algorithms are the most computationally intensive class of dosimetric calculations that solve the radiation transport equation in a way that explicitly models all the important physical

processes without any simplifications other than the discretization of space and/or time[4, 43]. Type C algorithms provide dosimetric accuracies close to that of measurements and allow for calculation times suitable for the clinic through advancements in both numerical techniques and hardware acceleration[45, 7].

Monte Carlo Simulation Monte Carlo methods historically have provided the gold standard for radiation dose calculations. They provide stochastic solutions to the Linear Boltzmann Transport Equation (LBTE) through the explicit simulation of the history of individual particles emitted from the source of radiation[62, 63]. The history of each particle includes all subsequent particles generated during interactions with the surrounding environment including Compton scattered photons, photoelectrons, Compton recoil electrons, pair-produced electrons and positrons, and bremsstrahlung photons[62, 64].

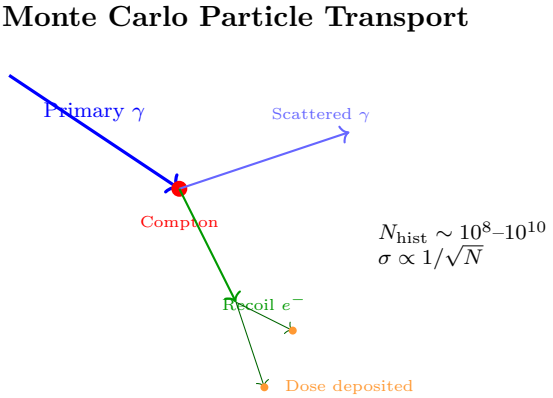


Figure 3.11: Monte Carlo simulation: stochastic tracking of individual particle histories with energy deposition scoring.

The steps involved in a Monte Carlo simulation of photon transport include: (1) Sampling the distance to the next interaction site using the exponential attenuation law $s = -\ln(\xi)/\mu_t$, where ξ is a uniform random number on $(0, 1)$ and μ_t is the total linear attenuation coefficient; (2) selecting the interaction type (Compton scattering, photoelectric absorption, pair production) via weighted sampling from partial cross-sections $\mu_{\text{Compton}}/\mu_t$, μ_{photo}/μ_t, μ_{pair}/μ_t; (3) Generating secondary particles with energies and directions determined by sampling from the Klein-Nishina differential cross-sections (Compton), photoelectric shell binding energies, or pair-production kinematics; and (4) Tracking all secondary particles through repeated applications of steps 1–3 until the energies of the secondary particles fall below predefined threshold values (usually 10 keV to 100 keV), at which point the remaining energy is deposited locally[62, 64].

The absorbed dose in each voxel is determined by summing up all the energy depositions from $N_{\text{hist}} \sim 10^8$–$10^{10}$ histories of individual particles[64, 63]. The statistical error in the voxel dose follows Poisson statistics and decreases with the square root of the number of histories: $\sigma \propto 1/\sqrt{N_{\text{hist}}}$, thus requiring quadrupling of the computation time to reduce the uncertainty by half[62]. Many modern Monte Carlo codes utilize various variance reduction techniques including particle splitting, Russian roulette, and forced detection to increase the statistical efficiency of the computation, allowing for the reduction of the statistical uncertainty in each voxel to less than $<2\,\%$ while reducing the computational[63, 64].

Validation studies demonstrate that Monte Carlo simulation is an accurate method of assessing dose calculation in complex media. The accuracy of Monte Carlo simulations are confirmed by comparisons to experimentally measured doses in anthropomorphic models of humans made with materials similar to bone, lung and soft tissue; these comparisons result in agreement within 2 % to 3 %, llargely due to uncertainty in phantom composition and detector placement rather than limitations inherent to the Monte Carlo simulation[42, 9].Furthermore, Monte Carlo

simulations can accurately model both the magnitude and distribution of dose perturbations at tissue-lung interfaces, such as charged particle equilibrium in tissue adjacent to areas of low density in lung, phenomena which cannot be adequately modeled by type-B algorithms[3].

The primary historical drawback to the use of Monte Carlo methods for routine clinical planning has been the computational time required to perform such calculations. Historically, CPU based implementations of Monte Carlo methods have taken anywhere from 2 h to 6 h per clinical plan to achieve <2 % statistical uncertainty[12, 7].however, advancements in graphics processing units (GPUs) have recently greatly diminished this barrier to entry. Currently available GPU accelerated versions of Monte Carlo methods (RayStation MC and Monaco MC) have decreased the time to calculate a treatment plan to 10 minutes to 10 min to 20 min while achieving the same level of statistical uncertainty 2 %, thereby making full Monte Carlo calculation possible for routine clinical application[7, 12].

Acuros XB: Deterministic Linear Boltzmann Transport Equation Solver Acuros XB (Varian Medical Systems, now Siemens Healthineers) has revolutionized clinical dose calculation with deterministic solutions of the Linear Boltzmann Transport Equation[4, 3]. In contrast to Monte Carlo simulation, which uses stochastic particle sampling to solve the linear Boltzmann transport equation (LBTE), Acuros XB numerically discretizes and solves the LBTE using numerical techniques used in nuclear reactor physics and astrophysical radiative transfer[4, 43].

Dose Reporting Conventions Acuros XB provides two dose reporting modes: dose-to-medium (D_m) and dose-to-water (D_w)[3, 11]. Dose-to-medium reports the energy deposited per unit mass of the actual tissue material present in each voxel, representing the physically absorbed dose:

$$D_m(\mathbf{r}) = \sum_{g=1}^{G} \sum_{n=1}^{N_{\text{dir}}} w_n \left(\frac{\mu_{\text{en},g}}{\rho} \right)_{\text{medium}} E_g \Psi_{g,n}(\mathbf{r}). \tag{3.14}$$

Dose-to-water applies stopping-power corrections to express dose as though each voxel were composed of water, facilitating comparison with traditional dosimetric measurements calibrated in water phantoms[3]:

$$D_w(\mathbf{r}) = D_m(\mathbf{r}) \times s_{\text{water,medium}}, \tag{3.15}$$

where $s_{\text{water,medium}}$ is the ratio of mass collision stopping powers. For cortical bone, $s_{\text{water,bone}} \approx 1.09$–$1.15$, resulting in $10--14\,\%$ higher D_w relative to D_m; for lung tissue, $s_{\text{water,lung}} \approx 1.0$, yielding minimal difference between reporting modes[3]. Contemporary clinical practice increasingly favors dose-to-medium reporting as more biologically relevant[11, 12].

Validation and Clinical Performance Validations of extensive studies have confirmed that Acuros XB is able to achieve a Monte Carlo equivalent level of accuracy but eliminates statistical noise and reduces computational time by a factor of 10–50×[3, 7]. Bush et al. compared Acuros XB to VMC++ Monte Carlo simulations in heterogeneous slab phantoms containing normal lung ($0.198\,\text{g cm}^{-3}$), very low-density lung ($0.035\,\text{g cm}^{-3}$), and cortical bone ($1.798\,\text{g cm}^{-3}$) inserts, reporting gamma pass rates (3%/3 mm criteria) of 100 %, 86 %, and 100 %, respectively[3]. By comparison, AAA achieved only 86 %, 11 %, and 100 %gamma passing (same configurations) indicating Acuros XB has significantly better performance in the extreme low-density media [3].

Validation of clinical lung SBRT using PRIMO Monte Carlo code as a reference demonstrated superior accuracy of Acuros XB over AAA. Sarin et al., calculated dose differences (mean) of $0.6--1.7\,\%$ between Acuros XB and Monte Carlo across ten lung SBRT patients, compared to 4.1 % for AAA[42]. The dose-volume histogram (DVH) data for both planning target volumes and organs-at-risk showed closer agreement to Monte Carlo for Acuros XB than AAA[42, 12].Additionally, Acuros XB effectively modelled the "rebuild-up" effect following

low-density lung inserts, as it captured the gradual restoration of charged-particle equilibrium occurring several millimeters distal to the tissue-air interface - a characteristic missing from AAA calculations[3].

The increased computational efficiency of Acuros XB (substantially larger than Monte Carlo) represents an approximate increase of 10 times compared to Type B algorithms. Calculation times for lung SBRT VMAT plans ranged from 42 s to 94 s (mean 66 s) for Acuros XB, compared to 5 s to 9 s for AAA[12, 11]. Although there is a 10 fold increase in the time to calculate Acuros XB, the time difference is clinically acceptable and is compensated for by avoiding the need to adjust dose prescriptions based on specific algorithms or post-processing dose correction[11].

3.3 Comparative Algorithm Performance and Clinical Selection Criteria

3.3.1 Quantitative Accuracy Benchmarks

Table 3.2 summarizes dosimetric accuracy metrics for representative photon dose calculation algorithms based on validation studies employing Monte Carlo simulation or experimental measurements as reference standards.

Table 3.2: Comparative dosimetric accuracy of photon dose calculation algorithms for lung radiotherapy. Lung accuracy represents mean dose difference within planning target volume relative to Monte Carlo reference. Gamma analysis employs 3%/3 mm criteria with 10% low-dose threshold. Calculation times are for volumetric modulated arc therapy (VMAT) lung stereotactic body radiotherapy plans on modern workstations.

Algorithm	Type	Lung Accuracy	Gamma Pass Rate	Calculation Time	Statistical Noise	Reference
Pencil Beam	A	−10 to −46%	60–75%	<5 s	None	[8, 9]
EPL/Batho	A	−5 to −15%	70–80%	<5 s	None	[49]
AAA	B	−1.3 to −3.3%	86–92%	5–10 s	None	[11, 3]
CCC	B	−0.8 to −2.5%	91–96%	10–30 s	None	[10, 48]
Acuros XB	C	+0.6 to −2.0%	98.9–100%	40–95 s	None	[3, 11, 12]
Monte Carlo	C	Reference	100%	10–360 min	<2%	[7, 12]

Multiple important conclusions result from this comparison of the algorithms. Firstly, type-A algorithms show a consistent error in underestimating the dose to lower density lung tissue and therefore, they cannot be used in modern clinical practices due to their inability to accurately model the radiation in heterogeneously structured tissues in the treatment of tumors with Stereotactic Body Radiation Therapyl[9, 46]. Secondly, type-B algorithms (AAA and CCC) appear to be sufficient for many soft tissue treatment sites; however, these algorithms also consistently overestimate the dose delivered to the lungs by an average of 1 % to 3 %, which can lead to target underdosing and compromise local control when prescription doses are not appropriately reduced[11, 46]. Thirdly, Acuros XB is capable of achieving Monte-Carlo equivalent dosimetry (average dose differences of (0.6 % to 2 % dose difference) with clinically practical calculation times (40 s to 95 s), linically acceptable calculation times, providing the optimal combination of dosimetric rigor and computational efficiency for the routine use in clinical settings[3, 12]. Fourthly, GPU accelerated Monte-Carlo has become a viable option for institutions that prioritize accuracy and are willing to tolerate slightly longer(10 min to 20 min)than typical calculation times[7].

3.3.2 Clinical Selection Guidelines by Treatment Site

The selection of dose calculation algorithm for clinical treatment planning should be informed by anatomical site-specific considerations of tissue heterogeneity, target-to-organ-at-risk proximity,

dose prescription magnitude, and regulatory or clinical trial requirements[46, 49].

For lung cancer treated with stereotactic body radiotherapy or conventionally fractionated intensity-modulated radiotherapy, Acuros XB or Monte Carlo algorithms are strongly recommended due to the extreme tissue heterogeneity characteristic of the thorax[3, 42, 46]. Validation studies have conclusively demonstrated that AAA overestimates planning target volume dose by 1.3 % to 8 % depending on tumor size, location, and beam arrangement, creating risk of geographic tumor miss and compromised local control[11, 42]. Major cooperative group protocols (NRG Oncology RTOG0813, RTOG0915) currently mandate Type B algorithms as minimum standard but increasingly recommend or require Type C algorithms for improved accuracy[46]. The use of Type A algorithms (pencil beam, equivalent path length) is explicitly contraindicated for lung radiotherapy[9, 49].

For head-and-neck cancer, AAA or collapsed cone convolution algorithms provide acceptable accuracy for most treatment volumes, with mean dose differences <2 % relative to Monte Carlo[48]. However, Acuros XB is preferred for cases involving extensive low-density regions (air-filled sinuses, nasopharynx, larynx) or when high-dose regions abut critical organs at risk requiring stringent dose constraints (optic chiasm, brainstem, spinal cord)[3, 46].

For prostate cancer, the choice among AAA, collapsed cone convolution, or Acuros XB produces minimal dosimetric differences (<1 % in planning target volume) due to the relatively homogeneous soft-tissue pelvic anatomy[11, 48]. Algorithm selection in this context is primarily driven by institutional preference and optimization of treatment planning workflow efficiency[46].

For breast cancer, particularly partial-breast irradiation or chest wall treatments, Acuros XB is preferred over AAA due to the presence of lung-breast interface and chest wall curvature introducing heterogeneity effects that Type B algorithms inadequately model[41, 46]. Systematic underprediction of chest wall dose by 2 % to 5 % has been documented for AAA, with potential implications for local control in post-mastectomy radiotherapy[41].

3.4 Computational Efficiency and Clinical Workflow Integration

The practical utility of dose calculation algorithms in contemporary radiotherapy is determined not solely by dosimetric accuracy but equally by computational efficiency and seamless integration into iterative treatment planning workflows[45, 46]. Modern inverse planning optimization for intensity-modulated radiotherapy and volumetric modulated arc therapy requires hundreds to thousands of dose calculations during gradient-based or stochastic optimization of beam aperture shapes and intensities[65]. This necessitates dose calculation algorithms capable of completing full three-dimensional dose distributions within seconds to minutes to maintain clinically acceptable total planning times[46].

Type A algorithms, despite their dosimetric inadequacy, achieve calculation times below 5 s, enabling real-time dose visualization during interactive planning[49].Type B Algorithms (AAA, collapsed cone convolution) are able to achieve the same level of computational efficiency that is characteristic of Type A Algorithms, but at a much higher degree of accuracy. Thus, they have served as the clinical workhorse for most of the treatment locations from the 1990s to date[48, 47]. The relatively short 5 s to 30 s calculation times allow dozens of iterations in an optimization cycle within a total optimization time of 5 min to 15 min[46].

The main computational hurdle associated with the development of Type C Algorithms has been finding a way to implement rigorous physics models within the constraints of time typical of a clinical environment. The first Monte Carlo based implementations required 2 h to 6 h per plan,which made it impossible to utilize them for iterative optimization. As such, Monte Carlo was relegated to final dose verification and/or research applications[64, 62]. Acuros XB's ability to achieve Monte Carlo equivalent accuracy in 40 s to 95 s—representing only 5–10× increase over AAA—constitutes a paradigm shift enabling routine clinical deployment of Type C algorithms[3, 11]. n addition to its ability to provide Monte Carlo equivalent accuracy, the

deterministic nature of Acuros XB (i.e., the lack of statistical noise) provides an advantage for optimization applications compared to Monte Carlo. This is because the stochastic dose fluctuations inherent in Monte Carlo make it difficult to accurately calculate gradients and assess convergence[45].

Recent advances in graphics processing unit (GPU) hardware and specialized tensor processing units (TPUs) promise further dramatic reductions in calculation time for both deterministic and stochastic transport solvers[7, 66]. GPU-accelerated implementations of Acuros-class algorithms targeting <10 s calculation times are under active development, potentially enabling real-time dose visualization with Type C accuracy[43]. Simultaneously, GPU-accelerated Monte Carlo codes have reduced calculation times to 10 min to 20 min, rendering full Monte Carlo viable for final dose calculation following optimization with faster surrogate models[7].

An emerging paradigm leverages deep learning to train neural networks as ultrafast dose prediction surrogates, achieving <1 s inference time with 2 % to 3 % accuracy relative to ground-truth Acuros XB or Monte Carlo calculations[67, 68]. These models enable thousands of dose evaluations during optimization, with final dose verification performed using the rigorous Type C algorithm employed for training data generation[68]. This hybrid approach synergistically combines the speed of learned surrogates with the accuracy of physics-based transport solvers, representing a promising direction for next-generation treatment planning systems[68].

3.5 Current Limitations and Future Directions

Although there has been significant advancement in the accuracy and efficiency of photon dose calculations, many limitations remain that restrict how precise dose can be delivered during radiotherapy. The calculation of doses for small fields (<2 cm presents a challenge to all types of algorithms as it results from lateral electronic disequilibrium, partial source occlusion and the effect of detector volume averaging which obscure the actual dose distribution[69]. Systematic errors in CT to Electron Density calibration curves range from 1 % to 2 % , and these errors are propagated into all dose calculation algorithms[70]. Temporal changes in patients anatomy between simulation imaging and treatment delivery (inter-fraction motion, tumor shrinkage, weight loss) make the static dose distributions calculated during planning invalid[71].

Therefore, future advancements will most likely involve the integration of dose calculation into online adaptive radiotherapy processes, where treatment plans are re-optimize daily using pre-treatment imaging acquired immediately prior to each fraction[72]. For this type of process to occur, dose calculations must be accomplished at an extremely fast rate (<30 s) , so that the entire plan can be adapted to within clinically acceptable time frames for treatment couch times. Using deep learning techniques to develop dose predictive models that were trained on comprehensive data sets developed with Acuros XB or Monte Carlo methods represent a promising avenue to achieve both the speed and accuracy required by such an application [68].

The development of large scale, high quality training data sets consisting of thousands of treatment plans with accurately calculated dose distributions is key to furthering the applications of Artificial Intelligence in radiotherapy treatment planning[68]. Recent efforts have focused on automated generation of such datasets using scripted treatment planning pipelines integrating Acuros XB for dose calculation[68]. The HMM-RT lung cancer dataset, comprising 1,622 clinically approved plans calculated with Acuros XB version 18.0.1, represents the largest publicly available repository of physics-accurate dose distributions for algorithm development and validation[68]. Continued expansion of such resources will catalyze development of next-generation planning systems combining human expertise with machine intelligence to optimize therapeutic outcomes.

3.6 Summary

The evolution of photon dose calculation algorithms from empirical correction-factor methods to rigorous radiation transport solvers represents one of the signal achievements of modern radiation oncology physics. Type A algorithms are obsolete for clinical use except in resource-constrained settings. Type B algorithms (Anisotropic Analytical Algorithm, collapsed cone convolution) remain widely deployed and provide adequate accuracy for many treatment sites, but systematically overestimate dose in heterogeneous lung tissue by 1 % to 8 %.TYPE C, the accuracy of the algorithms - Acuros XB and Monte Carlo — used in radiation treatment planning is verified at the 0.6 % to 2 % level (accuracy) when compared to measured data in experiments and also have computation time that can be performed by a computer in a time frame that is considered to be "acceptable" for clinical use.

Therefore, the use of Acuros XB or Monte Carlo algorithms is highly recommended for all lung cancer patients undergoing radiotherapy because these two types of algorithms provide the highest degree of precision in calculating the dose distribution that will occur throughout the body due to heterogeneity of tissues in the thorax; and, the deterministic solution provided by Acuros XB to the Linear Boltzmann Transport Equation provides an equivalent degree of accuracy to Monte Carlo methods with no statistically induced "noise", thus making it the current standard of care for thoracic radiotherapy dose calculations.

3.7 Research Gap and Thesis Contribution

Despite significant advances in dose calculation methodology, several critical gaps remain in both the scientific literature and clinical practice.

3.7.1 Lack of Open-Source Deterministic Transport Implementation

Acuros XB has demonstrated that deterministic LBTE solvers can achieve Monte Carlo-equivalent accuracy with clinically acceptable computation times[4, 3]. However, Acuros XB is a proprietary closed-source system. Academic researchers cannot independently verify dose calculations, cannot modify algorithms for research applications, and cannot integrate the solver with emerging machine learning frameworks. This lack of transparency fundamentally limits reproducibility and innovation in radiation dose calculation research.

3.7.2 Computational Cost Prohibits Inverse Planning Integration

Computational Cost Makes Direct Integration of Inverse Planning Impossible For each Monte Carlo simulation, it takes about 5-20 minutes to calculate a full resolution dose[5, 6]. nverse planning for optimization of IMRT and VMAT usually involves approximately 500-5000 dose calculations during the iterative optimization process[1, 32], 40-1600 hours of computational time per treatment plan if MC were used directly, obviously impossible for clinical use. Currently, all clinical optimizations are done using fast but less accurate algorithms during optimization, and then the final dose is calculated by either Acuros XB or MC. This two step method will introduce systematic errors since the optimizer has never seen the true dose distribution.

3.7.3 No Integration with Physics-Informed Machine Learning

There have been recent breakthroughs in the area of physics-informed machine learning (PINNs), where hybrid approaches that combine deep learning with physical governing equations can speed up traditionally numerical solutions by 10-100[73, 74]. However, application of PINNs to radiation dose calculation requires a differentiable, physically interpretable dose calculation framework. Closed-source commercial solvers cannot be differentiated. Monte Carlo codes are inherently non-differentiable due to stochastic sampling. An open-source deterministic LBTE

solver with transparent mathematical formulation is therefore a prerequisite for physics-informed machine learning research in radiation oncology.

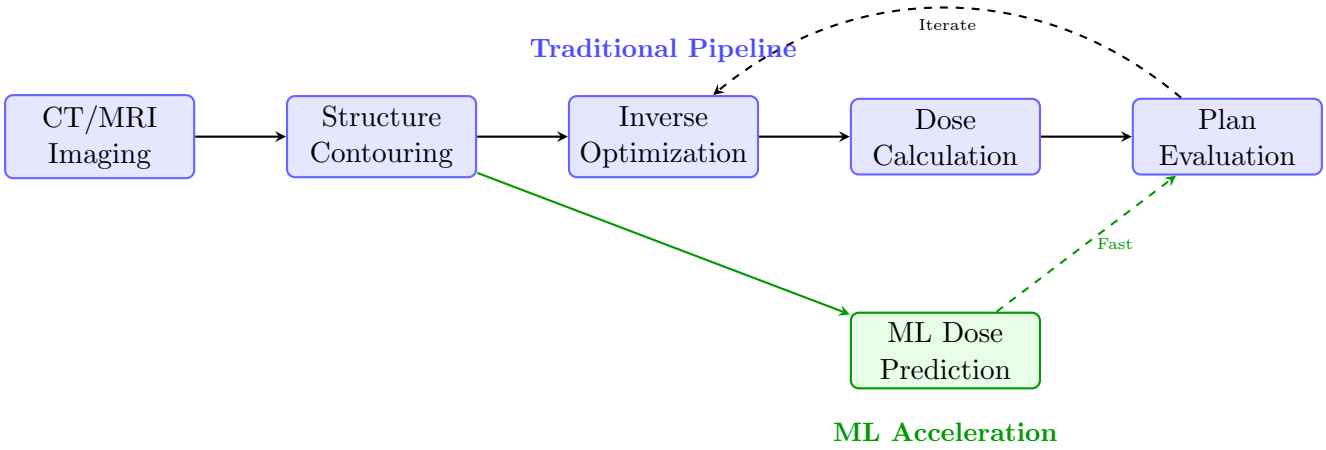


Figure 3.12: Integration of machine learning dose prediction into the radiotherapy treatment planning pipeline. The ML model can bypass iterative optimization by directly predicting clinically acceptable dose distributions from patient anatomy.

Chapter 4 Machine Learning for Dose Prediction

4.1 Introduction

External beam radiotherapy is an essential modality in cancer treatment, with approximately half of all cancer patients requiring radiation at some point during their disease course [75]. The overarching goal of radiotherapy is to deliver a tumoricidal dose to the planning target volume (PTV) while minimizing irradiation of surrounding organs at risk (OARs). Over the last decades, technical advances such as three-dimensional conformal radiotherapy, intensity-modulated radiotherapy (IMRT), volumetric modulated arc therapy (VMAT), and image-guided and adaptive radiotherapy have significantly increased the conformality and complexity of treatment plans [1, 32]. However, these advances come with a substantial increase in the dimensionality of the optimization problem that must be solved during treatment planning.

Radiotherapy treatment planning depends upon iterative optimization processes, where the dose distribution is calculated through use of dose engines, which include pencil beam convolution, collapsed cone convolution, or Monte Carlo simulations [76, 77]. Although Monte Carlo simulations are thought of as the "gold standard" regarding accuracy of dose calculation, Monte Carlo simulations are very computationally intensive, making them difficult to utilize in real-time or near-real-time in clinical settings. Even though the cost of other dose engine calculations is lower than that of Monte Carlo simulation, complex treatments of intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) may take tens of minutes to many hours to compute for each individual patient, and repeated re-optimizations will be required during online adaptive treatments to account for anatomical changes [78, 79]. Therefore, an obvious need exists for methods to speed up the dose calculation process, and, more broadly, the entire process of predicting dose and generating treatment plans. Machine learning and, specifically, deep learning, using Convolutional Neural Networks (CNNs), and similar architectures have been identified as possible solutions to this issue [80].

Rather than repeatedly solving the full optimization problem from scratch, a machine learning model can learn mappings from patient specific input (e.g., computed tomography images, Planning Target Volume (PTV) and Organ at Risk (OAR) contours, etc.), to dose distributions, fluence maps, or machine parameters, utilizing large databases of approved treatment plans as the training data. These models can then be utilized to rapidly predict dose in a small portion of a second, essentially acting as learned dose engines or knowledge-based planning systems. Simultaneously, efforts are being made to incorporate physics-based constraints[81, 73], uncertainty quantification [82], and end-to-end differentiable pipelines [83], with the aim of closing the gap between data-driven predictions and physically deliverable, clinically acceptable treatment plans.

In this review, we primarily focus on techniques that either use machine learning to speed up dose calculation directly, or use machine learning to make dose prediction as part of a larger knowledge-based or physics-guided planning system. We place special emphasis on the studies including the voxel-based dose prediction by McIntosh and Purdie [84], the analysis of the OpenKBP Grand Challenge Dataset by Babier and colleagues [85], the DiffDP model based on diffusion [86], the flexible-cm GAN for heterogeneous radiotherapy scenarios[86], the flexible-cm GAN for heterogeneous radiotherapy scenarios [67], the 3D U-Net-based method by Hien et al. using a DVH-based loss function [87], the physics-informed beam dose prediction by Sun et al. [88], the two-stage physics-guided planning framework of Achlatis et al. [83], the simultaneous dose and fluence prediction model by Li et al. [89], and the evidential learning framework by Tan et al. for uncertainty-aware dose prediction [82]. We position these studies within the broader literature on deep learning for dose prediction, provide insight into their respective advantages and disadvantages, and outline a path forward for future work.

4.2 Foundations of Dose Prediction and Accelerated Dose Calculation

The task of dose prediction in radiotherapy can be viewed as learning a mapping from patient specific information and, possibly, planning parameters to a three-dimensional dose distribution. In a traditional supervised learning environment, one is given CT images, defined structures (PTVs and OARs), and true dose distributions from previously approved treatment plans, and the objective is to develop a model capable of predicting the dose distributions of new patients without having to run the full optimization and dose calculation chain.

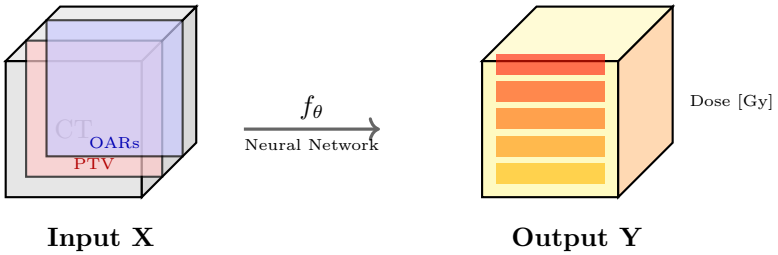


Figure 4.13: Dose prediction as a supervised learning problem: the neural network f_θ maps multi-channel input (CT, structure masks) to a 3D dose distribution.

The earliest approaches in this direction were atlas-based or statistical models rather than purely deep learning-based. McIntosh and Purdie proposed a voxel-based dose prediction framework that uses multi-patient atlas selection and a conditional random field (CRF) prior to predict dose distributions directly from anatomical information [84]. Their model exploited a library of previously treated patients, selecting similar "atlases" based on anatomical similarity and using a conditional random field to propagate dose distributions in a voxel-wise manner. Across six treatment sites and hundreds of patients (633 training, 113 testing), they achieved significant improvements in mean absolute error (MAE) over baseline approaches, demonstrating that meaningful dose prediction from anatomical features alone was feasible even before the current wave of deep learning. Their work established fundamental principles for knowledge-based planning and demonstrated that historical clinical data could be effectively leveraged to inform future treatment plans.

The paradigm shift occurred with the availability of large, standardized datasets and deep neural network architectures capable of learning high-dimensional mappings. The OpenKBP (Open Knowledge-Based Planning) grand challenge, introduced by Babier and co-workers in 2020, provided a benchmark dataset of 340 head-and-neck cancer patients treated with nine-beam IMRT [85]. For each patient, CT images, PTV and OAR masks, and the corresponding ground-truth dose distribution were made available, along with standardized training (200 patients), validation (40 patients), and test (100 patients) splits and well-defined performance metrics. Two metrics became particularly important: a voxel-wise dose score (mean absolute error of dose per voxel) and a DVH score that summarizes the discrepancy in dose-volume histogram (DVH) criteria over PTVs and OARs. The challenge attracted 195 participants from 28 countries, with 73 teams forming in the validation phase and receiving 1,750 total submissions [85]. The best-performing methods were based on 3D U-Net architectures and achieved DVH scores around 1.478 Gy, showing that deep learning could approximate clinically realistic dose distributions with high fidelity.

The difference between Dose Prediction and Dose Calculation should be clarified. Dose Calculation is traditionally done by using Physics-Based Models of Radiation Transport, typically using Monte Carlo or Deterministic Algorithms to solve the equations of the model; in contrast, Dose Prediction Models utilize Deep Learning to model the outputs of those algorithms based on the input parameters provided during Planning. Therefore, Dose Prediction using Deep Learning can be considered a form of Accelerated Surrogate Modeling for Dose Calculation,

provided that the Training Data has been generated by Accurate Physical Dose Computation. In practice, nearly all DL Models have been trained using Clinical Dose Distributions which are the result of both Dose Calculation and Human Planning/Optimization Choices; therefore, they inherently contain a combination of Physics and Planning Preferences.

Additionally, there are other lines of research that are utilizing Machine Learning at the Fluence Optimization Level and/or at the Level of Predicting Machine Parameters. For example, Multi-Agent Reinforcement Learning was used for Leaf Sequencing[90], and Models such as Flexible-CmGAN or Physics-Guided Planners were used to predict Multileaf Collimator (MLC) Apertures and Monitor Units (MU)[67, 83], while a Differentiable Dose Engine will then use those predicted MLC Apertures/MU to predict the corresponding 3D dose distribution and therefore integrate the dose prediction into a Closed Planning Pipeline. These approaches blur the distinction between dose prediction and treatment planning and create an opportunity for a full learned planning system capable of replacing a significant portion of the traditional optimization process.

4.3 Deep Learning Architectures for Dose Prediction

4.3.1 U-Net and Convolutional Neural Networks

The large majority of dose prediction methods rely on convolutional neural networks which take advantage of 3D volumetric input data. Specifically, the U-Net architecture [91] and all of its variants have been widely adopted because of their encoder-decoder structure with skip connections. The skip connections allow for the preservation of the spatial detail of the input data while allowing for the aggregation of context across multiple scales.An example of a dose prediction method based on a 3D convolutional neural network is that of Hien and colleagues

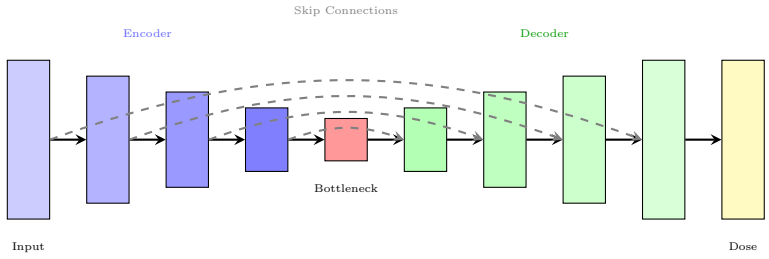


Figure 4.14: U-Net architecture for dose prediction showing encoder path (downsampling), bottleneck, decoder path (upsampling), and skip connections that preserve spatial resolution.

who created a custom 3D CNN specifically designed to predict doses in head and neck radiotherapy using the OpenKBP dataset [87]. Their architecture is composed of two cascading 3D U-Net-like models. The first model creates a coarse dose distribution and the second refines that coarse dose distribution. Both models take as input 128 x 128 x 128 volumes containing CT images and binary masks for OARs and PTVs. They introduce resi-dual connections [92] in the early layers to reduce vanishing gradients and to maintain low-level structural features, and use skip connections between the same encoder and decoder levels to recover the fine spatial detail during upsampling. A rectified linear unit (ReLU) activation is applied to the final layer to ensure that the output dose values are physically meaningful by pre-venting negative dose values.

A key contribution of Hien and colleagues is the development of a dose-volume histogram-based loss function to train the final model. Rather than optimizing solely for voxel-wise mean squared error or mean absolute error, Hien and colleagues compute a set of DVH-derived metrics, such as D_1, D_{95}, and D_{99} for PTVs (doses received by 1%, 95%, and 99% of the target volume) and mean dose (D_{mean}) and near-maximum dose ($D_{0.1cc}$) for OARs, and define a loss function

that penalizes discrepancies between these metrics for the predicted and ground-truth doses [87]. The DVH-based loss is formulated as:

$$\mathcal{L}_{DVH} = \frac{1}{n_p} \sum_{c} |DVH_c(D_{pred}) - DVH_c(D_{true})| \tag{4.16}$$

where n_p is the number of possible DVH criteria for criteria p, and DVH_c represents one of the clinically relevant metrics. By doing so, Hien and colleagues ensured that the predictions are both numerically close to the ground truth and clinically meaningful in terms of dose-volume constraints. On the OpenKBP test set, their model yielded a DVH score of approximately 1.444 Gy, outperforming several baseline models, including 3D U-Net variants submitted to the original challenge [87]. This result demonstrates that careful architectural design and loss function engineering can lead to state-of-the-art dose prediction performance.

Some other U-Net-based approaches have explored different design choices. Some of these approaches incorporated dense connections to promote feature reuse and to enhance gradient flow [93], while others employed dilated convolutions or hierarchical attention mechanisms to increase the receptive field without excessive parameter growth. Other works used cascaded or multi-stage architectures where an initial network produces a coarse dose, which is then refined by a second network, similar to coarse-to-fine strategies commonly found in image restoration [94].

4.4 Multi-Task Learning: Simultaneous Dose and Fluence Prediction

In addition to solely dose-focused models, some studies extended dose prediction to multiple modalities. Li et al. proposed a new method for jointly predicting three-dimensional dose distributions and two-dimensional fluence maps from a shared encoder, particularly for nasal pharynx carcinoma (NPC) intensity-modulated radiation therapy (IMRT) [89]. n that case, the encoder learns a common representation of the CT and the structure masks, which is then decoded to a volumetric dose map along one branch and to 2D fluence maps per beam along another branch, after applying a geometric projection operation. The multi-task setup takes advantage of the high correlation between dose and fluence and enhances overall prediction performance.

Li et al.'s model was trained and tested on 340 NPC patients (260 training, 40 validation, 40 testing). The fluence prediction resulted in a mean absolute error of 0.53 ± 0.13 (normalized to 2000 MU) and structural similarity index (SSIM) of 0.96 ± 0.02, with gamma passing rates exceeding 92% using 3%/3 mm criteria [89]. Most of the structures showed less than 1 Gy differences in DVH indices compared to ground truth. Moreover, Li et al. showed that the predicted fluence could be imported into the Eclipse treatment planning system for MLC leaf motion calculation and final dose verification, demonstrating the clinical applicability of the predicted plans. Therefore, the work of Li et al. shows that multi-task learning may help to bridge the gap between dose prediction and practical treatment planning implementation.

4.5 Transformer-Based Architectures

BTransformer architectures can be used to go beyond the use of purely convolutional architectures. Achlatis et al., developed a version of the UNETR [95] architecture for a vision transformer (ViT), where the encoder consists of a vision transformer processing 3D patches and the decoder reconstructs spatial feature maps through de-convolution and skip connections[83]. This was termed as "Deep RT Planner" in a physics guided pipeline that determines VMAT MLC aperture and MU. The architecture divides the 3D input volume into non-overlapping 16 x 16 x 16 patches, which are embedded in a latent space using positional encoding. A simplified version of the ViT, with four transformer blocks processes the sequence, producing multi-scale representation features.

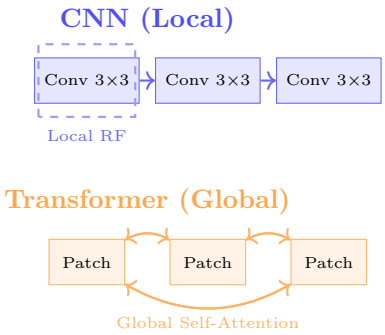


Figure 4.15: Comparison of CNN (local receptive field through stacked convolutions) and Transformer (global attention between all patches) paradigms for feature extraction.

While they were able to show that the transformer-based model had performance comparable to a well-tuned 3D U-Net, it did not appear to have an overwhelming advantage over the 3D U-Net, most likely because of the small size of the dataset used (104 training cases, 16 validation cases, 13 test cases). Generally, transformers [96] tend to be very hungry for data, which could have contributed to their somewhat limited success. It appears that, at least based on the current state-of-the-art research, CNNs continue to be the most reliable option for the type of task described above; that is, hundreds of patients. The 3D U-Net version employed in the study had average differences in PTV D95 of 0.42±1.83 Gy and V95 of 0.22 ± 1.87%, post-physics-guided training, which indicates that the results from this version were strongly aligned with clinical goals [83].

4.6 Generative and Physics-Informed Models

4.6.1 Generative Adversarial Networks

A large number of studies go beyond the point-wise regression and employ generative frameworks such as Generative Adversarial Networks (GANs) to create dose distributions with improved realism and sharper structural characteristics. Due to the fact that GANs can create realistic samples that are representative of the distribution of the training data [97] they have been successful in various medical imaging applications.

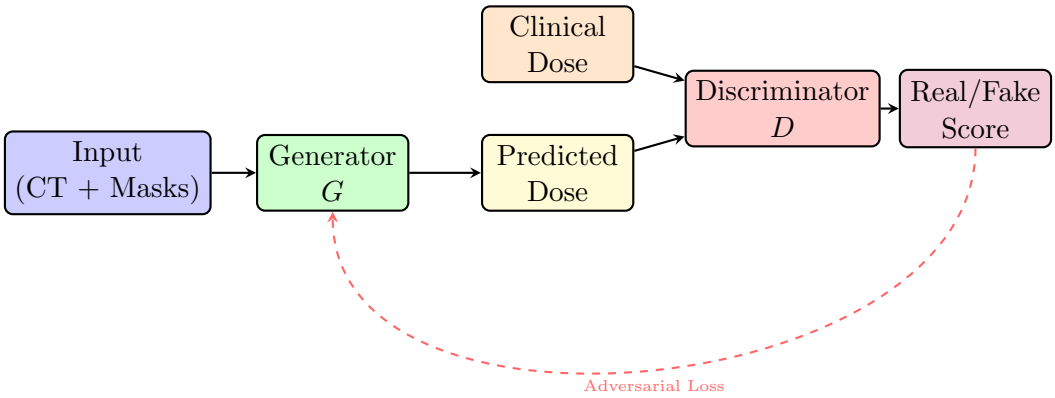


Figure 4.16: GAN architecture for dose prediction: the generator produces dose from anatomy, while the discriminator learns to distinguish real from predicted doses, providing adversarial feedback.

One example of a Conditional GAN (CGAN) is the Flexible-Cm GAN proposed by Gao et

al. This CGAN model is intended to predict three dimensional dose distributions [67], however unlike previous GAN based dose prediction models that focused on relatively homogeneous planning situations (i.e., single beam modality and fixed beam configurations) [98, 99], Flexible-Cm GAN is capable of handling a variety of clinical situations with multiple planning modes, including static-beam IMRT and dynamic-beam VMAT, as well as variable beam angles. As part of generating doses, the model uses not only the CT images and the PTV/OAR masks as conditionals, but also discrete variables representing planning mode and continuous representations of beam geometry.

One of the main challenges addressed by this work is the presence of missing conditions within clinical data. For instance, in many cases beam configuration information will be either unknown or unavailable for some patients. The authors develop a miss-consistency loss to address this issue. The loss function is defined as follows:

$$\mathcal{L}_{mc} = \sum_{j \neq i, m_j = 0} \| E_j(G|m_i = 0) - E_j(G) \| \tag{4.17}$$

where E_j extracts features corresponding to condition j, and the loss enforces consistency between predictions when condition i is missing versus present [67].

Flexibly optimizing CmGAN incorporates both an adversary loss for image quality as well as two additional loss functions; a reconstruction loss and a unique shift-dose-volume (SDV) loss which are based on dose volume histogram (DVH) principles. The SDV loss is formulated such that it maps the loss function of errors in DVH space to voxel space such that voxels with large contributions to the discrepancy of the DVH in high-dose regions have a larger effect than voxels with small contributions:

$$\mathcal{L}_{SDV} = \sum_{t=1}^{T} \sum_{j \in M_t} |Y_j - \hat{Y}_j| \cdot w \tag{4.18}$$

where T is the number of dose bins, M_t is the set of voxels in bin t, and w is the bin width. This provides a principled, differentiable surrogate for DVH-based planning objectives [67].

In experiments on two lung cancer datasets-Dataset 1 with angle configuration available (365 patients: 256 training, 30 validation, 70 test) and Dataset 2 without angle configuration (793 patients: 613 training, 60 validation, 120 test)Flexible-Cm GAN achieved better performance than many other baselines that are very competitive with the best 3D U-Net models for all three DVH based scores as well as the quality of the dose distribution predictions. The model produced a total dose MAE of 2.92 Gy (median) on Dataset 1; it produced a beam dose MAE of 0.96 Gy; it produced a DVH score of 1.444 Gy [67]. And importantly, it also allows clinicians to dynamically adjust planning parameters (such as beam angles or treatment mode) and view, in real-time, how their adjustments affect the predicted dose distribution.

4.7 Diffusion Models

Diffusion Models Offer a New Generative Paradigm With Some Benefits. The model called DiffDP developed by Feng et al. appears to be the first diffusion model that was created for predicting radiotherapy doses[86]and that was based on diffusion processes. In developing the DiffDP model, the authors of the paper identified a key limitation of traditional DL approaches when used for training dose predictions. Specifically, DL models that were trained using L1 or L2 loss functions will generally produce smoothed dose distributions with limited detail in terms of higher frequency components, including the ray-like patterns, that result from the beam penetration and attenuation characteristics of radiation therapy. These details are important in clinical practice because they represent directional and modulating effects of the radiation beams and how the dose delivered to the PTV compares to that delivered to the OARs.

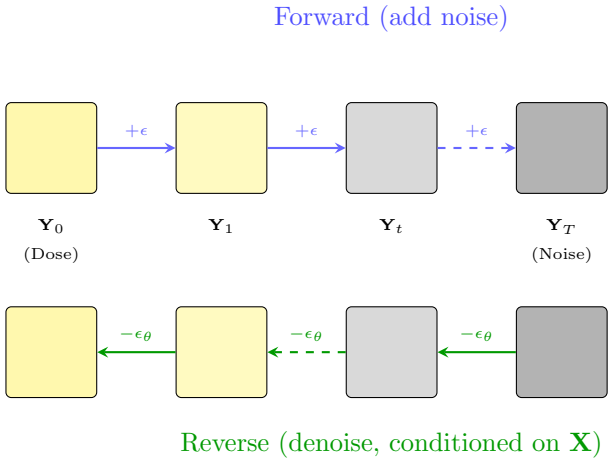


Figure 4.17: Diffusion model for dose prediction: the forward process adds noise to clinical dose; the reverse process learns to denoise conditioned on patient anatomy $\mathbf{X}$.

DiffDP is built on the framework of denoising diffusion probabilistic models (DDPM) [100]. During the forward process, a Markov chain progressively corrupts the ground-truth dose distribution by adding Gaussian noise according to:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{4.19}$$

where β_t is the noise schedule, ultimately yielding white noise. A neural network is trained to predict the noise added at each step, conditioned on anatomical information extracted by a separate structure encoder, which processes CT images and segmentation masks for PTV and OARs. During the reverse process, starting from pure noise $x_T \sim \mathcal{N}(0, I)$, the model iteratively denoises the sample:

$$p_\theta(x_{t-1}|x_t, c) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, c), \sigma_t^2 I) \tag{4.20}$$

guided by the learned noise predictor and the structural features, until a dose distribution is produced [86].

The structure encoder is designed to encode multi-scale anatomical context from CT and structure masks, and its features are fused with those of the noisy dose via residual and cross-attention mechanisms in the noise predictor network. The objective function for training is:

$$\mathcal{L} = \mathbb{E}_{x_0,\epsilon,t}[\|\epsilon - \epsilon_\theta(x_t, t, c)\|^2] \tag{4.21}$$

where ϵ is the true noise and ϵ_θ is the predicted noise.

In an in-house database of 130 patients with rectal cancer who were treated with VMAT (98 training, 10 validation, 22 testing), DiffDP has been compared to many good baselines, including U-net, DeepLabV3, a GAN based model and a cascaded 3D U-net (C3D) [86]. Among those, DiffDP performed better than all of them on various dosimetrical metrics. For example, for PTV metrics, DiffDP obtained the D98 = 0.0413 ± 0.0045 and D2 = 0.0008 ± 0.00001 these represent 16% and 78% improvement, respectively, over the second best method (DeepLabV3 with D98 = 0.0448 and D2 = 0.0036). Visually, DiffDP had sharper and more realistic high-frequency structures and lower error maps. Ablation studies have also shown that the contribution of the structural encoder is significant, which further emphasizes the role of anatomical conditioning in the dose prediction using diffusion models[86].

4.8 Physics-Informed Neural Networks

Physics-based methods form an additional category of methods that can be considered as part of AI-based dose prediction, by directly integrating elements of dose calculation into the machine learning paradigm[81, 101]. Sun and colleagues have recently presented an example of a

physics-based method that uses both physics-based data preprocessing and network architecture to predict individual beam-wise dose distributions in nasopharyngeal carcinoma intensity-modulated radiotherapy (IMRT) treatment plans[88]. Instead of predicting the total dose based upon conventional input variables, this method defines a one-to-many mapping where for each beam angle, a set of physics-based features is calculated.

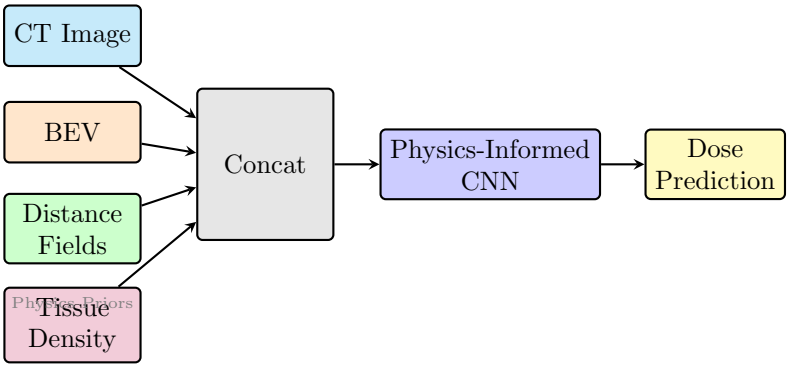


Figure 4.18: Physics-informed architecture incorporating beam-eye-view (BEV), distance fields, and tissue density as explicit physics-based input features alongside CT imaging.

In developing this method, two different types of features were generated for each beam angle. The first type of feature was a set of beam contours whose pixels represented the percent depth dose (PDD) values for the given beam angle; binary mask representations of beam geometry were populated with normalized 6 MV photon PDD values along the direction in which the beam enters the body. Thus, dose fall-off was encoded into the input representation. The second type of feature consisted of ROI projections, i.e., weighted projection matrices representing the PTV and OAR projected onto the beam's eye view. These projection matrices were calculated as follows: $W_i^{PTV} = N_{PTV}/N_{body}$ and $W_i^{OAR} = N_{OAR}/N_{body}$, where N denotes the number of voxels.

The network —a 2D U-Net applied to each beam individually—predicted individual beam doses that could be combined to determine the total dose. Standard mean absolute error (MAE) and gradient-based loss functions were used to evaluate both the beam dose and the total dose predictions. To introduce additional physics-based regularization, a "masked" PTV loss function was also implemented so that the sum of all predicted beam doses within the PTV would meet the requirements of clinical practice[88].

Physically meaningful results were obtained through the application of Spearman correlation analysis to assess the relationship between each of the input channels and the corresponding predicted dose value along and perpendicular to the direction of the incident beam. Correlation coefficients ranged from 0.69 to 0.79 when correlating predicted beam dose values with the corresponding PDD features in the longitudinal direction, and 0.71-0.82 when correlating predicted beam dose values with the corresponding PTV/OAR projections in the transverse direction, thus demonstrating that physically meaningful dose deposition patterns were being accurately captured by the model[88]. A quantitative assessment of the performance of PPIM-BDMG —the authors' best-performing model— showed that it had an average total dose MAE of 2.98 $\pm$ 0.41 Gy, an average beam-wise MAE of 0.96 $\pm$ 0.43 Gy, and a DVH score of 1.44 Gy over the 340 NPC patients included in the OpenKBP-style evaluation. Importantly, the authors' method provides greater interpretability than previously available dose prediction methods and has the potential to enable the use of predicted beam doses to perform treatment planning and quality assurance at the beam level.

4.9 Two-Stage Physics-Guided Planning

Achlatis and colleagues developed a complementary physics-guided framework focusing on the direct prediction of VMAT treatment plan parameters [83]. In their two-stage pipeline, a "Deep RT Planner" based on either a 3D U-Net or a UNETR transformer variant is trained to predict, from CT and structure masks, the multileaf collimator (MLC) apertures and monitor units (MU) for all control points in a dual-arc VMAT plan.

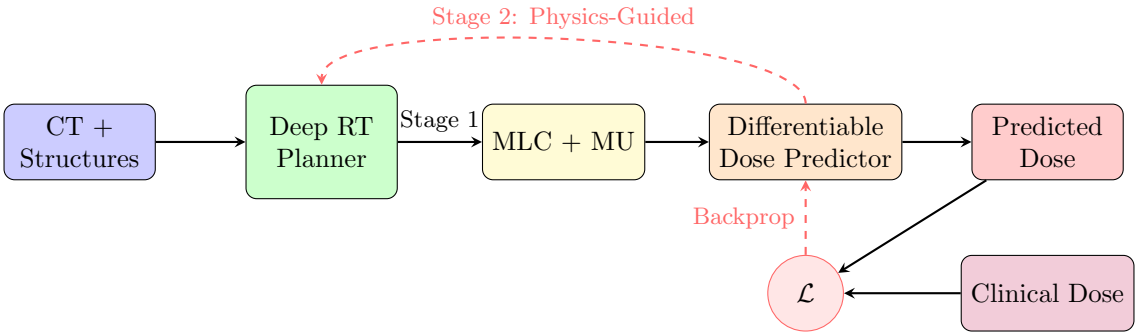


Figure 4.19: Two-stage physics-guided planning pipeline. Stage 1 trains the Deep RT Planner to predict MLC apertures and monitor units. Stage 2 refines the planner by backpropagating dose errors through a differentiable dose predictor.

In the first stage, the network is trained with purely data-driven supervision on these plan parameters using a loss function combining binary cross-entropy for MLC apertures and L_1 loss for monitor units:

$$\mathcal{L}_1 = \mathcal{L}_{BCE}(M_{pred}, M_{true}) + \lambda_1 \|MU_{pred} - MU_{true}\|_1 \tag{4.22}$$

Recognizing that multiple different MLC/MU configurations can generate clinically equivalent dose distributions, the authors then introduce a second training stage in which the plan predicted by the Deep RT Planner is passed through a differentiable RT Dose Predictor—a trained convolutional gated recurrent unit (GRU) network that approximates a Monte Carlo dose engine [102]. The dose predicted from the learned plan is compared to the clinical ground-truth dose using an L_2 loss, which is added to the original parameter-based loss:

$$\mathcal{L}_2 = \mathcal{L}_{BCE}(M_{pred}, M_{true}) + \lambda_1 \|MU_{pred} - MU_{true}\|_1 + \lambda_2 \|D_{pred} - D_{true}\|_2 \tag{4.23}$$

where D_{pred} is computed by the frozen, differentiable dose predictor from the predicted MLC and MU parameters [83].

The gradients are propagated throughout the differentiable dose engine to further refine the Deep RT Planner and as such guide the planner toward plans that both closely resemble the reference MLC/MU patterns as well as generate accurate dose distributions. The physics guided learning method is able to address the fundamental non-uniqueness issue in supervised plan parameter learning by utilizing clinically relevant dose information as an anchor for predictions.

In the case of a dataset of 133 prostate cancer patients who were all treated with a standardized 2-arc VMAT protocol (104 for training, 16 for validation, 13 for testing), which delivered a total dose of 62 Gy to the PTV, the physics-guided, two-stage training process resulted in substantial improvements in both target coverage metrics and OAR sparing compared to the first-stage only models [83]. After physics-guided training the 3D U-Net based Deep RT Planner produced mean differences in the PTV of $D_{95} = 0.42 \pm 1.83$ Gy, $D_{98} = -0.71 \pm 2.12$ Gy, and $V_{95} = -0.22 \pm 1.87\%$ after physics-guided training. For OARs, the rectum D_{mean} was reduced by 0.82 ± 4.45 Gy when comparing to the ground truth, showing that the OAR sparing has been improved. The gamma pass rate at 3%/3 mm criteria at the 90% dose threshold increased significantly from 56.46% (first stage) to 90.50% (second stage) for the 3D U-Net model[83].

In addition to the significant improvement in gamma pass rates, the time required for each inference was on the order of one second per patient when using GPU acceleration and CUDA-based preprocessing, making this approach suitable for use in online adaptive workflows. These results illustrate how the integration of physics-based constraints, differentiable dose engines and deep learning can be used to resolve ambiguity in plan parameters and to place learned models into clinically relevant dose space.

4.10 Uncertainty Quantification and Evidential Learning

A crucial element for clinical use of machine learning models in radiotherapy is the ability to quantify uncertainty. Due to the safety-implications of treatment plans, clinicians require not only an estimate of dose from a predictive model, but also a measure of confidence that predictive model has in its estimate. As traditional deep learning models are typically used to generate point estimates with no reasonable methods to provide uncertainty estimates, these models are largely unusable in clinical practice.

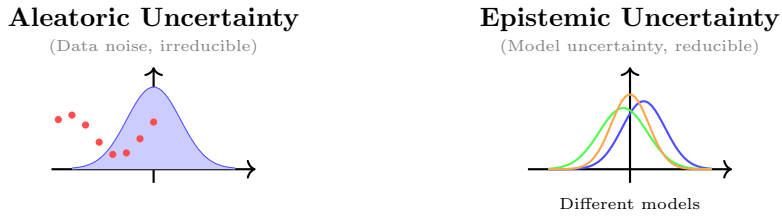


Figure 4.20: Two types of uncertainty: aleatoric (inherent data variability) and epistemic (model uncertainty due to limited training data or model capacity).

To address this problem, an and collaborators introduced a deep evidential learning framework for dose prediction on the OpenKBP dataset[82]. and employed an evidential method based on subjective logic and Dempster-Shafer theory[103], The evidential method provides the network with the parameters of a higher-order probability distribution over the possible doses the predictive model could output. The authors then defined their objective function so that both epistemic uncertainty (model uncertainty resulting from either insufficient training data, or inputting data outside the range of their training data), and aleatoric uncertainty (noise or ambiguity within the data, i.e., due to variations among planners, or random errors in measurements), could be separated.

Within their evidence-based framework, the authors defined epistemic uncertainty as the variance associated with the model's predictions when the model is trained on limited data, and/or the variance associated with model predictions when the model is given input data that lie outside the boundaries of the data used during training. They defined aleatoric uncertainty as the variance associated with the inherent noise or ambiguity present in the data; for example, inter-planner variability, or the presence of random errors in measurements. For a regression task, the evidence-based network parameterized a normal-inverse-gamma (NIG) distribution:

$$NIG(\mu, \alpha, \beta, \lambda) = \mathcal{N}(\mu|\gamma, \sigma^2) \cdot \Gamma^{-1}(\sigma^2|\alpha, \beta) \tag{4.24}$$

where μ is the predicted mean, and the uncertainty is quantified through α (shape) and β (scale) parameters. The epistemic uncertainty is given by $\beta/(\alpha - 1)$ for $\alpha > 1$, and aleatoric uncertainty by $\beta/(\lambda(\alpha - 1))$ [82].

Their experiments show that there is a very strong relationship between epistemic uncertainty and prediction error for both voxel-level (Spearman's Rho Correlation = 0.83) and DVH-level validation, showing that it can serve as a proxy measure for the reliability of a prediction [82]. n contrast to epistemic uncertainty, aleatoric uncertainty is much more sensitive to Gaussian noise perturbation of inputs than to random fluctuations in the inputs, and is therefore

consistent with its interpretation as representing data noise. The authors also show how to establish a linear relationship between median voxel-wise error and uncertainty thresholds, enabling the establishment of calibrated error/uncertainty relationships. As an added benefit, the authors also show how to develop confidence intervals for DVHs that are patient-specific and structure-specific, in terms of the expected variation in dose, that can be overlaid onto the point predictions to give a clinician a sense of the reliability of a particular prediction.

The evidential model performed similarly to the other models (MAE=3.09 Gy) but had the advantage of providing robust statistical uncertainty estimation when applied to the OpenKBP test set[82]. From a clinical perspective, these uncertainty estimates could be used to prompt a clinician to review the treatment plan again, to create uncertainty-aware QA plans, or to choose among several planning options based upon their predicted reliability. While the evidence learning framework can be combined with many of the architectures mentioned earlier and potentially be used to integrate into U-net, GAN, or diffusion-based models, this has yet to be explored.

4.11 Comparative Performance and Clinical Impact

It is beneficial when comparing the performance of machine-learning based dose prediction versus accelerated dose calculation to assess both raw error measures and their respective clinical implications and computational efficiencies. A summary of the performance measures of the key approaches presented in this review is provided in Table 4.3.

Table 4.3: Comparative performance of dose prediction methods across different studies and datasets. DVH scores and dose MAE values are reported where available.

Study	**Dataset**	**Architecture**	**Dose MAE (Gy)**	**DVH Score (Gy)**	**Time**
McIntosh 2017 [84]	633 train, 113 test (multi-site)	Atlas+CRF	1.91	-	Minutes
OpenKBP Winner [85]	340 HN (OpenKBP)	C3D U-Net	2.429	1.478	Seconds
Hien 2025 [87]	340 HN (OpenKBP)	3D U-Net + DVH loss	-	1.444	¡5s
Gao 2023 [67]	365 lung	Flexible-Cm GAN	2.92	1.444	Seconds
Feng 2024 [86]	130 rectum	DiffDP (diffusion)	-	Best DVH	Seconds
Sun 2025 [88]	340 NPC	Physics-informed CNN	2.98	1.44	Seconds
Achlatis 2025 [83]	133 prostate	UNETR + physics	-	D_{95}: 0.42±1.83	¡1s
Li 2023 [89]	340 NPC	Shared encoder	¡1 (DVH)	-	Seconds
Tan 2024 [82]	OpenKBP	Evidential learning	3.09	-	Seconds
Gao 2024 [90]	Lung (MARL)	Multi-agent RL	3.09	-	¡20s

The best reported DVH values on the OpenKBP dataset are converging towards 1.4 to 1.5

Gy, while voxel-wise mean absolute errors on this dataset are generally found to range between 2.4 and 3.1 Gy. These include the 3D U-Net with DVH-based loss proposed by Hien et al. [87] and the flexible-cm GAN[67] On the other hand, typical total dose MAE's on site-specific datasets such as rectal cancer, nasopharyngeal carcinoma, lung cancer and prostate cancer are typically found to range between 2.6 and 3.5 Gy depending on the case complexity, beam geometries, chosen metric(s), etc., whereas the majority of the clinically relevant DVH endpoints are found to differ from the original clinical plans by no more than 1-2 Gy or a few percent in volume coverage.

Therefore, these differences are small enough that predicted doses can serve as good initializations for conventional optimization processes, or alternatively, may serve as surrogates for rapid evaluation of the planned treatment.Several studies have demonstrated that automatic plans created through use of deep learning-based predictions and subsequent optimization match or slightly exceed manually optimized treatments with respect o target coverage and OAR sparing.For example, the simultaneous dose and fluence prediction model described by Li et al. demonstrate that using predicted fluence as an initialization for IMRT optimization results in final plans with DVH values equivalent to, or better than those achieved through traditional forward planning [89]. Similarly, the physics-guided models proposed by Achlatis et al. result in deliverable VMAT plans with high gamma pass rates (90.50% at 90% dose threshold) and DVH objectives that meet clinical tolerance[83].

The improvements in computation are also large scale. Traditional planning processes can take tens of minutes to hours per patient while deep learning-based predictor's will take anywhere from a fraction of a second to 5 seconds to run an inference using a modern GPU. This speed advantage is very important for adaptive radiotherapy as daily or weekly updates to CT or MRI images are common and the time available for replanning is small. Rapidly generated dose prediction data can be used to evaluate how good the current plan is, what changes should be made to the plan or serve as a starting point for further optimization with fewer iterations.

However, while many advantages exist with utilizing deep learning-based dose prediction systems, several disadvantages exist. Most of the deep learning-based dose prediction models have been trained on a dataset of patients' past treatments (retrospective) that are reflective of a particular institution's planning practice, beam model, and dose calculation algorithm; therefore, the predictive value of such a model may not be the same across different institutions or treatment machines, and potentially contains the same planning biases of the training institution. Also, due to the fact that there is no unique solution to a dose distribution problem, there is ambiguity in the supervised learning target when it comes to determining plan parameters (e.g., MLC positions, MU). Several techniques, including physics-guided and uncertainty aware models, are being developed to address some of these limitations, however, prior to widespread adoption of these models, additional prospective validations and careful clinical integration are required.

4.12 Challenges, Limitations, and Future Directions

4.12.1 Data Scarcity and Heterogeneity

Several challenges must be addressed to fully realize the potential of machine learning for accelerated dose calculation and dose prediction. A first challenge is dataset size and diversity. Many current studies use datasets comprising a few hundred patients from one or a small number of institutions. While this is sufficient to demonstrate feasibility and to train reasonably robust U-Net-like models, it may not be adequate for transformer-based architectures or for models designed to generalize across institutions, beam models, and treatment modalities [96].

The OpenKBP challenge represented a significant step forward by providing a standardized dataset and evaluation framework [85], but the dataset is limited to head-and-neck cancer treated with nine-beam IMRT. Collaborative efforts to extend such benchmarks to more anatom-

ical sites (lung, prostate, breast, gastrointestinal), treatment modalities (VMAT, SBRT, proton therapy), and institutions will be essential to build the foundation for truly generalizable models. Data augmentation strategies such as random flipping, translation, and rotation are commonly applied but have limited effectiveness for 3D medical imaging, as they can introduce unrealistic anatomical configurations or violate physical constraints [87].

4.12.2 Non-Uniqueness of Optimal Solutions

A second challenge concerns the representation of planning objectives and the non-uniqueness of optimal solutions. It is common practice to train models to reproduce the dose distributions of historical plans, but these plans may not be unique or even optimal in hindsight. Multiple different plans can achieve similar DVH metrics and clinical acceptability, particularly when trading off between PTV coverage and OAR sparing [67, 83]. This ambiguity creates a fundamental challenge for supervised learning, as the model may be penalized for predicting clinically equivalent but numerically different solutions.

Including more specific clinical goals (e.g., dose-volume constraints; toxicity modeling; etc.) in the model development will provide direction toward more universally acceptable models. To this end, physics informed loss functions, such as DVH-based [87], hift-Dose-Volume [67], or Moment-based [104] are being explored; however, additional research is required to include NTCP & TCP models or to allow direct optimization of multiple objective tradeoffs. Training in dose space (physics guided training), rather than parameter space offers an alternative approach [83],However, they depend on having differentiable dose calculations, which are also approximate representations of actual physical dose.

4.13 Interpretability and Clinical Trust

A fourth issue is that clinicians require trust and understanding about the predictions made by a deep-learning system. While physics-based models and evidential learning can give insight on how the models function and how well they do [88, 82], most DL architectures are "black box" regarding the behavior of traditional planning systems which have behavior that can be broken down into transparent optimization and dose calculations. Clinicians need to know why a model made a certain prediction, especially if that prediction does not follow the usual pattern of expected results or previous history.

Understanding and developing ways to make clinicians aware of what variables in the input contribute to a prediction for dose; identifying out-of-distribution examples; and providing a simple visual representation of uncertainty will be key to obtaining clinician acceptance of a deep-learning system for radiation therapy treatment planning. Tools such as, beam-wise feature analysis (demonstrated by Sun et al.) [88]; saliency mapping; gradient-based attribution methods; and local surrogate modeling may help in this regard. The Spearman correlation analysis between input features and output doses demonstrated by Sun and colleagues[88], demonstrates a significant step in the direction of creating interpretable physics-based models and demonstrate that the learned representations match the principles of physics[88].

Evidential learning [82] or ensemble based methods, to create confidence bounds, will enable clinicians to evaluate the reliability of predictions and determine when a human review should be performed. The confidence intervals for DVH's, developed by Tan and colleagues, represent a clinically relevant method of presenting the uncertainty in dose-volume metrics[82].

4.14 Deliverability and Physical Constraints

There is a 4th challenge in the deliverability (physical realizability) of delivered treatment plans using predicted fluence maps/dose distributions. These predicted fluence maps/dose distributions may violate physical constraints (e.g., Maximum Leaf Position Limitations, Leaf Travel

Speed Limits, Monitor Unit Rate Limits, Jaw Position Constraints). Li et al. used an existing planning system (Eclipse Treatment Planning System), to import predicted fluence and use it for leaf motion calculations and dose verification[89]. Achlatis et al. also explicitly predicted MLC aperture and monitor units (MU) so they were deliverable based on their formulation [83].

Post processing techniques, e.g., Ridge Regression for MU refinement[88] or constrained optimization to project predictions onto the space of deliverable plans, can be applied to improve the deliverability of predicted plans. Alternatively, incorporation of deliverability constraints directly into the learning objectives (e.g., via a differentiable MLC sequencer or a physics-based regularization) would allow predictions to be guaranteed to be physically realizable. The multi-agent reinforcement learning approach developed by Gao et al[90] illustrates one possible method of developing learning-based leaf sequencing methods that respect MLC mechanical constraints.

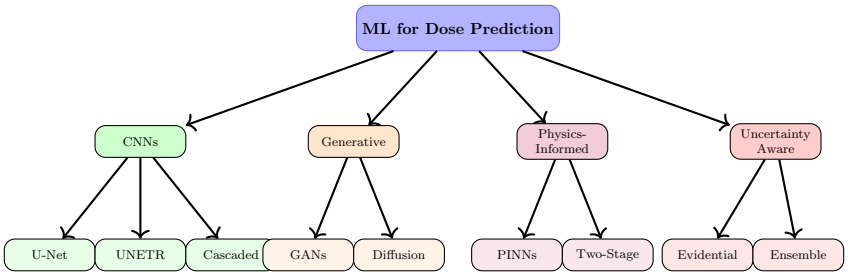


Figure 4.21: Taxonomy of machine learning methods for dose prediction in radiotherapy. Major categories include convolutional neural networks (U-Net, transformers, cascaded architectures), generative models (GANs, diffusion), physics-informed approaches (PINNs, two-stage physics-guided), and uncertainty-aware frameworks (evidential learning, ensembles).

4.15 Summary

Machine learning is advancing at an extraordinary rate, from being an experimental methodology to a solid methodology able to generate dose calculations which are both clinically plausible and of high quality in radiotherapy. Techniques include convolutional networks (such as 3D U-Net variants[87, 94], generative models (Flexible-Cm GAN [67] and DiffDP [86], physics-guided methodologies that incorporate differentiable dose engines[83], physics-informed methodologies [88] and methodologies based on evidential learning[82] which take into account uncertainties. Collectively these methodologies form a highly dynamic and complex ecosystem of methodologies. Each of these methodologies achieves dose and DVH errors of the order of 1.4–3.1 Gy when compared to clinical plans, perform inference in seconds or less and offer exciting opportunities for accelerating dose calculation, supporting automatic and adaptive planning, and reducing variability in the quality of treatment across institutions.

The OpenKBP grand challenge [85] has provided a clear benchmark for the community of researchers and practitioners, demonstrating how standardizing data and methods can lead to rapid progress and enable fair comparisons of methodologies. The top performing methodologies evaluated against this benchmark have converged to a common performance metric of approximately 1.4–1.5 Gy in terms of their DVH scores. In many instances, this represents dose distributions which are clinically indistinguishable from those generated by manual planning.

In addition to the development of methodologies which converge towards common goals, several key advances in methodology have been identified across complementary areas. Specifically, the use of DVH-based and physics-informed loss functions enables optimization of clinically meaningful objectives rather than arbitrary voxel-wise errors[87, 67, 88]. Furthermore, the use of generative models including GANs and diffusion architectures[67, 86]. have resulted in the generation of dose distributions that are sharper and more realistic with preserved high-

frequency features. The use of physics-guided training frameworks incorporating differentiable dose engines[83], enables predictions to be anchored in dose space and resolves ambiguity in plan parameters. Finally, methodologies for uncertainty quantification[82]. provide confidence estimates which enable risk-aware clinical decision making. Finally, multi-task learning methodologies which simultaneously predict dose and fluence[89]bridge the gap between prediction and deliverable planning.

Despite the significant advancements in methodologies described above, there are also significant challenges that remain with regard to generalization, interpretability, deliverability and clinical validation. The majority of current models are trained using small, single-institution datasets that do not capture the breadth of clinical practice. As previously discussed, the non-uniqueness of optimal dose distributions presents a fundamental ambiguity in supervised learning, and as such the majority of models remain "black boxes" that are difficult to interpret and validate from first principles.

The field has reached an important point at which the emphasis should be placed on making machine learning systems as robust and reliable as possible (in order to create widespread acceptance) for use in day-to-day clinical applications, rather than merely establishing technical feasibility of the systems. To achieve this goal, larger, more diverse, multi-institutional datasets will need to be developed to include data on many different areas of the body, including different treatment modalities. In addition, standardized evaluation procedures will have to be developed that extend beyond OpenKBP to additional disease sites and techniques.

Prospective clinical trials assessing efficiency, plan quality, and patient outcomes will need to accompany the development of regulatory frameworks for approval and deployment of AI-based planning systems. Additionally, the continued development of models that can provide both interpretability and measures of uncertainty, while also being informed by physical laws will be required to establish both clinical trust and safety in the application of machine learning based systems. The convergence of deep learning with domain knowledge from radiation oncology, medical physics, and radiobiology to create treatment planning systems that are faster, more consistent, and potentially superior to current manual approaches is within reach, and holds significant potential to improve outcomes for cancer patients world-wide. A detailed description of our proposed physics-informed architecture that combines 3D U-Net and Fourier Neural Operators is presented in sectionChapter 11.

Chapter 5 Lung Cancer Radiotherapy: Characteristics and Challenges

Lung cancer is still the most common reason for cancer-related deaths worldwide, and radiotherapy plays a significant role in the treatment of both early-stage and locally advanced non-small cell lung cancer (NSCLC)[105]. Radiotherapy applied to thoracic tumors has several unique technical and dosimetrical challenges due to extreme variability of tissue density throughout the thorax, respiratory motion, and close proximity of critical radiosensitive structures such as the esophagus, heart, and spinal cord[106, 107].

5.1 Clinical Applications and Treatment Modalities

For medically inoperable stage I NSCLC, Stereotactic Body Radiotherapy (SBRT) - also referred to as Stereotactic Ablative Radiotherapy (SABR) - has become the gold standard treatment method, providing local control rates of over 90%, and 5-year overall survival rates of 43-83% [106, 108]. SBRT utilizes ablative radiation doses (usually 48-60 Gy) delivered in 5 or fewer fractions over a period of 1-2 weeks, which contrasts with conventional fractionation schemes requiring 6-7 weeks[106, 109]. Success of SBRT is contingent upon accurate delivery of high doses to the tumor target area while limiting the impact to normal tissues surrounding it - a requirement that is heavily reliant on accurate dose calculation in heterogeneous media[106].

Peripheral lung tumors typically have SBRT prescriptions based on single-fraction regimens of 25-34 Gy, three-fraction schemes of 18-20 Gy per fraction (for a total of 54-60 Gy), or five-fraction protocols of 10-12 Gy per fraction (for a total of 50-60 Gy)[107, 108, 110]. Randomized Phase II trial data indicate that there was no difference in terms of toxicity, local control, progression free survival, and overall survival between 30 Gy delivered in one fraction and 60 Gy in three fractions[110]. The biological effective dose with an α/β ratio of 10 Gy (BED_{10}) should be above 100 Gy to ensure a tumor control probability of 90%[107, 108].

Central and ultra-central tumors (those located less than 2 cm from the proximal bronchial tree or those adjacent to mediastinal structures) require reduced fractionation schemes due to increased risk of fatal toxicity including hemoptysis and fistula development[111, 107, 108]. For central and ultra-central tumors, fractionated schemes of 4-8 fractions (e.g., 8×7.5 Gy $= 60$ Gy; 5×10–12 Gy $= 50$–60 Gy) are recommended to minimize the maximum dose to OARs and to respect the low tolerance of late responding tissues to high fractional doses[111, 107].

Intensity modulated radiotherapy (IMRT) and Volumetric modulated Arc Therapy (VMAT) are commonly used in the delivery of both SBRT and conventionally fractionated lung radiotherapy[105, 112, 113, 114]. Both IMRT and VMAT employ MLC modulation to create conformational shaped radiation beams around the tumor with steep dose gradients that allow sparing of adjacent OARs [109, 112]. Planning for IMRT and VMAT includes IGRT employing onboard imaging and 4DCT to accommodate respiratory induced tumor motion[106, 108].

5.2 Unique Dosimetric Challenges in Lung Radiotherapy

Extreme heterogeneity exists among tissues within the thoracic region; this is due to the presence of structures that are physically different by three orders of magnitude in terms of density: air-filled lung parenchyma (0.035–0.3 g/cm^3), soft tissue (0.9–1.1 g/cm^3), and cortical bone (1.4–1.8 g/cm^3)[115, 116]. These variations affect photon scattering, and secondary electron equilibrium. As photons travel through less dense lung tissue, they will experience increased scatter in the forward direction and decreased scatter laterally compared to soft tissue. Secondary electrons generated by Compton interaction will travel farther than those generated in soft tissue. Therefore, charged particle equilibrium (CPE) will be lost at tissue/lung interfaces[117, 118, 119].

The clinical consequences of these physical properties are: (i) dose will accumulate and decrease at tissue/air interfaces; (ii) dose will vary laterally at the field edge; (iii) the dose delivered by the beam will increase in the forward direction[115, 120, 118]. Failure to model these physical properties results in errors in dose calculation. Errors in dose calculation can result in underdosing of the target or overdosing of critical organs-at-risk (OAR), both of which can negatively impact treatment outcome.

In addition to the physical properties described above, respiratory motion also impacts the dosimetry of lung tumors. Lung tumors have been shown to move up to 30mm during a single respiratory cycle, and lower lobe tumors tend to move more than upper lobe tumors[106]. Several methods have been proposed to manage respiratory motion including respiratory gated radiation therapy (the beam is turned on when the patient's breathing is within a specified phase); active breathing control (the patient holds his/her breath for a set period while the radiation is being delivered); and internal target volume (ITV) concepts that account for the movement of the tumor throughout the respiratory cycle [106, 108]. Each method has unique dosimetric implications since the dose accumulated will occur over varying geometries.

5.3 Comparative Evaluation of Dose Calculation Algorithms

The accuracy of dose calculation in lung radiotherapy is critically dependent on the algorithm employed. Commercial treatment planning systems (TPS) implement algorithms spanning a spectrum of sophistication, from simple correction-based methods to advanced radiation transport solvers. The following analysis compares four principal classes of algorithms: Pencil Beam (PB), Anisotropic Analytical Algorithm (AAA), Acuros XB (AXB), and Monte Carlo (MC) simulation.

5.3.1 Pencil Beam and Correction-Based Algorithms

Pencil beam convolution algorithms, including variants employing density-scaling corrections (e.g., modified Batho, equivalent path length), represent the simplest class of heterogeneity correction[118, 119]. These methods apply 1D density corrections along beam rays, neglecting lateral photon scatter and secondary electron transport perturbations[117, 118]. In homogeneous media, PB algorithms achieve acceptable accuracy; however, their performance degrades dramatically in the presence of low-density materials.

5.3.2 Accuracy Limitations in Lung

A number of studies have shown that PB algorithms consistently overestimate dose in lung tissues. For photons, the degree of error increases as the energy of the photon beam increases; this is shown by an average 3% for 4 MV, 8% for 6 MV, and 14% for 18 MV X-rays[118]. The reason for this trend is due to increasing electron disequilibrium and lateral scatter range as photon energy increases. A comparison of PB algorithms with Monte Carlo in a mediastinum geometry phantom study resulted in discrepancies of 10-15% within lung volumes for megavoltage beams[118].

Proton therapy has limitations compared to the others. A multi-institutional phantom study was done to compare commercial pencil beam algorithms used for proton lung treatment to Monte Carlo; they were found to be "catastrophically" inaccurate: dose overestimated by 31% within the internal gross tumor volume (iGTV) and 46% within the planning target volume (PTV)[117]. Monte Carlo recalculation showed much better agreement with measured values than did the commercial pencil beam algorithms. They felt that "the pencil beam algorithms for proton dose calculation in the lung are unacceptable"[117] and recommend that clinically, MC-based algorithms should be mandatory for all thoracic malignancies.

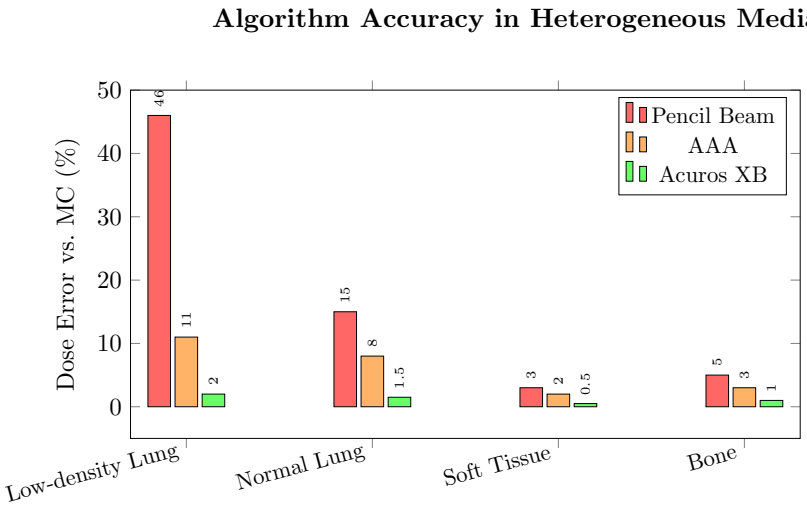


Figure 5.22: Dose calculation errors relative to Monte Carlo reference for different algorithms in various tissue types. Pencil Beam algorithms show catastrophic failures in low-density lung tissue (up to 46% overestimation), while Acuros XB maintains <2% agreement with Monte Carlo across all tissue types. Data synthesized from multiple validation studies[115, 117, 116, 118].

5.4 Anisotropic Analytical Algorithm (AAA)

The AAA is a development of PB algorithms, as it utilizes anisotropic photon scatter kernels developed using pre-calculated Monte Carlo (MC) simulations[121, 122]. to model both primary photon interactions, as well as convoluted energy deposition, based on the convolution of the TERMA with photon and electron scatter kernels that take into account changes in tissue density[123, 122].

5.4.1 Performance in Lung SBRT

The accuracy of AAA is also well documented as being poor in low-density areas. A recent comparison of AAA with Monte Carlo (MC) simulations (PRIMO Code), for lung SBRT plans, reported that AAA consistently overestimated dose at the periphery of the tumour and produced inaccurate Percent Depth Dose (PDD) profiles and lateral dose distributions[120]. Gamma analysis (3%/3 mm criteria) yielded pass rates of only 86% for AAA versus 100% for Acuros XB when benchmarked against MC[115].

IIn clinical VMAT lung SBRT plans, the mean PTV doses calculated by AAA were higher than those calculated by Acuros XB by 1.33–3.34%, depending on the beam energy (6 MV versus 10 MV Flattening Filter Free [FFF] beams) and the calculation grid size[124].The difference in the dose calculation results was even greater in very low density lung (0.035 g/cm^3): gamma agreement for AAA decreased to 11% compared to 86% for Acuros XB[115]. The overestimation of the dose in this area can create an incorrect perception of adequate target coverage, which can lead to geographic tumour miss and compromised local control.

Additionally, AAA has been shown to be unable to accurately model thin inhomogeneity layers and interfaces between materials of vastly different densities (i.e. bone-air or lung-soft tissue boundaries)[115, 120]. Furthermore, the rebuild-up curve following low-density inserts displayed spatial discrepancies and initiated at the material interface, rather than at a location 2-3 mm within the lung as predicted by MC[115].

5.5 Acuros XB (AXB): Grid-Based Boltzmann Transport Solver

Acuros XB represents a significant departure from current clinical dose calculation techniques and uses a deterministic solution to the Linear Boltzmann Transport Equation (LBTE) to explicitly model photon and electron transport through heterogeneous media[115, 120, 116]. Unlike convolution-based algorithms that use energy deposition kernels to approximate radiation transport, AXB directly solves the coupled transport equations on a spatial grid using discrete ordinates methods; this approach is similar to Monte Carlo simulation but does not have stochastic noise[115, 125]. The LBTE formulation and discretization methods are discussed in detail in Part III of this thesis.

5.5.1 Validation Against Monte Carlo and Measurements

The accuracy of AXB has been validated through multiple studies that independently compare AXB to full MC simulation and experimental measurements in lung phantoms. In a heterogenous lung phantom made up of normal lung (0.198 g/cm^3), light lung (0.035 g/cm^3), and bone (1.798 g/cm^3) inserts, AXB passed gamma pass rates of 100%, 86%, and 100%, respectively when compared to the MC, using 3%/3 mm criteria with MC VMC++ as reference[115]. while AAA only had a gamma pass rate of 86%, 11%, and 100% respectively for the same configurations[115].

A study validating SBRT treatment planning for lungs utilizing PRIMO as the MC reference, showed AXB to be superior than Gafchromic EBT film measurements in terms of PDD curve matching, lateral profile matching, and 2D gamma analysis[120]. AXB demonstrated better agreement with MC for DVH metrics in both target volume and OARs[120, 116] for clinical VMAT treatment plans. The mean dose difference between AXB and MC in ten lung SBRT patients ranged from 0.6-1.7%, while the dose difference between MC and AAA ranged from 4.1%[116].

AXB also accurately simulated the rebuild-up effect seen after inserting low density lung, simulating the build-up of charged particles in the air gap before reaching the point where charged particles are in equilibrium with the surrounding tissue [115]. Additionally, AXB correctly simulated dose perturbations at material interfaces, which included small dose peaks (2-3mm) proximal and distal to high-density bone inserts due to secondary electron backscatter, which were not calculated by AAA[115].

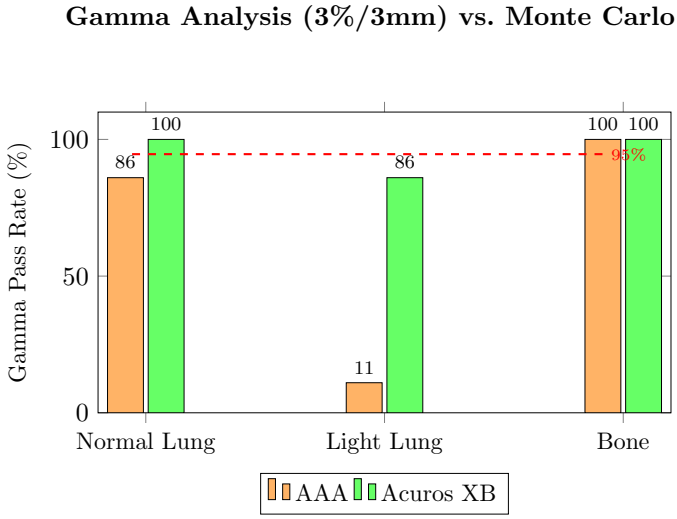


Figure 5.23: Gamma pass rates for AAA and Acuros XB compared to Monte Carlo reference (VMC++) using 3%/3mm criteria. In very low-density lung (0.035 g/cm^3), AAA achieves only 11% pass rate while Acuros XB maintains 86%. Data from Bush et al.[115]

5.5.2 Computational Efficiency

The AXB algorithm is physically correct, yet it can be calculated at reasonable speeds using the GPUs and optimization of an iterative solver. In comparison, reported computational times for lung SBRT plans vary among 42 – 94 s (mean = 66 s) for AXB [116] as opposed to 5 – 9 s for AAA and 124 – 358 s for a full MC (XVMC)[116]. As such, this trade-off between accuracy and computational speed has enabled AXB to become commonly used in the clinic for routine treatment planning.

5.6 Monte Carlo Simulation: The Gold Standard

Monte Carlo (MC) simulation is the gold standard for radiation dose calculation because it provides the most accurate representation of radiation transport physics through stochastic sampling of individual particle histories[115, 120, 117]. Modern MC algorithms (e.g., EGSnrc/VMC++, GEANT4, MCNP, PRIMO) are capable of simulating many aspects of radiation interactions including primary photon interaction (photoelectric effect, Compton scattering, pair production), secondary electron and positron transport, bremsstrahlung, and atomic relaxation processes [115, 120].

n the case of lung dose calculations, MC simulations have shown consistent agreement with experimental measurements for clinically relevant phantoms and geometries with differences less than 2 – 3 %[120, 117, 126].However, MC simulations inherently have statistical noise that results in significantly longer computation times for clinically acceptable uncertainties (less than 2% per voxel). Therefore, even when using variance reduction techniques, MC simulations require very large numbers of particle histories (10^8–10^{10}) and therefore result in long computation times of typically 2 – 6 hours per plan[116, 127].

Recent improvements in MC simulation using GPU acceleration (e.g., RayStation MC) have resulted in significant reductions in computational time to approximately 10 – 20 min while maintaining statistical uncertainties of less than 2 %[126, 127]. A recent study comparing the performance of RayStation MC, Acuros XB, and AAA for lung cancer patients with high tissue heterogeneity showed that both MC and AXB provided similar levels of accuracy as each other but were significantly better than AAA; and further, RayStation MC was superior to AXB in the case of highly heterogeneous anatomical structures[126, 127].

5.7 Summary of Algorithm Performance in Lung

Table 5.4 summarizes the comparative performance of the four algorithm classes for lung radiotherapy dose calculation.

Table 5.4: Comparative performance of dose calculation algorithms for lung cancer radiotherapy.

Algorithm	Physical Basis	Lung Accuracy	Calculation Time	Clinical Use
Pencil Beam	1D density scaling	Poor (10–46% error)	<5 s	**Unacceptable**
AAA	Anisotropic kernels	Moderate (3–8% error)	5–10 s	Limited
Acuros XB	LBTE solver	Excellent (0.6–2% vs. MC)	40–95 s	**Recommended**
Monte Carlo	Stochastic transport	Gold standard (reference)	10–360 min	Research/QA

The overwhelming evidence shows that Pencil Beam algorithms are inappropriate for lung radiotherapy and therefore should never be used clinically[117, 118]. While AAA has been shown to be satisfactory for soft tissue sites, this method consistently underestimates dose in lung by 3–8% and can compromise tumor control[124, 115, 120, 116]. However, Acuros XB provides MC-equivalent accuracy (within 0.6–2%) with clinically practicable computation times and is clearly the preferred algorithm for lung SBRT and IMRT planning[115, 120, 116, 126]. Full MC simulation is indispensible for the purposes of commissioning, validation and research,

however with GPU acceleration, this becomes increasingly feasible as an option for routine clinical use[126, 127].

5.8 The Uniqueness and Significance of Acuros XB Implementation

Acuros XB marks a significant step forward in the field of computational radiation dosimetry, by providing a means of linking the theoretically rigorous nature of Monte Carlo simulation with the practical limitations associated with clinical workflows. It represents a new paradigm in terms of the accuracy of treatment planning particularly for complex anatomical structures such as the thorax.

5.8.1 Scarcity of LBTE-Based Implementations

While there are many theoretical benefits to deterministic radiation transport solver methods, very few have been implemented clinically. To date, Acuros XB (Varian Medical Systems, now Siemens Healthineers) is one of the only commercially available grid-based Boltzmann solvers that is cleared for clinical use [115, 125]. Other advanced algorithms include Collapsed Cone Convolution used in RayStation and Pinnacle TPS, which is more accurate than AAA, however, also uses kernel- based convolution and does not solve the full transport equation[124, 116], Finally, there are Monte Carlo TPS modules such as RayStation MC and Monaco MC that provide equivalent or higher accuracy to AXB, however they take significantly longer to compute and present with statistical noise [126, 127].

Historically, the high degree of computational complexity associated with solving the LBTE—a 7-dimensional integro-differential equation, where three dimensions are associated with position, two dimensions are associated with direction, one dimension is associated with energy, and one dimension is associated with time—has precluded the determination of radiation transport from a deterministic perspective in a clinically relevant timescale[115, 125].Acuros XB overcomes this barrier using discrete ordinates (SN) discretization that divides angular flux into finite directions, multigroup energy treatment that divides the continuous energy spectrum into 13–15 energy groups, iterative Gauss-Seidel solvers that utilize the spatial coherence of dose distributions to achieve rapid convergence and GPU parallelization that allows the massively parallel computation of voxel- wise transport.

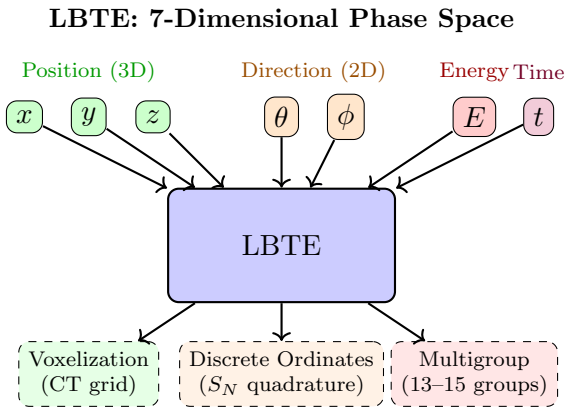


Figure 5.24: The Linear Boltzmann Transport Equation operates in a 7-dimensional phase space: 3 spatial dimensions, 2 angular dimensions, energy, and time. Acuros XB discretizes this space using voxelization, discrete ordinates (S_N), and multigroup energy treatment to achieve clinically feasible computation times.

Chapter 6 Theoretical Foundation: Radiation Transport Physics

6.1 Historical Context and Clinical Importance

External beam radiotherapy (EBRT) uses ionizing radiation to destroy cancer cells while minimizing damage to surrounding healthy tissue. The primary goal of treatment planning is to deliver a prescribed dose to the target volume with high conformity and minimal toxicity to organs at risk (OARs) [13].

Modern radiotherapy relies on megavoltage (MV) photon beams produced by medical linear accelerators (linacs). Typical energies range from 4 MV to 25 MV, with 6 MV being the most common for external beam treatments [14].

6.2 Radiotherapy Treatment Planning Overview

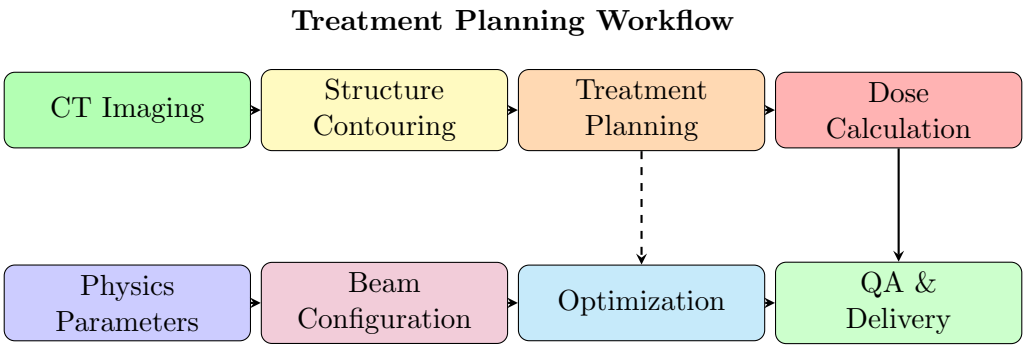


Figure 6.25: Overview of the external beam radiotherapy treatment planning workflow. This work focuses on the dose calculation component using the Linear Boltzmann Transport Equation (LBTE).

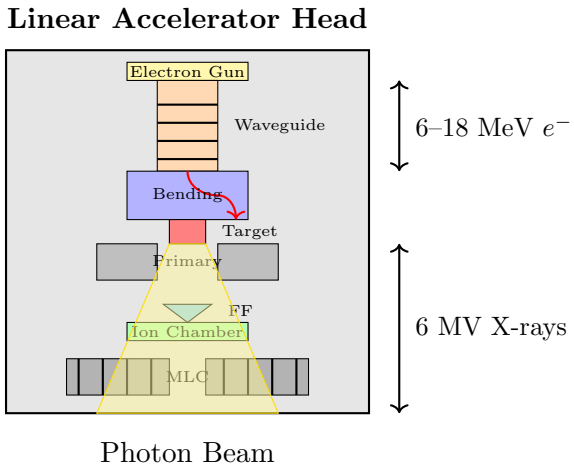


Figure 6.26: Schematic of a medical linear accelerator (linac) head. Electrons are accelerated in the waveguide, bent by magnets, and strike the target to produce bremsstrahlung X-rays. The MLC shapes the beam to conform to the target volume.

6.3 Types of Ionizing Radiation

Ionizing radiation is classified by its ability to remove electrons from atoms, creating ions. In radiotherapy, the relevant types are:

In radiotherapy, the relevant types include photons (X-rays and γ-rays), i.e. electromagnetic radiation with $E > 10$ keV: X-rays are produced primarily via bremsstrahlung in linac targets, while γ-rays originate from nuclear transitions. Electron radiation consists of charged particles with rest mass $m_e = 0.511$ MeV/c^2; primary electrons are delivered directly by linac beams, whereas secondary electrons are generated by photon interactions in tissue and are chiefly responsible for local energy deposition. Positrons are produced in pair-production events when $E_\gamma > 1.022$ MeV and subsequently annihilate with electrons, yielding two 511 keV photons.

6.4 Energy Ranges in Medical Physics

Table 6.5: Energy classifications in medical radiation physics [128].

Classification	**Energy Range**	**Primary Application**
Superficial X-rays	50–150 kVp	Skin lesions
Orthovoltage X-rays	150–500 kVp	Shallow tumors
Megavoltage X-rays	1–25 MV	Deep-seated tumors
Electron beams	4–25 MeV	Superficial/skin boost

6.5 Megavoltage Photon Beam Production in Medical Linear Accelerators

The linac represents the foundation of modern external beam radiotherapy; it generates MV photons from the electron beam produced by a highly complex process involving multiple stages[14, 13]. n order to understand the requirements for the input data used in advanced dose calculation methods like Acuros XB, the basic physics of how the MV photon beam is generated—i.e., electron beam acceleration and absorption of the photon beam in tissues —must be understood.

6.5.1 Electron Acceleration and Waveguide Transport

The first stage in producing a MV photon beam begins with electron beam generation in a thermionic cathode located in the electron gun. The thyratron is triggered simultaneously with the electron gun and the microwave power source (magnetron or klystron) that produces radiofrequency (RF) waves at3 GHz (S-band), to the accelerator waveguide[13].

These RF waves are then transmitted down the accelerator waveguide to accelerate electrons to relativistic speeds (typically 6 to 18 MeV)using electromagnetic standing waves in the waveguide. The accelerating gradient is determined by the strength of the electric field in the RF waves, and thus the kinetic energy gained by the electrons is directly related to the strength of the electric field. The accelerated electron beam exits the waveguide and travels via bending magnets to the treatment head[13].

6.6 Bremsstrahlung Target and Beam Modulation

After passing through the bending magnets, the high energy electron beam strikes a high atomic number target (e.g., tungsten, Z = 74). Upon striking the tungsten target, the electrons rapidly decelerate due to interaction with the tungsten target nuclei and convert their kinetic energy into bremsstrahlung photons via Coulombic interactions[13, 129]. Thus, the resulting photon beam is polyenergetic with an energy range limited by the initial energy of the incident electron beam (maximum energy equal to the incident electron energy) and a mean energy of approximately 30% of the maximum energy.

The forward peaked photon beam is further modified by passage through a flattening filter. The purpose of the flattening filter is to reduce the intensity of the central axis of the photon

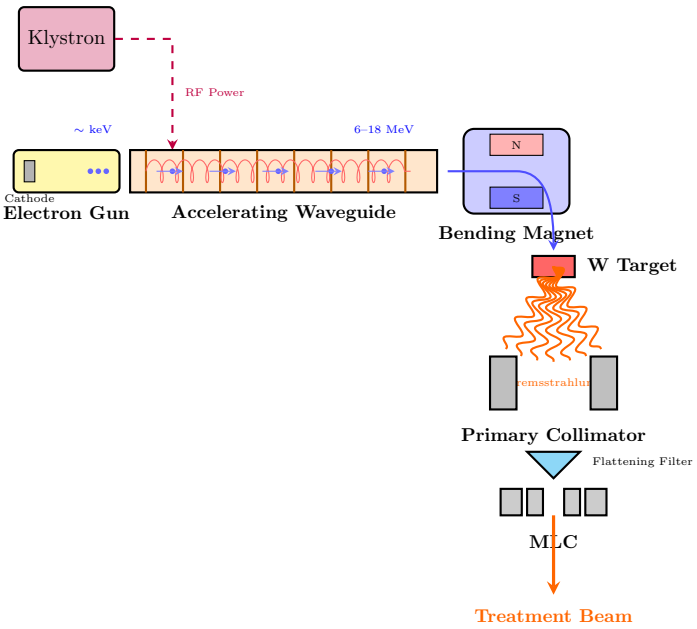


Figure 6.27: Schematic diagram of a medical linear accelerator (linac) showing the complete photon beam production chain. Electrons are generated at the cathode, accelerated through the RF waveguide, bent by the magnet toward the tungsten target where bremsstrahlung occurs. The resulting photon beam is shaped by the primary collimator, flattening filter, and multileaf collimator (MLC) before delivery to the patient.

beam while maintaining lateral homogeneity of the dose distribution. After modification by the flattening filter, the beam is shaped by a combination of a primary collimator and a multileaf collimator (MLC), where the MLC defines the specific shape of the field aperture for patient treatments[13].

6.7 Photon Interactions in Tissue

Upon entering the patient, megavoltage photons interact with tissue primarily through Compton scattering, which dominates at therapeutic energies (1 to 25 MeV). In this process, an incident photon transfers part of its energy to an outer-shell electron, ejecting it as a recoil electron. These secondary electrons are responsible for the actual dose deposition in the medium. The probability of Compton interaction depends primarily on electron density (ρ_e), making it the fundamental interaction modeled by dose calculation algorithms[13, 13].

6.8 Radiobiological Basis of Radiotherapy

The therapeutic index of radiotherapy relies on the differential ability of ionizing radiation to induce lethal damage in tumor cells compared to normal tissues. This biological effect is mediated primarily through DNA damage and the cellular response to that damage.

6.8.1 Mechanisms of DNA Damage Induction

Ionizing radiation disrupts cellular integrity through two distinct mechanisms. In direct action, the radiation track directly ionizes the DNA molecule, causing sugar-phosphate backbone rupture; this is the dominant mechanism for high-LET radiation but accounts for only approximately 30% of damage from standard photon therapy. In indirect action, the radiation interacts with intracellular water molecules through radiolysis, generating highly reactive free radicals, primarily the hydroxyl radical ($^{\bullet}$OH), which diffuse nanometer distances to chemically

attack DNA; this mechanism accounts for approximately 70% of the biological effect of low-LET photons[130, 131].

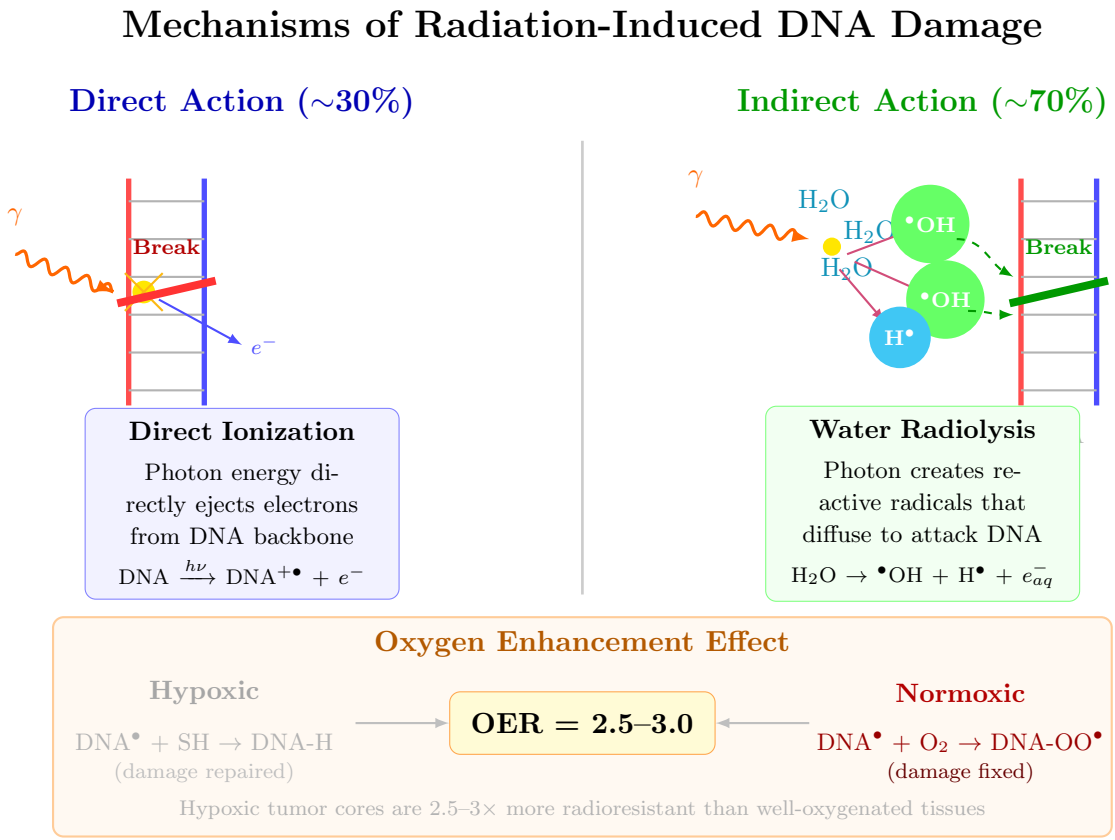


Figure 6.28: Schematic representation of direct and indirect mechanisms of radiation-induced DNA damage. Direct action involves ionization of the DNA molecule itself (∼30% of damage), while indirect action proceeds through water radiolysis and hydroxyl radical ($^{\bullet}$OH) attack (∼70% of damage). Molecular oxygen enhances damage fixation, explaining the increased radiosensitivity of well-oxygenated tissues (OER ≈ 2.5–3.0).

The efficacy of indirect action is strongly modulated by the presence of molecular oxygen. Oxygen "fixes" the damage produced by free radicals, making it permanent (Oxygen Fixation Hypothesis). Consequently, hypoxic tumor regions, often found in necrotic centers, may be 2.5 to 3 times more resistant to radiation than well-oxygenated tissues, a phenomenon quantified by the Oxygen Enhancement Ratio (OER)[132].

6.9 DNA Double-Strand Breaks and Cellular Fate

Among the various lesions induced (base damage, single-strand breaks), the DNA double-strand break (DSB) is the critical lethal event. A DSB involves the severance of both strands of the DNA helix within close proximity, disrupting genomic continuity. If unrepaired or misrepaired, DSBs lead to chromosomal aberrations (e.g., dicentrics, rings) that physically prevent successful cell division, resulting in mitotic catastrophe—the primary mode of radiation-induced cell death in solid tumors[133, 134].

Cells possess repair pathways, principally Non-Homologous End-Joining (NHEJ) and Homologous Recombination (HR), to mitigate this damage. However, the high flux of damage during fractionated radiotherapy is designed to overwhelm the repair capacity of rapidly dividing tumor cells while allowing sublethal damage repair in normal tissues between fractions. The precision of dose calculation algorithms like Acuros XB is therefore critical to ensure that the

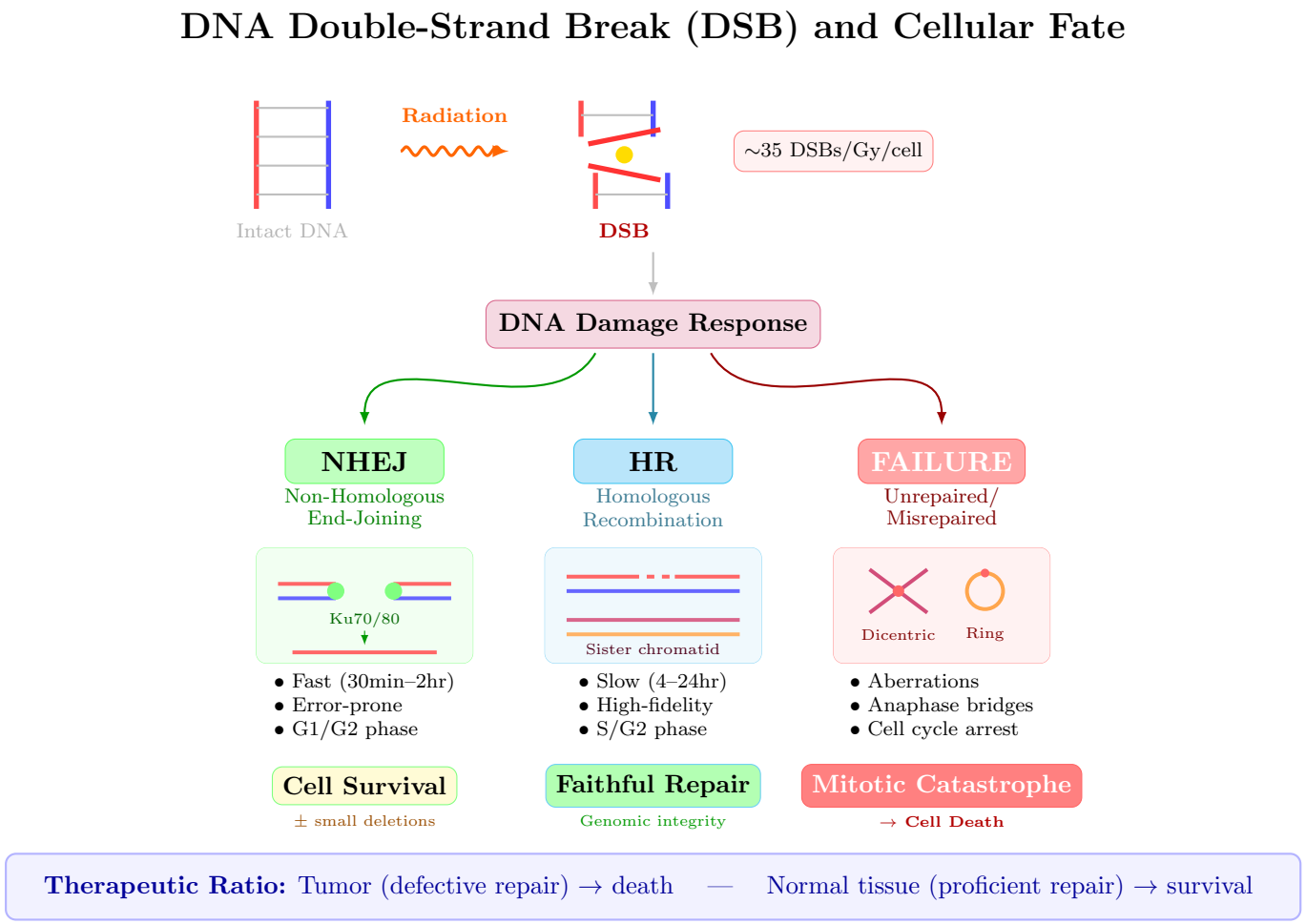


Figure 6.29: DNA double-strand break (DSB) formation and cellular fate pathways. Following radiation-induced DSB, cells employ NHEJ (fast, error-prone) or HR (slow, high-fidelity) repair mechanisms. Unrepaired or misrepaired DSBs lead to chromosomal aberrations and mitotic catastrophe—the primary mechanism of radiation-induced tumor cell death. The therapeutic ratio exploits differential repair capacities between tumor and normal tissues.

physical dose delivered correlates accurately with the predicted biological control probability.

6.10 Interaction Cross Section Units

Interaction cross-sections represent the likelihood of particles interacting with one another and are measured using a variety of equivalent units based upon an application's context. A barn is defined as the microscopic interaction cross-section for each atomic interaction and is represented by the equation (1 b = 10^{-24} cm^{2}The Mass Attenuation Coefficient (cm^2/g) or macroscopic cross-section per unit mass, and the Linear Attenuation Coefficient(cm^{-1}) or macroscopic cross-section per unit length, are the two primary representations of macroscopic cross-sections. Conversion between these representations is done through the equations:

$$\sigma_{\text{macro}}[\text{cm}^{-1}] = \frac{N_A \rho}{A} \sigma_{\text{micro}}[\text{cm}^2] = \rho \cdot \frac{\sigma}{\rho} \tag{6.25}$$

6.11 Photon Interactions with Matter

Photons interact with matter through four primary mechanisms. The relative importance of each process depends on photon energy E and atomic number Z of the absorbing material [135, 136].

6.11.1 Overview of Interaction Processes

Figure 6.30 presents a comprehensive overview of the photon interaction processes and their resulting secondary particles. An incident high-energy photon can undergo Rayleigh (coherent) scattering, photoelectric absorption, Compton (incoherent) scattering, or pair production. Each interaction produces characteristic secondary particles that ultimately deposit energy in the medium.

6.12 Photoelectric Effect

In the photoelectric effect, a photon is completely absorbed by an atom, ejecting a bound electron (photoelectron) with kinetic energy:

$$T_e = E_\gamma - E_B \tag{6.26}$$

where E_B is the binding energy of the electron shell.

Cross section dependence. The photoelectric cross section varies approximately as [135]:

$$\sigma_{\text{pe}} \propto \frac{Z^n}{E^m} \tag{6.27}$$

where $n \approx 4$–5 and $m \approx 3$ for photon energies above the K-edge.

The exact K-shell photoelectric cross section is [137]:

$$\sigma_K = \sigma_0 \frac{32}{3} \alpha^4 Z^5 \left(\frac{m_e c^2}{E} \right)^{7/2} \phi(E/E_K) \tag{6.28}$$

where $\sigma_0 = 8\pi r_e^2/3$ is the Thomson cross section, $\alpha = 1/137$ is the fine structure constant, and ϕ is a correction factor.

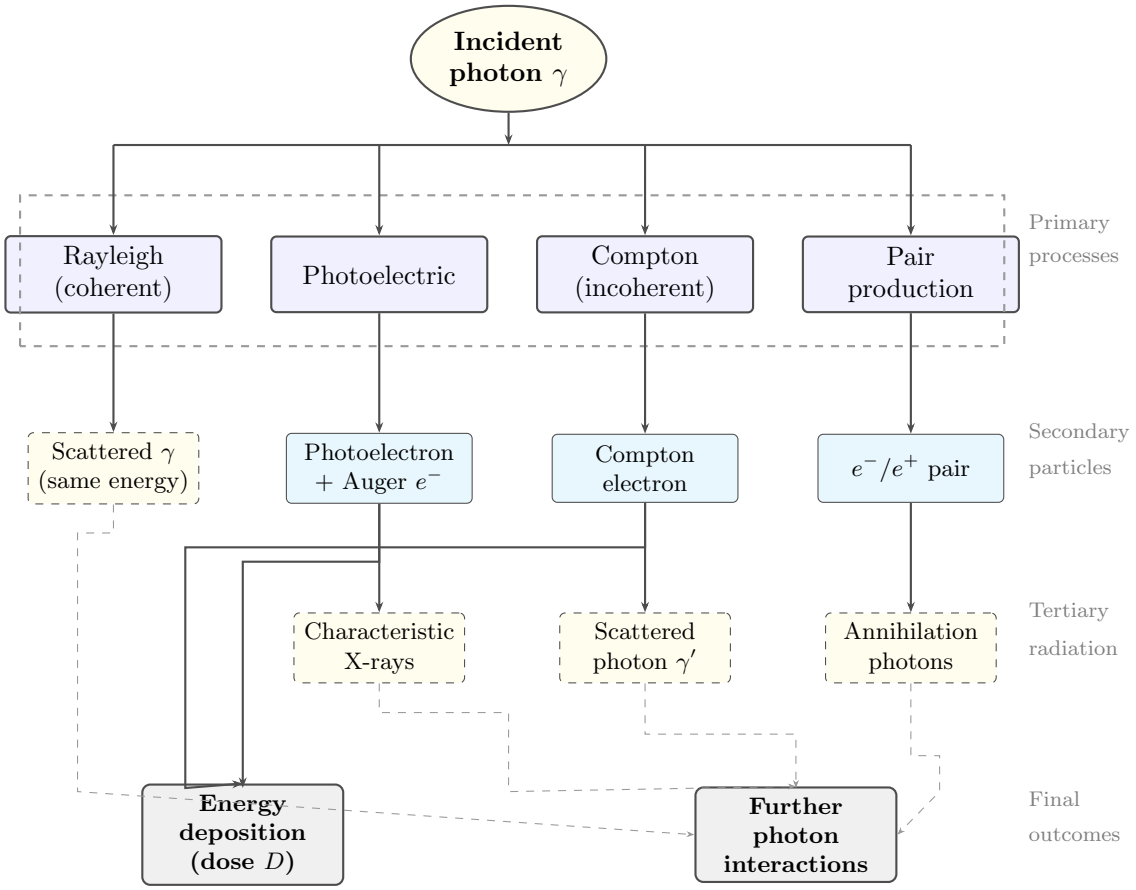


Figure 6.30: Photon interaction processes in matter. An incident high-energy photon can undergo Rayleigh scattering (elastic, no energy transfer), photoelectric absorption (full absorption, ejecting an electron), Compton scattering (partial energy transfer to an electron), or pair production (conversion to e^-/e^+, $E_\gamma > 1.022$ MeV). Secondary electrons deposit dose through ionization; tertiary photons may undergo further interactions. Adapted from [14].

Clinical relevance. The strong Z dependence of the photoelectric effect has significant clinical implications. This interaction mechanism is primarily responsible for the contrast between bone and soft tissue in diagnostic imaging, provides the dominant absorption mechanism in kilovoltage radiation therapy, and contributes to dose enhancement phenomena observed near high-Z materials such as metal implants. At 6 MV therapeutic photon energies in soft tissue, however, photoelectric interactions contribute less than 1% of total attenuation [13].

6.13 Total Photon Attenuation

The total mass attenuation coefficient is the sum of all processes:

$$\frac{\mu}{\rho} = \frac{\sigma_{\text{pe}}}{\rho} + \frac{\sigma_{\text{inc}}}{\rho} + \frac{\sigma_{\text{coh}}}{\rho} + \frac{\sigma_{\text{pair}}}{\rho} \tag{6.29}$$

The macroscopic attenuation coefficient is $\mu = (\mu/\rho) \cdot \rho$ [cm^{-1}].

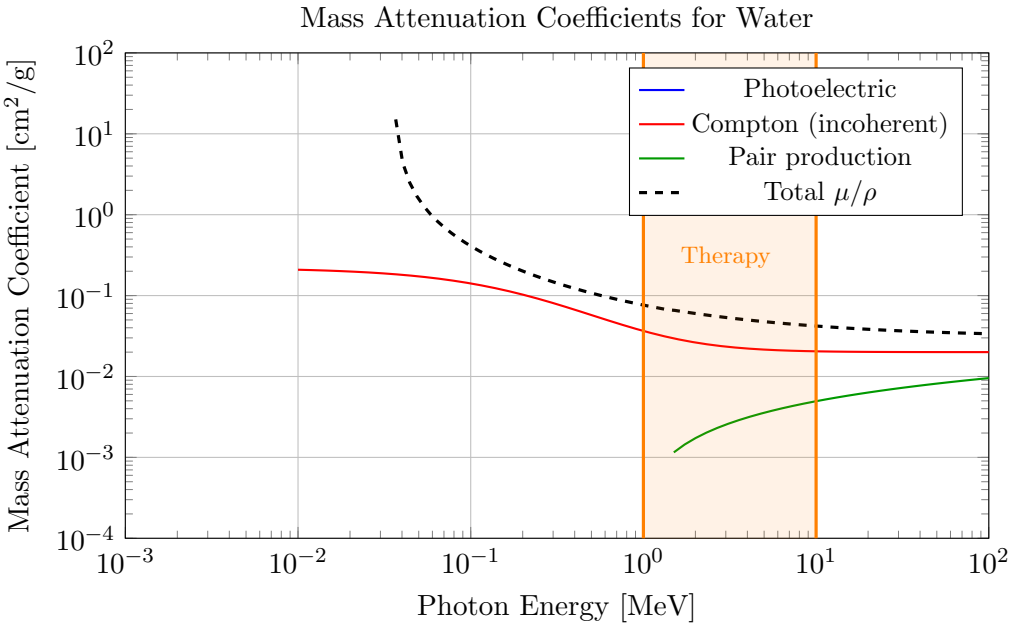


Figure 6.31: Mass attenuation coefficients for water showing dominance of Compton scattering in the therapeutic energy range (1–10 MeV, shaded). Data based on NIST XCOM database [138].

6.14 Electron Interactions with Matter

Electrons interact continuously with matter through Coulomb forces. Unlike photons, electrons undergo many small interactions rather than discrete large-angle events. The interactions are classified as **collisional** (with atomic electrons) and **radiative** (with nuclei) [139, 135].

Figure 6.32 presents a comprehensive overview of electron interaction processes. Unlike photons which undergo discrete interactions, electrons experience continuous energy loss combined with occasional hard collisions.

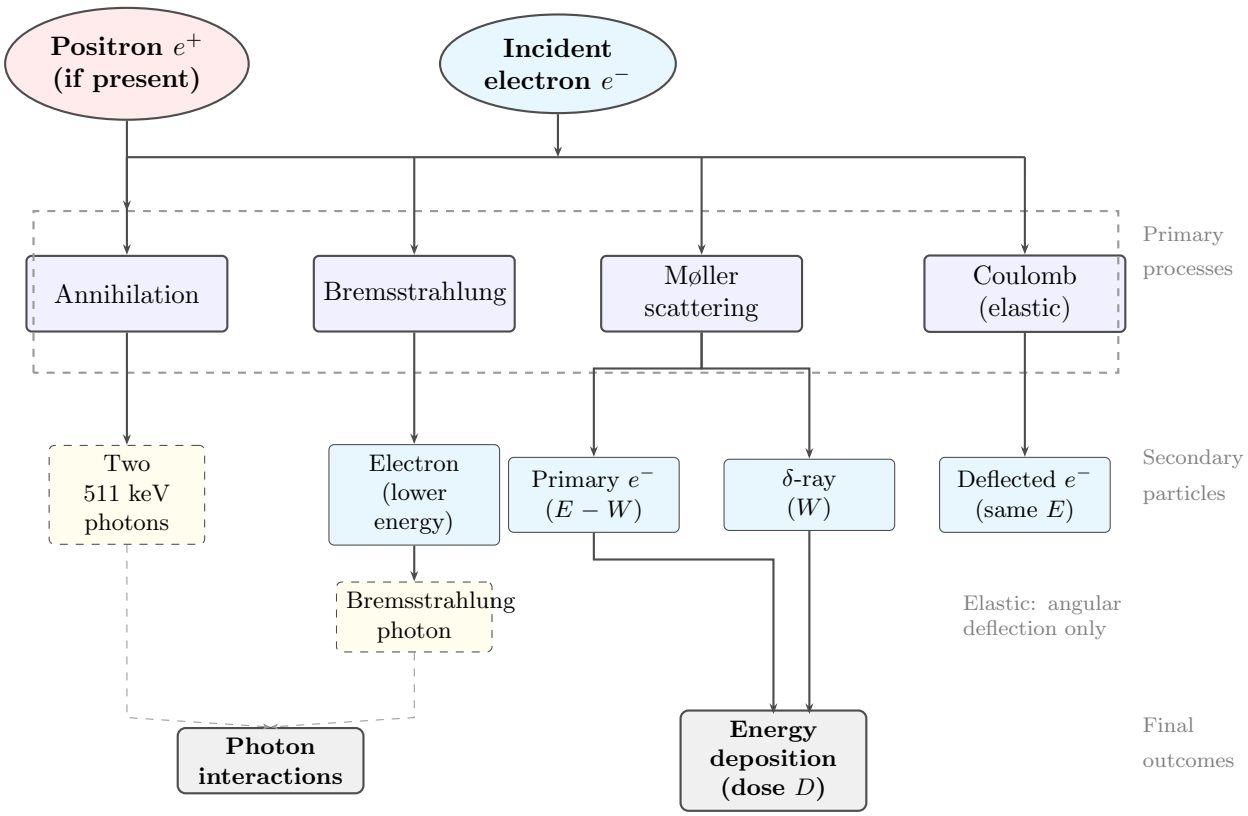


Figure 6.32: Electron interaction processes in matter. An incident electron can undergo: (1) annihilation with a positron (producing two 511 keV photons), (2) bremsstrahlung in the nuclear Coulomb field (emitting a photon), (3) Møller scattering with atomic electrons (producing δ-rays), or (4) Coulomb elastic scattering (angular deflection without energy loss). Secondary electrons deposit dose; emitted photons undergo further interactions. Adapted from [135].

6.15 Dose Quantities and Units

6.15.1 Absorbed Dose

Absorbed dose D is the fundamental dosimetric quantity, defined as the mean energy imparted per unit mass [128]:

$$D = \frac{d\bar{\epsilon}}{dm} \tag{6.30}$$

SI Unit: Gray (Gy) = J/kg = 100 rad

6.16 Kerma

Kerma (Kinetic Energy Released per unit MAss) is the sum of initial kinetic energies of all charged particles liberated by uncharged particles:

$$K = \frac{d\bar{E}_{\mathrm{tr}}}{dm} \tag{6.31}$$

For photons in the MV range [135]:

$$K = \Phi \cdot \frac{\mu_{\mathrm{tr}}}{\rho} \tag{6.32}$$

where μ_{tr}/ρ is the mass energy-transfer coefficient.

Chapter 7 The Boltzmann Transport Equation

The Linear Boltzmann Transport Equation (LBTE) describes the transport of neutral and charged particles through matter. It is the fundamental equation underlying deterministic dose calculation algorithms [140].

7.1 General Form

For a particle type p with angular flux $\Psi^p(\mathbf{r}, E, \mathbf{\Omega})$:

$$\mathbf{\Omega} \cdot \nabla \Psi^p + \sigma_t^p \Psi^p = Q^p \tag{7.33}$$

where Q^p includes scattering, production, and external sources.

7.2 Coupled Photon-Electron System

For photon beams, the system couples photon and electron transport [4]:

Photon equation:

$$\mathbf{\Omega} \cdot \nabla \Psi^\gamma + \sigma_t^\gamma \Psi^\gamma = \int_{E'}^{\infty} \int_{4\pi} \sigma_s^{\gamma\gamma}(E' \to E, \mathbf{\Omega}' \to \mathbf{\Omega}) \Psi^\gamma \, d\Omega' dE' + q^\gamma \tag{7.34}$$

Electron equation (with CSD):

$$\mathbf{\Omega} \cdot \nabla \Psi^e + \sigma_t^e \Psi^e - \frac{\partial}{\partial E}(S_R \Psi^e) = q^{ee} + q^{\gamma e} + q^e \tag{7.35}$$

The term $-\partial(S_R \Psi^e)/\partial E$ is the Continuous Slowing Down (CSD) operator, which accounts for continuous energy loss.

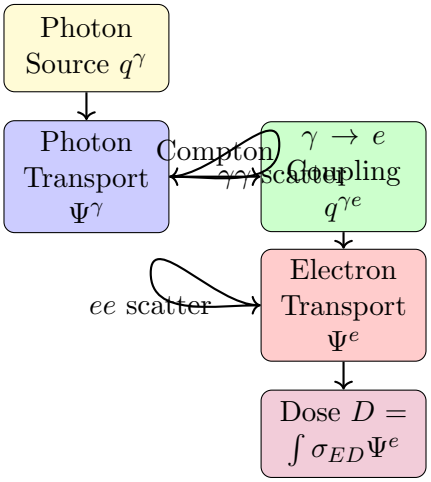


Figure 7.33: Coupled photon-electron transport scheme used in Acuros XB and similar deterministic dose calculation algorithms.

Chapter 8 The Acuros XB Algorithm

Acuros XB (eXternal Beam) is a deterministic dose calculation algorithm that solves the LBTE using the discrete ordinates (S_N) method [4, 141, 48].

8.1 Algorithm Overview

At a high level, Acuros XB proceeds through source modeling, uncollided transport, first-collision source construction, collided transport, and dose computation. First, the photon fluence emerging from the linac head is characterized using a multi-source model derived from Monte Carlo simulations or commissioning measurements. Primary (uncollided) photons are then transported analytically by ray tracing with Beer–Lambert attenuation, and the resulting uncollided fluence is used to form first-collision scattering sources and photon-to-electron production sources. The coupled LBTE is subsequently solved iteratively for the collided (scattered) photon and electron components using S_N discrete ordinates. Finally, absorbed dose is computed from the electron fluence through the energy-deposition coefficient.

Table 8.6: Key Acuros XB algorithm parameters as specified in Eclipse [142].

Parameter	Value	Description
G_γ	25	Number of photon energy groups
G_e	49	Number of electron energy groups
L_γ	≤ 7	Legendre order for photon scattering
L_e	1	Legendre order for electron scattering (CSD handles forward peak)
N	4–16	S_N quadrature order (energy-dependent)
M	32–512	Number of discrete directions ($M = 2N^2$)
E_γ^{cut}	1 keV	Photon transport cutoff
E_e^{cut}	200 keV	Electron transport cutoff
Spatial	DD	Diamond difference with step fixup

8.2 Discrete Ordinates (S_N) Method

The angular domain is discretized into M directions $\{\mathbf{\Omega}_m, w_m\}$ using Gaussian quadrature [140]:

$$\int_{4\pi} f(\mathbf{\Omega})\, d\Omega \approx \sum_{m=1}^{M} w_m f(\mathbf{\Omega}_m) \tag{8.36}$$

Acuros XB uses Square Tchebychev-Legendre (STL) quadrature with $N \in \{4, 6, 8, 10, 12, 14, 16\}$ giving $M = 2N^2$ directions [4].

8.3 Energy Groups (Multigroup)

Energy is discretized into groups with edges $\{E_0, E_1, \ldots, E_G\}$:

$$\Psi_g(\mathbf{r}, \mathbf{\Omega}) = \int_{E_g}^{E_{g+1}} \Psi(\mathbf{r}, E, \mathbf{\Omega})\, dE \tag{8.37}$$

Typical configurations: In typical clinical configurations, photons are represented with approximately 10–25 energy groups spanning the range from 1 keV to 25 MeV, whereas electrons are commonly represented with approximately 12–50 groups spanning 0.2 MeV to 25 MeV.

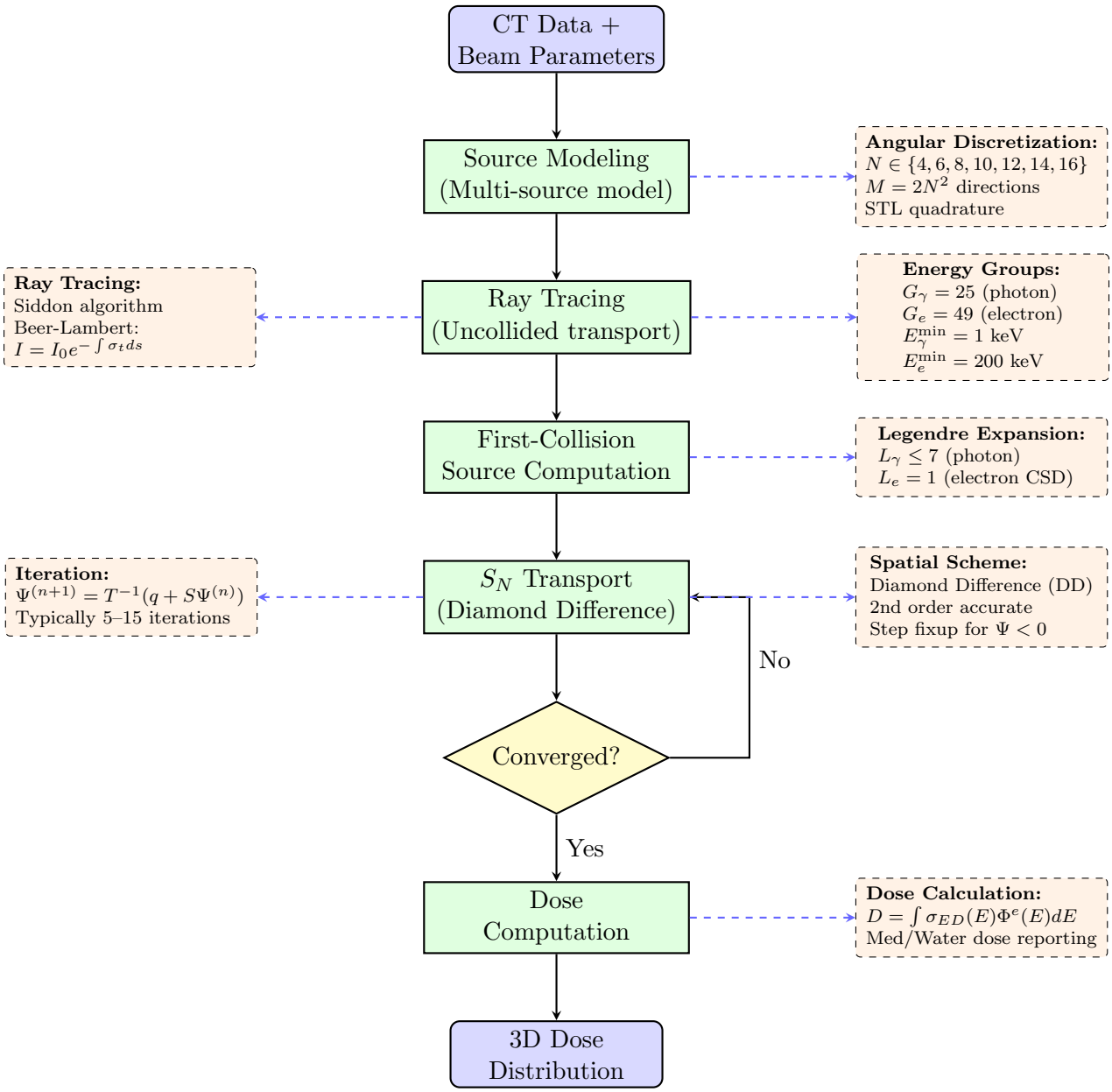


Figure 8.34: Complete flow chart of the Acuros XB deterministic dose computation algorithm using key numerical values. The algorithm with key numerical parameters. The algorithm utilizes the G_γ photon and G_e electron energy groups, $M = 2N^2$ discrete ordinate direction schemes utilizing STL (Simple Triangular Linear) quadratures, and Legendre expansion orders L_γ and L_e to represent the scattering kernels. Spatially, the DD (diamond difference) scheme is used for spatial discretization with step fix-up for positivity.

8.4 Spatial Discretization

The spatial domain is discretized on the CT grid. Each voxel is assigned material properties (density, composition) from the CT Hounsfield Units (HU).

The transport equations are solved by discretizing them via finite differences of the space domain, that is, via a spatial differencing scheme: the diamond-difference scheme is second-order accurate, but it can produce non-physical negative fluxes in strongly heterogeneous regions; step characteristics are first order (with respect to time) but they enforce positivity; and hybrid schemes are combinations of the diamond-difference with local step-characteristic based fix-up methods, which maintains positivity without completely losing the accuracy.

8.5 Legendre Expansion of Scattering

The scattering kernel is expanded in Legendre polynomials:

$$\sigma_s(\mathbf{\Omega}' \cdot \mathbf{\Omega}) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi} \sigma_{s,\ell} P_\ell(\mathbf{\Omega}' \cdot \mathbf{\Omega}) \tag{8.38}$$

Acuros XB uses $L \leq 7$ for photons (forward-peaked Compton) and $L = 1$ for electrons (highly forward-peaked elastic scattering handled by CSD).

8.6 Mathematical Framework: Spherical Harmonics and Angular Discretization

The angular dependence of radiation fluence is naturally described using spherical harmonics, which form a complete orthonormal basis on the unit sphere.

8.6.1 Coordinate System and Notation

The direction of particle travel $\mathbf{\Omega}$ is specified using spherical coordinates (θ, φ) on the unit sphere, where $\theta \in [0, \pi]$ is the polar angle measured from the x_3-axis and $\varphi \in [0, 2\pi)$ is the azimuthal angle in the x_1-x_2 plane.

The direction vector in Cartesian coordinates is:

$$\mathbf{\Omega} = (\Omega_1, \Omega_2, \Omega_3) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \tag{8.39}$$

with the commonly used notation $\mu = \cos\theta = \Omega_3$ for the direction cosine.

8.7 Angular Fluence Expansion

The angular fluence can be expanded in spherical harmonics:

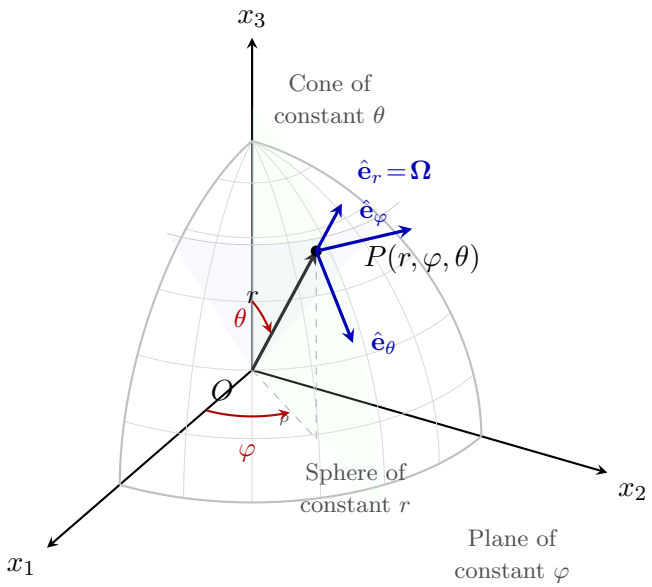
$$\Psi(\mathbf{r}, E, \mathbf{\Omega}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \phi_{\ell m}(\mathbf{r}, E) Y_\ell^m(\mathbf{\Omega}) \tag{8.40}$$

The moments $\phi_{\ell m}$ are computed as:

$$\phi_{\ell m}(\mathbf{r}, E) = \int_{4\pi} \Psi(\mathbf{r}, E, \mathbf{\Omega}) Y_\ell^{m*}(\mathbf{\Omega})\, d\Omega \tag{8.41}$$

The scalar fluence is the zeroth moment:

$$\phi(\mathbf{r}, E) = \phi_{00}(\mathbf{r}, E)\sqrt{4\pi} = \int_{4\pi} \Psi(\mathbf{r}, E, \mathbf{\Omega})\, d\Omega \tag{8.42}$$



$$\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\theta = \hat{\mathbf{e}}_\varphi$$
$$\hat{\mathbf{e}}_\theta \times \hat{\mathbf{e}}_\varphi = \hat{\mathbf{e}}_r$$
$$\hat{\mathbf{e}}_\varphi \times \hat{\mathbf{e}}_r = \hat{\mathbf{e}}_\theta$$

Figure 8.35: Spherical polar coordinates (r, φ, θ) for $\mathbf{x} = (x_1, x_2, x_3)$. The unit direction vector $\mathbf{\Omega} = \hat{\mathbf{e}}_r$ is parameterized by polar angle θ (from x_3-axis) and azimuthal angle φ (in x_1-x_2 plane). The orthonormal basis $(\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_\theta, \hat{\mathbf{e}}_\varphi)$ spans the local coordinate system at each point on the unit sphere.

8.8 Discrete Ordinates Quadrature

The discrete ordinates (S_N) method replaces angular integrals with weighted sums:

$$\int_{4\pi} f(\mathbf{\Omega})\, d\Omega \approx \sum_{m=1}^{M} w_m f(\mathbf{\Omega}_m) \tag{8.43}$$

Level-symmetric (LS) quadrature: Directions are placed on "level" lines of constant $|\mu|$, with octant symmetry. For order N, there are $N(N+2)/8$ directions per octant, giving $M = N(N+2)$ total.

Square Legendre-Chebyshev (SLC) quadrature: Used in Acuros XB, with $M = 2N^2$ directions. Provides better integration of products of spherical harmonics up to order $2N-1$.

Table 8.7: Number of discrete ordinates directions for different quadrature orders.

Order N	Level-Symmetric M	Square LC M
4	24	32
6	48	72
8	80	128
10	120	200
12	168	288

8.9 Scattering Kernel Theory

8.9.1 General Scattering Operator

The scattering source term involves integration over all incoming angles and energies:

$$q_s(\mathbf{r}, E, \mathbf{\Omega}) = \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s(\mathbf{r}, E' \to E, \mathbf{\Omega}' \cdot \mathbf{\Omega})\, \Psi(\mathbf{r}, E', \mathbf{\Omega}') \tag{8.44}$$

For azimuthally symmetric scattering (depending only on $\mu_0 = \mathbf{\Omega}' \cdot \mathbf{\Omega}$), the kernel is expanded in Legendre polynomials:

$$\sigma_s(E' \to E, \mu_0) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi} \sigma_{s,\ell}(E' \to E) P_\ell(\mu_0) \tag{8.45}$$

8.10 Legendre Moments of Scattering

The Legendre moments are computed from the differential cross section:

$$\sigma_{s,\ell}(E' \to E) = 2\pi \int_{-1}^{1} \sigma_s(E' \to E, \mu_0) P_\ell(\mu_0)\, d\mu_0 \tag{8.46}$$

The zeroth moment is the total scattering cross section:

$$\sigma_{s,0}(E' \to E) = \sigma_s^{\text{total}}(E' \to E) \tag{8.47}$$

The first moment relates to the mean cosine of scattering:

$$\bar{\mu}_0 = \frac{\sigma_{s,1}}{\sigma_{s,0}} \tag{8.48}$$

8.11 Cross-Section Generation Pipeline

The computation of multigroup cross sections constitutes a critical prerequisite for the deterministic LBTE solver, as the accuracy of the transport solution depends directly on the fidelity of the underlying interaction data. This section covers the entire cross-section generation process from the basic physics-based models to the radiative-transfer ready data structure. All the components are built on top of the Radiant.jl framework [**?**].

Radiant.jl is an open-source Julia package which is focused on generating the cross-section data for the multi-group radiation-transport calculations from the first principles physics- based models. It uses ENDF/B-VIII.0 nuclear data evaluations. In this context, it implements all the channels for the photon interactions — Rayleigh coherent scattering, Compton incoherent scattering, photoelectric absorption (with the corresponding fluorescence and Auger emission) and pair production. Additionally, Radiant.jl also includes the electron interactions — elastic Coulomb scattering, inelastic Møller scattering and bremsstrahlung. For every possible material composition, Radiant.jl performs a multigroup energy condensation within the user-defined energy groups and computes the Legendre expansion moments of the differential scattering cross-sections to any order. As such, Radiant.jl generates the entire set of the data required for running the SN codes.

The cross-section extraction pipeline is constructed of three stages. The first stage is the Julia/Radiant.jl physics generation based on 44 different material compositions given as elemental mass fractions. This produces the raw cross-section data. The second stage is the schema normalization and unit harmonization by Python. Here, the output format produced by Radiant is converted into the legacy solver schema. The third stage consists of YAML conversion with 12 integrity checks. At the end of these stages, the files are ready for the solvers. The pipeline generates all the necessary multigroup cross-sections (Σ_t, Σ_s with Legendre moments), scattering transfer matrices ($\gamma \to \gamma$, $\gamma \to e$, $e \to e$), restricted stopping powers (S_R), energy-deposition coefficients (σ_{ED}), and soft-scatter cross sections for the complete set of 44 tissue-equivalent materials spanning from air ($\rho = 0.00121\,\mathrm{g/cm^3}$) to stainless steel ($\rho = 8.0\,\mathrm{g/cm^3}$).

8.11.1 Multigroup Energy Discretization

The energy domain is divided into $G_\gamma = 10$ photon groups and $G_e = 12$ electron groups with edges:

$$E_\gamma: \quad 0.001,\ 0.00239,\ 0.00572,\ \ldots,\ 2.536,\ 6.0\ \mathrm{MeV} \tag{8.49}$$

$$E_e: \quad 0.200,\ 0.266,\ 0.353,\ \ldots,\ 4.525,\ 6.0\ \mathrm{MeV} \tag{8.50}$$

The group-averaged angular flux for photon group g is:

$$\psi_g(\vec{r}, \hat{\Omega}) = \frac{1}{\Delta E_g} \int_{E_g^-}^{E_g^+} \psi(\vec{r}, E, \hat{\Omega})\, dE \tag{8.51}$$

The multigroup photon equation becomes:

$$\hat{\Omega} \cdot \nabla \psi_g^\gamma + \Sigma_{t_g}^\gamma \psi_g^\gamma = \sum_{g'=1}^{G_\gamma} \sum_{\ell=0}^{L_\gamma} \frac{2\ell+1}{4\pi} \Sigma_{s\ell}^{\gamma\gamma}(g' \to g) \sum_{m=-\ell}^{\ell} Y_\ell^m(\hat{\Omega})\, \phi_{\ell,m}^{g'} + Q_g^{\mathrm{ext}} \tag{8.52}$$

8.11.2 Cross-Section Physics Models

This section presents the complete first-principles physics models used by the Radiant.jl framework for computing multigroup cross-sections. All models are derived from fundamental quantum mechanical principles and implemented with rigorous attention to numerical accuracy.

Photon Interactions The following subsections describe the four primary photon interaction mechanisms implemented in Radiant.jl: Rayleigh (coherent) scattering, Compton (incoherent) scattering, the photoelectric effect, and pair production.

Rayleigh Scattering Rayleigh scattering, also often called coherent scattering, consists of the elastic change of direction of an incoming photon by atomic electrons. The Rayleigh scattering model from PENELOPE [143], which is based on atomic form factors [144] and anomalous scattering factors [145], is used.

The Rayleigh cross-sections, which describe the elastic scattering of photons ($p = p' = \gamma$), are given by:

$$\sigma_s^i(E,\mu) = \pi r_e^2 \left(1+\mu^2\right) \left[\left(F_i(E,\mu) + f_i'(E)\right)^2 + \left(f_i''(E)\right)^2\right] \tag{8.53}$$

where $F_i(E,\mu)$ is the atomic form factor for the ith-element, and $f_i'(E)$ and $f_i''(E)$ are respectively the real and imaginary parts of the anomalous scattering factors, tabulated by JENDL-5 [146] based on the EPDL library [147]. The double differential cross-sections are:

$$\sigma_s^i(E \to E', \mu) = \sigma_s^i(E,\mu)\delta(E'-E) \tag{8.54}$$

The Rayleigh Legendre moments of the differential scattering cross-sections are:

$$\Sigma_{s,\ell}^{\gamma\to\gamma}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_{-1}^{1} P_\ell(\mu)\sigma_s^i(E,\mu)\, d\mu \tag{8.55}$$

which are integrated using numerical quadrature. The Rayleigh total cross-sections are:

$$\Sigma_t^\gamma(E) = \Sigma_{s,0}^{\gamma\to\gamma}(E) \tag{8.56}$$

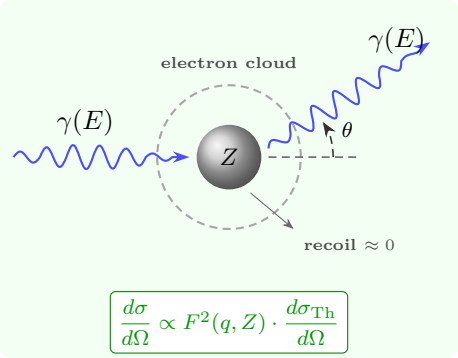


Figure 8.36: **Rayleigh scattering**: Coherent elastic scattering off the whole atom. Energy unchanged ($E'=E$), only direction changes.

Compton Scattering Compton scattering, often called incoherent scattering, consists of the ionization of an atomic electron by an incoming photon, which loses energy in the process. This interaction is dominant for intermediate energies ranging from hundreds of keV to a few MeV. The proposed model uses the Waller-Hartree incoherent function applied to the Klein-Nishina cross-section [148] to encompass electron binding effects [149].

The Compton cross-section describes the interaction of an incoming photon ($p' = \gamma$) with atomic electrons, resulting in a scattered photon ($p = \gamma$) and a produced electron ($p = \texttt{e-}$). The Klein-Nishina differential cross-section in the energy of the scattered photon is:

$$\sigma_s(E \to E') = \frac{\pi r_e^2}{E^2}\left[\frac{E}{E'} + \frac{E'}{E} - 2\left(\frac{1}{E'} - \frac{1}{E}\right) + \left(\frac{1}{E'} - \frac{1}{E}\right)^2\right] \tag{8.57}$$

where E is the incoming photon energy, E' is the scattered photon energy, and $W = E - E'$ is the produced electron energy. The scattering angles for the scattered photon and produced electron are respectively:

$$\mu_\gamma = 1 + \frac{1}{E} - \frac{1}{E'} \quad \text{and} \quad \mu_e = \frac{1+E}{E}\left[1 + \frac{2}{W}\right]^{-1/2} \tag{8.58}$$

The double differential cross-sections with incoherent scattering factor $S_i(E,\mu)$ from JENDL-5 [146] are:

$$\sigma_s^i(E \to E', \mu) = \frac{1}{2\pi} S_i(E,\mu)\sigma_s(E \to E')\delta(\mu - \mu_\gamma) \tag{8.59}$$

$$\sigma_s^i(E \to W, \mu) = \frac{1}{2\pi} S_i(E,\mu)\sigma_s(E \to W)\delta(\mu - \mu_e) \tag{8.60}$$

The Compton Legendre moments of the differential scattering cross-sections for the incoming photon are:

$$\sigma_{s,\ell}^i(E \to E') = P_\ell(\mu_\gamma) S_i(E,\mu_\gamma)\sigma_s(E \to E') \tag{8.61}$$

The multigroup Legendre moments of the scattering cross-sections are:

$$\Sigma_{s,\ell,g'}^{\gamma\to\gamma}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_{\max\left\{E'_{g+1/2}, \frac{E}{1+2E}\right\}}^{\min\left\{E'_{g-1/2}, E\right\}} dE'\, \sigma_{s,\ell}^i(E \to E') \tag{8.62}$$

For the produced electron:

$$\Sigma_{s,\ell,g'}^{\gamma\to e^-}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_{E'_{g+1/2}}^{\min\left\{E'_{g-1/2}, \frac{2E^2}{1+2E}\right\}} dW\, \sigma_{s,\ell}^i(E \to W) \tag{8.63}$$

The Compton total cross-sections are:

$$\Sigma_t^\gamma(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_{\frac{E}{1+2E}}^{E} dE'\, S_i(E,\mu_\gamma)\sigma_{s,0}(E \to E') \tag{8.64}$$

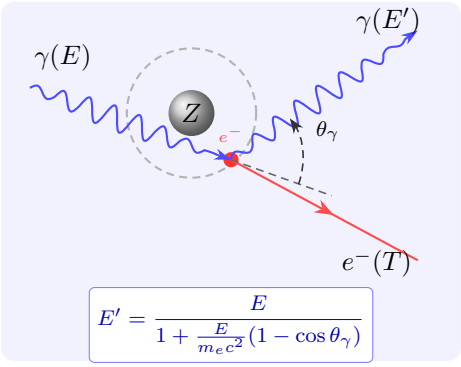


Figure 8.37: **Compton scattering**: Incoherent scattering off a quasi-free atomic electron (Z). Photon loses energy; electron recoils with $T = E - E'$.

Photoelectric Effect The photoelectric effect consists of the emission of electrons by absorption of photons. Tables from the subshell-dependent Evaluated Photon Data Library (EPDL) [150] and the Sauter distribution are used.

The photoelectric cross-section describes the absorption of an incoming photon ($p' = \gamma$) and the emission of an atomic electron ($p = \texttt{e-}$). The microscopic absorption cross-sections are:

$$\sigma_a^i(E) = \sum_{k=1}^{N_{\texttt{shells}}} \sigma_a^{i,k}(E) \tag{8.65}$$

where $\sigma_a^{i,k}(E)$ is given by linear interpolation of the JENDL-5 library data [146], available for $Z_i \leq 100$ and energies up to 100 GeV.

The photoelectric scattering cross-sections are:

$$\sigma_s^{i,k}(E \to E', \mu) = \sigma_a^{i,k}(E)\delta(E' - E + U_{i,k})\Theta(E, \mu) \tag{8.66}$$

where E' is the photo-electron energy, $U_{i,k}$ is the binding energy of the kth shell, and the normalized Sauter cross-section for the K-shell is [151]:

$$\Theta(E, \mu) = \Gamma(E)\frac{1-\mu^2}{(1-\beta\mu)^4}\left[1 + \frac{\gamma(\gamma-1)(\gamma-2)}{2}(1-\beta\mu)\right] \tag{8.67}$$

with normalization factor:

$$\Gamma(E) = \left\{\frac{4}{3(1-\beta^2)^2} + \frac{\gamma(\gamma-1)(\gamma-2)}{2\beta^3}\left[\frac{2\beta}{1-\beta^2} - \ln\left(\frac{1+\beta}{1-\beta}\right)\right]\right\}^{-1} \tag{8.68}$$

The Legendre moments of the normalized Sauter cross-section are computed analytically:

$$\Theta_\ell(E) = \frac{\Gamma(E)}{2^\ell}\sum_{k=0}^{\lfloor \ell/2 \rfloor} C_{\ell,k}\sum_{j=0}^{1}(-1)^j\left[I^{\ell-2k+2j,4}(\beta) + \frac{\gamma(\gamma-1)(\gamma-2)}{2}I^{\ell-2k+2j,3}(\beta)\right] \tag{8.69}$$

The scattering cross-sections are:

$$\Sigma_{s,\ell,g'}^{\gamma\to\texttt{e}^-}(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\sum_{k=1}^{N_{\texttt{shells}}}\sigma_a^{i,k}(E)\Theta_\ell(E) \times \begin{cases} 1 & E - U_{i,k} \in [E'_{g+1/2}, E'_{g-1/2}] \\ 0 & \text{otherwise} \end{cases} \tag{8.70}$$

The photoelectric total cross-sections are:

$$\Sigma_t^\gamma(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\sigma_a^i(E) \tag{8.71}$$

Pair Production An electron-positron pair is produced from the incoming photon energy near the atom. The threshold is approximately 1.022 MeV (twice the electron rest mass). This phenomenon dominates at high energies (>10 MeV). A variation of the model of Baró et al. [152], based on the screening function of Tsai [153], is employed.

The differential cross-section in a monoelemental material i is:

$$\sigma_s^i(E \to E') = \begin{cases} A(Z_i, E)\left[2\left(\frac{1}{2} - \frac{E'+1}{E}\right)^2\phi_{i,1}(E') + \phi_{i,2}(E')\right] & E > 2 \\ 0 & \text{otherwise} \end{cases} \tag{8.72}$$

where E is the incoming photon energy and E' is the outgoing electron or positron energy. The screening functions derived from Tsai [153] are:

$$\phi_{i,1}(E') = \max\left\{g_{i,1}(E') + g_{i,0}, 0\right\}, \quad \phi_{i,2}(E') = \max\left\{g_{i,2}(E') + g_{i,0}, 0\right\} \tag{8.73}$$

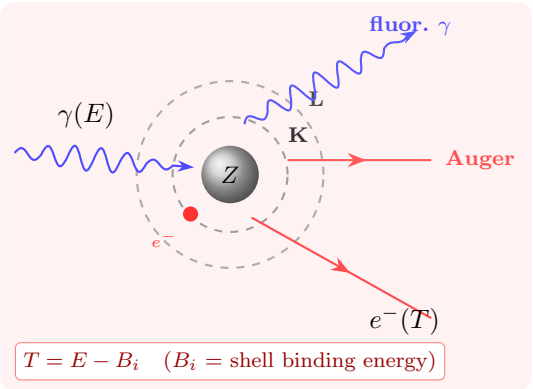


Figure 8.38: **Photoelectric effect**: Photon absorbed; inner-shell electron ejected with $T = E - B_i$. Vacancy filled by fluorescence or Auger emission.

where:

$$g_{i,1}(E') = \frac{7}{3} - 2\ln(1+b_i^2) - 6b_i \arctan\left(\frac{1}{b_i}\right) - b_i^2\left[4 - 4b_i \arctan\left(\frac{1}{b_i}\right) - 3\ln\left(1+\frac{1}{b_i^2}\right)\right] \tag{8.74}$$

$$g_{i,2}(E') = \frac{11}{6} - 2\ln(1+b_i^2) - 3b_i \arctan\left(\frac{1}{b_i}\right) - \frac{b_i^2}{2}\left[4 - 4b_i \arctan\left(\frac{1}{b_i}\right) - 3\ln\left(1+\frac{1}{b_i^2}\right)\right] \tag{8.75}$$

with $g_{i,0} = 4\ln(r_{s,i}) - 4f_{C,i}$ and $b_i = \frac{r_{s,i}}{2}\frac{E}{(E'+1)(E-E'-1)}$.

The reduced screening radius $r_{s,i}$ is tabulated in Baró [152] for $Z_i \leq 92$ and extended to $Z_i \leq 99$ in PENELOPE [143]. The high-energy Coulomb correction is [154]:

$$f_{C,i} = \alpha^2 Z_i^2 \sum_{k=1}^{\infty} \frac{1}{k(k^2+\alpha^2 Z_i^2)} \tag{8.76}$$

The normalization factor $A(Z_i, E)$ is the ratio between JENDL-5 tabulated values and the model with $A = 1$ [155, 146]:

$$A(Z_i, E) = \frac{\sigma_t^{\texttt{JENDL-5}}(Z_i, E)}{\sigma_t(Z_i, E)|_{A=1}} \tag{8.77}$$

The double differential cross-section uses the Sommerfield distribution [156, 153]:

$$\sigma_s^i(E \to E', \mu) = \frac{1}{2\pi}\sigma_s^i(E \to E')\Theta(E, \mu) \tag{8.78}$$

The multigroup Legendre moments are:

$$\Sigma_{s,\ell,g'}^{\gamma\to e^{\pm}}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_{E'_{g+1/2}}^{\min\left\{E'_{g-1/2}, E-2\right\}} dE'\, \sigma_{s,\ell}^i(E \to E')\mathcal{H}_b \tag{8.79}$$

The pair production total cross-sections are:

$$\Sigma_t^{\gamma}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_0^{E-2} dE'\, \sigma_{s,0}^i(E \to E')\mathcal{H}_b \tag{8.80}$$

The absorption cross-sections required for annihilation calculations are:

$$\Sigma_a^{\texttt{PP}}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \int_0^{\min\left\{E-2, E_{G+1/2}\right\}} dE'\, \sigma_{s,0}^i(E \to E')\mathcal{H}_b \tag{8.81}$$

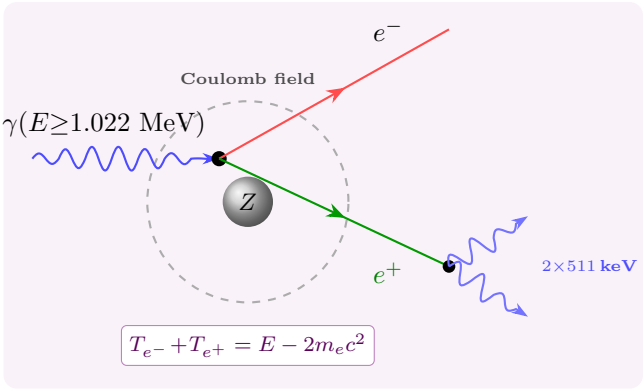


Figure 8.39: **Pair production**: Photon converts into e^-e^+ pair in nuclear Coulomb field ($E{\geq}1.022\,\text{MeV}$). Positron annihilates $\rightarrow 2{\times}511\,\text{keV}$ γ.

Electron and Positron Interactions The following subsections describe the four primary electron/positron interaction mechanisms: impact ionization, elastic scattering, bremsstrahlung, and annihilation.

Impact Ionization When travelling through the medium, electrons or positrons ionize atomic electrons in impact ionization. This interaction is the dominant energy-loss phenomenon up to a few MeV [157]. The Boltzmann-Fokker-Planck formalism is used. Catastrophic cross-sections are defined by subshell-dependent Møller [158] and Bhabha [159] models for electrons and positrons respectively, based on Seltzer [160, 161] and Salvat et al. [162, 157]. Soft stopping powers are based on the state-of-the-art models of Salvat and Andreo [163].

The collisional inelastic differential cross-section in a monoelemental material i is:

$$\sigma_s^i(E \rightarrow W) = \sum_{k=1}^{N_{\text{shells}}} Z_{i,k}\sigma_s^{i,k}(E \rightarrow W) \tag{8.82}$$

where $Z_{i,k}$ is the mean number of electrons in subshell k, E is the incoming energy, E' is the scattered energy, and $W = E - U_{i,k} - E'$ is the knock-on electron energy with binding energy $U_{i,k}$. The close interaction cross-section is:

$$\sigma_s^{i,k}(E \rightarrow W) = \frac{2\pi r_e^2}{\beta^2} F_{i,k}^{\pm}(E, W) \tag{8.83}$$

The subshell-dependent Møller factor for electrons is [158, 160]:

$$F_{i,k}^{-}(E, W) = \frac{1}{(W + U_{i,k})^2} + \frac{1}{(E - W)^2} + \frac{1}{(E + 1)^2} - \frac{(2E + 1)}{(E + 1)^2(E - W)(W + U_{i,k})} \tag{8.84}$$

The subshell-dependent Bhabha factor for positrons is [159]:

$$F_{i,k}^{+}(E, W) = \frac{1}{(W + U_{i,k})^2}\left[1 - b_1\frac{W + U_{i,k}}{E} + b_2\left(\frac{W + U_{i,k}}{E}\right)^2 - b_3\left(\frac{W + U_{i,k}}{E}\right)^3 + b_4\left(\frac{W + U_{i,k}}{E}\right)^4\right] \tag{8.85}$$

with coefficients:

$$b_1 = \left(\frac{\gamma - 1}{\gamma}\right)^2 \frac{2(\gamma + 1)^2 - 1}{\gamma^2 - 1}, \quad b_2 = \left(\frac{\gamma - 1}{\gamma}\right)^2 \frac{3(\gamma + 1)^2 + 1}{(\gamma + 1)^2} \tag{8.86}$$

$$b_3 = \left(\frac{\gamma - 1}{\gamma}\right)^2 \frac{2\gamma(\gamma - 1)}{(\gamma + 1)^2}, \quad b_4 = \left(\frac{\gamma - 1}{\gamma}\right)^2 \frac{(\gamma - 1)^2}{(\gamma + 1)^2} \tag{8.87}$$

The scattering angles are [156]:

$$\mu_p = \sqrt{\frac{E'(E+2)}{E(E'+2)}}, \quad \mu_s = \sqrt{\frac{W(E+2)}{E(W+2)}} \tag{8.88}$$

The maximum knock-on energy for electrons and positrons is:

$$W^{-}_{\mathtt{max}} = \frac{E - U_{i,k}}{2}, \quad W^{+}_{\mathtt{max}} = E - U_{i,k} \tag{8.89}$$

The multigroup Legendre moments for particle scattering are:

$$\Sigma^{\mathtt{e}^{\pm}\to\mathtt{e}^{\pm}}_{s,\ell,g'}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \int_{\max\left\{E'_{g+1/2}, W^{\pm}_{\mathtt{max}}\right\}}^{\min\left\{E'_{g-1/2}, E_c(E), E-U_{i,k}\right\}} dE'\, \sigma^{i,k}_{s,\ell}(E \to E')\mathcal{H}_b \tag{8.90}$$

For knock-on electrons:

$$\Sigma^{\mathtt{e}^{\pm}\to\mathtt{e}^{-}}_{s,\ell,g'}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \int_{E'_{g+1/2}}^{\min\left\{E'_{g-1/2}, W^{\pm}_{\mathtt{max}}\right\}} dW\, \sigma^{i,k}_{s,\ell}(E \to W)\mathcal{H}_b \tag{8.91}$$

The catastrophic total cross-sections are:

$$\Sigma^{\mathtt{e}^{\pm}}_{t}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \int_{E-E_c(E)}^{W^{\pm}_{\mathtt{max}}} dW\, \sigma^{i,k}_{s,0}(E \to W)\mathcal{H}_b \tag{8.92}$$

The total collisional stopping power for any compound is given by the density- and shell-corrected Bethe formula [164]:

$$S_t(E) = \frac{2\pi r_e^2}{\beta^2} \mathcal{N}_e^{\mathtt{eff}} \left[\ln\left\{ \frac{E+2}{2} \left(\frac{E}{I^{\mathtt{eff}}} \right)^2 \right\} + f^{(\pm)} - \delta_F - 2\mathcal{C}(E) \right] \tag{8.93}$$

where $\mathcal{N}_e^{\mathtt{eff}} = \sum_i \mathcal{N}_{n,i} f_i Z_i$ is the effective electron density, δ_F is the Fermi density effect, and $\mathcal{C}(E)$ is the shell correction from SBETHE [164]. The effective mean excitation energy is [143]:

$$I^{\mathtt{eff}} = \exp\left\{ \frac{1}{Z^{\mathtt{eff}}} \sum_{i=1}^{N_e} f_i Z_i \log(I_i) \right\}, \quad Z^{\mathtt{eff}} = \sum_{i=1}^{N_e} f_i Z_i \tag{8.94}$$

The electron factor $f^{(-)}$ and positron factor $f^{(+)}$ are [165]:

$$f^{(-)} = 1 - \beta^2 - \frac{(2E+1)}{(E+1)^2}\ln(2) + \frac{1}{8}\left(\frac{E}{E+1}\right)^2 \tag{8.95}$$

$$f^{(+)} = 2\ln(2) - \frac{\beta^2}{12}\left[23 + \frac{14}{E+2} + \frac{10}{(E+2)^2} + \frac{4}{(E+2)^3}\right] \tag{8.96}$$

The Fermi density effect is [166, 167]:

$$\delta_F = \frac{1}{Z^{\mathtt{eff}}} \sum_{i=1}^{N_e} \sum_{k=1}^{N_{\mathtt{shells}}} f_i Z_{i,k} \ln\left(1 + \frac{L^2}{W_{i,k}^2}\right) - \frac{L^2}{\Omega_p^2}(1-\beta^2) \tag{8.97}$$

where L is determined by solving:

$$(1-\beta^2) = \frac{\Omega_p^2}{Z^{\mathtt{eff}}} \sum_{i=1}^{N_e} \sum_{k=1}^{N_{\mathtt{shells}}} \frac{f_i Z_{i,k}}{W_{i,k}^2 + L^2} \tag{8.98}$$

The soft stopping power is:

$$S^{\mathtt{e}^{\pm}}(E) = S_t(E) - \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \int_{E-E_c(E)}^{W^{\pm}_{\mathtt{max}}} dW\, (W + U_{i,k}) \sigma^{i,k}_{s,0}(E \to W) \mathcal{H}_b \tag{8.99}$$

The inelastic catastrophic absorption cross-sections for positrons, required for annihilation calculations, are:

$$\Sigma_a^{\mathtt{inel}}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \int_{E-\min\{E_{G+1/2}, E_c\}}^{W^{+}_{\mathtt{max}}} dW\, \sigma^{i,k}_{s,0}(E \to W) \mathcal{H}_b \tag{8.100}$$

The plasma energy is:

$$\Omega_p^2 = 4\pi \mathcal{N}_e^{\mathtt{eff}} \hbar^2 c^2 r_e^2 \tag{8.101}$$

and the resonance energy in the $k^{\mathtt{th}}$-subshell is [168]:

$$W_{i,k} = \sqrt{(aU_{i,k})^2 + \frac{2}{3}\frac{Z_{i,k}}{Z_i}\Omega_p^2} \tag{8.102}$$

where a is determined by solving:

$$Z_i \ln\left(I^{\mathtt{eff}}\right) = \sum_{k=1}^{N_{\mathtt{shells}}} Z_{i,k} \ln\left(\sqrt{(aU_{i,k})^2 + \frac{2}{3}\frac{Z_{i,k}}{Z_i}\Omega_p^2}\right) \tag{8.103}$$

using the Newton-bisection method.

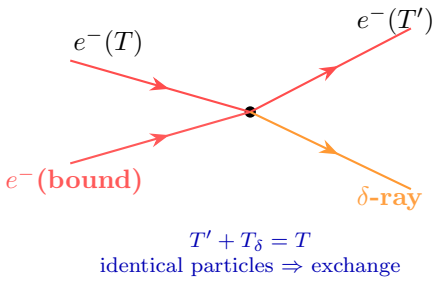


Figure 8.40: **Møller (inelastic) scattering**: Energy transferred to a bound atomic electron (δ-ray). Indistinguishable-particle exchange contributes to the cross section.

Elastic Scattering Elastic interaction of electrons and positrons with matter is of utmost importance for accurate particle distribution modeling up to hundreds of MeV. A combined Molière screened Mott-Rutherford cross-section is employed, based on the interpolation formula of Lijian et al. [169, 170], with the knock-on correction from Kawrakow [171] and Seltzer's adjustment to the Molière screening factor [172].

The Mott cross-section for the elastic interaction of an incoming electron or positron ($p = p' = \mathtt{e}^{\pm}$) with the Coulomb field of an atom is:

$$\sigma_s^i(E,\mu) = \frac{2\pi Z_i(Z_i + \xi_i^{\pm}) r_e^2}{\beta^2 E(E+2)} \frac{1}{(1-\mu+2\eta_i)^2} \mathcal{R}^i_{\mathrm{Mott}} \tag{8.104}$$

The Molière screening factor with Seltzer's adjustment is [156, 173, 172]:

$$\eta_i = \frac{\alpha^2 Z_i^{2/3}\left(1.13 + 3.76\left(\frac{Z_i\alpha}{\beta}\right)^2 \sqrt{\frac{E}{E+1}}\right)}{4\left(\frac{9\pi^2}{128}\right)^{2/3} E(E+2)} \tag{8.105}$$

where $\alpha \approx 1/137$ is the fine structure constant.

The Mott ratio $\mathcal{R}^i_{\text{Mott}}$ uses the interpolation formula of Lijian et al. [169]:

$$\mathcal{R}^i_{\text{Mott}} = \sum_{j=0}^{4} a_{i,j}(1-\mu)^{j/2}, \quad a_{i,j} = \sum_{k=1}^{6} b_{k,j}(Z_i)\left(\beta - \bar{\beta}\right)^{k-1} \tag{8.106}$$

with $\bar{\beta} = 0.7181287$. Parameters $b_{k,j}(Z_i)$ are available for $Z_i \leq 118$ and energies 1 keV to 900 MeV [170].

The Kawrakow correction for knock-on double counting is [171]:

$$\xi^{\pm}_i = 1 - \frac{g^{\pm}_{\texttt{inel},i}}{g_{\texttt{el},i}} \tag{8.107}$$

with:

$$g_{\texttt{el},i} = \frac{\tilde{\sigma}^i_{s,0}(E) - \tilde{\sigma}^i_{s,2}(E)}{3} \tag{8.108}$$

$$g^{\pm}_{\texttt{inel},i} = \sum_{k=1}^{N_{\texttt{shells}}} \frac{Z_{i,k}}{Z_i} E(E+2) \int_{E_{G+1/2}}^{W^{\pm}_{\max}} dW \left(1-\mu_p^2\right) F^{\pm}_{i,k}(E,W)\mathcal{H}_b \tag{8.109}$$

The Legendre moments of the elastic cross-sections are:

$$\Sigma^{\texttt{e}^{\pm}\rightarrow\texttt{e}^{\pm}}_{s,\ell}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i \frac{2\pi Z_i(Z_i+\xi^{\pm}_i) r_e^2}{\beta^2 E(E+2)} \tilde{\sigma}^i_{s,\ell}(E) \tag{8.110}$$

where:

$$\tilde{\sigma}^i_{s,\ell}(E) = \int_{-1}^{1} d\mu\, P_\ell(\mu) \frac{\mathcal{R}^i_{\text{Mott}}}{(1-\mu+2\eta_i)^2} \tag{8.111}$$

These integrals are computed analytically using the Legendre polynomial expansion [174]:

$$P_\ell(\mu) = \frac{1}{2^\ell} \sum_{k=0}^{\lfloor \ell/2 \rfloor} C_{\ell,k}\mu^{\ell-2k}, \quad C_{\ell,k} = \frac{(-1)^k(2\ell-2k)!}{k!(\ell-k)!(\ell-2k)!} \tag{8.112}$$

The Legendre moments can therefore be expressed analytically as:

$$\tilde{\sigma}^i_{s,\ell}(E) = \frac{1}{2^\ell} \sum_{k=0}^{\lfloor \ell/2 \rfloor} C_{\ell,k} \sum_{r=1}^{2} I_r^{\ell,k} \tag{8.113}$$

where $I_1^{\ell,k}$ is given by:

$$I_1^{\ell,k} = \sum_{j=0}^{2} \alpha_{i,j} \left(\mathcal{G}_1^{\ell-2k+j}(1+2\eta_i,-1,1) - \mathcal{G}_1^{\ell-2k+j}(1+2\eta_i,-1,-1) \right) \tag{8.114}$$

and $I_2^{\ell,k}$ is given by:

$$I_2^{\ell,k} = 2\sum_{j=0}^{1} \alpha_{i,j+3} \sum_{g=0}^{\ell-2k+j} (-1)^g \frac{(\ell-2k+j)!}{g!(\ell-2k+j-g)!} \mathcal{G}_2^{2+2g}(2\eta_i, 1, \sqrt{2}) \tag{8.115}$$

with $\alpha_{i,0} = a_{i,0} + a_{i,2} + a_{i,4}$, $\alpha_{i,1} = -(a_{i,2} + 2a_{i,4})$, $\alpha_{i,2} = a_{i,4}$, $\alpha_{i,3} = a_{i,1} + a_{i,3}$ and $\alpha_{i,4} = -a_{i,3}$.

The first integral function from Gradshteyn et al. [175] (Eq. 4, Sect. 2.111) is:

$$\mathcal{G}_1^n(a,b,x) = \sum_{g=1}^{n-1} (-1)^{g-1} \frac{g a^{g-1} x^{n-g}}{(n-g)b^{g+1}} + (-1)^{n-1} \frac{a^n}{b^{n+1}(a+bx)} + (-1)^{n+1} \frac{n a^{n-1}}{b^{n+1}} \ln(a+bx) \tag{8.116}$$

and the second integral function from Gradshteyn et al. (Sect. 2.172–2.174) is:

$$\mathcal{G}_2^n(a,b,x) = \begin{cases} \dfrac{x}{2aR} + \dfrac{1}{2a\sqrt{ab}} \arctan\left(\dfrac{bx}{\sqrt{ab}}\right) & n = 0 \\ -\dfrac{x^{n-1}}{(3-n)bR} + \dfrac{(n-1)a}{(3-n)b}\mathcal{G}_2^{n-2}(a,b,x) & n \text{ even} > 0 \end{cases} \tag{8.117}$$

where $R = a + bx^2$.

This fully analytical solution can be numerically unstable due to catastrophic cancellations for high-order Legendre terms when η_i is large (isotropic scattering regime at low energies). The following stability correction is applied:

$$\textbf{if} \quad \left|\sigma^i_{s,\ell_x}(E)\right| > \left|\sigma^i_{s,0}(E)\right| \quad \textbf{then} \quad \sigma^i_{s,\ell}(E) = 0 \quad \forall\, \ell \geq \ell_x \tag{8.118}$$

The elastic total cross-sections are:

$$\Sigma_t^{\mathtt{e}^\pm}(E) = \Sigma_{s,0}^{\mathtt{e}^\pm \to \mathtt{e}^\pm}(E) \tag{8.119}$$

The transport correction decomposes elastic scattering into soft and catastrophic components [176]. The momentum transfer for the Fokker-Planck operator is:

$$T_g = \frac{\Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,L_{\mathtt{max}}-1,g\to g} - \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,L_{\mathtt{max}},g\to g}}{2L_{\mathtt{max}}} \tag{8.120}$$

$$\Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm,\mathtt{soft}}_{s,0,g\to g} = \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,L_{\mathtt{max}},g\to g} + T_g L_{\mathtt{max}}(L_{\mathtt{max}}+1) \tag{8.121}$$

The Legendre moments of the soft component are given by [177]:

$$\Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm,\mathtt{soft}}_{s,\ell,g\to g} = \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm,\mathtt{soft}}_{s,0,g\to g} - T_g\ell(\ell+1) \tag{8.122}$$

for $\ell \in \{1,\ldots,L_{\mathtt{max}}\}$. The soft component should be withdrawn from cross-sections, since the AFP operator handles soft scattering. The total elastic cross-sections are redefined as:

$$\Sigma_{t,g} \leftarrow \Sigma_{t,g} - \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,L_{\mathtt{max}},g\to g} - T_g L_{\mathtt{max}}(L_{\mathtt{max}}+1) \tag{8.123}$$

and the ℓ-order Legendre moments of the scattering cross-sections are redefined as:

$$\Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,\ell,g\to g} \leftarrow \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,\ell,g\to g} - \Sigma^{\mathtt{e}^\pm\to\mathtt{e}^\pm}_{s,L_{\mathtt{max}},g\to g} - T_g\left[L_{\mathtt{max}}(L_{\mathtt{max}}+1) - \ell(\ell+1)\right] \tag{8.124}$$

This extended transport correction [178, 179, 180] transfers the handling of forward-peaked scattering from the Boltzmann operator to the Fokker-Planck operator, which can be tackled by finite-difference discretization schemes [181]. This amplitude reduction aims to reduce the scattering ratio, which is very close to one in charged particle transport, thereby accelerating the convergence of source iterations [182].

Bremsstrahlung Bremsstrahlung occurs when an incoming electron or positron interacts with the field of the atomic nucleus and its electrons, producing a photon. This interaction is vital at high energies. The tables of Seltzer and Berger [183] are used, which are considered the most accurate bremsstrahlung data available [163]. A modified dipole distribution is used for emitted photons [184, 185].

The bremsstrahlung differential cross-section for monoelemental material i is [183]:

$$\sigma^i_s(E \to E_\gamma) = \begin{cases} \dfrac{F^\pm_{p,i} Z_i^2 \tilde{\sigma}_s(Z_i, E, E_\gamma)}{E_\gamma \beta^2} & E_\gamma \leq E \\ 0 & \text{otherwise} \end{cases} \tag{8.125}$$

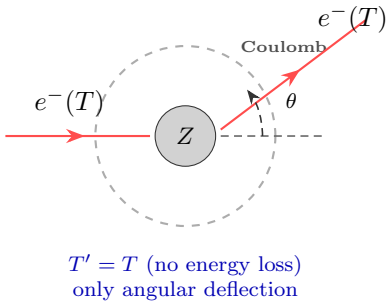


Figure 8.41: **Elastic (Coulomb) scattering**: Energy unchanged, screened Rutherford cross section. Extremely forward-peaked ($\bar{\mu} \approx 0.85$).

where $\tilde{\sigma}_s(Z_i, E, E_\gamma)$ is the scaled energy-weighted cross-section from interpolation of Seltzer-Berger tables, defined for $Z_i \leq 100$ and energies 1 keV to 10 GeV.

The positron factor is [164]:

$$F_{p,i}^{+} = 1 - \exp\left\{-1.2359 \times 10^{-1} t_i + 6.1274 \times 10^{-2} t_i^2 - 3.1516 \times 10^{-2} t_i^3 \right. \\ \left. + 7.7446 \times 10^{-3} t_i^4 - 1.0595 \times 10^{-3} t_i^5 + 7.0568 \times 10^{-5} t_i^6 - 1.8080 \times 10^{-6} t_i^7 \right\} \quad (8.126)$$

where $t_i = \ln\left(1 + \frac{10^6 E}{Z_i^2}\right)$, based on the tabulated positron-to-electron ratios from Kim et al. [186], and $F_{p,i}^{-} = 1$ for electrons.

The double differential cross-sections are [156]:

$$\sigma_s^i(E \to E_e, \mu) = \frac{1}{2\pi} \sigma_s^i(E \to E_e) \delta(\mu - 1) \quad (8.127)$$

$$\sigma_s^i(E \to E_\gamma, \mu) = \frac{1}{2\pi} \sigma_s^i(E \to E_\gamma) \Theta(E, \mu) \quad (8.128)$$

where it is assumed that the incoming electron does not deflect from its trajectory ($\mu_e = 1$).

The photon angular distribution is [184, 187]:

$$\Theta(E, E_\gamma, \mu) = \frac{3(1 - C^2)}{4(2A + B)(1 - \mu C)^2} \left[(A + B) + (A - B) \left(\frac{\mu - C}{1 - \mu C} \right)^2 \right] \quad (8.129)$$

where parameters $A, B, C \in (0, 1)$ are fit to the shape function from Poškus [188] for $E \leq 3$ MeV, and give the dipole distribution ($A = 1$, $B = 0$, $C = \beta$) for $E > 3$ MeV.

The Legendre moments of the angular distribution are computed analytically:

$$\Theta_\ell(E, E_\gamma) = \frac{3(1 - C^2)}{4(2A + B)} \frac{1}{2^\ell} \sum_{k=0}^{\lfloor \ell/2 \rfloor} C_{\ell,k} \left[(A + B) I^{\ell - 2k, 2}(C) + (A - B) \sum_{j=0}^{2} \alpha_j I^{\ell - 2k + j, 4}(C) \right] \quad (8.130)$$

where $\alpha_0 = C^2$, $\alpha_1 = -2C$ and $\alpha_2 = 1$. The remaining integrals are:

$$I^{n,m}(a) = \int_{-1}^{1} d\mu \frac{\mu^n}{(1 - a\mu)^m} = \mathcal{G}_3^{n,m}(a, 1) - \mathcal{G}_3^{n,m}(a, -1) \quad (8.131)$$

solved using integral expressions from Gradshteyn et al. [175] (Sect. 2.153):

$$\mathcal{G}_3^{n,m}(a,x) = \begin{cases} -\dfrac{1}{a}\log|1-ax| & n=0,\ m\neq 1 \\ \dfrac{(1-ax)^{1-m}}{1-m} & n=0,\ m=1 \\ -\dfrac{1}{a(n-m+1)}\left[(1-ax)^{1-m}x^n - n\mathcal{G}_3^{n-1,m}(a,x)\right] & n>0,\ n-m\neq -1 \\ -\dfrac{1}{(1-m)}\left[(1-ax)^{1-m}x^{n+1} - (n-m+2)\mathcal{G}_3^{n,m+1}(a,x)\right] & n>0,\ n-m=-1 \end{cases} \tag{8.132}$$

The multigroup Legendre moments for the scattered lepton are:

$$\Sigma_{s,\ell,g'}^{\mathtt{e}^\pm\to\mathtt{e}^\pm}(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\int_{E'_{g+1/2}}^{\min\left\{E'_{g-1/2},E_c(E)\right\}} dE_e\,\sigma_{s,\ell}^i(E\to E_e)\mathcal{H}_b \tag{8.133}$$

For the produced photon:

$$\Sigma_{s,\ell,g'}^{\mathtt{e}^\pm\to\gamma}(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\int_{E'_{g+1/2}}^{\min\left\{E'_{g-1/2},E\right\}} dE_\gamma\,\sigma_{s,\ell}^i(E\to E_\gamma)\mathcal{H}_b \tag{8.134}$$

The radiative stopping power is:

$$S_t(E) = \alpha r_e^2(E+1)\sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_iF_{p,i}^\pm Z_i^2\phi_{\mathtt{rad}}(Z_i,E) \tag{8.135}$$

where $\phi_{\mathtt{rad}}(Z_i,E)$ is from Seltzer-Berger tables [183].

The catastrophic bremsstrahlung total cross-sections are:

$$\Sigma_t^{\mathtt{e}^\pm}(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\int_0^{E_c(E)} dE_e\,\sigma_{s,0}^i(E\to E_e) \tag{8.136}$$

The bremsstrahlung catastrophic absorption cross-sections, required for annihilation calculations, are:

$$\Sigma_a^{\mathtt{brem}}(E) = \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\int_0^{E_{G+1/2}} dE_e\,\sigma_{s,0}^i(E\to E_e) \tag{8.137}$$

The soft radiative stopping power is:

$$S^{\mathtt{e}^\pm}(E) = S_t(E) - \sum_{i=1}^{N_e}\mathcal{N}_{n,i}f_i\int_0^{E_c(E)} dE_e\,(E-E_e)\sigma_{s,0}^i(E\to E_e) \tag{8.138}$$

Annihilation Annihilation consists of an incoming positron annihilating with an atomic electron to produce two photons. The combined photon energy equals the sum of kinetic and mass energies of the annihilated pair. As the positron kinetic energy approaches zero, this interaction becomes inevitable. The model of Nelson et al. [189] is employed.

The differential cross-section for the lowest energy photon is [190, 148]:

$$\sigma_s(E\to E_{\gamma_-}) = \frac{\pi r_e^2}{(\gamma+1)^2(\gamma^2-1)}\left[S(\zeta)+S(1-\zeta)\right] \tag{8.139}$$

with:

$$S(\zeta) = -(\gamma+1)^2 + (\gamma^2+4\gamma+1)\frac{1}{\zeta} - \frac{1}{\zeta^2} \tag{8.140}$$

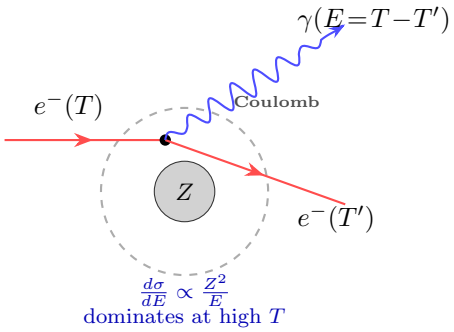


Figure 8.42: **Bremsstrahlung**: Electron decelerates in nuclear Coulomb field, emitting a photon. Cross section scales as Z^2 and dominates at high energies.

where $\zeta = E_{\gamma_-}/(E+2)$ is the ratio of lowest photon energy to total energy. The bounds are:

$$E^{\texttt{max}} = \frac{\gamma+1}{2}, \quad E^{\texttt{min}} = \frac{\gamma+1}{\gamma+1+\sqrt{\gamma^2-1}} \tag{8.141}$$

The scattering angles for lowest and highest energy photons are:

$$\mu_- = \frac{1}{\sqrt{\gamma^2-1}}\left[\gamma+1-\frac{1}{\zeta}\right], \quad \mu_+ = \frac{1}{\sqrt{\gamma^2-1}}\left[\gamma+1-\frac{1}{1-\zeta}\right] \tag{8.142}$$

The multigroup Legendre moments for the lowest energy photons are:

$$\Sigma^{\texttt{e+}\to\gamma}_{s,\ell,g'}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i Z_i \int_{\max\left\{E^\gamma_{g+1/2}, E^{\texttt{min}}\right\}}^{\min\left\{E^\gamma_{g-1/2}, E^{\texttt{max}}\right\}} dE_{\gamma_-}\, \sigma_{s,\ell}(E \to E_{\gamma_-}) \mathcal{H}_b \tag{8.143}$$

The multigroup Legendre moments for the highest energy photons are:

$$\Sigma^{\texttt{e+}\to\gamma}_{s,\ell,g'}(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i Z_i \int_{\max\left\{E^\gamma_{g+1/2}, E^{\texttt{max}}\right\}}^{\min\left\{E^\gamma_{g-1/2}, (\gamma+1)-E^{\texttt{min}}\right\}} dE_{\gamma_+}\, \sigma_{s,\ell}(E \to E_{\gamma_+}) \mathcal{H}_b \tag{8.144}$$

The annihilation total cross-sections are:

$$\Sigma^{\texttt{e+}}_t(E) = \sum_{i=1}^{N_e} \mathcal{N}_{n,i} f_i Z_i \int_{E^{\texttt{min}}}^{E^{\texttt{max}}} dE_{\gamma_-}\, \sigma_{s,0}(E \to E_{\gamma_-}) \tag{8.145}$$

When positrons scatter below the cutoff energy $E_{G+1/2}$, they annihilate producing two 511 keV photons with isotropic distribution [191]:

$$\Sigma^{\texttt{e+}\to\gamma}_{s,0,g'\to g} = 2\left[\Sigma^{\texttt{inel}}_{a,g'} + \Sigma^{\texttt{brem}}_{a,g'}\right] \times \begin{cases} 1 & 1 \in [E'_{g+1/2}, E'_{g-1/2}] \\ 0 & \text{otherwise} \end{cases} \tag{8.146}$$

For photon-initiated pair production with positrons below cutoff energy, the photon-to-photon annihilation contribution is:

$$\Sigma^{\gamma\to\gamma}_{s,0,g'\to g} = 2\Sigma^{\texttt{pp}}_{a,g'} \times \begin{cases} 1 & 1 \in [E'_{g+1/2}, E'_{g-1/2}] \\ 0 & \text{otherwise} \end{cases} \tag{8.147}$$

where the $\ell \geq 1$ moments are zero since scattering is isotropic.

Atomic Relaxation The ionization of an atomic electron in shell k by an incoming particle leaves a vacancy in the atom's electronic structure. This vacancy is filled by an outer-shell electron, leading to the emission of a fluorescence photon or ejection of an Auger electron. This process triggers additional vacancies, resulting in a relaxation cascade [156, 192]. Only transitions with probability $> 0.1\%$ and producing particles above the cutoff energy are included.

The differential cross-section for production of fluorescence ($p' = \gamma$) or Auger electron ($p' =$ `e-`) following cascade transition j is:

$$\sigma_s^{i,k,j,p\to p'}(E \to E') = \eta_{i,k,j}^{p'} \delta(E' - \Delta E_{i,k,j}^{p'}) \sigma_t^{i,k,p}(E) \tag{8.148}$$

where $\sigma_t^{i,k,p}(E)$ is the k-shell cross-section for the initiating interaction, $\Delta E_{i,k,j}^{p'}$ is the produced particle energy, and $\eta_{i,k,j}^{p'}$ is the cascade probability from JENDL-5 [146] based on EADL [193], as implemented by Hébert and Naceur [194]. Data is available for $Z \in [6, 100]$ for subshells K, L1–L3, M1–M5, N1–N7, O1–O7, P1–P3, and Q1. Production of fluorescence and Auger radiation is assumed isotropic.

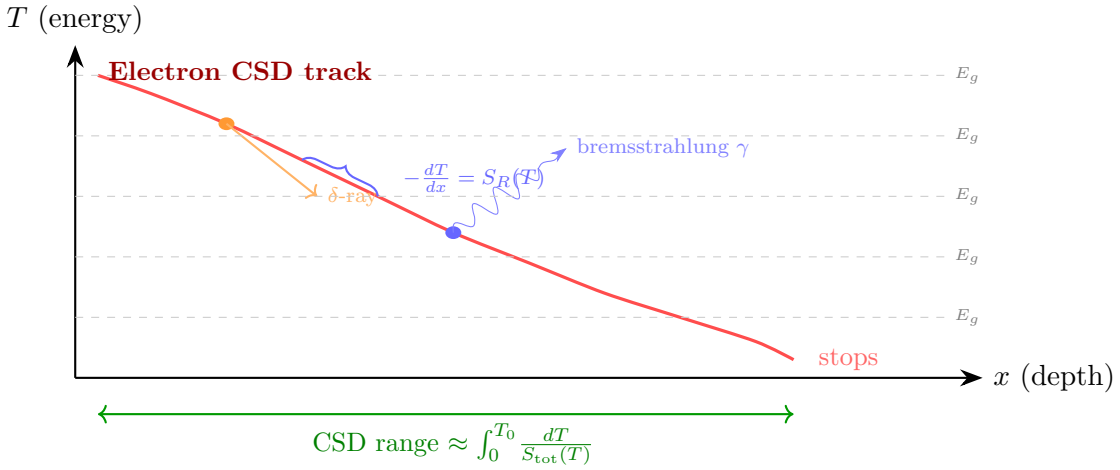


Figure 8.43: Conceptual diagram of the Continuously Slowing Down (CSD) approximation. The electron loses energy continuously along its track (governed by the restricted stopping power S_R), with occasional catastrophic events (Møller δ-rays, bremsstrahlung). Horizontal dashed lines indicate multigroup energy boundaries; the total range is given by integrating the reciprocal of the total stopping power.

Polynomials and Quadratures Mathematical Framework This section explains the mathematical principles for angular discretization and calculation of cross-section moments in Radiant.jl.

Legendre Polynomials The Legendre polynomials, denoted $P_\ell(x)$ for $\ell \geq 0$, are a complete and orthogonal system defined over $x \in [-1, 1]$ satisfying:

$$\int_{-1}^{1} dx\, P_\ell(x) P_{\ell'}(x) = \frac{2}{2\ell + 1} \delta_{\ell,\ell'} \tag{8.149}$$

where $\delta_{\ell,\ell'}$ is the Kronecker delta. They are also eigenfunctions of the Legendre differential equation:

$$\frac{d}{dx}\left[(1 - x^2)\frac{d}{dx}\right] P_\ell(x) = -\ell(\ell + 1) P_\ell(x) \tag{8.150}$$

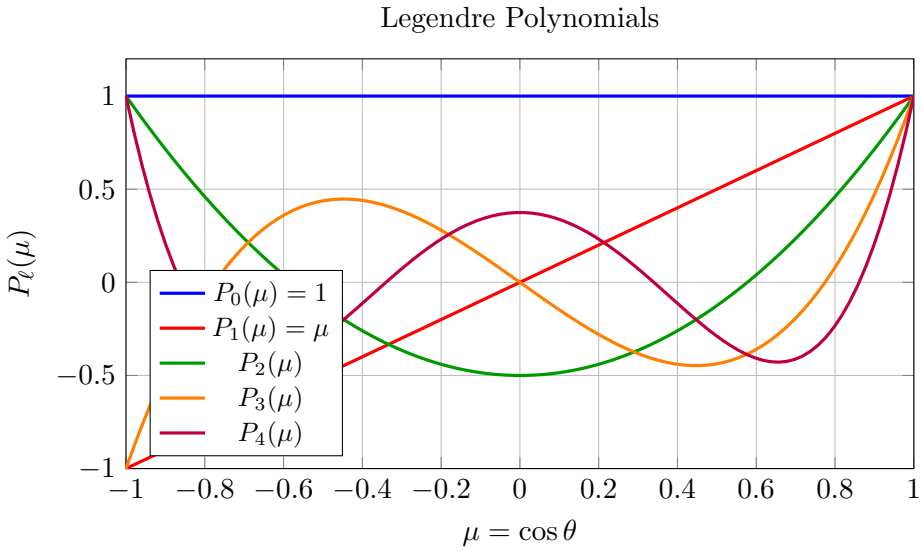


Figure 8.44: Legendre polynomials $P_\ell(\mu)$ for $\ell = 0$ to 4. Higher orders capture increasingly rapid angular variations.

The Legendre polynomials are computed using Bonnet's recursion formula:

$$P_\ell(x) = \begin{cases} 1 & \ell = 0 \\ x & \ell = 1 \\ \left[\dfrac{2\ell-1}{\ell}\right] x P_{\ell-1}(x) - \left[\dfrac{\ell-1}{\ell}\right] P_{\ell-2}(x) & \ell \geq 2 \end{cases} \tag{8.151}$$

The Legendre polynomials are bounded by [195]:

$$|P_\ell(x)| \leq 1 \tag{8.152}$$

The ℓ-order Legendre moments of a function $f(x)$ are defined as:

$$f_\ell = \int_{-1}^{1} dx\, P_\ell(x) f(x) \tag{8.153}$$

For a positive function $f(x)$, using the triangle inequality and the Legendre polynomial upper bound, the following inequality holds for $\ell \geq 1$:

$$\frac{|f_\ell|}{f_0} = \frac{\left|\int_{-1}^{1} dx\, P_\ell(x) f(x)\right|}{f_0} \leq \frac{\int_{-1}^{1} dx\, |P_\ell(x)|\, f(x)}{f_0} \leq 1 \tag{8.154}$$

Application to Scattering Cross Sections. The differential scattering cross section is expanded in Legendre polynomials:

$$\Sigma_s(g' \to g, \mu_0) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi}\, \Sigma_{s\ell}(g' \to g)\, P_\ell(\mu_0) \tag{8.155}$$

where $\mu_0 = \hat{\Omega}' \cdot \hat{\Omega}$ is the scattering cosine.

Storage Format. The Legendre moments are stored in three-dimensional arrays: $(G_\gamma \times G_\gamma \times L_\gamma)$ with $L_\gamma = 4$ for photon-to-photon scattering, $(G_e \times G_e \times L_e)$ with $L_e = 8$ for electron-to-electron scattering, and $(G_\gamma \times G_e \times L_{\gamma e})$ with $L_{\gamma e} = 4$ for photon-to-electron coupling (see Figs. 8.45 and 8.46).

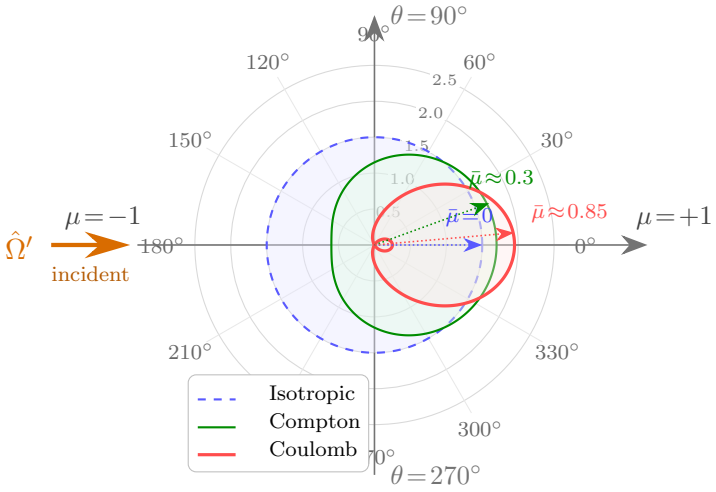


Figure 8.45: Polar diagram of Legendre-expanded differential scattering cross section $f(\mu_0) = \sum_\ell (2\ell{+}1)/(4\pi)\sigma_{s,\ell} P_\ell(\mu_0)$. **Dashed blue**: isotropic ($\ell{=}0$). **Green**: photon Compton ($\bar{\mu}^\gamma{\approx}0.3$, $L_\gamma{=}4$). **Red**: electron Coulomb ($\bar{\mu}^e{\approx}0.85$, $L_e{=}8$).

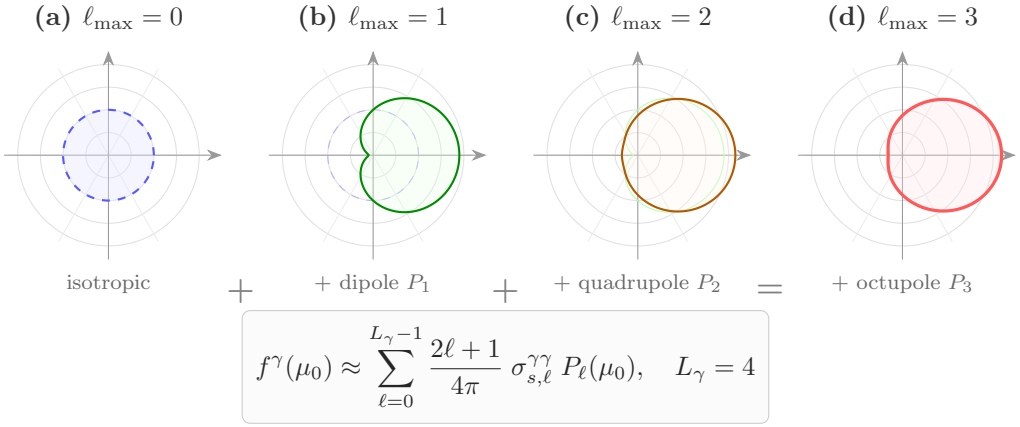


Figure 8.46: Progressive Legendre expansion convergence for photon Compton scattering. **(a)** P_0: isotropic circle. **(b)** $+P_1$: forward lobe emerges. **(c)** $+P_2$: forward peak sharpens. **(d)** Full $L_\gamma{=}4$ expansion: converged shape.

Normalized Legendre Polynomials The normalized Legendre polynomials, denoted $\tilde{P}_\ell(x)$ for $\ell \geq 0$, are a complete and orthogonal system defined over $x \in [-1/2, 1/2]$ satisfying:

$$\int_{-1/2}^{1/2} dx\, \tilde{P}_\ell(x)\tilde{P}_{\ell'}(x) = \delta_{\ell,\ell'} \tag{8.156}$$

They are computed using the following recursion formula [196]:

$$\tilde{P}_\ell(x) = \begin{cases} 1 & \ell = 0 \\ 2\sqrt{3}x & \ell = 1 \\ 2\sqrt{\dfrac{2\ell+1}{2\ell-1}}\left[\dfrac{2\ell-1}{\ell}\right] x\tilde{P}_{\ell-1}(x) - \sqrt{\dfrac{2\ell+1}{2\ell-3}}\left[\dfrac{\ell-1}{\ell}\right]\tilde{P}_{\ell-2}(x) & \ell \geq 2 \end{cases} \tag{8.157}$$

An upper bound for the normalized Legendre polynomial is:

$$\left|\tilde{P}_\ell(x)\right| \leq \sqrt{2\ell+1} \tag{8.158}$$

For a positive function $f(x)$, the normalized Legendre moment inequality for $\ell \geq 1$ is:

$$\frac{|f_\ell|}{f_0} \leq \sqrt{2\ell+1} \tag{8.159}$$

Associated Legendre Polynomials The associated Legendre polynomials (Ferrer definition, without the $(-1)^m$ factor) are:

$$P_\ell^m(x) = (1-x^2)^{m/2}\frac{d^m}{dx^m}P_\ell(x), \quad 0 \leq m \leq \ell \tag{8.160}$$

For numerical computation, the following identity is used [197]:

$$P_m^m(x) = (1-x^2)^{m/2}(2m-1)!! = (1-x^2)^{m/2}\prod_{k=1}^{m}(2k-1) \tag{8.161}$$

If $\ell = m+1$:

$$P_{m+1}^m(x) = (2m+1)xP_m^m(x) \tag{8.162}$$

For all other cases with $\ell > m+1$:

$$P_\ell^m(x) = \frac{(2\ell-1)xP_{\ell-1}^m(x) - (\ell+m-1)P_{\ell-2}^m(x)}{\ell-m} \tag{8.163}$$

Real Spherical Harmonics The real spherical harmonics, denoted $R_\ell^m(\mu,\phi)$ for $\ell \geq 0$ and $|m| \leq \ell$, are a complete and orthogonal system defined over $\mu \in [-1,1]$ and $\phi \in [0, 2\pi]$ satisfying:

$$\int_{4\pi} d^2\Omega\, R_\ell^m(\vec{\Omega})R_{\ell'}^{m'}(\vec{\Omega}) = \frac{4\pi}{2\ell+1}\delta_{\ell,\ell'}\delta_{m,m'} \tag{8.164}$$

They are eigenfunctions of the Laplace equation:

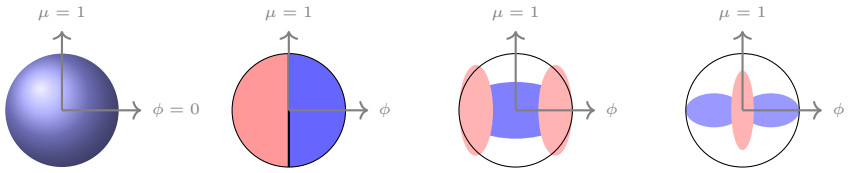
$$\left[\frac{\partial}{\partial\mu}\left[(1-\mu^2)\frac{\partial}{\partial\mu}\right] + \frac{1}{1-\mu^2}\frac{\partial^2}{\partial\phi^2}\right]R_\ell^m = -\ell(\ell+1)R_\ell^m \tag{8.165}$$

The real spherical harmonics are given by [180]:

$$R_\ell^m(\vec{\Omega}) = \sqrt{(2-\delta_{m,0})\frac{(\ell-|m|)!}{(\ell+|m|)!}}P_\ell^{|m|}(\mu)\mathcal{T}_m(\phi) \tag{8.166}$$

with:

$$\mathcal{T}_m(\phi) = \begin{cases} \cos m\phi, & m \geq 0 \\ \sin|m|\,\phi, & m < 0 \end{cases} \tag{8.167}$$



R_0^0 (isotropic) R_1^0 (dipole) R_2^0 (quadrupole) R_3^0 (octupole)

Figure 8.47: Real spherical harmonics $R_\ell^0(\mu, \phi)$ for $m = 0$. Blue/red indicate positive/negative regions. The polar axis ($\mu = 1$) points upward; higher ℓ captures finer angular structure.

Gauss-Legendre Quadrature The N-order Gauss-Legendre quadrature is defined over $x \in [-1, 1]$ and gives exact results for polynomials of degree $2N - 1$ or less. The i-th node x_i corresponds to the i-th root of $P_N(x)$, found using Newton iteration with the initial estimate of Tricomi [198, 199]:

$$x_i^0 = \left[1 - \frac{N-1}{8N^3} - \frac{1}{384N^4}\left(39 - \frac{28}{\sin^2(\phi_i)}\right)\right]\cos(\phi_i) \tag{8.168}$$

with:

$$\phi_i = \frac{\pi}{2}\left(\frac{4i-1}{2N+1}\right) \tag{8.169}$$

The derivative of the Legendre polynomial needed for Newton iteration is:

$$P'_N(x) = \frac{N\left(P_{N-1}(x) - xP_N(x)\right)}{1-x^2} \tag{8.170}$$

The weights are:

$$\omega_i = \frac{2}{\left(1-x_i^2\right)\left(P'_N(x_i)\right)^2} \tag{8.171}$$

Gauss-Lobatto Quadrature The N-order Gauss-Lobatto quadrature is defined over $x \in [-1, 1]$ and gives accurate results for polynomials of degree $2N - 3$ or less. The i-th node x_i corresponds to the $(i-1)$-th root of $P'_{N-1}(x)$, with $x_1 = -1$ and $x_N = 1$. A good initial value is:

$$x_i = \cos\left((i-1)\frac{\pi}{N-1}\right) \tag{8.172}$$

The second derivative of the Legendre polynomial is:

$$P''_N(x) = \frac{2xP'_N(x) - N(N+1)P_N(x)}{1-x^2} \tag{8.173}$$

The weights are:

$$\omega_i = \begin{cases} \dfrac{2}{N(N-1)\left(P_{N-1}(x_i)\right)^2} & x_i \neq \pm 1 \\ \dfrac{2}{N(N-1)} & \text{otherwise} \end{cases} \tag{8.174}$$

Product Quadrature with Chebyshev Azimuthal Quadrature The N-order product quadrature with Chebyshev azimuthal quadrature is defined over the unit sphere. This method is symmetric over 8 octants and has no evaluation points along the reference frame axes. The quadrature has N levels with $2N$ points each, for a total of:

$$N_d = 2N^2 \tag{8.175}$$

For level $n \in \{1, \ldots, N\}$ and point $m \in \{1, \ldots, 2N\}$ on that level, the direction cosines are:

$$\mu_i = \mu_n \tag{8.176}$$

$$\eta_i = \sqrt{1 - \mu_n^2} \cos(\phi_m) \tag{8.177}$$

$$\xi_i = \sqrt{1 - \mu_n^2} \sin(\phi_m) \tag{8.178}$$

where:

$$\phi_m = \frac{(2m-1)\pi}{2N} \tag{8.179}$$

and the corresponding weight is:

$$w_i = w_n \frac{\pi}{N} \tag{8.180}$$

Carlson Angular Quadrature The Carlson quadrature [200] is defined over the unit sphere and requires fewer evaluation points than product quadrature. This method is symmetric over 8 octants with no evaluation points along the reference frame axes. In the octant with positive cosines, the N-order weights are:

$$\omega_m = \frac{4(N - 2n + 2)}{N(N+2)} \tag{8.181}$$

with $m \in \{1, \ldots, N/2\}$. The main direction cosine is:

$$\mu_m = \bar{\mu}_m + f\mu_{m-1/2} \tag{8.182}$$

with:

$$\bar{\mu}_m = 1 - \frac{(N - 2n + 2)^2}{N(N+2)} \tag{8.183}$$

and:

$$\mu_{m-1/2} = 1 - \frac{(N - 2n + 2)(N - 2n + 4)}{N(N+2)} \tag{8.184}$$

The factor f is determined by solving:

$$\sum_{m=1}^{N/2} \omega_m \mu_m^2 = \frac{1}{3} \tag{8.185}$$

using the Newton-bisection method. The two other direction cosines are:

$$\eta_{m,k} = \sqrt{1 - \mu_m^2} \sin\theta_{m,k}, \quad \xi_{m,k} = \sqrt{1 - \mu_m^2} \cos\theta_{m,k} \tag{8.186}$$

where:

$$\theta_{m,k} = \frac{\pi}{2}\left[\frac{2m-1}{N - 2k + 2} A_n + \frac{1 - A_n}{2}\right] \tag{8.187}$$

with $k \in \{1, \ldots, N/2 - m + 1\}$. The A_n factors are determined by solving:

$$\omega_0 \sum_{m=1}^{N/2} \sum_{k=1}^{N/2-m+1} \eta_{m,k} = \sum_{m=1}^{N/2} \omega_m \mu_m \tag{8.188}$$

using the Newton-bisection method.

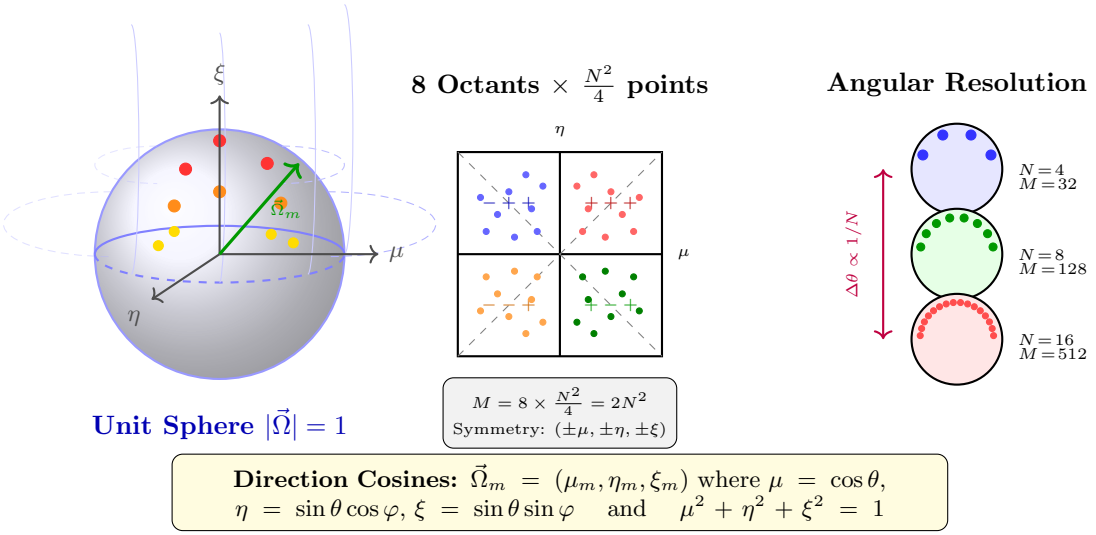


Figure 8.48: S_N discrete ordinates angular quadrature with STL points. **Left:** Unit sphere with quadrature points at different polar levels and the direction vector $\vec{\Omega}_m$. **Center:** Octant decomposition—points are reflected to all 8 octants by sign symmetry ($M = 2N^2$ directions). **Right:** Angular resolution comparison for various N orders.

8.11.3 Material Library: 44 Materials

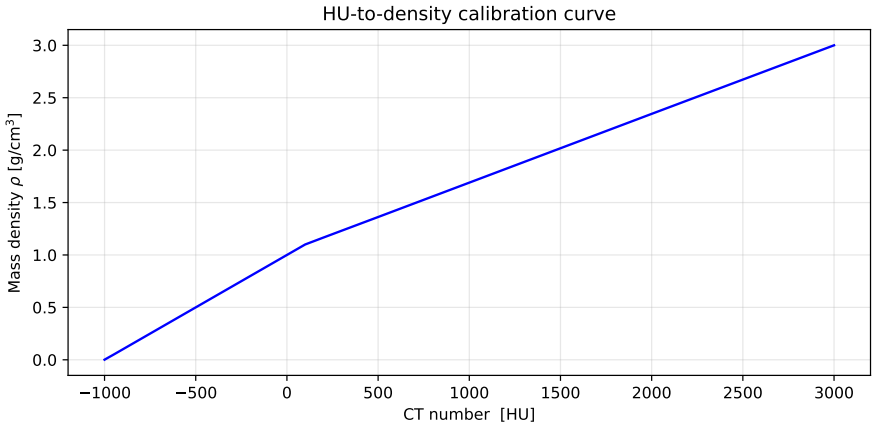


Figure 8.49: Piecewise-linear HU-to-density calibration curve used for CT→ ρ mapping.

HU-to-Density Calibration

Material Density Thresholds The 44 materials span from air ($\rho = 0.00121\,\mathrm{g/cm^3}$) to stainless steel ($\rho = 8.0\,\mathrm{g/cm^3}$). Assignment uses **overlap-linear** interpolation between adjacent material bins.

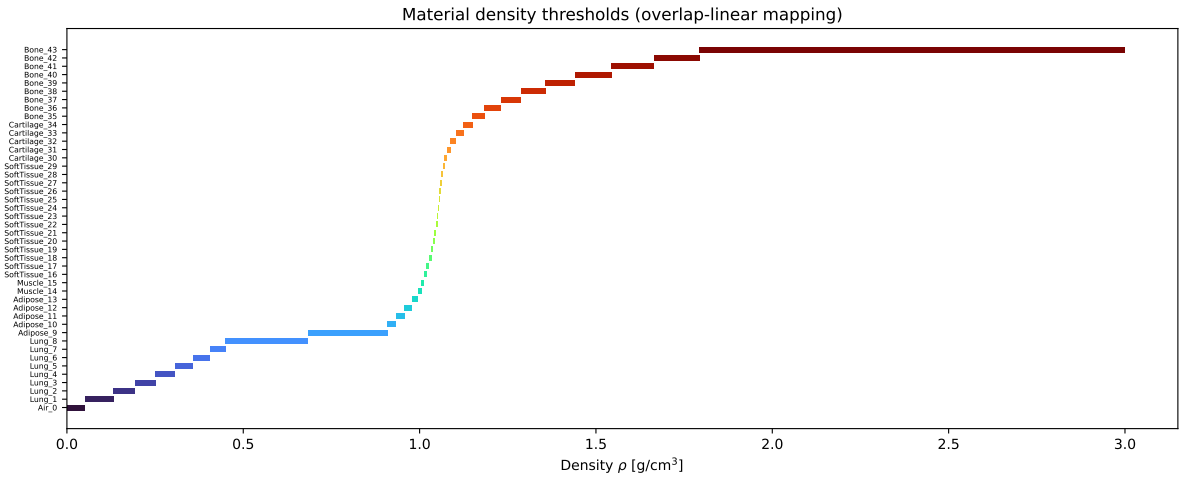


Figure 8.50: Density ranges for all 44 materials. Overlap-linear mapping blends cross sections in transition zones.

The overlap-linear mixing formula for a voxel with density ρ falling between materials m and $m+1$:

$$\Sigma_t(\vec{r}) = w\,\Sigma_t^{(m)} + (1-w)\,\Sigma_t^{(m+1)}, \qquad w = \frac{\rho_{\max}^{(m)} - \rho}{\rho_{\max}^{(m)} - \rho_{\min}^{(m+1)}} \tag{8.189}$$

8.11.4 Cross-Section Visualizations

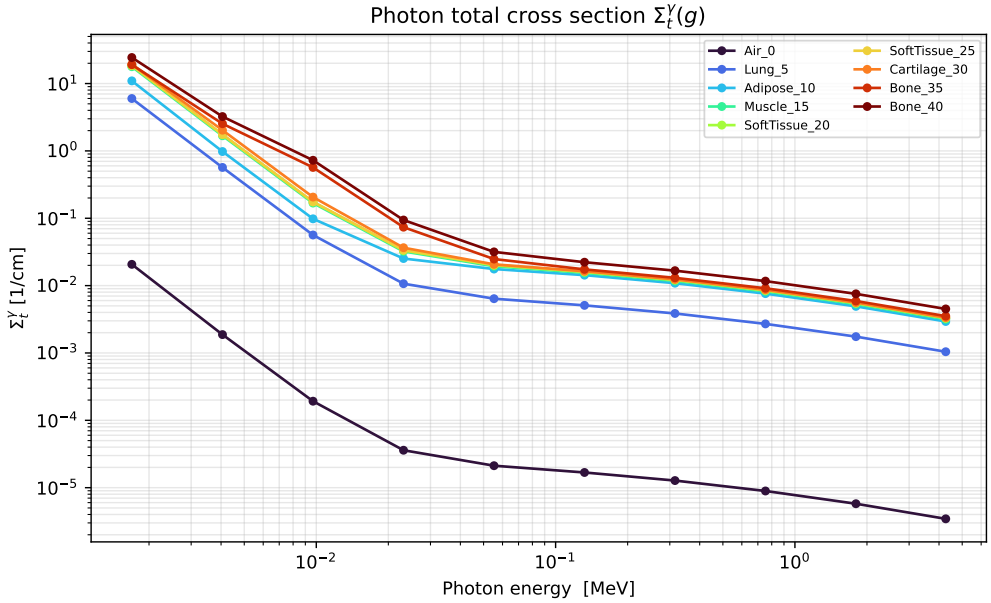


Figure 8.51: Photon total macroscopic cross section $\Sigma_t^\gamma(g)$ for representative materials. Low-energy groups dominated by photoelectric absorption; high-energy by Compton scattering. High-Z materials (titanium, steel) show dramatically higher Σ_t^γ.

Photon Total Cross Section

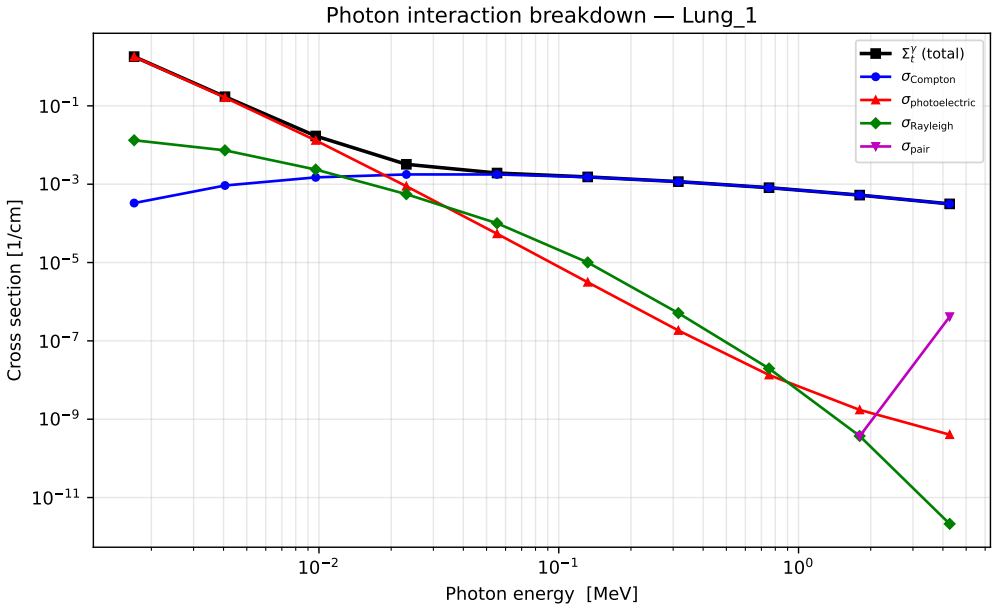


Figure 8.52: Decomposition of Σ_t^γ into individual interaction channels for a tissue-like material. The crossover from photoelectric dominance to Compton dominance occurs near 30–80 keV. Pair production contributes only at the highest energies (>1 MeV).

Photon Interaction Breakdown

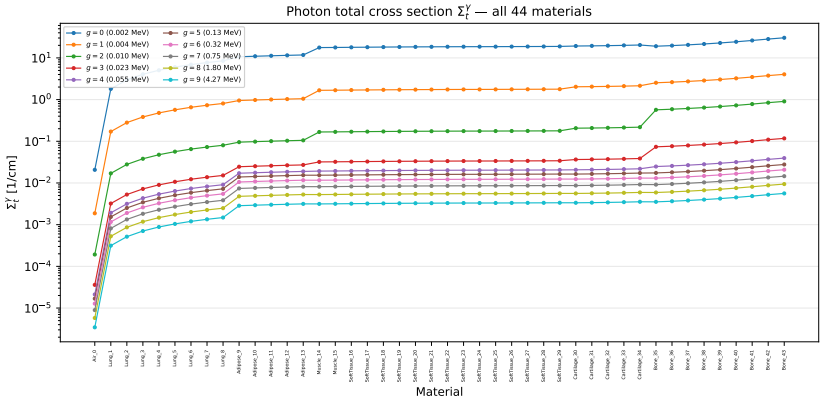


Figure 8.53: Heatmap of $\log_{10} \Sigma_t^\gamma$ across all 44 materials (rows) and 10 photon energy groups (columns). The dramatic increase for high-Z materials at low energies is clearly visible.

All 44 Materials: Photon Σ_t^γ Heatmap

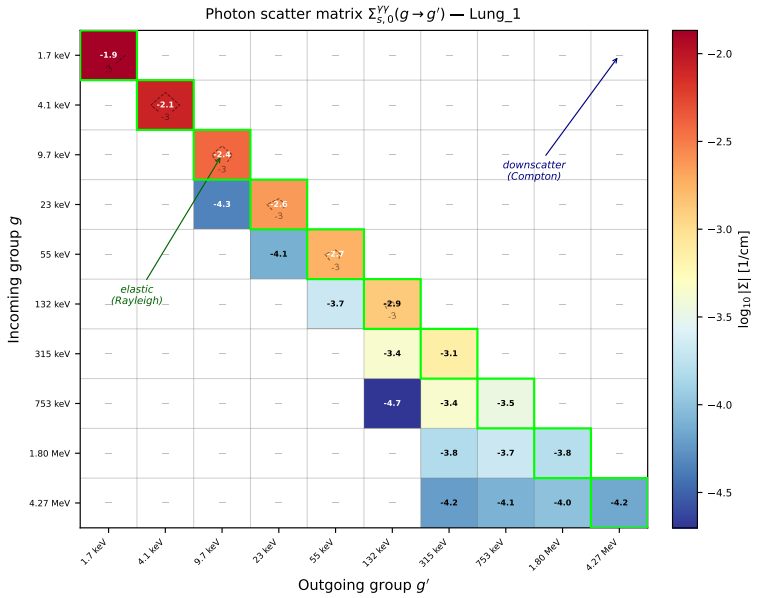


Figure 8.54: Photon-photon scattering matrix (zeroth Legendre moment $\ell = 0$) on a log scale. The dominant diagonal reflects elastic (Rayleigh) and within-group Compton scatter. Off-diagonal elements represent Compton downscatter.

Photon Scattering Transfer Matrix

Photon Scattering: Legendre Moment Analysis A key physics feature of the Acuros XB implementation is the full Legendre expansion of the differential scattering cross section, rather than the isotropic ($\ell = 0$ only) approximation common in simpler solvers. The anisotropy is characterised by the ratios $\sigma_{s,\ell}/\sigma_{s,0}$ and the mean scattering cosine:

$$\bar{\mu}^{\gamma}(g) = \frac{\sum_{g'} \Sigma_{s1}^{\gamma\gamma}(g \to g')}{\sum_{g'} \Sigma_{s0}^{\gamma\gamma}(g \to g')} \tag{8.190}$$

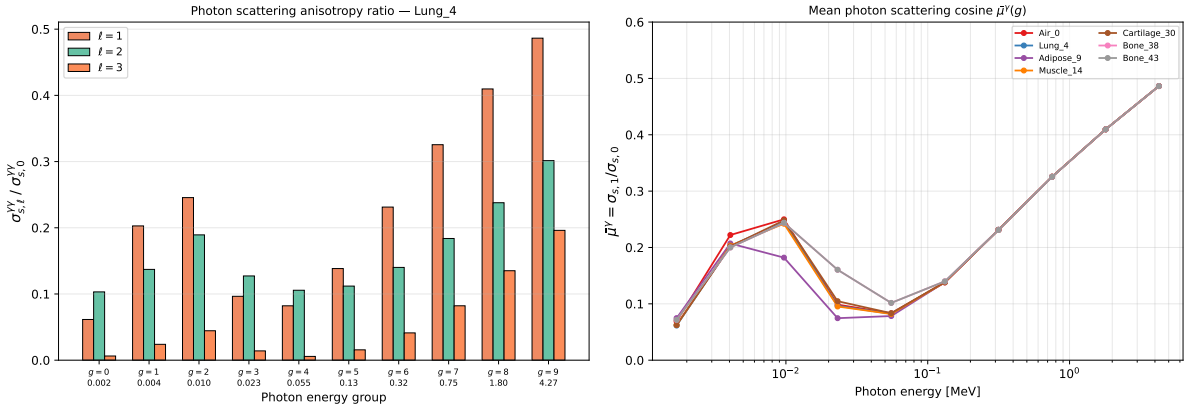


Figure 8.55: **Left:** Ratio $\sigma^{\gamma\gamma}_{s,\ell}/\sigma^{\gamma\gamma}_{s,0}$ for $\ell = 1, 2, 3$ per photon energy group (Lung_4). Higher-energy groups exhibit larger $\ell = 1$ components, reflecting forward-peaked Compton scattering. **Right:** Mean scattering cosine $\bar{\mu}^{\gamma}$ vs energy for several materials.

The differential scattering cross section is reconstructed from its Legendre expansion:

$$f^{\gamma}(g,\mu) = \sum_{\ell=0}^{L_{\gamma}-1} \frac{2\ell+1}{4\pi}\, \sigma_{s,\ell}^{\gamma}(g)\, P_{\ell}(\mu) \tag{8.191}$$

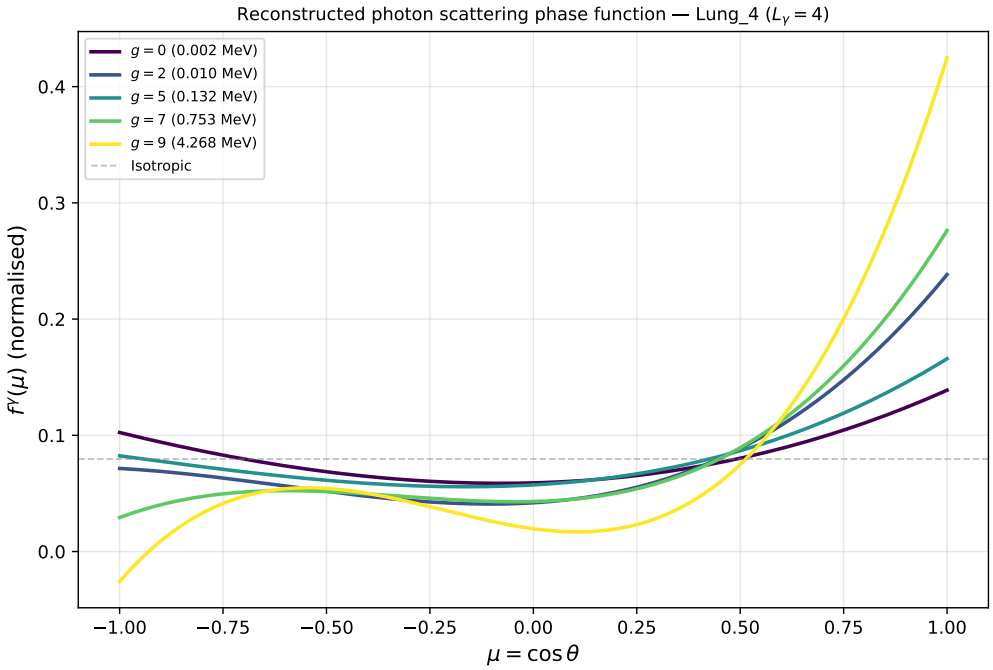


Figure 8.56: Reconstructed photon scattering phase function $f^{\gamma}(\mu)$ for selected energy groups (Lung_4, $L_{\gamma} = 4$). Low-energy groups are nearly isotropic; higher-energy groups show pronounced forward peaking consistent with Compton kinematics.

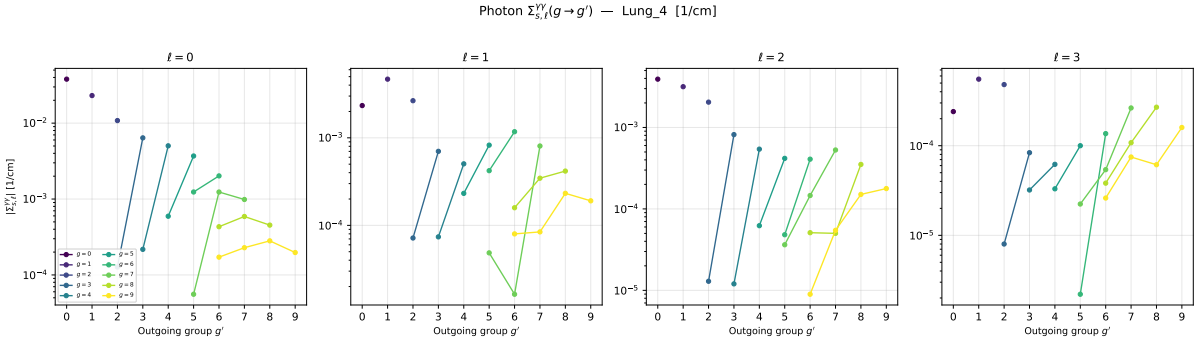


Figure 8.57: Photon scattering matrix $\Sigma_{s\ell}^{\gamma\gamma}(g \to g')$ shown for all four Legendre orders $\ell = 0, 1, 2, 3$ (Lung_4). The $\ell = 0$ panel reproduces the standard isotropic matrix; higher orders reveal the angular structure. Note that the $\ell = 3$ moments are 1–2 orders of magnitude smaller than $\ell = 0$, confirming convergence.

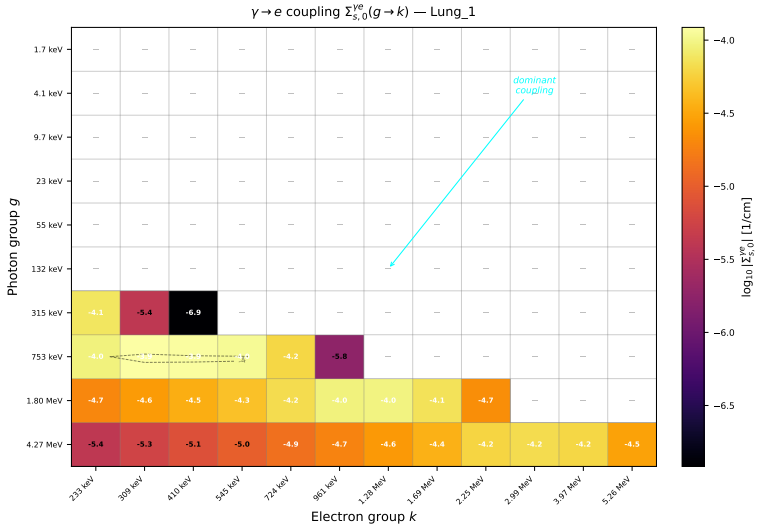


Figure 8.58: Photon-to-electron coupling matrix, zeroth moment $\Sigma_{s0}^{\gamma e}(g \to k)$ on a log scale. This matrix drives the energy transfer from photon to secondary-electron kinetic energy through Compton, photoelectric, and pair-production channels.

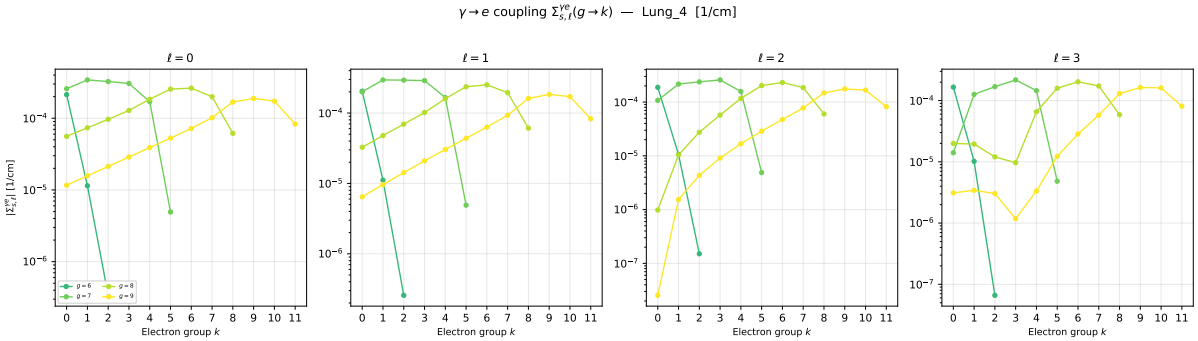


Figure 8.59: Photon→electron coupling $\Sigma_{s\ell}^{\gamma e}(g \to k)$ for all four orders $\ell = 0, 1, 2, 3$ (Lung_4). The $\ell = 0$ panel shows the standard energy-transfer matrix. Higher-order panels reveal that the $\gamma \to e$ angular distribution is more isotropic than photon-photon scattering.

Photon-to-Electron Coupling: Legendre Moments

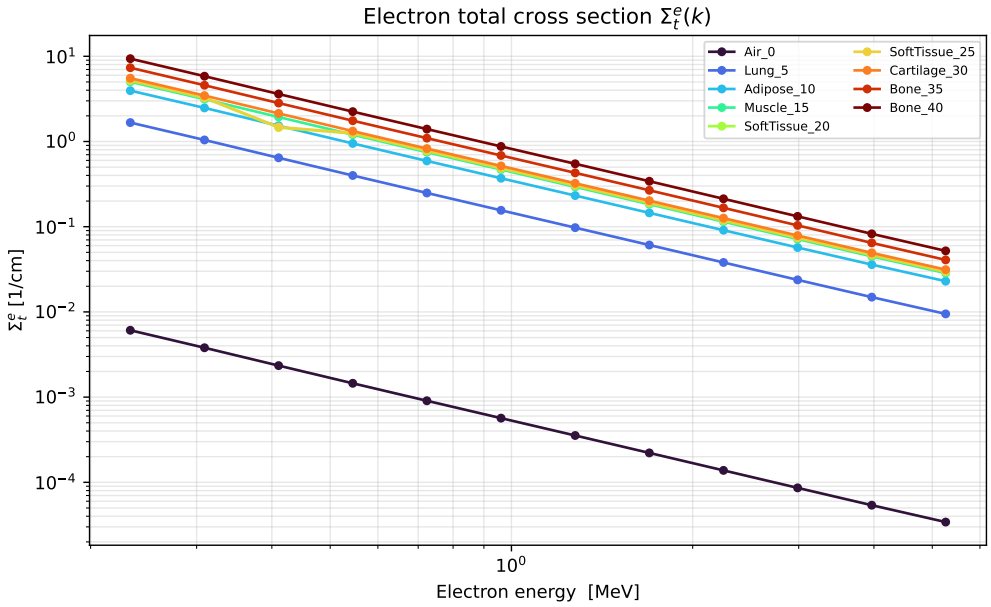


Figure 8.60: Electron total macroscopic cross section $\Sigma_t^e(k)$. Shows characteristic $1/E$ dependence. Dense materials have proportionally higher cross sections due to density scaling.

Electron Total Cross Section

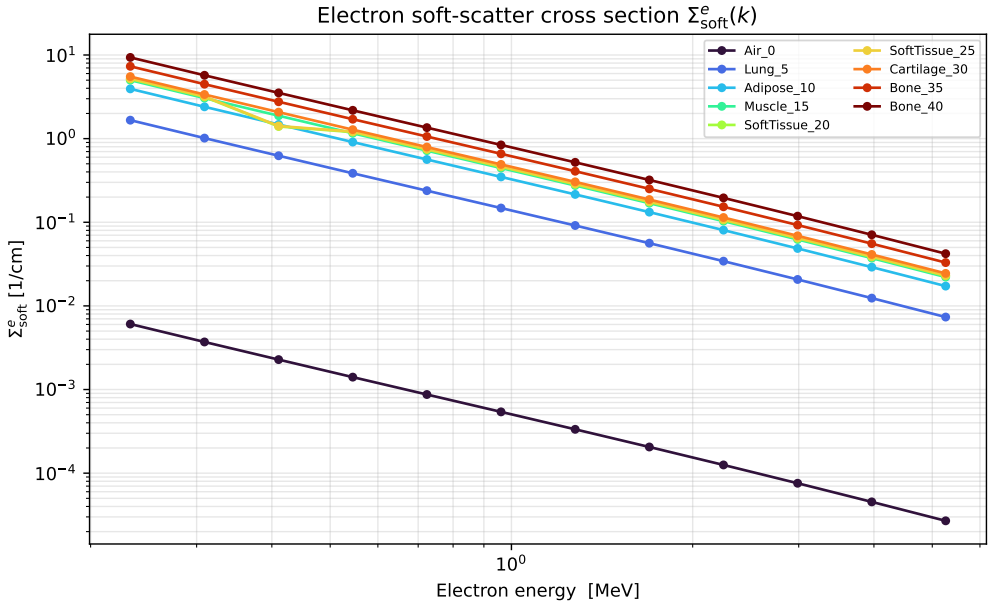


Figure 8.61: Electron soft-scatter cross section $\Sigma_{\text{soft}}^e(k)$. This component represents small-angle elastic scattering and is used in the Fokker–Planck angular diffusion treatment.

Electron Soft-Scatter Cross Section

Electron Scattering: Legendre Moment Analysis Electron–electron scattering is extremely forward-peaked due to the long-range Coulomb interaction. This demands a higher Legendre order ($L_e = 8$) than photon scattering ($L_\gamma = 4$). The mean scattering cosine for

electrons is:

$$\bar{\mu}^e(k) = \frac{\sum_{k'} \Sigma_{s1}^{ee}(k \to k')}{\sum_{k'} \Sigma_{s0}^{ee}(k \to k')} \approx 0.84\text{–}0.85 \tag{8.192}$$

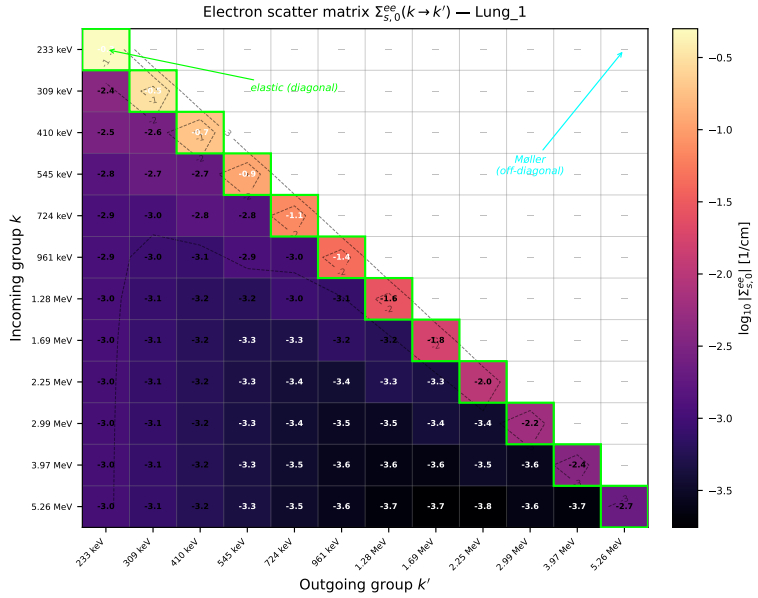


Figure 8.62: Electron-electron scattering matrix ($\ell = 0$) on a log scale. The strong diagonal represents elastic scattering; off-diagonal elements represent inelastic (Møller) energy transfer.

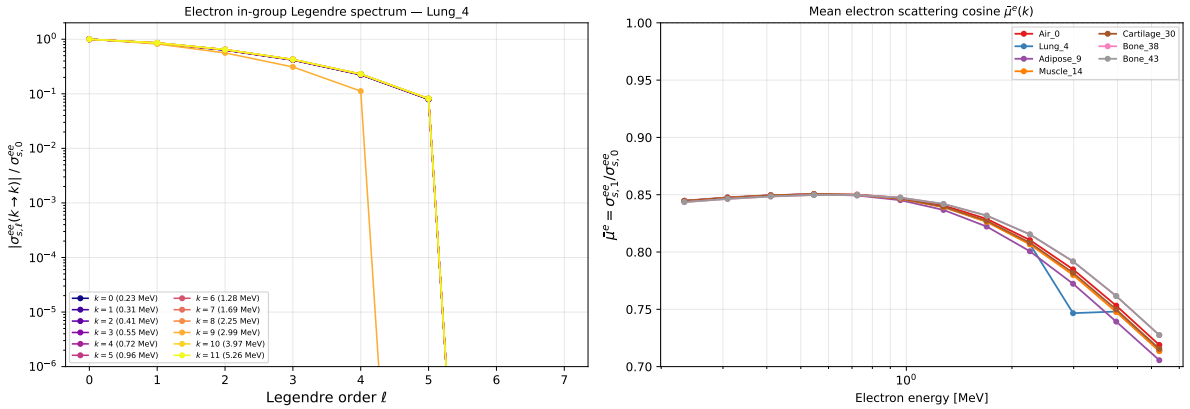


Figure 8.63: **Left:** In-group Legendre spectrum $|\sigma_{s,\ell}^{ee}(k \to k)|/\sigma_{s,0}^{ee}$ vs ℓ for each electron energy group (Lung_4). The moments decay slowly—even $\ell = 6$ retains ∼10% of $\ell = 0$—confirming the need for 8 Legendre orders. **Right:** Mean electron scattering cosine $\bar{\mu}^e(k)$ across materials.

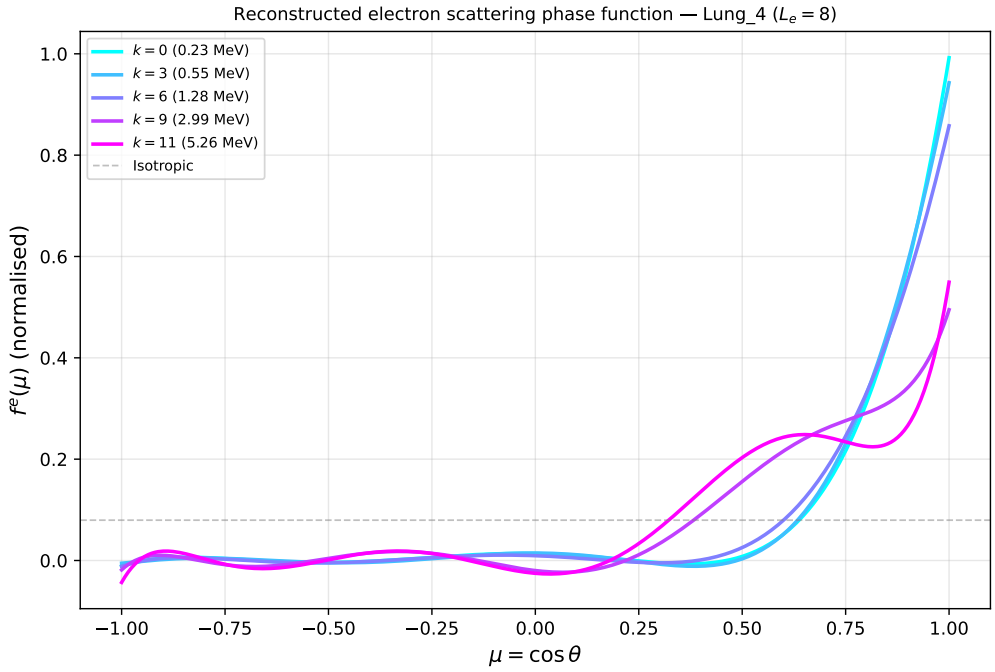


Figure 8.64: Reconstructed electron scattering phase function $f^e(\mu)$ for selected energy groups (Lung_4, $L_e = 8$). All groups exhibit pronounced forward peaking with $f^e(+1) \gg f^e(-1)$, characteristic of small-angle Coulomb scattering. Compare with the isotropic reference (dashed).

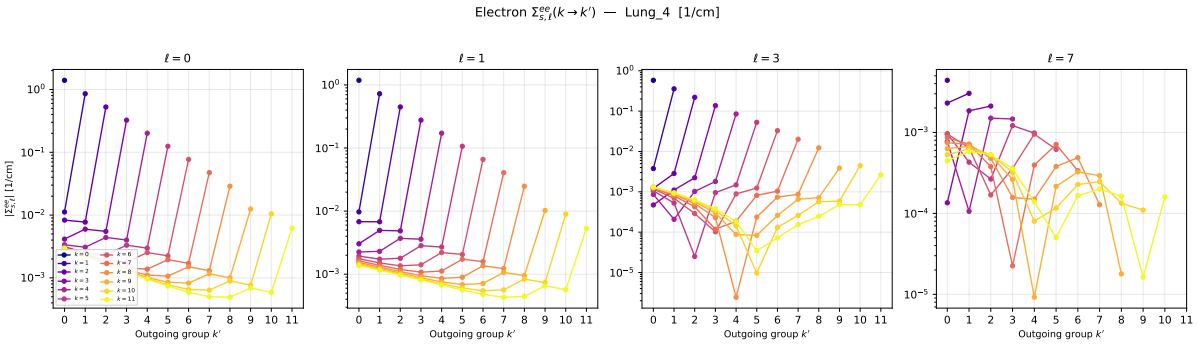


Figure 8.65: Electron scattering matrix $\Sigma^{ee}_{s\ell}(k \to k')$ for selected Legendre orders $\ell = 0, 1, 3, 7$ (Lung_4). The $\ell = 7$ panel is zero, providing a clean truncation. The CSD structure is visible as the strong near-diagonal band.

Transport Correction Analysis The transport cross section $\Sigma_t^{tr} = \Sigma_t - \Sigma_{s1}$ quantifies the effective attenuation after accounting for forward scattering anisotropy:

$$\Sigma_{tr}(g) = \Sigma_t(g) - \sum_{g'} \Sigma_{s1}(g \to g') \tag{8.193}$$

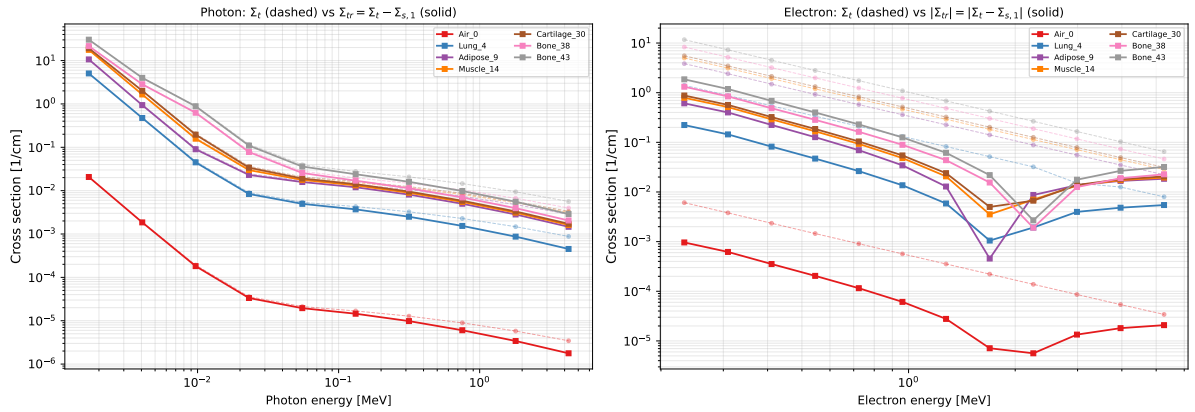


Figure 8.66: **Left:** Photon total Σ_t^γ (dashed) vs transport Σ_{tr}^γ (solid) for representative materials. The small separation confirms modest photon anisotropy. **Right:** Electron counterpart. The large gap between Σ_t^e and $|\Sigma_{tr}^e|$ reflects the extreme forward-peaking ($\bar{\mu}^e \approx 0.85$): the transport cross section is roughly 6× smaller than the total.

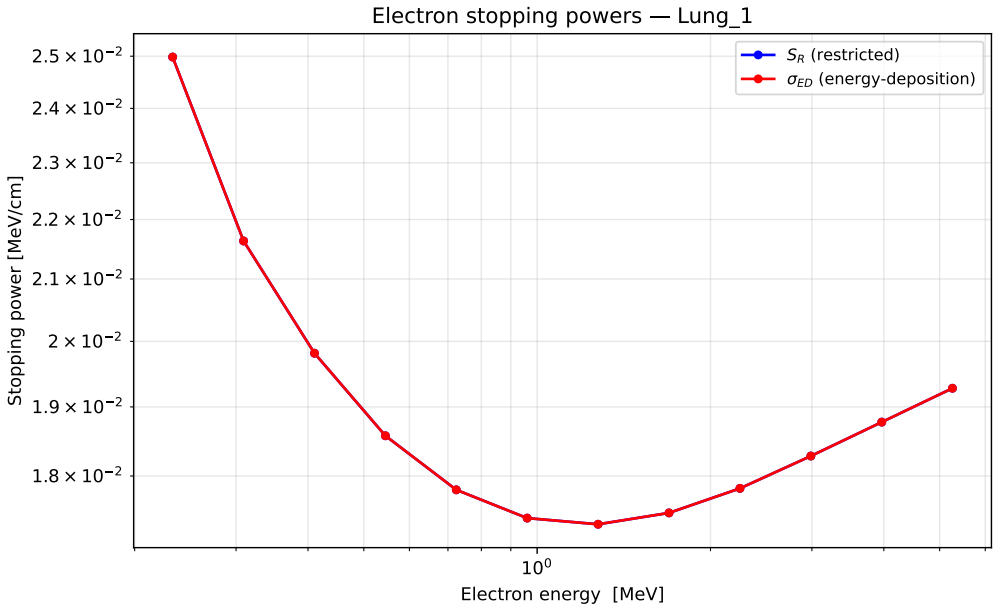


Figure 8.67: Restricted stopping power S_R and energy-deposition coefficient σ_{ED} for a tissue-equivalent material. These quantities govern the CSD energy-loss term and the dose deposition respectively.

Electron Stopping Powers

8.11.5 Solver Consumption: From YAML to Voxel Fields

The solver file `ct_to_coeffs.py` loads the YAML library and constructs per-voxel cross-section fields:

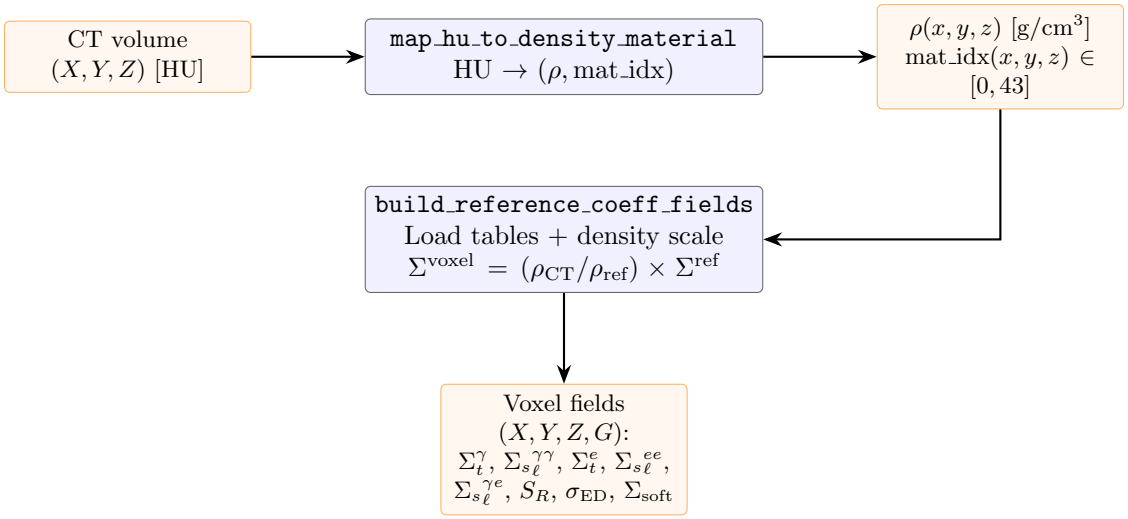


Figure 8.68: CT-to-coefficients pipeline in the solver. Each voxel gets density-scaled cross sections from the YAML material library.

The density scaling formula for voxel at position $\vec{r}$:

$$\Sigma_g(\vec{r}) = \frac{\rho_{\text{CT}}(\vec{r})}{\rho_{\text{ref}}^{(m)}} \times \Sigma_g^{(m),\text{ref}} \tag{8.194}$$

where $\rho_{\text{ref}}^{(m)}$ is the reference density of material m from `materials_library.yaml` and $\Sigma_g^{(m),\text{ref}}$ is the tabulated reference cross section. During loading, `ct_to_coeffs.py` converts from cm-units to mm-units:

$$\Sigma \text{ [1/mm]} = \Sigma \text{ [1/cm]} \ / \ 10 \tag{8.195}$$

$$S \text{ [MeV/mm]} = S \text{ [MeV/cm]} \ / \ 10 \tag{8.196}$$

8.11.6 Per-Material Data Structure

Each of the 44 materials in `materials_library.yaml` contains the data fields listed in Table 8.8.

Table 8.8: Data fields per material in the YAML library.

Section	Key	Shape	Description
photon	`sigma_t`	(G_γ)	Total cross section
	`sigma_comp`	(G_γ)	Compton (incoherent)
	`sigma_phot`	(G_γ)	Photoelectric
	`sigma_pair`	(G_γ)	Pair production
	`sigma_rayl`	(G_γ)	Rayleigh (coherent)
	`sigma_s_leg`	$(G_\gamma, G_\gamma, L_\gamma)$	Scattering moments
electron	`sigma_t`	(G_e)	Total cross section
	`S_R`	(G_e)	Restricted stopping power
	`sigma_ED`	(G_e)	Energy-deposition coefficient
	`sigma_soft`	(G_e)	Soft-scatter cross section
	`sigma_ee_leg`	(G_e, G_e, L_e)	Scattering moments
$\gamma \rightarrow e$	`sigma_gammae_leg`	$(G_\gamma, G_e, L_{\gamma e})$	Coupling moments
	`sigma`	(G_γ)	Total coupling (scalar)

8.12 Energy Deposition and Dose Calculation

8.13 Energy Deposition Mechanisms

Dose is ultimately deposited by charged particles (electrons and positrons) through collisional interactions with atomic electrons [135].

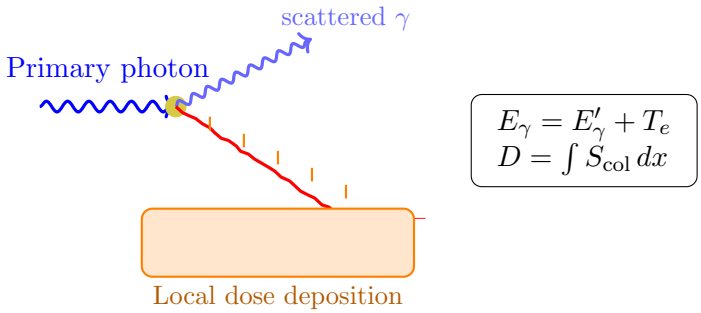


Figure 8.69: Energy flow in photon dose deposition: photon energy is transferred to electrons via Compton scattering; electrons deposit energy locally through collisional stopping power.

8.14 Restricted vs. Unrestricted Stopping Power

The stopping power can be decomposed based on energy transfer:

$$S_{\text{col}}(E) = S_{\text{col}}^{\Delta}(E) + S_{\text{col}}^{>\Delta}(E) \tag{8.197}$$

Here S_{col}^{Δ} denotes the restricted stopping power, i.e. he collisional energy loss associated with energy transfers below a cutoff Δ that are treated as locally deposited, while $S_{\text{col}}^{>\Delta}$ denotes the complementary (unrestricted) contribution associated with the production of energetic δ-rays that must be transported explicitly.

In Acuros XB, Δ is typically the electron cutoff energy (0.2–0.5 MeV).

8.15 Energy Deposition Cross Section

The energy deposition cross section σ_{ED} relates electron fluence to dose:

$$\sigma_{ED}(E) = S_{\text{col}}(E) = -\left(\frac{dE}{dx}\right)_{\text{col}} \tag{8.198}$$

In multigroup form:

$$\sigma_{ED,k} = \frac{\int_{E_k}^{E_{k+1}} S_{\text{col}}(E)\, dE}{E_{k+1} - E_k} \approx S_{\text{col}}(\bar{E}_k) \tag{8.199}$$

8.16 Dose Calculation Formula

The absorbed dose at position $\mathbf{r}$ is:

$$\boxed{D(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \sum_{k=1}^{G_e} \phi_k^e(\mathbf{r}) \cdot \sigma_{ED,k}(\mathbf{r}) \cdot C_{\text{unit}}} \tag{8.200}$$

where $\phi_k^e(\mathbf{r})$ is the scalar electron fluence in group k (particles/cm^2), $\sigma_{ED,k}(\mathbf{r})$ is the corresponding energy-deposition cross section (MeV/cm), $\rho(\mathbf{r})$ is the mass density (g/cm^3), and $C_{\text{unit}} = 1.602 \times 10^{-10}$ converts MeV/g to Gy.

8.17 Numerical Methods for Transport

8.18 Spatial Discretization

The computational domain is discretized on a Cartesian grid aligned with the CT voxels. For a voxel with indices (i, j, k) and dimensions $(\Delta x, \Delta y, \Delta z)$:

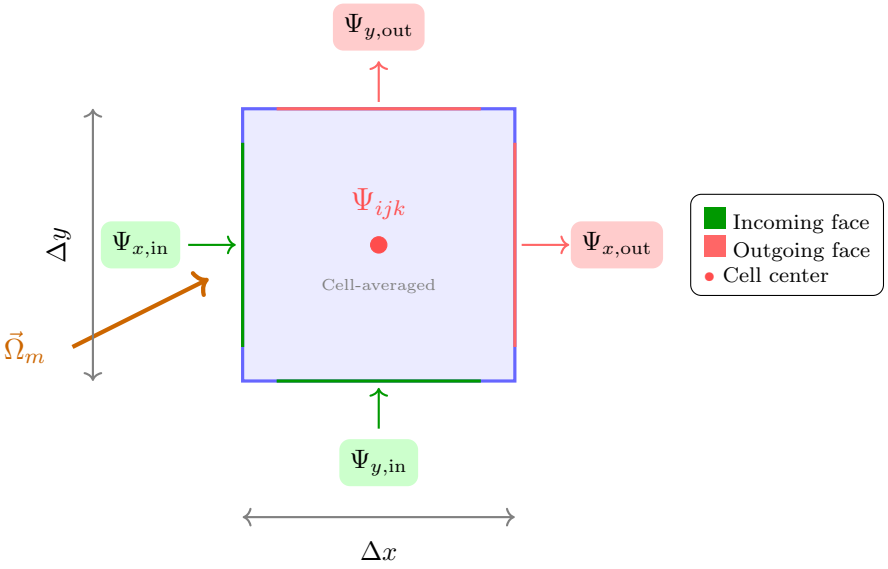


Figure 8.70: Voxel geometry for spatial discretization (2D cross-section view). The incoming face fluences Ψ_{in} (green) and outgoing face fluences Ψ_{out} (red) are related to the cell-averaged fluence Ψ_{ijk} (center) through the differencing scheme. The particle direction $\vec{\Omega}_m$ determines which faces are incoming vs. outgoing.

8.19 Transport Sweep Algorithm

The discrete ordinates equations are solved by "sweeping" through the grid in the direction of particle flow:

For each discrete direction $\mathbf{\Omega}_m$, the solver first determines the appropriate sweep ordering from the signs of the direction cosines (μ, η, ξ), then traverses voxels from upstream to downstream. At every voxel, the cell-average fluence is obtained from the incoming face values according to the chosen differencing scheme, after which outgoing face values are computed and passed to downstream voxels.

8.20 Diamond Difference Scheme

The diamond difference (DD) scheme assumes linear variation:

$$\Psi = \frac{1}{2}(\Psi_{\text{in}} + \Psi_{\text{out}}) \tag{8.201}$$

Leading to:

$$\Psi = \frac{q + 2a\Psi_{\text{in}}}{\sigma_t + 2a} \tag{8.202}$$

Properties: The diamond-difference scheme is second-order accurate, but it may produce negative fluences in optically thin cells or in the presence of strong sources, which motivates the use of fixup procedures when negativity arises.

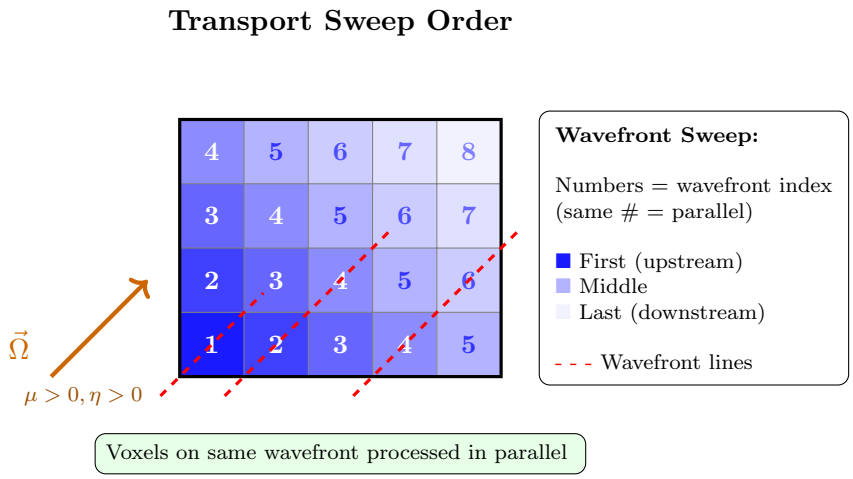


Figure 8.71: 2D illustration of the transport sweep algorithm for direction $\vec{\Omega}$ with positive direction cosines ($\mu > 0$, $\eta > 0$). Voxels are processed along diagonal wavefronts (red dashed lines) from upstream (dark blue, wavefront 1) to downstream (light blue, wavefront 8). Numbers indicate wavefront index—voxels with the same number can be processed in parallel, enabling efficient parallelization.

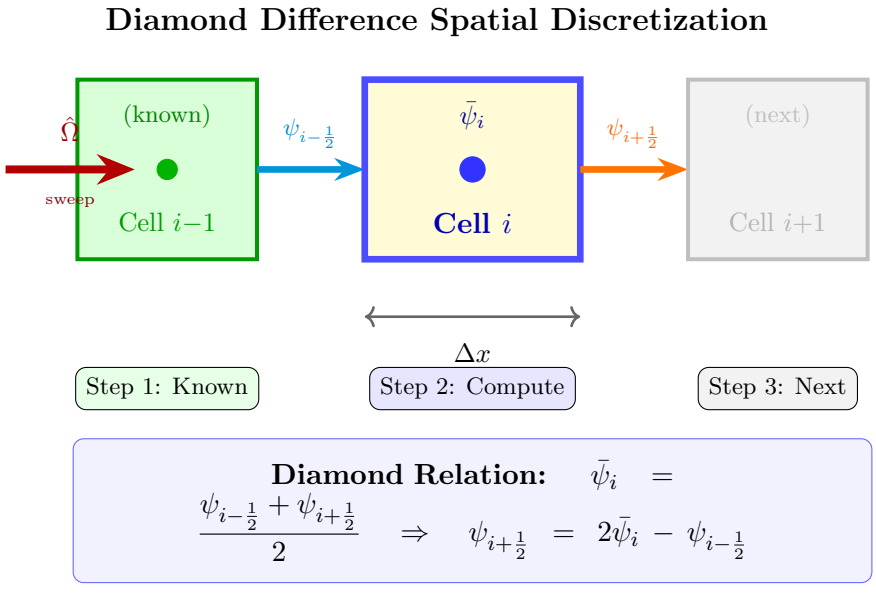


Figure 8.72: Diamond Difference sweep process: particles enter cell i with known incoming flux $\psi_{i-1/2}$ (cyan arrow), the cell-averaged flux $\bar{\psi}_i$ is computed at the cell center, and the outgoing flux $\psi_{i+1/2}$ (orange arrow) propagates to the next cell.

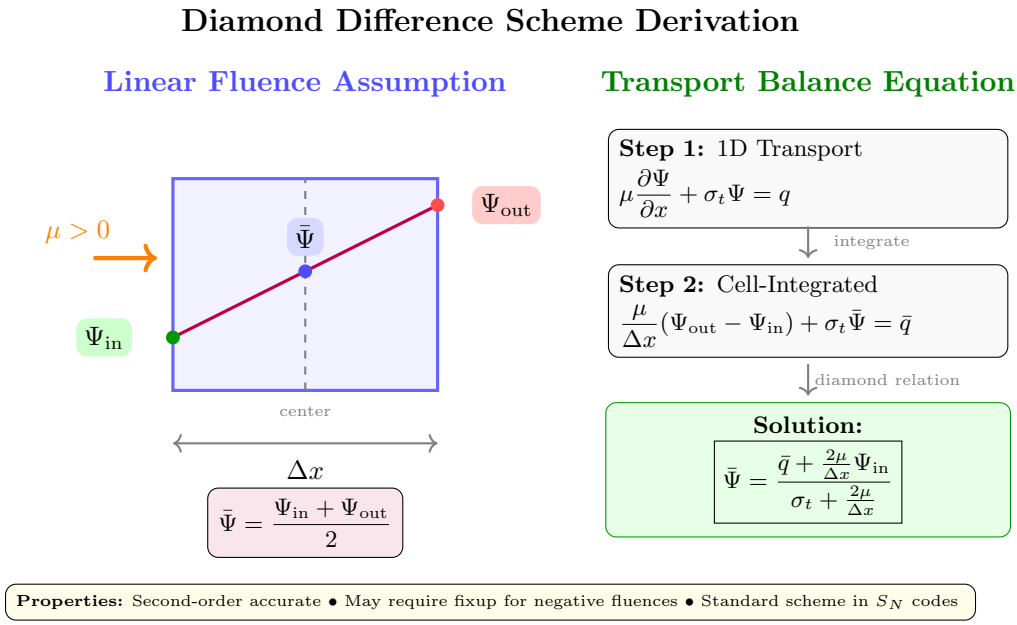


Figure 8.73: Diamond difference scheme derivation. **Left:** The scheme assumes linear variation of fluence within each cell, so the cell-averaged fluence $\bar{\Psi}$ equals the midpoint of the incoming and outgoing values. **Right:** Combining this assumption with the cell-integrated transport equation yields an explicit formula for $\bar{\Psi}$ in terms of the incoming fluence Ψ_{in}, source $\bar{q}$, and total cross-section σ_t.

8.21 Iterative Solution Strategy

The coupled photon-electron system is solved iteratively:

The coupled photon–electron system is solved iteratively. An initial guess is formed by setting $\Psi^{(0)} = \Psi_{\text{unc}}$ (uncollided contribution only), after which a source-iteration update is applied according to

$$\Psi^{(n+1)} = T^{-1}\left(q + S\Psi^{(n)}\right), \tag{8.203}$$

where T is the transport operator and S is the scattering operator. Convergence is assessed by requiring

$$\max_{\mathbf{r},g,m} \left| \frac{\Psi^{(n+1)} - \Psi^{(n)}}{\Psi^{(n+1)}} \right| < \epsilon. \tag{8.204}$$

Electrons are normally coupled using an iterative downscatter-only process, in which each group of electrons is marched from high to lower energy levels; photon-to-electron coupling is also performed in this same iteration loop.

It may take between 5 and 15 iterations to achieve acceptable convergence for photons; it should require at most one pass for electrons because they use a downscatter-only energy coupling mechanism.

8.22 Required Cross Sections

Table 8.9: Cross-section data required for coupled photon-electron transport.

Particle	Cross Section	Description
Photon	$\sigma_t^{\gamma}(E)$	Total attenuation
	$\sigma_{s,\ell}^{\gamma\gamma}(E' \to E)$	Legendre moments of Compton scatter
	$\sigma_{s,\ell}^{\gamma e}(E' \to E)$	Compton electron production
Electron	$\sigma_t^{e}(E)$	Total interaction (elastic + inelastic)
	$\sigma_{s,\ell}^{ee}(E' \to E)$	Electron-electron scattering
	$S_R(E)$	Restricted stopping power
Dose	$\sigma_{ED}(E)$	Energy deposition coefficient

8.23 Data Sources

8.24 Multigroup Condensation

Continuous cross sections are condensed to multigroup form:

$$\sigma_g = \frac{\int_{E_g}^{E_{g+1}} \sigma(E) w(E)\, dE}{\int_{E_g}^{E_{g+1}} w(E)\, dE} \tag{8.205}$$

Common weighting functions: Common choices include a flat weighting $w(E) = 1$ (simple averaging), a $1/E$ weighting appropriate for approximate slowing-down spectra, and flux-weighted condensation using a representative fluence spectrum.

8.25 Key Parameters of the Acuros XB Algorithm

This section provides a comprehensive overview of the key numerical and physical parameters used in the Acuros XB algorithm, as documented in the official Eclipse Photon and Electron Algorithms Reference Guide [142] and supporting literature [4, 48, 3].

8.26 Transport Cutoff Energies

Transport cutoff energies define the threshold below which particles are no longer tracked; instead, their remaining energy is deposited locally. These cutoffs balance computational efficiency against accuracy.

Table 8.10: Transport cutoff energies in Acuros XB [142].

Particle	Cutoff Energy	CSDA Range	Description
Photon (γ)	1 keV	–	Photons below 1 keV deposit all remaining energy locally
Electron (e^-)	200 keV	≈0.4 mm	Electrons below 200 keV deposit residual energy locally; earlier Eclipse versions used 500 keV

Remark Chapter 8.1. *Earlier versions of Acuros XB (prior to Eclipse v11) used a 500 keV electron cutoff. The reduction to 200 keV improves accuracy in superficial dose regions and near air-tissue interfaces [3].*

8.27 Energy Group Structure

Acuros XB employs a multigroup discretization of the energy domain, with cross-section libraries generated using the Sandia CEPXS code [201].

Table 8.11: Energy group structure in Acuros XB [142].

Particle	Number of Groups	Notes
Photon (γ)	25	Groups from ∼1 keV to >20 MeV; higher groups inactive for lower-energy beams
Electron (e^-)	49	Groups from cutoff (∼200 keV) to maximum beam energy

The group boundaries follow a logarithmic-like spacing to capture rapid cross-section variations at lower energies, with finer resolution near key interaction thresholds.

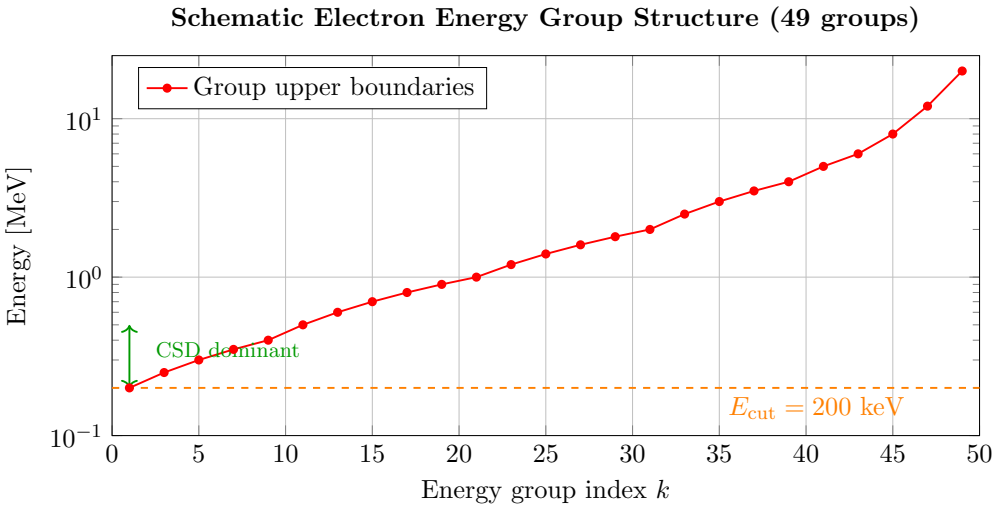


Figure 8.74: Schematic of the 49 electron energy groups, wherein groups with lower energy are closer spaced to capture the rapidly-varying stopping power and scattering cross sections accurately. The energy loss between groups is handled by the CSD operator.

8.28 Angular Quadrature Order (Discrete Ordinates)

Acuros XB uses Square Tchebychev-Legendre (STL) discrete ordinates quadrature for angular discretization. The quadrature order N varies with particle energy to balance accuracy and computational cost [142, 4].

Table 8.12: Angular quadrature orders and number of directions in Acuros XB.

Order N	Directions $M = 2N^2$	Typical Application
4	32	Very low energy particles, buildup region
6	72	Low energy photons and electrons
8	128	Medium energy range
10	200	Higher energies
12	288	High energy with forward peaking
14	392	Very forward-peaked scattering
16	512	Maximum resolution for MeV-range particles

The energy-dependent angular refinement has physical justification: at higher energies particles move much larger mean free paths and scatter more in a forward peaked manner - thus requiring a finer angular grid to accurately sample the associated geometrically thin forward scattering cone.

8.29 Scattering Expansion (Legendre Order)

The differential scattering cross section is expanded in Legendre polynomials:

$$\sigma_s(\mathbf{r}, E' \to E, \mu_0) = \sum_{\ell=0}^{\ell_{\max}} \frac{2\ell+1}{4\pi} \sigma_{s,\ell}(\mathbf{r}, E' \to E) P_\ell(\mu_0) \tag{8.206}$$

where $\mu_0 = \mathbf{\Omega} \cdot \mathbf{\Omega}'$ is the cosine of the scattering angle.

Table 8.13: Legendre expansion parameters in Acuros XB [142].

Parameter	Value	Description
$\ell_{\max}$	7	Maximum Legendre order for all interactions
Range	$0 \le \ell \le 7$	8 moments total
Application	Fixed	Same $\ell_{\max}$ for $\gamma\gamma$, γe, ee

8.30 Continuous Slowing Down (CSD) and Stopping Power

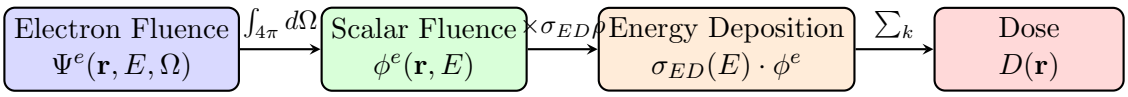
For electron transport, Acuros XB employs the Boltzmann-Fokker-Planck formulation with a Continuous Slowing Down (CSD) operator to handle the numerous soft Coulomb collisions [142, 4].

8.30.1 Dose Calculation Formula

The dose is computed by integrating the electron fluence against the energy deposition coefficient:

$$\boxed{D(\mathbf{r}) = \int_0^\infty dE \int_{4\pi} d\Omega \, \sigma_{ED}^e(\mathbf{r}, E) \, \rho(\mathbf{r}) \, \Psi^e(\mathbf{r}, E, \Omega)} \tag{8.207}$$

where σ_{ED}^e is the energy deposition cross section (effectively the linear stopping power S_R) with units [MeV/cm], and ρ is the mass density [g/cm^3].



$$D \quad = \quad \tfrac{1}{\rho} \textstyle\sum_{k=1}^{G_e} \phi_k^e \;\cdot\; \sigma_{ED,k} \;\cdot\; \Delta E_k \;\cdot\; C_{\text{unit}} \quad \text{[Gy]}$$

Figure 8.75: Flowchart of dose calculation from electron fluence in the multigroup formulation. The scalar fluence is weighted by the energy deposition coefficient and summed over all energy groups.

8.31 Spatial Discretization and Numerical Stability

8.31.1 Linear Discontinuous Galerkin Method

Acuros XB uses a **Linear Discontinuous Galerkin** (LDG) finite-element method for spatial discretization [142, 202]. This is equivalent to the classic diamond-difference scheme but formulated with linear basis functions.

In the LDG formulation, the angular flux is represented as a piecewise linear function within each voxel while allowing discontinuities at voxel faces. This yields second-order accuracy in sufficiently smooth regions, but it can also generate unphysical negative flux values in the presence of sharp gradients or strong material heterogeneities, motivating the use of positivity-preserving safeguards.

8.31.2 Positivity Fix and Step Fallback

To maintain non-negative flux and numerical stability, Acuros XB includes a safeguard mechanism:

Step Fallback Criterion:
If the ratio of linear stopping power between adjacent voxels exceeds a threshold:

$$\frac{S_R^{\text{voxel}}}{S_R^{\text{neighbor}}} > 1.5 \tag{8.208}$$

the solver reverts to first-order (step) differencing at that interface, terminating the electron and re-emitting it in the next voxel to prevent oscillations or negative flux [142].

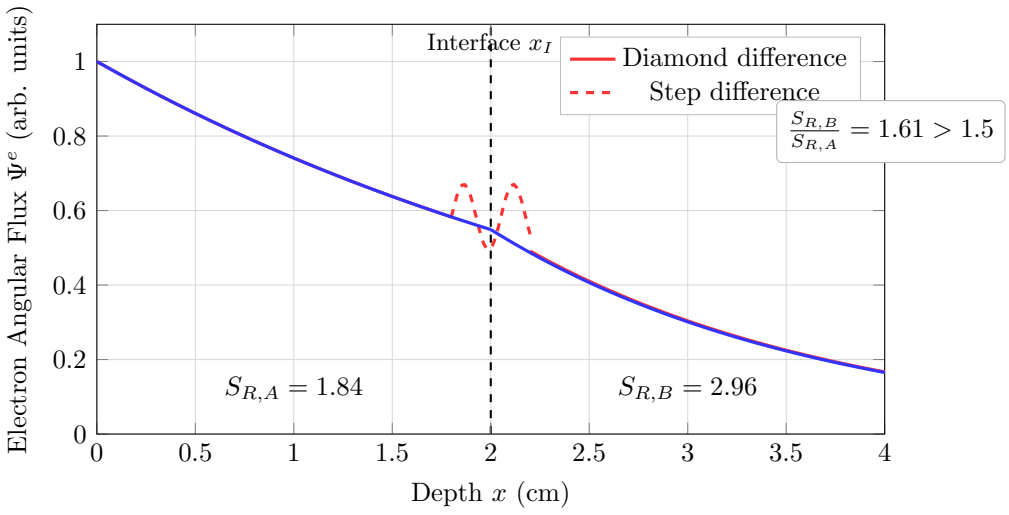


Figure 8.76: Numerical behavior of electron flux solutions at a material interface with stopping power ratio exceeding the stability threshold ($S_{R,B}/S_{R,A} > 1.5$). The diamond difference scheme (red) exhibits oscillatory behavior near the interface due to numerical instability, while the step difference scheme (blue) maintains monotonic, physically meaningful flux distributions. The Acuros XB solver automatically activates the step fallback when the stopping power ratio criterion is violated.

8.32 Summary of Key Parameters

Table 8.14: Summary of key Acuros XB algorithm parameters.

Parameter	**Value**	**Reference**
Photon cutoff	1 keV	[142]
Electron cutoff	200 keV	[142]
Photon energy groups	25	[142]
Electron energy groups	49	[142]
Angular quadrature	$N = 4$–16 (S_N)	[142, 4]
Legendre order	$\ell_{\max} = 7$	[142]
W_{cut}	$\sim$1 keV	[142]
Step fallback threshold	1.5	[142]
Low-density threshold	$\rho < 0.12$ g/cm^3	[142]
Cross-section source	CEPXS code	[201]

Chapter 9 Implementation Details

This section provides implementation-specific details of the LBTE solver, including equation mappings, discretization choices, and code architecture.

9.1 Scope and Goal

This subsection provides a strict, equation-by-equation mapping between the Acuros XB algorithm (as stated in the "Eclipse Photon and Electron Algorithms Reference" PDF) and the current solver implementation in this repository. It includes the full set of Acuros equations (Eqs. 19–38 in the PDF excerpt) and the corresponding solver equations and discretizations, with explicit notes on where the implementation matches or deviates.

9.2 Notation

We denote spatial position by $\mathbf{r} = (x, y, z)$ and particle direction by $\Omega = (\mu, \eta, \zeta)$. Particle energy is denoted by E, with subscripts γ and e referring to photons and electrons, respectively. The angular fluence (angular flux) is written as $\Psi^\gamma(\mathbf{r}, E, \Omega)$ for photons and $\Psi^e(\mathbf{r}, E, \Omega)$ for electrons, and the corresponding scalar fluence is $\phi(\mathbf{r}, E) = \int_{4\pi} \Psi(\mathbf{r}, E, \Omega)\, d\Omega$. The macroscopic total cross section is denoted by σ_t, and $\sigma_{s,\ell}$ denotes the ℓ-th Legendre moment of the scattering operator. The restricted collisional plus radiative stopping power is denoted by S_R. External sources from the beam model are written as q^γ and q^e, while the scattering and production source terms are written as $q^{\gamma\gamma}$, $q^{\gamma e}$, and q^{ee}.

9.3 Acuros XB Equations from Reference Documentation

LBTE and Boundary Conditions

$$\Omega \cdot \nabla \Psi^\gamma + \sigma_t^\gamma \Psi^\gamma = q^{\gamma\gamma} + q^\gamma \tag{A19}$$

$$\Omega \cdot \nabla \Psi^e + \sigma_t^e \Psi^e - \frac{\partial}{\partial E}\left(S_R \Psi^e\right) = q^{ee} + q^{\gamma e} + q^e \tag{A20}$$

$$\mathbf{r} \in V,\ \Omega \in 4\pi,\ E > 0 \tag{A21}$$

$$\Psi^\gamma = 0,\ \Omega \cdot \mathbf{n} < 0 \tag{A22}$$

$$\Psi^e = 0,\ \Omega \cdot \mathbf{n} < 0 \tag{A23}$$

Scattering and production sources

$$q^{\gamma\gamma}(\mathbf{r}, E, \Omega) = \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{\gamma\gamma}(\mathbf{r}, E' \to E, \Omega \cdot \Omega')\, \Psi^\gamma(\mathbf{r}, E', \Omega') \tag{A24}$$

$$q^{\gamma e}(\mathbf{r}, E, \Omega) = \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{\gamma e}(\mathbf{r}, E' \to E, \Omega \cdot \Omega')\, \Psi^\gamma(\mathbf{r}, E', \Omega') \tag{A25}$$

$$q^{ee}(\mathbf{r}, E, \Omega) = \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{ee}(\mathbf{r}, E' \to E, \Omega \cdot \Omega')\, \Psi^e(\mathbf{r}, E', \Omega') \tag{A26}$$

Legendre expansion of scattering and angular fluence

$$\sigma_s^{\gamma\gamma/\gamma e/ee}(\mathbf{r}, E' \to E, \Omega \cdot \Omega') = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi}\, \sigma_{s,\ell}^{\gamma\gamma/\gamma e/ee}(\mathbf{r}, E' \to E)\, P_\ell(\mu_0) \tag{A27}$$

$$\Psi(\mathbf{r}, E', \Omega') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \phi_{\ell m}(\mathbf{r}, E')\, Y_{\ell m}(\Omega') \tag{A28}$$

$$q^{\gamma\gamma/\gamma e/ee}(\mathbf{r}, E, \Omega) = \sum_{\ell=0}^{7} \sum_{m=-\ell}^{\ell} \left[\int_0^{\infty} dE'\, \sigma_{s,\ell}^{\gamma\gamma/\gamma e/ee}(\mathbf{r}, E' \to E)\, \phi_{\ell m}(\mathbf{r}, E') \right] Y_{\ell m}(\Omega) \tag{A29}$$

Acuros XB caps the Legendre order at $\ell \leq 7$.

Point-source decomposition and uncollided analytic solution

$$\Omega \cdot \nabla \Psi^{\gamma} + \sigma_t^{\gamma} \Psi^{\gamma} = q^{\gamma\gamma} + q^{\gamma}(E, \Omega)\, \delta(\mathbf{r} - \mathbf{r}_p) \tag{A30}$$

$$\Omega \cdot \nabla \Psi^{e} + \sigma_t^{e} \Psi^{e} - \frac{\partial}{\partial E}\left(S_R \Psi^{e}\right) = q^{ee} + q^{\gamma e} + q^{e} \tag{A31}$$

$$\Psi^{\gamma} = \Psi^{\gamma}_{\text{unc}} + \Psi^{\gamma}_{\text{coll}} \tag{A32}$$

$$\Omega \cdot \nabla \Psi^{\gamma}_{\text{unc}} + \sigma_t^{\gamma} \Psi^{\gamma}_{\text{unc}} = q^{\gamma}(E, \Omega)\, \delta(\mathbf{r} - \mathbf{r}_p) \tag{A33}$$

$$\Omega \cdot \nabla \Psi^{\gamma}_{\text{coll}} + \sigma_t^{\gamma} \Psi^{\gamma}_{\text{coll}} = q^{\gamma\gamma}_{\text{coll}} + q^{\gamma\gamma}_{\text{unc}} \tag{A34}$$

$$\Omega \cdot \nabla \Psi^{e} + \sigma_t^{e} \Psi^{e} - \frac{\partial}{\partial E}\left(S_R \Psi^{e}\right) = q^{ee} + q^{\gamma e}_{\text{coll}} + q^{\gamma e}_{\text{unc}} + q^{e} \tag{A35}$$

$$\Psi^{\gamma}_{\text{unc}}(\mathbf{r}, E, \Omega) = \delta(\Omega - \Omega_{\mathbf{r} \leftarrow \mathbf{r}_p})\, \frac{q^{\gamma}(E, \Omega)}{4\pi \left\|\mathbf{r} - \mathbf{r}_p\right\|^2}\, e^{-\tau(\mathbf{r}, \mathbf{r}_p)} \tag{A36}$$

Dose conversion

$$D_i = \int_0^{\infty} dE \int_{4\pi} d\Omega\, \frac{\sigma^{e}_{ED}(\mathbf{r}, E)}{\rho(\mathbf{r})}\, \Psi^{e}(\mathbf{r}, E, \Omega) \tag{A37}$$

Macroscopic cross section

$$\sigma = \frac{N_A\, \rho}{M}\, \sigma_{\text{micro}} \tag{A38}$$

9.4 Solver Equations (Implemented Form)

This section expresses the equations as implemented in the code, using the same physics but with explicit multigroup and discrete-ordinates discretization.

Energy-group discrete ordinates form

For photons in group g and direction m:

$$\Omega_m \cdot \nabla \Psi^{\gamma}_{g,m} + \sigma^{\gamma}_{t,g} \Psi^{\gamma}_{g,m} = q^{\gamma}_{g,m} + \left(S_{\gamma\gamma} \Psi^{\gamma}\right)_{g,m} \tag{S1}$$

For electrons in group k:

$$\Omega_m \cdot \nabla \Psi^{e}_{k,m} + \sigma^{e}_{t,k} \Psi^{e}_{k,m} - \frac{\partial}{\partial E}\left(S_R \Psi^{e}\right)_{k,m} = \left(S_{ee} \Psi^{e}\right)_{k,m} + \left(S_{\gamma e} \Psi^{\gamma}\right)_{k,m} + q^{e}_{k,m} \tag{S2}$$

Source decomposition and uncollided field (implemented)

The solver builds $\Psi^{\gamma}_{\text{unc}}$ via ray tracing from the beam model. The implemented formula is:

$$\Psi^{\gamma}_{\text{unc}}(\mathbf{r}, g, m) = \frac{q^{\gamma}_g}{4\pi} \frac{e^{-\tau(\mathbf{r},\mathbf{r}_p)}}{\|\mathbf{r}-\mathbf{r}_p\|^2} \frac{1}{w_m} w_{\text{scale}}\, \delta_{m\approx\Omega_{\mathbf{r}\leftarrow\mathbf{r}_p}} \tag{S3}$$

with beamlet strength per control point:

$$q^{\gamma}_g = (\texttt{norm_const}) \times (\Delta\text{MU}) \times A(u,v) \times dA \times \text{shares}_g \tag{S4}$$

where $A(u,v)$ is the BEV aperture transmission, dA is the BEV cell area, and shares_g are energy-spectrum weights. Mapping to discrete directions uses one of: single-angle binning, k-nearest binning, or angular Gaussian spreading.

Legendre scattering operator (discrete ordinates)

The photon scattering operator is implemented as:

$$(S_{\gamma\gamma}\Psi)_{g_{out},m} = \sum_{g_{in}} \sum_{\ell=0}^{L_g-1} (2\ell+1)\, \Sigma^{\gamma\gamma}_{s,\ell}(g_{in}\to g_{out}) \left[K^{(\ell)} \circ_w \Psi_{g_{in}}\right]_m \tag{S5}$$

with

$$\left[K^{(\ell)} \circ_w \Psi\right]_m = \sum_{m'} K^{(\ell)}_{mm'}\, w_{m'}\, \Psi_{m'} \tag{S6}$$

The $1/(4\pi)$ factor is incorporated into the precomputed Legendre kernel weights during solver initialization. The same structure is used for $S_{\gamma e}$ and S_{ee}.

CSD energy discretization (implemented)

Energy marching uses an upwind scheme from high to low energy:

$$\left(\sigma^e_{t,k} + \frac{S_k}{\rho\,\Delta E_k}\right)\Psi^e_k = q_k + (S_{ee}\Psi^e)_k + \frac{S_{k+1}}{\rho\,\Delta E_k}\Psi^e_{k+1} \tag{S7}$$

where S_k is the stopping power at group k. The solver uses S/ρ when density is available (Acuros-style mass stopping power).

Dose from electron fluence (implemented)

The implementation computes scalar fluence and then:

$$\phi^e_k(\mathbf{r}) = \sum_m w_m\, \Psi^e_{k,m}(\mathbf{r}) \tag{S8}$$

$$D(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \sum_k \sigma^e_{ED}(\mathbf{r},k)\, \phi^e_k(\mathbf{r})\, C_{\text{unit}} \tag{S9}$$

with $C_{\text{unit}} = 1.602\times10^{-7}$ to convert MeV/mm^3 to Gy. (There is an optional ΔE_k weighting in the code path when provided.)

9.5 Discretization and Numerical Methods

The numerical solution of the Linear Boltzmann Transport Equation requires careful discretization in angle, energy, and space. This section summarizes the key discretization choices that align with the Acuros XB algorithm.

9.5.1 Angular Quadrature

Acuros XB uses Square-Tchebyshev-Legendre (ST) quadrature of orders $N \in \{4, 6, 8, 10, 12, 14, 16\}$ providing $M = 2N^2$ discrete directions. A global ST grid is generated by the implementation at maximum order $N_{\max}$, with energy-dependent subsampling applied to each group and subsequent normalization to 4π solid angle. An energy-dependent order schedule is followed, which enforces monotonicity and nondecreasing behavior over all groups:

$$N(E) = \underset{N\in\{4,6,8,\ldots,16\}}{\text{nearest}} \left(N_{\min} + t(E)(N_{\max} - N_{\min})\right), \quad t(E) = \frac{\log E - \log E_{\min}}{\log E_{\max} - \log E_{\min}} \tag{Q1}$$

9.5.2 Spatial Transport Sweep

The transport sweep algorithm processes voxels in wavefront order to respect the upwind dependency of the discrete ordinates equations. For sweep direction signs $s_x, s_y, s_z \in \{-1, +1\}$, wavefronts are defined as planes of constant $w = i_s + j_s + k_s$ where i_s, j_s, k_s are sweep-reordered indices. Wavefronts are processed sequentially within each octant, with eight octants covering all possible propagation directions.

Both the first order-step and second order-diamond difference schemes can be used for the spatial balance equations. Diamond difference has better accuracy than the step difference scheme for optically thin cells; it also includes a positivity fixup that switches to the step difference characteristics if there is an outgoing flux through a face from a cell where the flux would otherwise be negative.

9.5.3 Scattering Source Evaluation

The scattering source is evaluated using precomputed Legendre folding matrices $K^{(\ell)}$:

$$F_m^{(\ell)} = \sum_n K_{mn}^{(\ell)}\, w_n\, \Psi_n, \quad S_{g\to g'} = \sum_\ell (2\ell + 1)\Sigma_{s,\ell}\, F^{(\ell)}. \tag{LF1}$$

This formulation directly aligns with the Legendre-expanded scattering kernel structure employed by Acuros XB (equations A27–A29 in the reference documentation).

9.6 Mathematical Derivation: From LBTE to Iterative Solver

This section provides a rigorous mathematical derivation showing how the continuous Linear Boltzmann Transport Equations (A19–A36) are transformed into the discrete iterative solver scheme. The derivation proceeds through four stages: (1) operator formulation, (2) discretization, (3) source decomposition, and (4) fixed-point iteration construction.

9.6.1 Stage 1: Operator Formulation of the Continuous LBTE

Photon transport operator. Starting from equation (A19), the photon LBTE is:

$$\Omega \cdot \nabla\Psi^\gamma + \sigma_t^\gamma \Psi^\gamma = q^{\gamma\gamma} + q^\gamma \tag{D1}$$

We define the *streaming-collision operator* $\mathcal{L}^\gamma$ and *scattering operator* $\mathcal{S}^{\gamma\gamma}$:

$$\mathcal{L}^\gamma[\Psi] \equiv \Omega \cdot \nabla\Psi + \sigma_t^\gamma \Psi \tag{D2}$$

$$\mathcal{S}^{\gamma\gamma}[\Psi] \equiv \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{\gamma\gamma}(E' \to E, \Omega\cdot\Omega')\, \Psi(E', \Omega') \tag{D3}$$

The operator $\mathcal{S}^{\gamma\gamma}$ corresponds directly to equation (A24). Using these operators, equation (D1) becomes:

$$\boxed{\mathcal{L}^\gamma[\Psi^\gamma] = \mathcal{S}^{\gamma\gamma}[\Psi^\gamma] + q^\gamma} \tag{D4}$$

Electron transport operator with CSD. From equation (A20), the electron LBTE includes the continuous slowing-down (CSD) term:

$$\Omega \cdot \nabla \Psi^e + \sigma_t^e \Psi^e - \frac{\partial}{\partial E}\left(S_R \Psi^e\right) = q^{ee} + q^{\gamma e} + q^e \tag{D5}$$

Define the extended electron operator that includes CSD:

$$\mathcal{L}^e[\Psi] \equiv \Omega \cdot \nabla \Psi + \sigma_t^e \Psi - \frac{\partial}{\partial E}\left(S_R \Psi\right) \tag{D6}$$

$$\mathcal{S}^{ee}[\Psi^e] \equiv \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{ee}(E' \to E, \Omega \cdot \Omega')\, \Psi^e(E', \Omega') \tag{D7}$$

$$\mathcal{S}^{\gamma e}[\Psi^\gamma] \equiv \int_0^\infty dE' \int_{4\pi} d\Omega'\, \sigma_s^{\gamma e}(E' \to E, \Omega \cdot \Omega')\, \Psi^\gamma(E', \Omega') \tag{D8}$$

The electron equation becomes:

$$\boxed{\mathcal{L}^e[\Psi^e] = \mathcal{S}^{ee}[\Psi^e] + \mathcal{S}^{\gamma e}[\Psi^\gamma] + q^e} \tag{D9}$$

9.6.2 Stage 2: Angular and Energy Discretization

Discrete ordinates approximation. The angular domain is discretized using a quadrature set $\{(\Omega_m, w_m)\}_{m=1}^M$ where $\sum_{m=1}^M w_m = 4\pi$. The continuous angular flux $\Psi(\mathbf{r}, E, \Omega)$ is approximated by discrete values:

$$\Psi(\mathbf{r}, E, \Omega) \approx \Psi_m(\mathbf{r}, E) \quad \text{for } \Omega \approx \Omega_m \tag{D10}$$

Angular integrals are replaced by quadrature sums:

$$\int_{4\pi} f(\Omega)\, d\Omega \approx \sum_{m=1}^M w_m\, f(\Omega_m) \tag{D11}$$

Multigroup energy discretization. The energy domain is partitioned into G groups for photons and K groups for electrons:

$$\text{Photons:} \quad [E_{\gamma,\min}, E_{\gamma,\max}] = \bigcup_{g=1}^G [E_{g+1}, E_g], \quad E_1 > E_2 > \cdots > E_{G+1} \tag{D12}$$

$$\text{Electrons:} \quad [E_{e,\min}, E_{e,\max}] = \bigcup_{k=1}^K [E_{k+1}, E_k] \tag{D13}$$

The group fluence is defined as:

$$\Psi_{g,m}(\mathbf{r}) = \int_{E_{g+1}}^{E_g} \Psi(\mathbf{r}, E, \Omega_m)\, dE \tag{D14}$$

Discretized photon transport equation. Applying discrete ordinates to equation (D4), for each group g and direction m:

$$\Omega_m \cdot \nabla \Psi^\gamma_{g,m} + \sigma^\gamma_{t,g} \Psi^\gamma_{g,m} = \left(S_{\gamma\gamma} \Psi^\gamma\right)_{g,m} + q^\gamma_{g,m} \tag{D15}$$

This is the discrete form of equation (S1).

Discretized scattering operator. From equation (A27), the differential scattering cross section is expanded in Legendre polynomials:

$$\sigma_s(E' \to E, \mu_0) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi}\, \sigma_{s,\ell}(E' \to E)\, P_\ell(\mu_0) \tag{D16}$$

where $\mu_0 = \Omega' \cdot \Omega$.

Substituting into the scattering integral and applying multigroup discretization:

$$(S_{\gamma\gamma}\Psi^\gamma)_{g,m} = \sum_{g'=1}^{G} \sum_{\ell=0}^{L} (2\ell+1)\, \sigma_{s,\ell}^{\gamma\gamma}(g' \to g) \sum_{n=1}^{M} \frac{w_n}{4\pi}\, P_\ell(\Omega_n \cdot \Omega_m)\, \Psi_{g',n}^{\gamma} \tag{D17}$$

Define the Legendre folding matrix:

$$K_{mn}^{(\ell)} = P_\ell(\Omega_m \cdot \Omega_n) \tag{D18}$$

The scattering source becomes:

$$\boxed{(S_{\gamma\gamma}\Psi^\gamma)_{g,m} = \sum_{g'=1}^{G} \sum_{\ell=0}^{L} (2\ell+1)\, \sigma_{s,\ell}^{\gamma\gamma}(g' \to g) \left[\frac{1}{4\pi} \sum_{n=1}^{M} w_n\, K_{mn}^{(\ell)}\, \Psi_{g',n}^{\gamma} \right]} \tag{D19}$$

This directly corresponds to equations (S5) and (S6).

9.6.3 Stage 3: Source Decomposition (Uncollided/Collided Split)

Point-source representation. From equation (A30), the external photon source is represented as a delta function at source positions:

$$q^\gamma(\mathbf{r}, E, \Omega) = \sum_p q_p^\gamma(E, \Omega)\, \delta(\mathbf{r} - \mathbf{r}_p) \tag{D20}$$

Uncollided/collided decomposition. Following equation (A32), decompose the photon fluence:

$$\Psi^\gamma = \Psi_{\text{unc}}^\gamma + \Psi_{\text{coll}}^\gamma \tag{D21}$$

Substituting into the transport equation (D4) and separating terms:
Uncollided equation (from A33):

$$\mathcal{L}^\gamma[\Psi_{\text{unc}}^\gamma] = q^\gamma \tag{D22}$$

Collided equation (from A34):

$$\mathcal{L}^\gamma[\Psi_{\text{coll}}^\gamma] = \mathcal{S}^{\gamma\gamma}[\Psi_{\text{coll}}^\gamma] + \mathcal{S}^{\gamma\gamma}[\Psi_{\text{unc}}^\gamma] \tag{D23}$$

Analytic solution for uncollided flux. The uncollided equation (D22) has the analytic solution given by equation (A36). Along the ray from source point $\mathbf{r}_p$ to field point $\mathbf{r}$:

$$\Psi_{\text{unc}}^\gamma(\mathbf{r}, E, \Omega) = \delta(\Omega - \Omega_{\mathbf{r} \leftarrow \mathbf{r}_p})\, \frac{q^\gamma(E, \Omega)}{4\pi\, \|\mathbf{r} - \mathbf{r}_p\|^2}\, e^{-\tau(\mathbf{r}, \mathbf{r}_p)} \tag{D24}$$

where $\tau(\mathbf{r}, \mathbf{r}_p) = \int_{\mathbf{r}_p}^{\mathbf{r}} \sigma_t^\gamma(s)\, ds$ is the optical depth.

First-collision source. The uncollided flux drives the *first-collision source*:

$$q_{\text{fc}}^{\gamma\gamma} = \mathcal{S}^{\gamma\gamma}[\Psi_{\text{unc}}^\gamma] \tag{D25}$$

In discrete form, this corresponds to equation (FC1).

9.6.4 Stage 4: Fixed-Point Iteration Construction

Operator inversion. Define the discrete streaming-collision operator for group g, direction m:

$$L^{\gamma}_{g,m}[\Psi] = \Omega_m \cdot \nabla\Psi + \sigma^{\gamma}_{t,g}\Psi \tag{D26}$$

The *transport sweep operator* $T^{\gamma}_{g,m} = (L^{\gamma}_{g,m})^{-1}$ inverts this operator by solving the transport equation along characteristics. Given a source Q, the sweep computes:

$$\Psi = T^{\gamma}_{g,m}[Q] \quad \Leftrightarrow \quad L^{\gamma}_{g,m}[\Psi] = Q \tag{D27}$$

Derivation of the fixed-point iteration. Rewrite the collided photon equation (D23) in operator form:

$$L^{\gamma}_{g,m}[\Psi^{\gamma}_{\text{coll}}] = \left(S_{\gamma\gamma}\Psi^{\gamma}_{\text{coll}}\right)_{g,m} + q^{\gamma\gamma}_{\text{fc},g,m} \tag{D28}$$

Apply the inverse operator $T^{\gamma}_{g,m}$ to both sides:

$$\Psi^{\gamma}_{\text{coll}} = T^{\gamma}_{g,m}\left[\left(S_{\gamma\gamma}\Psi^{\gamma}_{\text{coll}}\right)_{g,m} + q^{\gamma\gamma}_{\text{fc},g,m}\right] \tag{D29}$$

This is an implicit equation for $\Psi^{\gamma}_{\text{coll}}$ because the scattering source $S_{\gamma\gamma}$ depends on $\Psi^{\gamma}_{\text{coll}}$ itself. The standard solution method is *source iteration* (Richardson iteration):

Source iteration algorithm.

Theorem: Fixed-Point Iteration for Photon Transport

Given the collided photon equation (D28), define the iteration map:

$$\Psi^{(n+1)} = T^{\gamma}\left[S_{\gamma\gamma}\Psi^{(n)} + q_{\text{fc}}\right] \tag{D30}$$

1. Initialize: $\Psi^{(0)} = 0$ (or from previous solution)
2. Iterate until convergence: $\|\Psi^{(n+1)} - \Psi^{(n)}\| < \epsilon$

This corresponds exactly to equation (ALG2) in the algorithmic workflow.

Convergence analysis. The iteration (D30) converges if the spectral radius $\rho(T^{\gamma}S_{\gamma\gamma}) < 1$. For subcritical systems (absorbing media), this is guaranteed because:

$$\rho(T^{\gamma}S_{\gamma\gamma}) \leq \sup_{\mathbf{r},g} \frac{\sigma^{\gamma}_{s,g}(\mathbf{r})}{\sigma^{\gamma}_{t,g}(\mathbf{r})} = c < 1 \tag{D31}$$

where c is the maximum scattering ratio (always < 1 due to absorption).

9.6.5 Stage 5: Electron Transport with CSD Energy Coupling

Energy marching derivation. For electrons, the CSD term $-\partial(S_R\Psi^e)/\partial E$ couples adjacent energy groups. Discretizing the energy derivative using an upwind (backward) difference:

$$-\frac{\partial}{\partial E}\left(S_R\Psi^e\right) \approx \frac{S_k\Psi^e_k - S_{k+1}\Psi^e_{k+1}}{\Delta E_k} \tag{D32}$$

Substituting into the discrete electron equation:

$$\Omega_m \cdot \nabla\Psi^e_{k,m} + \sigma^e_{t,k}\Psi^e_{k,m} + \frac{S_k}{\Delta E_k}\Psi^e_{k,m} = \left(S_{ee}\Psi^e\right)_{k,m} + \left(S_{\gamma e}\Psi^{\gamma}\right)_{k,m} + q^e_{k,m} + \frac{S_{k+1}}{\Delta E_k}\Psi^e_{k+1,m} \tag{D33}$$

Define the effective total cross section including CSD:

$$\tilde{\sigma}^e_{t,k} = \sigma^e_{t,k} + \frac{S_k}{\rho\,\Delta E_k} \tag{D34}$$

The electron equation becomes:

$$\boxed{\left(\Omega_m \cdot \nabla + \tilde{\sigma}^e_{t,k}\right) \Psi^e_{k,m} = \left(S_{ee}\Psi^e\right)_{k,m} + \left(S_{\gamma e}\Psi^\gamma\right)_{k,m} + q^e_{k,m} + \frac{S_{k+1}}{\rho\,\Delta E_k}\Psi^e_{k+1,m}} \tag{D35}$$

This is exactly equation (ALG4). The key insight is that *energy marching from high to low energy* eliminates the implicit energy coupling: when solving for group k, the flux Ψ^e_{k+1} in the higher-energy group is already known.

Algorithm 1 Electron Transport via Energy Marching with Source Iteration

Require: Converged photon flux Ψ^γ, cross-sections, stopping powers
Ensure: Electron flux Ψ^e in all energy groups

Compute $\gamma \to e$ production source: $(S_{\gamma e}\Psi^\gamma)_{k,m}$ for all k, m
for $k = 1$ to K **do** ▷ Energy march: high to low
  Initialize scattering source: $(S_{ee}\Psi^e)^{(0)}_k = 0$
  Set CSD inflow: $\Psi^e_{\text{in}} = \frac{S_{k+1}}{\rho\Delta E_k}\Psi^e_{k+1}$ (if $k > 1$, else 0)
  repeat ▷ Source iteration within group k
    Sweep: Solve $\left(\Omega_m \cdot \nabla + \tilde{\sigma}^e_{t,k}\right) \Psi^{e,(n+1)}_{k,m} = (S_{ee}\Psi^e)^{(n)}_{k,m} + (S_{\gamma e}\Psi^\gamma)_{k,m} + \Psi^e_{\text{in}}$
    Update scattering source: $(S_{ee}\Psi^e)^{(n+1)}_k = \sum_\ell (2\ell+1)\sigma^{ee}_{s,\ell}\left[K^{(\ell)}\Psi^{e,(n+1)}_k\right]$
  until $\|\Psi^{e,(n+1)}_k - \Psi^{e,(n)}_k\| < \epsilon_{ee}$
end for

Complete electron solution algorithm.

9.6.6 Summary: Complete Mapping from A19–A36 to Iteration Scheme

The following table summarizes the mathematical transformation from the continuous LBTE equations to the discrete iterative solver:

Table 9.15: Mapping from Acuros Equations (A19–A36) to Solver Iteration Scheme

Acuros Eq.	**Derivation Step**	**Solver Equation**
A19 (Photon LBTE)	→ Operator form (D4) → Discretize	S1, ALG1
A20 (Electron LBTE)	→ CSD discretization (D32-D35)	S2, ALG4
A24–A26 (Scattering sources)	→ Legendre expansion (D16-D19)	S5, S6, LF1
A27–A29 (Legendre expansion)	→ Folding matrices (D18)	S5, S6
A30–A35 (Source decomposition)	→ Unc/coll split (D21-D25)	S3, FC1, FC2
A36 (Uncollided solution)	→ Ray tracing (D24)	S3, S4, RT1–RT5
A37 (Dose computation)	→ Quadrature summation	S8, S9

The iterative solver thus emerges naturally from:

1. **Operator splitting**: Separating streaming/collision from scattering
2. **Discretization**: Angular quadrature + multigroup energy

3. **Source decomposition**: Analytic uncollided + iterative collided
4. **Fixed-point iteration**: Inverting the transport operator per iteration
5. **Energy marching**: Exploiting the upwind structure of CSD for electrons

9.7 Algorithmic Workflow

The coupled photon-electron transport solution follows a structured pipeline:

9.7.1 Photon Transport Iteration

Define the discrete-ordinates transport operator for photons as

$$\left(L^{\gamma}_{g,m}\Psi^{\gamma}\right)_{g,m} = \Omega_m \cdot \nabla\Psi^{\gamma}_{g,m} + \sigma^{\gamma}_{t,g}\Psi^{\gamma}_{g,m}. \tag{ALG1}$$

Let $T^{\gamma}_{g,m} = (L^{\gamma}_{g,m})^{-1}$ be the sweep operator. The fixed-point iteration is

$$\Psi^{\gamma,(n+1)}_{g,m} = T^{\gamma}_{g,m}\Big(q^{\gamma}_{g,m} + (S_{\gamma\gamma}\Psi^{\gamma,(n)})_{g,m}\Big). \tag{ALG2}$$

Convergence is assessed via the relative residual $r^{(n)} = \|\Psi^{(n)} - \Psi^{(n-1)}\|_2 / \|\Psi^{(n)}\|_2$.

9.7.2 Electron Transport with Energy Marching

Electron transport uses the same sweep operator with CSD energy coupling. For each energy group k (high to low), the implicit upwind CSD update is

$$\left(\sigma^{e}_{t,k} + \frac{S_k}{\rho\,\Delta E_k}\right)\Psi^{e}_{k} = q^{e}_{k} + (S_{ee}\Psi^{e})_k + (S_{\gamma e}\Psi^{\gamma})_k + \frac{S_{k+1}}{\rho\,\Delta E_k}\Psi^{e}_{k+1}. \tag{ALG4}$$

The downward energy march handles the continuous slowing-down coupling, while outer iterations update the electron-electron scattering source.

9.7.3 Solution Pipeline

The first stage in the solution of the entire CT-solving process involves the conversion of the raw data into a format for use by the CT planning system. During this time, the Hounsfield Units (HU) are converted to mass density values, and each voxel is given an appropriate classification based on its material properties. Once the initial pre-processing has been completed, the angular quadrature is created based upon the angular sub-sampling described by Equation (Q1). Subsequent to the creation of the angular quadrature, the uncollided photon fluence is determined via ray tracing from the beam apertures through the patient geometry. The results of this operation will be used as the source term to drive the remainder of the solution. Once the uncollided photon fluence has been determined, the first collision source will be determined. This operation establishes the relationship between the photons being emitted from the source, the scattered radiation field, and the gamma-to-electron conversion that will initiate the electron cascade. After determining the first collision source, the photon source iteration will continue until it converges to an acceptable solution as per Equations (ALG1 – ALG2). At that point, coupled electron transport with energy marching will be initiated using the implicit up-wind scheme provided by Equation (ALG4). Finally, the pipeline will conclude with the determination of the dose that was deposited into the patient as a result of the accumulation of electron energy according to Equations (S6 – S9).

9.7.4 Complete Coupled Photon-Electron Dose Calculation Algorithm

The following algorithm presents the complete, unified procedure for deterministic dose calculation using the Linear Boltzmann Transport Equation with coupled photon-electron transport. This algorithm consolidates all individual components into a single, executable workflow structured in six phases. All equations are presented in full analytical form.

Algorithm 2 Complete LBTE Solver — Part I: Initialization and Uncollided Transport

========== PHASE 1: INITIALIZATION AND PREPROCESSING ==========

Require: CT volume $\mathcal{V}$ with Hounsfield units $\mathrm{HU}(i,j,k)$, beam parameters, treatment plan
Ensure: 3D dose distribution $D(i,j,k)$ [Gy]

Step 1.1: CT Preprocessing and Material Assignment
for each voxel $(i,j,k) \in \mathcal{V}$ **do**
  Mass density: $\rho(i,j,k) \leftarrow \text{HU-to-Density}\big(\mathrm{HU}(i,j,k)\big)$ using calibration curve
  Material index: $m(i,j,k) \leftarrow \text{Material-Assignment}\big(\mathrm{HU}(i,j,k)\big)$ ▷ 44 tissue types
end for

Step 1.2: Cross-Section Data Loading from Radiant.jl Library
Load photon data ($G = 25$ energy groups, Legendre order $L_\gamma \le 7$):
  Total cross-section: $\sigma_t^\gamma(m,g)$ $[\mathrm{cm}^{-1}]$
  Scattering moments: $\sigma_{s,\ell}^{\gamma\gamma}(m, g' \to g)$ for $\ell = 0, 1, \ldots, L_\gamma$
Load electron data ($K = 49$ energy groups, $L_e = 1$):
  Total cross-section: $\sigma_t^e(m,k)$
  Scattering moments: $\sigma_{s,\ell}^{ee}(m, k' \to k)$, $\sigma_{s,\ell}^{\gamma e}(m, g \to k)$
  Restricted stopping power: $S_R(m,k)$ $[\mathrm{MeV{\cdot}cm^2/g}]$
  Energy deposition cross-section: $\sigma_{ED}(m,k)$

Step 1.3: Angular Quadrature Initialization
Generate Square-Tchebyshev-Legendre (STL) quadrature: $\{(\Omega_m, w_m)\}_{m=1}^{M}$ with $M = 2N_{\max}^2$
Energy-dependent quadrature order (Eq. Q1):
  $$N(E) = \underset{N \in \{4,6,8,\ldots,16\}}{\text{nearest}} \left(N_{\min} + \frac{\log E - \log E_{\min}}{\log E_{\max} - \log E_{\min}} (N_{\max} - N_{\min}) \right)$$
Precompute Legendre folding matrices (Eq. D18):
  $K_{mn}^{(\ell)} = P_\ell(\Omega_m \cdot \Omega_n)$ for $\ell = 0, 1, \ldots, L_{\max}$ and $m, n = 1, \ldots, M$

========== PHASE 2: UNCOLLIDED PHOTON TRANSPORT ==========

Step 2.1: Beam Source Model Construction
for each control point c in treatment plan **do**
  Compute IEC-61217 rotation matrices:
    $R_G(\theta_G)$ — gantry rotation about $+y$
    $R_C(\theta_C)$ — collimator rotation about $-z$
    $R_T(\theta_T)$ — patient support rotation about $-y$
  Combined transformation: $\mathbf{R} = R_G \cdot R_C \cdot R_T$
  Build beam eye view (BEV) aperture grid: $A(u_i, v_j) \in [0,1]$ ▷ MLC transmission
  for each BEV cell (u_i, v_j) with $A(u_i, v_j) > 0$ **do**
    Compute source position in patient coordinates: $\mathbf{r}_{\text{src}} \leftarrow \mathbf{R} \cdot \mathbf{r}_{\text{BEV}}$
    Compute ray direction: $\mathbf{d} \leftarrow (\mathbf{r}_{\text{iso}} - \mathbf{r}_{\text{src}}) / \|\mathbf{r}_{\text{iso}} - \mathbf{r}_{\text{src}}\|$
    Per-cell per-group fluence (Eq. RT4):
      $\Phi_{g,\text{cell}} = C_{\text{norm}} \cdot \Delta\text{MU} \cdot A(u_i, v_j) \cdot \Delta A \cdot s_g$
      where $\Delta A = \Delta u \cdot \Delta v$ is BEV cell area and s_g is spectrum share ($\sum_g s_g = 1$)
  end for
end for

Step 2.2: Siddon Ray Tracing for Uncollided Flux
Initialize: $\Psi_{\text{unc},g,m}^\gamma(i,j,k) \leftarrow 0$ for all voxels, groups, directions
for each ray $(\mathbf{r}_{\text{src}}, \mathbf{d}, \Phi_{g,\text{cell}})$ **do**
  Map continuous direction to discrete ordinate (Eq. RT5):
    $m^* = \arg\max_{m' \in \{1,\ldots,M\}} (\mathbf{d} \cdot \Omega_{m'})$
  Initialize optical depth: $\tau_g \leftarrow 0$ for all photon groups $g = 1, \ldots, G$
  for each voxel (i,j,k) along ray via Siddon's algorithm **do**
    Compute path length through voxel: Δs
    Distance from source: $R = \|\mathbf{r}_{i,j,k} - \mathbf{r}_{\text{src}}\|$
    Half-step attenuation: $\tau_g \leftarrow \tau_g + \sigma_{t,g}^\gamma\big(m(i,j,k)\big) \cdot \frac{\Delta s}{2}$
    Deposit uncollided flux at voxel center (Eq. A36, S3):
      $$\Psi_{\text{unc},g,m^*}^\gamma(i,j,k) \mathrel{+}= \frac{\Phi_{g,\text{cell}}}{4\pi \cdot w_{m^*} \cdot R^2} \cdot e^{-\tau_g}$$
    Remaining half-step: $\tau_g \leftarrow \tau_g + \sigma_{t,g}^\gamma\big(m(i,j,k)\big) \cdot \frac{\Delta s}{2}$
  end for
end for

Algorithm 3 Complete LBTE Solver — Part II: First-Collision Sources and Photon Transport

========== PHASE 3: FIRST-COLLISION SOURCE COMPUTATION ==========

Step 3.1: Compute Photon First-Collision Scattering Source (Eq. FC1)
for each voxel (i, j, k), outgoing group g', direction m **do**

$$q^{\gamma\gamma}_{\mathrm{fc},g',m}(i,j,k) = \sum_{g=1}^{G} \sum_{\ell=0}^{L_\gamma - 1} (2\ell+1)\, \sigma^{\gamma\gamma}_{s,\ell}\big(m(i,j,k), g \to g'\big) \left[\sum_{n=1}^{M} \frac{w_n}{4\pi} K^{(\ell)}_{mn} \Psi^{\gamma}_{\mathrm{unc},g,n}(i,j,k) \right]$$

end for

Step 3.2: Compute Electron First-Collision Production Source (Eq. FC2)
for each voxel (i, j, k), electron group k, direction m **do**

$$q^{\gamma e}_{\mathrm{fc},k,m}(i,j,k) = \sum_{g=1}^{G} \sum_{\ell=0}^{L_{\gamma e} - 1} (2\ell+1)\, \sigma^{\gamma e}_{s,\ell}\big(m(i,j,k), g \to k\big) \left[\sum_{n=1}^{M} \frac{w_n}{4\pi} K^{(\ell)}_{mn} \Psi^{\gamma}_{\mathrm{unc},g,n}(i,j,k) \right]$$

end for

========== PHASE 4: COLLIDED PHOTON TRANSPORT (SOURCE ITERATION) ==========

Step 4: Photon Source Iteration (Fixed-Point Method)
Initialize collided flux: $\Psi^{\gamma,(0)}_{\mathrm{coll},g,m}(i,j,k) \leftarrow 0$ for all voxels, groups, directions
Initialize iteration counter: $n \leftarrow 0$
repeat
  $n \leftarrow n + 1$

  Step 4.1: Update Scattering Source from Previous Iteration (Eq. S5, D19)
  for each voxel (i, j, k), group g, direction m **do**

$$(S_{\gamma\gamma}\Psi^{\gamma,(n-1)})_{g,m} = \sum_{g'=1}^{G} \sum_{\ell=0}^{L_\gamma} (2\ell+1)\, \sigma^{\gamma\gamma}_{s,\ell}(g' \to g) \left[\sum_{n'=1}^{M} \frac{w_{n'}}{4\pi} K^{(\ell)}_{mn'} \Psi^{\gamma,(n-1)}_{g',n'} \right]$$

  end for

  Step 4.2: Transport Sweep using Diamond Difference (8 Octants, Wavefront Parallel)
  for each octant $(s_x, s_y, s_z) \in \{-1, +1\}^3$ **do**
    Define sweep ordering: i_s, j_s, k_s based on octant signs
    for wavefront $w = 0$ to $N_x + N_y + N_z - 3$ **do**
      for all voxels (i, j, k) satisfying $i_s + j_s + k_s = w$ **in parallel** (Eq. PS2) **do**
        for each direction $m \in \mathcal{O}(s_x, s_y, s_z)$ **in parallel do**
          for each photon group $g = 1$ to G **do**
            Total source: $Q_{g,m} = q^{\gamma\gamma}_{\mathrm{fc},g,m} + (S_{\gamma\gamma}\Psi^{\gamma,(n-1)})_{g,m}$
            Solve transport equation (Eq. ALG1):
              $\Omega_m \cdot \nabla \Psi^{\gamma,(n)}_{\mathrm{coll},g,m} + \sigma^{\gamma}_{t,g} \Psi^{\gamma,(n)}_{\mathrm{coll},g,m} = Q_{g,m}$
            Using diamond difference with step fixup for positivity
          end for
        end for
      end for
    end for
  end for

  Step 4.3: Convergence Check
  Compute total flux: $\Psi^{\gamma,(n)} = \Psi^{\gamma}_{\mathrm{unc}} + \Psi^{\gamma,(n)}_{\mathrm{coll}}$
  Compute relative residual: $r^{(n)} = \dfrac{\|\Psi^{\gamma,(n)} - \Psi^{\gamma,(n-1)}\|_2}{\|\Psi^{\gamma,(n)}\|_2}$
until $r^{(n)} < \epsilon_\gamma$ **or** $n > n^{\gamma}_{\max}$

Output: Converged photon angular flux $\Psi^{\gamma} \leftarrow \Psi^{\gamma,(n)}$

Algorithm 4 Complete LBTE Solver — Part III: Electron Transport and Dose Computation

========== PHASE 5: COUPLED ELECTRON TRANSPORT (ENERGY MARCHING) ==========

Step 5.1: Compute Total Photon-to-Electron Production Source
for each voxel (i, j, k), electron group k, direction m **do**

$$(S_{\gamma e}\Psi^{\gamma})_{k,m} = \sum_{g=1}^{G}\sum_{\ell=0}^{L_{\gamma e}}(2\ell+1)\,\sigma_{s,\ell}^{\gamma e}(g\to k)\left[\sum_{n=1}^{M}\frac{w_n}{4\pi}K_{mn}^{(\ell)}\Psi_{g,n}^{\gamma}\right]$$

end for

Step 5.2: Energy Marching from High to Low Energy
Initialize electron flux: $\Psi_{k,m}^{e}(i,j,k) \leftarrow 0$ for all voxels, groups, directions
for electron group $k = 1$ to K **do** ▷ Energy ordering: $E_1 > E_2 > \cdots > E_K$

Step 5.2.1: Compute CSD Inflow from Higher Energy Group
if $k > 1$ **then**
CSD energy coupling term: $\Psi_{\mathrm{CSD},m}^{e}(i,j,k) = \dfrac{S_{k-1}\big(m(i,j,k)\big)}{\rho(i,j,k)\cdot\Delta E_k}\cdot\Psi_{k-1,m}^{e}(i,j,k)$
else
$\Psi_{\mathrm{CSD},m}^{e} \leftarrow 0$ ▷ No inflow at highest energy
end if

Step 5.2.2: Compute Effective Total Cross-Section Including CSD (Eq. D34)
for each voxel (i, j, k) **do**

$$\tilde{\sigma}_{t,k}^{e}(i,j,k) = \sigma_{t,k}^{e}\big(m(i,j,k)\big) + \frac{S_k\big(m(i,j,k)\big)}{\rho(i,j,k)\cdot\Delta E_k}$$

end for

Step 5.2.3: Source Iteration within Energy Group k
Initialize: $\Psi_k^{e,(0)} \leftarrow 0$, inner iteration $j \leftarrow 0$
repeat
$j \leftarrow j + 1$

Update electron-electron scattering source:
for each voxel, direction m **do**

$$(S_{ee}\Psi^{e})_{k,m}^{(j-1)} = \sum_{k'=1}^{k-1}\sum_{\ell=0}^{L_e}(2\ell+1)\,\sigma_{s,\ell}^{ee}(k'\to k)\left[\sum_{n=1}^{M}\frac{w_n}{4\pi}K_{mn}^{(\ell)}\Psi_{k',n}^{e}\right]$$

end for

Assemble total source for group k**:**
$Q_{k,m}^{e} = (S_{\gamma e}\Psi^{\gamma})_{k,m} + q_{\mathrm{fc},k,m}^{\gamma e} + (S_{ee}\Psi^{e})_{k,m}^{(j-1)} + \Psi_{\mathrm{CSD},m}^{e}$

Transport sweep with modified cross-section (Eq. ALG4):
for each octant, wavefront, voxel (i, j, k), direction m (as in Phase 4) **do**
Solve: $\left(\Omega_m\cdot\nabla + \tilde{\sigma}_{t,k}^{e}\right)\Psi_{k,m}^{e,(j)} = Q_{k,m}^{e}$
end for

Compute residual: $r_k^{(j)} = \|\Psi_k^{e,(j)} - \Psi_k^{e,(j-1)}\|_2 / \|\Psi_k^{e,(j)}\|_2$
until $r_k^{(j)} < \epsilon_{ee}$ **or** $j > n_{\max}^{ee}$
Store converged solution: $\Psi_k^{e} \leftarrow \Psi_k^{e,(j)}$
end for

========== PHASE 6: DOSE COMPUTATION ==========

Step 6: Compute Absorbed Dose from Electron Fluence
for each voxel (i, j, k) **do**

Step 6.1: Compute scalar electron fluence per energy group (Eq. S8):
for $k = 1$ to K **do**

$$\phi_k^{e}(i,j,k) = \sum_{m=1}^{M} w_m\cdot\Psi_{k,m}^{e}(i,j,k)$$

end for

Step 6.2: Energy-weighted dose deposition (Eq. S9, A37):

$$D(i,j,k) = \frac{C_{\mathrm{unit}}}{\rho(i,j,k)}\sum_{k=1}^{K}\sigma_{ED}\big(m(i,j,k),k\big)\cdot\phi_k^{e}(i,j,k)$$

where $C_{\mathrm{unit}} = 1.602\times10^{-7}$ converts MeV/mm^3 to Gy
end for

return $D(i, j, k)$ for all voxels $(i, j, k) \in \mathcal{V}$ ▷ Final 3D dose distribution [Gy]

Algorithm complexity and computational cost. The dominant computational cost resides in Phases 4 and 5 (transport sweeps). For a patient grid of $N_x \times N_y \times N_z$ voxels, M angular directions, G photon groups, and K electron groups:

- **Memory:** $\mathcal{O}((G+K) \cdot M \cdot N_x N_y N_z)$ for storing angular fluences
- **Time per sweep:** $\mathcal{O}(M \cdot N_x N_y N_z)$ with wavefront parallelism reducing wall-clock time
- **Total iterations:** Typically $n_\gamma \approx 5$–15 photon iterations, $n_{ee} \approx 2$–5 per electron group

Typical execution times for a 256^3 patient grid with clinical beam configurations range from 30 seconds to 3 minutes depending on hardware (GPU acceleration can achieve <10 seconds).

9.8 Transport Along Characteristic Lines and Quadrature Properties

9.8.1 Characteristic Line Formulation

The transport equation along a characteristic ray (the path of a particle travelling in direction Ω) reduces from a partial differential equation to an ordinary differential equation. Defining the path parameter s along direction Ω such that $\mathbf{r}(s) = \mathbf{r}_0 + s\,\Omega$, the streaming operator becomes a total derivative:

$$\frac{d\Psi}{ds} + \sigma_t(s)\,\Psi(s) = Q(s) \tag{CH1}$$

where $Q(s)$ represents the total source (scattering plus external) at position $\mathbf{r}(s)$.

This is a first-order linear ODE with the integrating factor $e^{\tau(s)}$, where $\tau(s) = \int_0^s \sigma_t(s')\,ds'$ is the optical depth. The general analytic solution is:

$$\boxed{\Psi(s) = \Psi(0)\,e^{-\tau(s)} + \int_0^s Q(s')\,e^{-[\tau(s)-\tau(s')]}\,ds'} \tag{CH2}$$

The first term represents the uncollided contribution (attenuation of the incoming flux), and the integral represents the collided contribution from sources distributed along the ray. For the special case of a homogeneous medium with constant σ_t and source Q:

$$\Psi(s) = \Psi(0)\,e^{-\sigma_t s} + \frac{Q}{\sigma_t}\left(1 - e^{-\sigma_t s}\right) \tag{CH3}$$

This characteristic solution is the mathematical basis for both the uncollided flux computation via ray tracing (Equation A36) and the sweep-based spatial discretization, where the upwind scheme in Equation (**??**) approximates the attenuation integral over each voxel segment.

9.8.2 Angular Quadrature Weight Properties

The discrete ordinates approximation replaces the continuous angular integral with a weighted sum over M discrete directions $\{\Omega_m\}_{m=1}^M$ with associated weights $\{w_m\}_{m=1}^M$. For the quadrature to be physically consistent, the weights must satisfy the following integral preservation conditions.

Zeroth moment (normalization): The total solid angle must be preserved:

$$\sum_{m=1}^{M} w_m = \int_{4\pi} d\Omega = 4\pi \tag{QW1}$$

First moment (current cancellation): For an isotropic field, the net particle current must vanish:

$$\sum_{m=1}^{M} w_m \, \Omega_m = \int_{4\pi} \Omega \, d\Omega = \mathbf{0} \tag{QW2}$$

This three-component vector equation requires that the quadrature set be symmetric with respect to each coordinate axis. The Square-Tchebychev (ST) quadrature used in Acuros XB satisfies this by construction through octant symmetry.

Second moment (diffusion limit): For accurate diffusion-limit behaviour:

$$\sum_{m=1}^{M} w_m \, \Omega_{m,\alpha} \, \Omega_{m,\beta} = \frac{4\pi}{3} \, \delta_{\alpha\beta}, \quad \alpha, \beta \in \{x, y, z\} \tag{QW3}$$

These conditions ensure that the scalar flux $\phi = \sum_m w_m \Psi_m$ correctly approximates $\int_{4\pi} \Psi \, d\Omega$, and that the Eddington tensor is correctly reproduced in the diffusion limit.

9.8.3 Multigroup Cross-Section Averaging

The multigroup cross-sections used in the discrete energy formulation are obtained by averaging the continuous-energy cross-sections over each energy group using a weighting spectrum $W(E)$:

$$\sigma_{t,g} = \frac{\int_{E_g^{\min}}^{E_g^{\max}} \sigma_t(E) \, W(E) \, dE}{\int_{E_g^{\min}}^{E_g^{\max}} W(E) \, dE} \tag{MG1}$$

Similarly, the group-to-group scattering cross-section moments are:

$$\sigma_{s,\ell}(g \to g') = \frac{\int_{E_{g'}^{\min}}^{E_{g'}^{\max}} dE' \int_{E_g^{\min}}^{E_g^{\max}} dE \; \sigma_{s,\ell}(E \to E') \, W(E)}{\int_{E_g^{\min}}^{E_g^{\max}} W(E) \, dE} \tag{MG2}$$

The weighting spectrum $W(E)$ is typically chosen as a $1/E$ spectrum for intermediate energies (reflecting the Compton-scattered photon spectrum) or a more physically motivated spectrum derived from a preliminary transport calculation.

9.9 Detailed Scattering Operator Formulation

9.9.1 Precomputed Weighted Legendre Matrices

To optimize runtime evaluation of the scattering source, the normalization and quadrature weights are combined with the Legendre folding matrix into a precomputed weighted matrix:

$$K_{\text{lw}}^{(\ell)} = K^{(\ell)} \cdot \frac{w}{4\pi} \tag{SC1}$$

where $K_{mm'}^{(\ell)} = P_\ell(\Omega_m \cdot \Omega_{m'})$ is the Legendre folding matrix and w is the diagonal matrix of quadrature weights. This absorbs the $1/(4\pi)$ normalization and weight multiplication into a single precomputed tensor, reducing the scattering source evaluation to:

$$(S\Psi)_m = \sum_{\ell=0}^{L-1} (2\ell + 1) \, \sigma_{s,\ell} \left[K_{\text{lw}}^{(\ell)} \, \Psi \right]_m \tag{SC2}$$

9.9.2 Full Photon-Photon Scattering Operator

The complete photon-photon scattering operator for multigroup discrete ordinates is:

$$\boxed{(S_{\gamma\gamma}\Psi^{\gamma})_{g_{\text{out}},m} = \sum_{g_{\text{in}}=1}^{G_\gamma} \sum_{\ell=0}^{L_\gamma-1} (2\ell+1)\, \sigma_{s,\ell}^{\gamma\gamma}(g_{\text{in}} \to g_{\text{out}}) \left[K_{\text{lw}}^{(\ell)}\, \Psi_{g_{\text{in}}}^{\gamma}\right]_m} \tag{SC3}$$

where L_γ is the maximum Legendre order for photon-photon scattering (typically 2–4).

9.9.3 Photon-to-Electron Coupling Operator

The photon-to-electron coupling source, which drives the electron cascade from photon interactions (Compton, photoelectric, pair production), is:

$$\boxed{(S_{\gamma e}\Psi^{\gamma})_{k,m} = \sum_{g=1}^{G_\gamma} \sum_{\ell=0}^{L_{\gamma e}-1} (2\ell+1)\, \sigma_{s,\ell}^{\gamma e}(g \to k) \left[K_{\text{lw}}^{(\ell)}\, \Psi_{g}^{\gamma}\right]_m} \tag{SC4}$$

where $L_{\gamma e}$ is the maximum Legendre order for photon-electron coupling (typically 1).

9.9.4 Electron-Electron Scattering Operator

The electron-electron (Møller) scattering operator is:

$$\boxed{(S_{ee}\Psi^{e})_{k_{\text{out}},m} = \sum_{k_{\text{in}}=1}^{G_e} \sum_{\ell=0}^{L_e-1} (2\ell+1)\, \sigma_{s,\ell}^{ee}(k_{\text{in}} \to k_{\text{out}}) \left[K_{\text{lw}}^{(\ell)}\, \Psi_{k_{\text{in}}}^{e}\right]_m} \tag{SC5}$$

where L_e is the maximum Legendre order for electron-electron scattering (typically 1–8).

9.9.5 Coupled Electron Scattering During Source Iteration

During outer/source iterations, the electron scattering operator remains fully coupled in energy:

$$(S_{ee}\Psi^{e})_{k_{\text{out}},m}^{(n)} = \sum_{k_{\text{in}}=1}^{G_e} \sum_{\ell=0}^{L_e-1} (2\ell+1)\, \sigma_{s,\ell}^{ee}(k_{\text{in}} \to k_{\text{out}}) \left[K_{\text{lw}}^{(\ell)}\, \Psi_{k_{\text{in}}}^{e,(n)}\right]_m \tag{SC6}$$

This will make both of elastic and inelastic (Møller) transfer terms consistent with the discretized LBTE / CSD formulation that is utilized by the solver. Vendor-public Acuros XB source documentation does not exist for an implementation switch to a post-first-iteration elastic-only operator; therefore elastic-only should be considered as an optional heuristic, as opposed to a mandatory physics requirement.

Legendre Order Schedule: The Legendre expansion is truncated at:

- L_γ: Maximum order for photon-photon scattering (typically 2–4)
- L_e: Maximum order for electron-electron scattering (typically 1–8)
- $L_{\gamma e}$: Maximum order for photon-electron coupling (typically 1)

9.10 Ray Tracing Geometry and First-Collision Sources

9.10.1 BEV Sampling and Ray Geometry

The isocenter plane is sampled on a grid (u_i, v_j) with spacing $\Delta u, \Delta v$. For each open cell where the aperture transmission $A(u_i, v_j) >$ threshold, a divergent ray is cast from the virtual source position:

$$\mathbf{r}_{\text{src}} = \mathbf{r}_{\text{iso}} - \text{SAD} \cdot \mathbf{e}_z \tag{RT1}$$

$$\mathbf{r}_{\text{focus}} = \mathbf{r}_{\text{iso}} + u_i\, \mathbf{e}_x + v_j\, \mathbf{e}_y \tag{RT2}$$

$$\mathbf{d} = \frac{\mathbf{r}_{\text{focus}} - \mathbf{r}_{\text{src}}}{|\mathbf{r}_{\text{focus}} - \mathbf{r}_{\text{src}}|} \tag{RT3}$$

where $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ are the beam coordinate system unit vectors computed from the gantry, collimator, and couch rotation sequence (Equation 9.212).

9.10.2 Siddon Raymarching Algorithm

For each ray, the uncollided flux is computed via Siddon's algorithm for efficient voxel traversal along a ray through the Cartesian grid:

Algorithm 5 Siddon Raymarching for Uncollided Flux

Require: Ray origin $\mathbf{r}_{\text{src}}$, direction $\mathbf{d}$, cross-sections $\sigma_t(\mathbf{r}, g)$
Ensure: Uncollided flux $\Psi_{\text{unc}}(\mathbf{r}, g, m)$
1: Initialize optical depth: $\tau_g \leftarrow 0$ for all groups g
2: Compute ray–grid intersection: find entry $t_{\min}$ and exit $t_{\max}$
3: $t \leftarrow t_{\min}$
4: **while** $t < t_{\max}$ **do**
5: Find current voxel indices (i, j, k) at $\mathbf{r} = \mathbf{r}_{\text{src}} + t\,\mathbf{d}$
6: Find next voxel boundary crossing t_{next}
7: Compute step length: $\Delta s = t_{\text{next}} - t$
8: **Half-step attenuation:** $\tau_g \leftarrow \tau_g + \sigma_{t,g}(i, j, k) \cdot (\Delta s/2)$
9: **Deposit at voxel center:** $\Psi_{\text{unc}}(i, j, k, g, m)\ +\!= \dfrac{\Phi_{g,\text{cell}}}{4\pi\, w_m\, R^2}\, e^{-\tau_g}$
10: **Remaining half-step:** $\tau_g \leftarrow \tau_g + \sigma_{t,g}(i, j, k) \cdot (\Delta s/2)$
11: Advance: $t \leftarrow t_{\text{next}} + \epsilon$
12: **end while**

The half-step attenuation scheme ensures that the optical depth is evaluated at the voxel *center* rather than at the entry face, preventing systematic bias in the uncollided flux.

9.10.3 Beam Cell Fluence Prefactor

The per-cell per-group fluence entering each ray is:

$$\Phi_{g,\text{cell}} = C_{\text{norm}} \cdot \Delta\text{MU} \cdot A(u_i, v_j) \cdot \Delta A \cdot s_g \tag{RT4}$$

where C_{norm} is the overall normalization constant, ΔMU is the monitor units for the control point, $A(u_i, v_j)$ is the soft aperture transmission (0–1), $\Delta A = \Delta u \cdot \Delta v$ is the BEV cell area [mm^2], and s_g is the energy spectrum share for group g (with $\sum_g s_g = 1$).

9.10.4 Angular Bin Mapping

The continuous ray direction $\mathbf{d}$ is mapped to the nearest discrete ordinates direction:

$$m = \arg\max_{m'} (\mathbf{d} \cdot \Omega_{m'}) \tag{RT5}$$

The uncollided flux contribution is weighted by the inverse quadrature weight $1/w_m$ to preserve the correct normalization when reconstructing the scalar flux via quadrature summation.

9.10.5 First-Collision Sources

The first-collision sources are computed by applying the scattering operators to Ψ_{unc}:

Photon First-Collision Source:

$$q^{\gamma\gamma}_{\text{unc}}(\mathbf{r}, g', m) = \sum_{g} \sum_{\ell=0}^{L_\gamma - 1} (2\ell + 1)\, \sigma^{\gamma\gamma}_{s,\ell}(g \to g') \left[K^{(\ell)}_{\text{lw}}\, \Psi^{\gamma}_{\text{unc},g} \right]_m \tag{FC1}$$

Electron First-Collision Source:

$$q^{\gamma e}_{\text{unc}}(\mathbf{r}, k, m) = \sum_{g} \sum_{\ell=0}^{L_{\gamma e} - 1} (2\ell + 1)\, \sigma^{\gamma e}_{s,\ell}(g \to k) \left[K^{(\ell)}_{\text{lw}}\, \Psi^{\gamma}_{\text{unc},g} \right]_m \tag{FC2}$$

A collision with a photon will be the initial collision to generate the colliding transport iterations (Equations ALG1–ALG2 for photons, ALG4 for electrons).

Figure 9.77 is an example of end-to-end validation of the LBTE Solver Pipeline in a representative lung cancer VMAT case. The four panes are summary views of the main quantities: the uncollided photon fluence Ψ_{unc} that was computed by the ray tracing module (lower-left pane); the dose distribution as solved by the solver (upper-left pane); the clinical reference dose distribution from the Eclipse Treatment Planning System (TPS) (upper-right pane); and the point-wise gamma index map at 3%/3 mm (bottom right). The solver has a 92.55% gamma pass rate with a total computation time of 34 seconds, indicating that the physics model of the transport solution is correct and that the features derived from it can be used appropriately in the neural network.

9.11 Parallel Sweep Algorithm and Domain Decomposition

9.11.1 Octant Decomposition for Parallel Processing

The M angular directions are grouped into 8 octants based on the signs of their direction cosines:

$$\mathcal{O}(s_x, s_y, s_z) = \{m : \text{sign}(\mu_{x,m}) = s_x,\ \text{sign}(\mu_{y,m}) = s_y,\ \text{sign}(\mu_{z,m}) = s_z\} \tag{PS1}$$

All directions within an octant share the same sweep ordering (upwind dependency pattern), enabling *intra-octant parallelism* where all directions in an octant can be swept simultaneously.

9.11.2 Wavefront Parallelism

Within each octant, the wavefront decomposition identifies planes of voxels that can be processed in parallel. For sweep signs (s_x, s_y, s_z), the wavefront index is:

$$w = i_s + j_s + k_s \tag{PS2}$$

where i_s, j_s, k_s are the sweep-reordered spatial indices. All voxels sharing the same wavefront index w have their upwind dependencies already satisfied and can be computed simultaneously, providing $\mathcal{O}(X \cdot Y)$ parallelism on a wavefront plane of an $X \times Y \times Z$ grid.

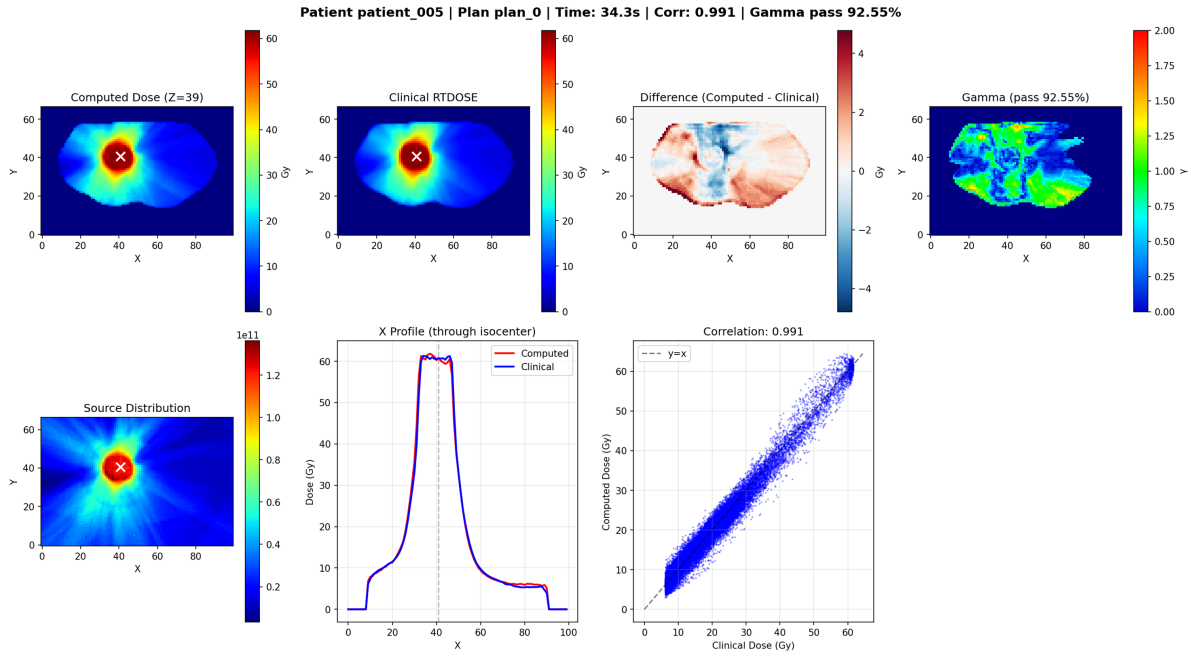


Figure 9.77: LBTE solver validation for a representative VMAT lung patient. **Top left:** computed dose from the deterministic Boltzmann transport solver. **Top right:** clinical reference dose (Eclipse Acuros XB). **Bottom left:** uncollided photon fluence Ψ_{unc} obtained via analytical ray-tracing through the patient CT. **Bottom right:** gamma index map (3%/3 mm, 10% threshold) showing a pass rate of 92.55%. Total solver wall-clock time: 34 s.

9.11.3 Spatial Domain Decomposition

For large-scale computations, the spatial domain is decomposed into P non-overlapping subdomains:

$$V = \bigcup_{p=1}^{P} V_p, \quad V_p \cap V_q = \emptyset \;\text{ for }\; p \neq q \tag{PS3}$$

The inter-subdomain communication consists of passing boundary face fluxes Ψ_{face} between adjacent subdomains in the sweep direction. The communication pattern is determined by the sweep direction: each subdomain must receive its upstream face fluxes before it can begin sweeping.

9.11.4 Load Balancing Metric

The parallel efficiency of the domain-decomposed sweep is quantified by:

$$\eta = \frac{\sum_{p=1}^{P} T_p}{P \cdot \max_p(T_p)} \tag{PS4}$$

where T_p is the computational time for subdomain p. The efficiency $\eta \in (0, 1]$ measures how evenly the workload is distributed. For optimal load balancing, subdomains should be partitioned such that each contains approximately equal numbers of non-void voxels, since void (air) voxels require no transport computation.

9.11.5 GPU Thread Mapping

In the GPU implementation, each wavefront is mapped to a kernel launch where individual threads process independent voxels within the wavefront. The thread mapping uses a fused

group-angle-voxel index:

$$\text{thread_id} = m \cdot N_{\text{voxels_in_wavefront}} + v \tag{PS5}$$

where m is the direction index within the octant and v is the voxel index within the wavefront, collapsing both parallelism dimensions into a single kernel launch.

9.12 Boundary Conditions, Convergence Criteria, and Error Propagation

9.12.1 Vacuum Boundary Conditions

The LBTE is solved on a finite domain V with boundary ∂V. Vacuum (free-space) boundary conditions specify that no particles enter the domain from outside:

$$\boxed{\Psi(\mathbf{r}_b, \Omega) = 0 \quad \text{for all } \Omega \cdot \hat{\mathbf{n}}(\mathbf{r}_b) < 0, \quad \mathbf{r}_b \in \partial V} \tag{BC1}$$

where $\hat{\mathbf{n}}(\mathbf{r}_b)$ is the outward unit normal to the boundary at position $\mathbf{r}_b$. The condition $\Omega \cdot \hat{\mathbf{n}} < 0$ selects incoming directions. This is the standard boundary condition for radiotherapy dose calculation, where the patient is surrounded by air and the only radiation source is the treatment beam.

For a Cartesian grid, the six faces of the computational domain have normals $\hat{\mathbf{n}} \in \{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$. The boundary condition is applied naturally during the transport sweep: at the upstream boundary of each sweep direction, the incoming face flux is set to zero.

Reflecting boundary conditions (used for symmetry exploitation):

$$\Psi(\mathbf{r}_b, \Omega) = \Psi(\mathbf{r}_b, \Omega - 2(\Omega \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}) \quad \text{for } \Omega \cdot \hat{\mathbf{n}} < 0 \tag{BC2}$$

These may be employed when the patient geometry exhibits symmetry planes, halving the computational domain in the symmetric dimension.

9.12.2 Convergence of Source Iteration

The source iteration scheme for photon transport (Equation ALG2) is a fixed-point iteration of the form:

$$\Psi^{(n+1)} = T\left(q_{\text{ext}} + S\Psi^{(n)}\right) \tag{CV1}$$

where T is the transport sweep operator and S is the scattering operator. The spectral radius of the iteration operator $T \circ S$ determines the convergence rate:

$$\rho(T \circ S) = \max_g \frac{\sigma_{s,0}^{\gamma\gamma}(g \to g)}{\sigma_t^{\gamma}(g)} = \max_g c_g \tag{CV2}$$

where c_g is the scattering ratio for group g. For photons in tissue at therapeutic energies, $c_g \approx 0.3$–0.7, ensuring rapid convergence (typically 3–8 iterations).

The convergence is monitored via the relative L^2 residual:

$$r^{(n)} = \frac{\left\|\Psi^{(n+1)} - \Psi^{(n)}\right\|_2}{\left\|\Psi^{(n+1)}\right\|_2} < \epsilon_{\text{tol}} \tag{CV3}$$

with typical tolerance $\epsilon_{\text{tol}} = 10^{-4}$ to 10^{-5}. Convergence is declared when $r^{(n)} < \epsilon_{\text{tol}}$ for all energy groups.

9.12.3 Error Propagation in Spatial Discretization

The spatial discretization error in the upwind (step) scheme is first-order in the mesh spacing:

$$\|\Psi - \Psi_h\| = \mathcal{O}(h) \quad \text{(step difference)} \tag{EP1}$$

where $h = \max(\Delta x, \Delta y, \Delta z)$ is the maximum voxel dimension. The diamond difference scheme achieves second-order accuracy:

$$\|\Psi - \Psi_h\| = \mathcal{O}(h^2) \quad \text{(diamond difference)} \tag{EP2}$$

subject to the positivity constraint (negativity fixup reverts to first-order step on affected voxels).

The angular discretization error depends on the quadrature order N and the angular complexity of the solution. For a smoothly varying angular flux:

$$\left\| \phi - \sum_m w_m \Psi_m \right\| = \mathcal{O}\left(\frac{1}{N^{L+1}} \right) \tag{EP3}$$

where L is the order of the Legendre expansion of the scattering kernel. In the multigroup discretization, the energy error scales with the group width:

$$\|\Psi_{\text{continuous}} - \Psi_{\text{multigroup}}\| = \mathcal{O}(\Delta E_g) \tag{EP4}$$

The total error in the computed dose is bounded by the combined contributions from all discretizations:

$$\frac{|D_{\text{computed}} - D_{\text{exact}}|}{D_{\text{exact}}} \leq C_1 h + C_2 N^{-(L+1)} + C_3 \Delta E + C_4 \epsilon_{\text{tol}} \tag{EP5}$$

where C_1, C_2, C_3, C_4 are problem-dependent constants. In practice, the spatial discretization dominates for typical clinical grid spacings (2.5 mm), while the angular and energy errors are secondary.

9.13 Preprocessing and Source Construction Algorithms

9.13.1 CT Data to Transport Coefficients

Algorithm 6 CT Data to Transport Coefficients (Preprocessing Pipeline)

Require: CT image (HU values), materials library, calibration tables
Ensure: Voxel-wise cross-sections $\{\sigma_t, \sigma_s, S_R\}$

1: **for** each voxel $\mathbf{r}$ **do**
2: Convert HU to density: $\text{HU}(\mathbf{r}) \to \rho(\mathbf{r})$ via calibration lookup table
3: Assign material: $\rho(\mathbf{r}) \to \text{mat_id}(\mathbf{r})$ based on density thresholds
4: Lookup reference cross-sections: $\sigma_{\text{ref}}(\text{mat_id}(\mathbf{r}))$ from library
5: Scale to local density: $\sigma(\mathbf{r}) = \dfrac{\rho(\mathbf{r})}{\rho_{\text{ref}}(\text{mat_id})} \cdot \dfrac{\sigma_{\text{library}}}{10}$
6: **end for**
7: Build scattering tensors: $\sigma_s^{\gamma\gamma}(\mathbf{r}, g, g', \ell)$, $\sigma_s^{ee}(\mathbf{r}, k, k', \ell)$, $\sigma_s^{\gamma e}(\mathbf{r}, g, k, \ell)$

The division by 10 in the scaling step converts from the materials library units (cm^{-1}) to the solver's spatial units (mm^{-1}).

9.13.2 RTPLAN to Uncollided Photon Flux

Algorithm 7 RTPLAN to Uncollided Photon Flux (Source Construction)

Require: RTPLAN (gantry angles, MLC positions, MU), isocenter, energy spectrum
Ensure: Uncollided photon flux $\Psi^{\gamma}_{\text{unc}}(\mathbf{r}, g, m)$

```
for each control point do
    Compute beam geometry: gantry, collimator, couch rotation transforms
    Construct beam basis vectors e_x, e_y, e_z (Eq. 9.212)
    for each BEV cell (u_i, v_j) with A(u_i, v_j) > 0.01 do
        Define ray: r_src → r_focus(u_i, v_j) (Eqs. RT1–RT3)
        Compute fluence: Φ_{g,cell} (Eq. RT4)
        Map to angular bin: m = arg max(d · Ω_{m'}) (Eq. RT5)
        Siddon raymarching: Compute Ψ_unc(r, g, m) (Algorithm 7)
    end for
end for
Build first-collision sources: q^{γγ}_unc, q^{γe}_unc (Eqs. FC1–FC2)
```

9.14 Physical Quantities and Units Reference

Table 9.16: Physical Quantities and Units in the LBTE Solver

Quantity	Symbol	Solver Units
Angular fluence	$\Psi(\mathbf{r}, E, \Omega)$	particles/(mm^2 sr)
Scalar fluence	$\phi(\mathbf{r}, E)$	particles/mm^2
Total cross-section	σ_t	1/mm
Scattering cross-section	σ_s	1/mm
Stopping power (linear)	S_R	MeV/mm
Energy deposition coefficient	σ_{ED}	MeV/mm
KERMA coefficient	K	MeV/mm
Density	ρ	g/cm^3
Dose	D	Gy (J/kg)
Energy	E	MeV
Spatial coordinates	$\mathbf{r} = (x, y, z)$	mm
Voxel size	$\Delta x, \Delta y, \Delta z$	mm
Source-Axis Distance	SAD	mm (typically 1000)
Monitor Units	MU	dimensionless

Critical Unit Conversions:

- Cross-sections: Library (1/cm) → Solver (1/mm): divide by 10
- Stopping power: Library (MeV/cm) → Solver (MeV/mm): divide by 10
- Dose conversion: MeV/mm^3 → Gy: multiply by $C_{\text{unit}} = 1.602 \times 10^{-7}$

9.15 Implementation Insights and Numerical Validation

9.15.1 Memory-Efficient Scattering via Material Indirection

The full scattering tensor $\sigma_s(\mathbf{r}, g, g', \ell)$ for a typical 256^3 grid with $G_e = 21$ electron groups and $L = 4$ Legendre moments would require approximately 60 GB of memory. Instead, the

implementation stores material-indexed tables and uses a voxel-to-material indirection map:

$$\text{Material tables: } \sigma_s^{\text{table}}(M, g, g', \ell), \quad M \approx 44 \text{ materials} \tag{MI1}$$

$$\text{Runtime lookup: } \sigma_s(\mathbf{r}, g, g', \ell) = \sigma_s^{\text{table}}(\text{mat_id}(\mathbf{r}), g, g', \ell) \tag{MI2}$$

This reduces memory from 60 GB to approximately 0.3 MB, a reduction factor of $\sim 2 \times 10^5$.

9.15.2 Half-Segment Attenuation Correction

In Siddon raymarching (Algorithm 7), the attenuation to a voxel *center* requires only half the segment length, not the full segment:

$$\tau_{\text{to_center}} = \tau_{\text{entry}} + \frac{\sigma_t \Delta s}{2} \tag{MI3}$$

Using the incorrect full-segment attenuation $\tau_{\text{entry}} + \sigma_t \Delta s$ would systematically over-attenuate the uncollided flux, producing artificially elevated dose at depth.

9.15.3 Energy Conservation Constraint

For all materials and energy groups, the scattering cross-sections must satisfy:

$$\sum_{g'=1}^{G_\gamma} \sigma_{s,\ell=0}^{\gamma\gamma}(g \to g') \leq \sigma_t^{\gamma}(g) \tag{VC1}$$

Only the $\ell = 0$ moment represents the total scattering probability. Higher moments ($\ell \geq 1$) describe the angular shape of scattering and can numerically exceed σ_t without violating physical conservation.

9.15.4 Photon Downscatter Constraint

For Compton scattering, photon energy can only decrease:

$$\sigma_s^{\gamma\gamma}(g \to g') = 0 \quad \text{if } E_{g'} > E_g \tag{VC2}$$

The scattering matrix is therefore lower-triangular in energy, which the source iteration exploits by processing energy groups in descending order.

9.15.5 Mean Free Path Validation

For soft tissue at therapeutic photon energies, the expected mean free paths are:

- At 1 MeV: $\sigma_t \approx 0.008$ mm^{-1} $\Rightarrow$ MFP ≈ 125 mm
- At 6 MeV: $\sigma_t \approx 0.003$ mm^{-1} $\Rightarrow$ MFP ≈ 330 mm

If the computed $\sigma_t > 0.1$ mm^{-1} (MFP < 10 mm), a unit conversion error (e.g., keV vs MeV energy scale) is likely present.

9.16 Computational Optimization

The solver achieves clinically acceptable processing times with the aid of a graphics processor unit (GPU) using optimization techniques that include fused kernel execution. Fused kernel execution reduces overhead costs by collapsing the group and angle dimensions into a single dimension which allows each wave front to be processed by one kernel launch per wave front. Memory-aware chunking applies the scattering operator in spatial chunks for large grids, limiting GPU memory requirements while maintaining mathematical exactness since scattering is pointwise in space. Mixed precision computation stores intermediate face fluxes in reduced

precision while maintaining full precision for the cell-average solution, balancing accuracy and memory bandwidth. Persistent buffer allocation ensures that working arrays for fluence and face fluxes are allocated once and reused across iterations, eliminating repeated memory allocation overhead.

9.17 Beam Model and Source Sampling

The uncollided source is built from the RTPLAN beam model using BEV sampling, aperture transmission, and discrete-ordinates angular mapping.

9.18 IEC-61217 Coordinate System and Mathematical Framework

The transformation of beam geometry parameters from the treatment machine coordinate system to the patient imaging coordinate system constitutes a fundamental preprocessing step in any dose calculation algorithm. The International Electrotechnical Commission standard IEC-61217 establishes the conventions for radiotherapy equipment coordinate systems, which must be reconciled with the DICOM standard used for computed tomography imaging data.

9.18.1 Coordinate System Definitions

The IEC Fixed Coordinate System is defined with its origin at the machine isocenter, where the three orthogonal basis vectors are oriented as follows: the X_f axis points toward the patient's left side (for a patient in supine head-first position), the Y_f axis points toward the gantry (anterior direction in IEC convention), and the Z_f axis points toward the patient's superior direction (head). This forms a right-handed coordinate system.

The DICOM Patient Coordinate System, commonly referred to as LPS (Left-Posterior-Superior), shares the same X and Z axis orientations but defines the Y axis in the opposite direction: the DICOM Y axis points toward the patient's posterior (back). This difference is captured by the transformation matrix

$$T_{\text{IEC}\to\text{DICOM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{9.209}$$

which negates the Y component when converting from IEC to DICOM coordinates.

9.18.2 Gantry Rotation Kinematics

The gantry angle θ_G parameterizes the rotation of the radiation source about the isocenter in the fixed coordinate system. At $\theta_G = 0°$, the source is positioned directly above the patient (anterior position), delivering an anterior-posterior (AP) beam. The gantry rotates in the clockwise direction when viewed from the patient's feet (along the positive Z axis), such that $\theta_G = 90°$ places the source on the patient's left side, $\theta_G = 180°$ positions the source inferior to the patient (posterior-anterior beam), and $\theta_G = 270°$ places the source on the patient's right side.

This clockwise rotation convention is opposite to the standard mathematical convention for positive rotation (counterclockwise). Consequently, when computing the rotation matrix, the gantry angle must be negated:

$$R_G(\theta_G) = R_z(-\theta_G) = \begin{pmatrix} \cos\theta_G & \sin\theta_G & 0 \\ -\sin\theta_G & \cos\theta_G & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9.210}$$

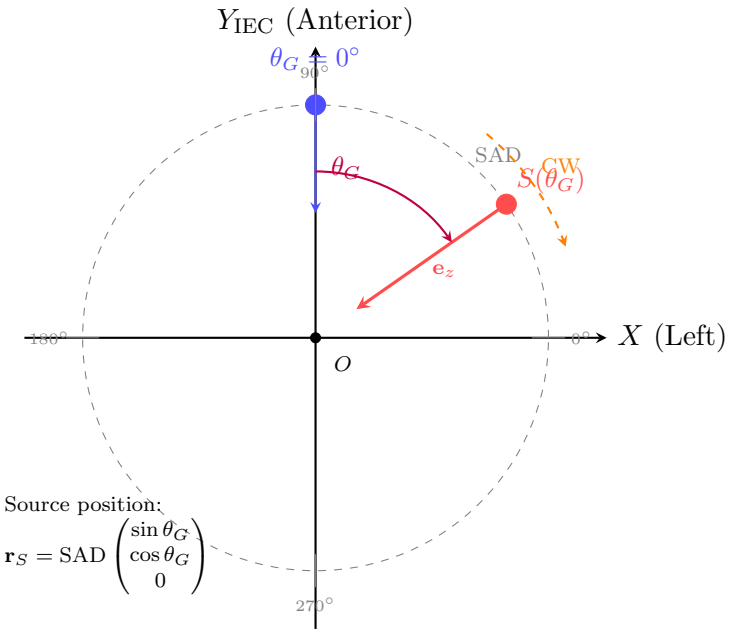


Figure 9.78: Gantry rotation geometry in the IEC fixed coordinate system (axial plane view from patient's feet). The source position $S(\theta_G)$ traces a circle of radius SAD centered at the isocenter O. The clockwise rotation convention requires negation of the angle when applying standard rotation matrices.

9.18.3 Collimator and Couch Rotations

The collimator rotates within the beam plane about the central beam axis. Denoting the collimator angle as θ_C, this rotation affects the orientation of the multi-leaf collimator (MLC) leaves and jaw positions relative to the patient anatomy. The collimator rotation is applied in the beam coordinate system and is mathematically equivalent to a rotation in the plane perpendicular to the beam direction.

The treatment couch (patient support system) can rotate about the vertical axis passing through the isocenter, parameterized by angle θ_T. This rotation follows the same clockwise convention as the gantry rotation when viewed from above. The combined effect of all three rotations on the beam basis vectors is computed through successive matrix multiplication.

9.18.4 Beam Basis Vector Derivation

The beam coordinate system is defined by three orthonormal vectors: $\mathbf{e}_x$ (cross-plane direction, parallel to MLC leaf motion), $\mathbf{e}_y$ (in-plane direction, perpendicular to leaf motion), and $\mathbf{e}_z$ (beam propagation direction, from source toward isocenter).

At the reference configuration ($\theta_G = \theta_C = \theta_T = 0$), these vectors in the IEC system are

$$\mathbf{e}_x^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_y^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}_z^{(0)} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}. \tag{9.211}$$

The beam direction $\mathbf{e}_z^{(0)}$ points in the negative Y_{IEC} direction because at gantry zero the source is above the patient (positive Y_{IEC}) and the beam propagates downward toward the isocenter.

The general beam basis vectors are obtained by applying the rotation sequence

$$\mathbf{e}_i = T_{\text{IEC}\to\text{DICOM}} \cdot R_T(\theta_T) \cdot R_G(\theta_G) \cdot R_C(\theta_C) \cdot \mathbf{e}_i^{(0)}, \quad i \in \{x, y, z\}, \tag{9.212}$$

where the IEC-to-DICOM transformation is applied last to express the result in the patient imaging coordinate system.

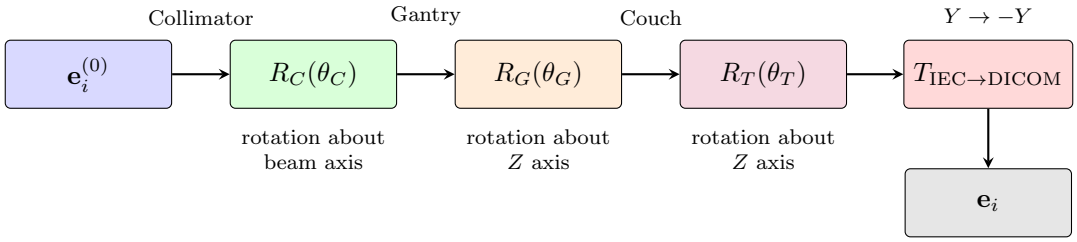


Figure 9.79: Sequential application of rotation transformations to compute beam basis vectors in DICOM LPS coordinates. Each transformation is applied from left to right, with the final IEC-to-DICOM conversion negating the Y component.

9.18.5 Explicit Rotation Matrices

The standard rotation matrix about the Z axis by angle α in the counterclockwise direction is

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9.213}$$

For the gantry and couch rotations, the IEC clockwise convention requires using $R_z(-\theta)$ rather than $R_z(\theta)$. The collimator rotation acts in the beam plane (perpendicular to $\mathbf{e}_z$), which after gantry rotation may no longer align with the Z axis. In the implementation, the collimator rotation is applied to the cross-plane and in-plane vectors as

$$\mathbf{e}_x = \cos\theta_C\,\tilde{\mathbf{e}}_x + \sin\theta_C\,\tilde{\mathbf{e}}_y, \tag{9.214}$$

$$\mathbf{e}_y = -\sin\theta_C\,\tilde{\mathbf{e}}_x + \cos\theta_C\,\tilde{\mathbf{e}}_y, \tag{9.215}$$

where $\tilde{\mathbf{e}}_x$ and $\tilde{\mathbf{e}}_y$ are the basis vectors after gantry and couch rotations but before collimator rotation.

9.19 Material Tables and Element Coverage

This section lists the materials and their elemental compositions as stored in the materials libraries, and summarizes element coverage across materials.

Eclipse materials library (materials_cross_sections_data_exact_eclipse/materials_library.yan

Material	Density (g/cm^3)	Composition (mass fractions)
Air_STP	0.0012	N:0.7553, O:0.2318, Ar:0.0128, C:0.0001
Lung_ICRP1975	0.2600	O:0.7571, C:0.1023, H:0.1013, N:0.0286, Cl:0.0027, S:0.0022, K:0.0019, Na:0.0018, P:0.0008, Mg:0.0007, Fe:0.0004, Ca:0.0001, Zn:0.0000
Adipose_ICRP1975	0.9200	C:0.6372, O:0.2323, H:0.1195, N:0.0080, Cl:0.0012, S:0.0007, Na:0.0005, K:0.0003, P:0.0002, Ca:0.0000, Fe:0.0000, Mg:0.0000, Zn:0.0000
Muscle_Skeletal_ICRP1975	1.0500	O:0.7548, C:0.1078, H:0.1006, N:0.0277, K:0.0030, S:0.0024, P:0.0018, Cl:0.0008, Na:0.0008, Mg:0.0002, Zn:0.0001, Fe:0.0000, Ca:0.0000
Cartilage_ICRP1975	1.1000	O:0.7440, C:0.0990, H:0.0960, N:0.0220, P:0.0220, S:0.0090, Na:0.0050, Cl:0.0030
Bone_ICRP1975	1.8500	O:0.4461, Ca:0.2099, C:0.1443, P:0.1050, H:0.0472, N:0.0420, S:0.0032, Mg:0.0022, Zn:0.0001

Material	Density (g/cm^3)	Composition (mass fractions)
Water	1.0000	O:0.8881, H:0.1119
Aluminum	2.7000	Al:1.0000
TitaniumAlloy_Ti6Al4V	4.4200	Ti:0.9000, Al:0.0600, V:0.0400
StainlessSteel	8.0000	Fe:0.6837, Cr:0.1900, Ni:0.0950, Mn:0.0200, Si:0.0100, C:0.0008, P:0.0004
Gold	19.3200	Au:1.0000
Wood	0.7000	O:0.4934, C:0.4445, H:0.0622
Cork	0.1900	O:0.4934, C:0.4445, H:0.0622
Polystyrene	1.0500	C:0.9226, H:0.0774
Epoxy	1.0400	C:0.7328, O:0.1446, S:0.0725, H:0.0501
PMMA	1.1900	C:0.7328, O:0.1446, S:0.0725, H:0.0501
Radel	1.3000	C:0.7328, O:0.1446, S:0.0725, H:0.0501
PEEK	1.3100	C:0.7328, O:0.1446, S:0.0725, H:0.0501
PVC	1.3800	C:0.7328, O:0.1446, S:0.0725, H:0.0501
Acetal	1.4200	C:0.7328, O:0.1446, S:0.0725, H:0.0501
PVDF	1.7700	C:0.7328, O:0.1446, S:0.0725, H:0.0501
PTFE	2.2000	C:0.7328, O:0.1446, S:0.0725, H:0.0501

Eclipse element coverage (by material count and weight range)

Element	Materials count	Weight range
Al	2	0.0600–1.0000
Ar	1	0.0128–0.0128
Au	1	1.0000–1.0000
C	18	0.0001–0.9226
Ca	4	0.0000–0.2099
Cl	4	0.0008–0.0030
Cr	1	0.1900–0.1900
Fe	4	0.0000–0.6837
H	17	0.0472–0.1195
K	3	0.0003–0.0030
Mg	4	0.0000–0.0022
Mn	1	0.0200–0.0200
N	6	0.0080–0.7553
Na	4	0.0005–0.0050
Ni	1	0.0950–0.0950
O	17	0.1446–0.8881
P	6	0.0002–0.1050
S	13	0.0007–0.0725
Si	1	0.0100–0.0100
Ti	1	0.9000–0.9000
V	1	0.0400–0.0400
Zn	4	0.0000–0.0001

Chapter 10 Clinical Dataset and Validation Result

This section presents a comprehensive description of the lung cancer radiotherapy dataset employed for validation of the LBTE-based dose calculation solver, along with the validation results demonstrating clinical accuracy.

10.1 Lung Cancer Clinical Dataset

The data set is part of a heterogeneous multi-cohort multi-site radiotherapy collection that was created using an automated iterative radiotherapy planning (AIRTP) pipeline in the Eclipse treatment planning system (Version 18.0.1 Varian Medical Systems)[68]. The AIRTP uses the Acuros XB dose calculation algorithm based on the linear boltzmann transport equation.

10.2 Dataset Generation and Characteristics

The lung cancer patient group consists of 776 patients with a total of 1,757 individualized treatment plans (1,450 for training, 132 for validating, and 175 for testing)[68]. Each patient has two independently optimized treatment plans: one is a Volumetric Modulated Arc Therapy (VMAT) plan and the other is an Intensity-Modulated Radiation Therapy (IMRT) plan. The treatment plans were developed on the TrueBeam linear accelerator with 6 MV photon beams. Treatment plans were created using the Photon Optimization module (PO version 18.0.1) for creating the plans and the Acuros XB algorithm (AXB 18.0.1) for calculating the doses for the treatment plans.

The planning process utilizes a fully automated process for creating treatment plans to reduce typical treatment planning time from 3–6 hours of manual processing to approximately 0.1–1 hour of automated processing. Automated planning includes OAR contouring utilizing the AI Rad Companion system (Siemens Healthineers), automated creation of helper structures based upon Rapid Plan guidelines, automated determination of beam configurations, determination of optimization objectives and automated iterative refinement of plan quality.

This dataset includes DICOM-format radiotherapy data. Specifically, it includes: CT (Imaging data), RTSTRUCT (Delineated structures, radiotherapy structures containing delineations of organs at risk and planning target volumes), RTPLAN (Radiotherapy treatment plans), and RTDOSE (Calculated dose distributions). The data originates from nine TCIA cohorts and covers patients with head-and-neck and lung cancer. This dataset can be connected to the GDP-HMM challenge at AAPM 2025. The task of the challenge is about generalizable dose prediction for radiotherapy. The HMM-RT dataset include the combined TCIA Datasets: NSCLC-Cetuximab, NSCLC-Radiomics, Head-Neck Cetuximab, Head-Neck-PET-CT, HNC-IMRT-70-33, HNSCC, HNSCC-3DCT-RT, and TCGA-HNSC[68].

Table 10.19: Summary statistics of the lung cancer dataset compared with the OpenKBP benchmark.

Attribute	**HMM-RT Lung**	**OpenKBP**
Number of patients	776	340
Number of plans	1,757	340
Plans per patient	2 (IMRT + VMAT)	1
Treatment planning system	Eclipse v.2023	CERR (2003)
Dose calculation algorithm	Acuros XB 18.0.1	MC
Prescribed doses	Variable	Fixed
Quality assessment	Scorecard-based	None

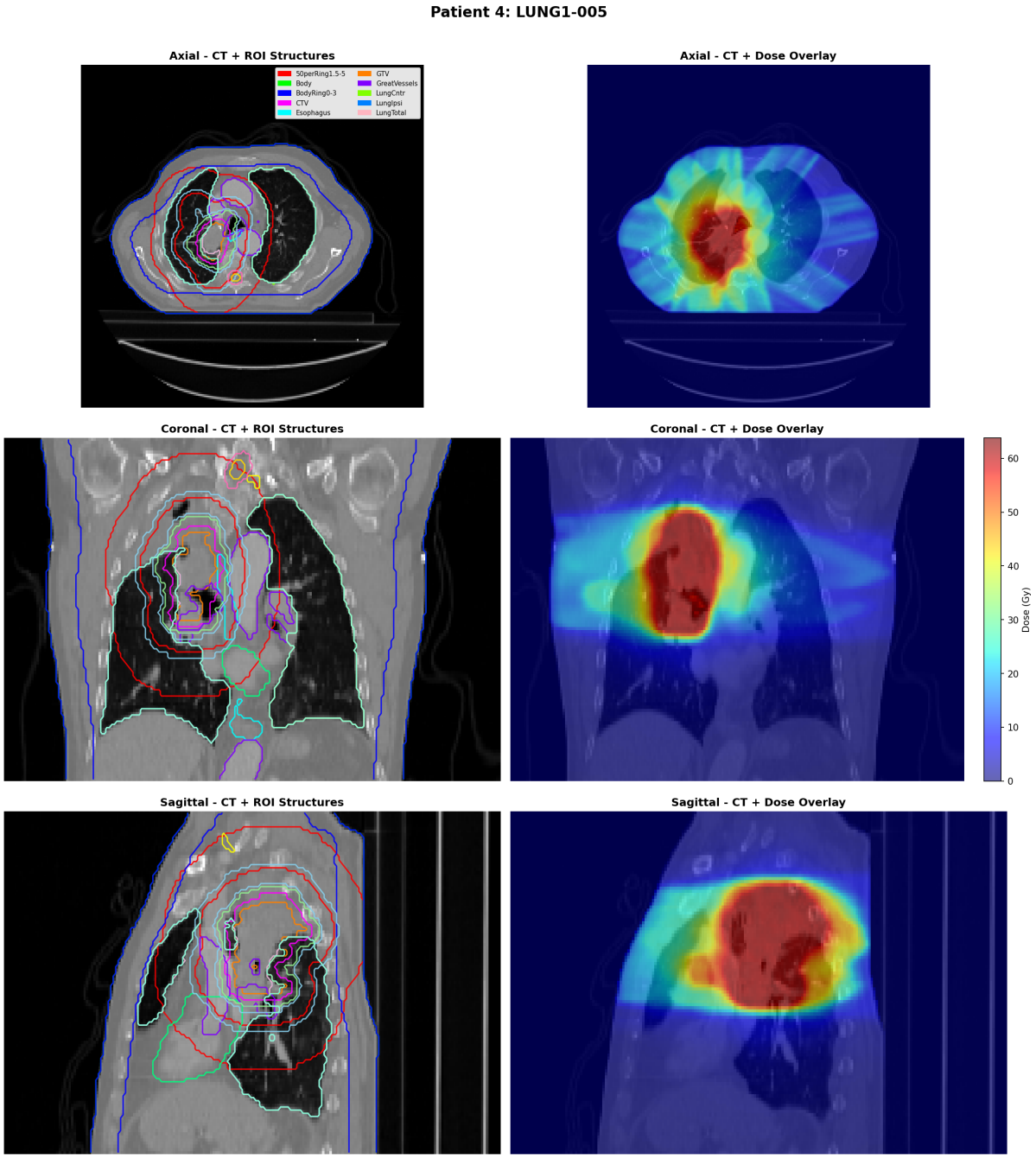


Figure 10.80: Multi-view visualization of patient anatomy with CT imaging and RTSTRUCT overlay. The figure demonstrates the spatial relationship between the planning target volume (PTV), clinical target volume (CTV), gross tumor volume (GTV), and critical organs at risk including the heart, spinal cord, esophagus, and lungs. The color-coded structure contours enable precise geometric verification of treatment planning constraints and anatomical coverage across axial, sagittal, and coronal planes.

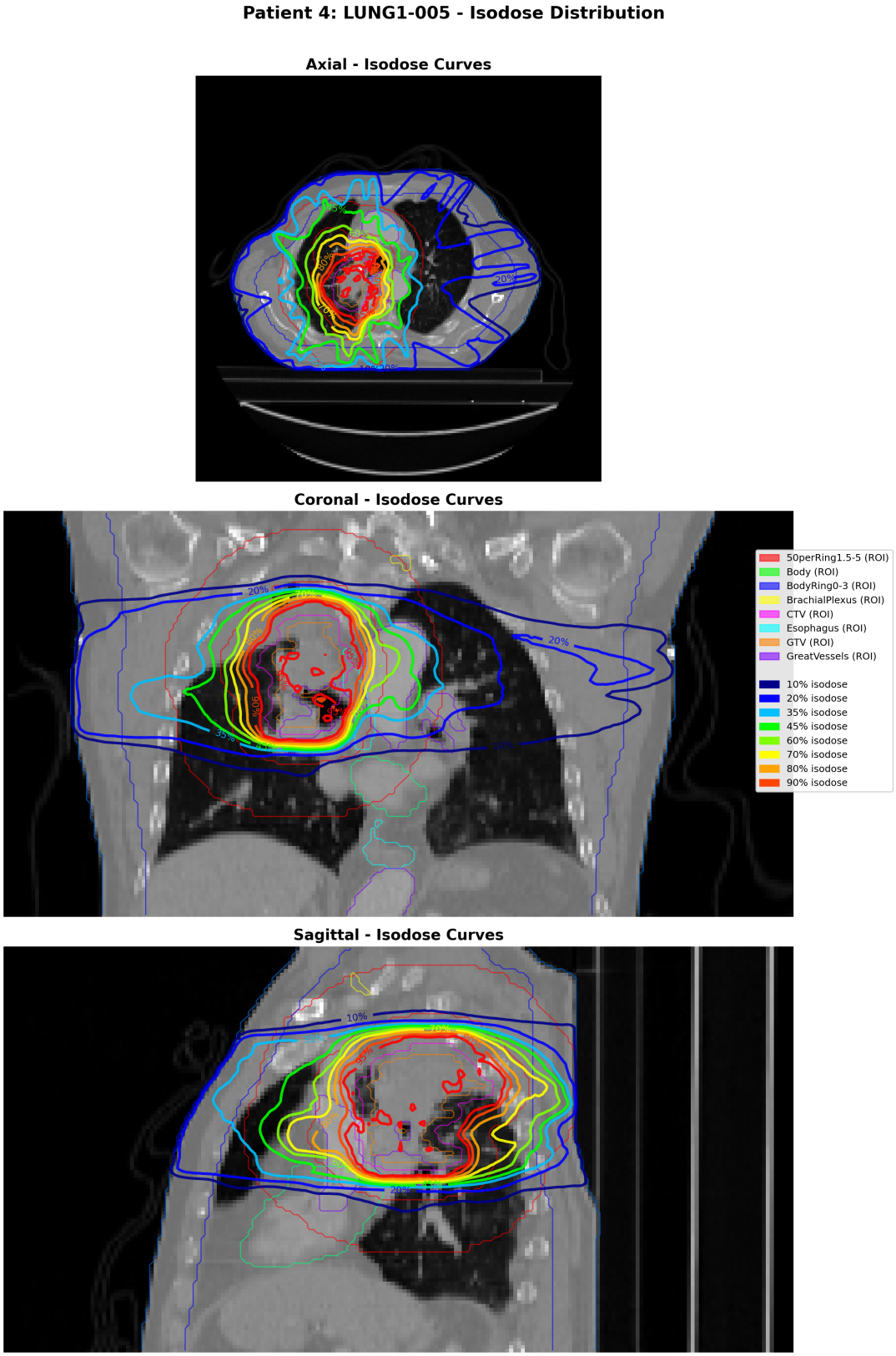


Figure 10.81: Isodose distributions are visualized as isodose lines, which represent points where radiation absorbed by tissue will be the same, illustrated with red representing a higher dose of radiation; approaching the prescribed dose, down to blue representing lower doses of radiation; or spillage. The isodose lines demonstrate how quickly dose drops off, near the PTV periphery; this indicates a very conformal treatment using modern VMAT techniques. The dose can vary considerably throughout the lung tissue due to the charged particle disequilibrium present in soft-tissue. The 95% isodose line (usually shown as green), represents the point of acceptable target coverage for clinical purposes, and the 50% and 20% isodose volumes represent the volume of intermediate and low dose "bath" delivered to surrounding normal tissues.

10.3 Beam Configuration and Delivery Parameters

The beam geometry for lung cancer treatments is adapted based on tumor laterality. The lateral distance from the isocenter to the patient's mid-sagittal plane determines the selection among three angle configurations: if this distance exceeds 5 cm, a lateralized template (left or right) is selected depending on tumor location; otherwise, a full-range angular configuration is employed. For VMAT delivery, two co-planar arcs are utilized with alternating gantry rotation directions (clockwise and counterclockwise). For IMRT delivery, 7 beam angles are used for lateralized tumors and 9 angles for centrally located tumors.

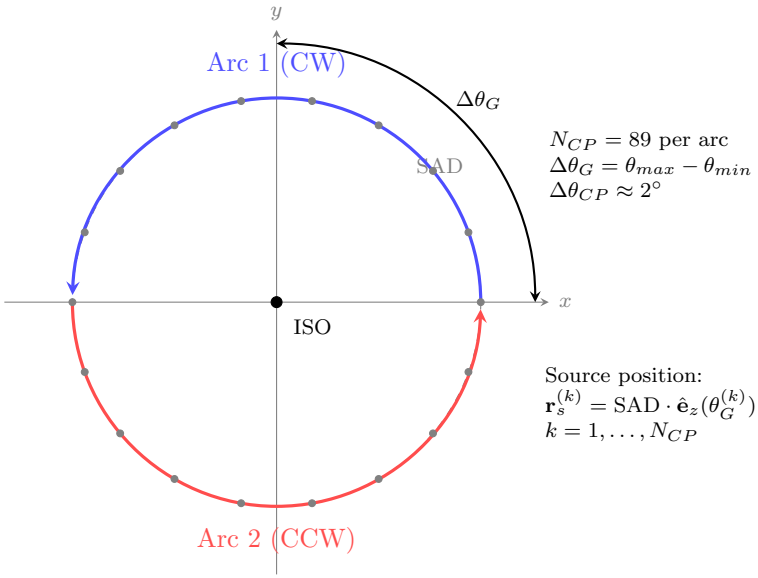


Figure 10.82: Dual-arc VMAT configuration for lung cancer treatment. Arc 1 rotates clockwise (CW) while Arc 2 rotates counterclockwise (CCW). Each arc comprises approximately 89 control points, yielding a total of 178 control points per plan. The angular spacing between consecutive control points is approximately 2°.

In addition to dynamically shaping the radiation field as it moves with the MLC, the MLC motion also allows for optimization of both OAR sparing and target coverage simultaneously at all gantry angles. The BEV perspective shows how the radiation field relates to the critical anatomy. For example at gantry angle 171° (Figure 10.84), the MLC are configured to block the cardiac silhouette and still deliver adequate PTV dose, which minimizes cardiac dose and reduces the risk of radiation-induced cardiac toxicity. Geometric avoidance is very important now, since there has been an increasing amount of data showing mean heart dose is associated with increased cardiac mortality after thoracic radiotherapy.

Similarly, at gantry angle 215° (Figure 10.85), the MLC have created a concave field shape that surrounds the spinal cord projection, taking advantage of the spatial separation between the PTV and this critical serial organ. The maximum dose constraint for the spinal cord (typically $D_{0.1cc} < 22$ Gy for conventional fractionation) necessitates stringent geometric avoidance, as exceeding tolerance doses risks permanent myelopathy. The collimator angle of 358.0° in this control point aligns the MLC leaf direction nearly parallel to the spinal cord axis, maximizing the shielding efficiency of the tungsten leaves.

Similarly, at gantry angle 275° (Figure 10.86), the MLC provides esophageal sparing. The esophagus, as a tubular organ traversing the mediastinum, presents a challenging optimization problem due to its proximity to the PTV in many lung cancer cases. The aperture at this angle creates a partial block over the esophageal projection, accepting reduced PTV coverage from this beam direction in exchange for maintaining esophageal mean dose below clinically acceptable

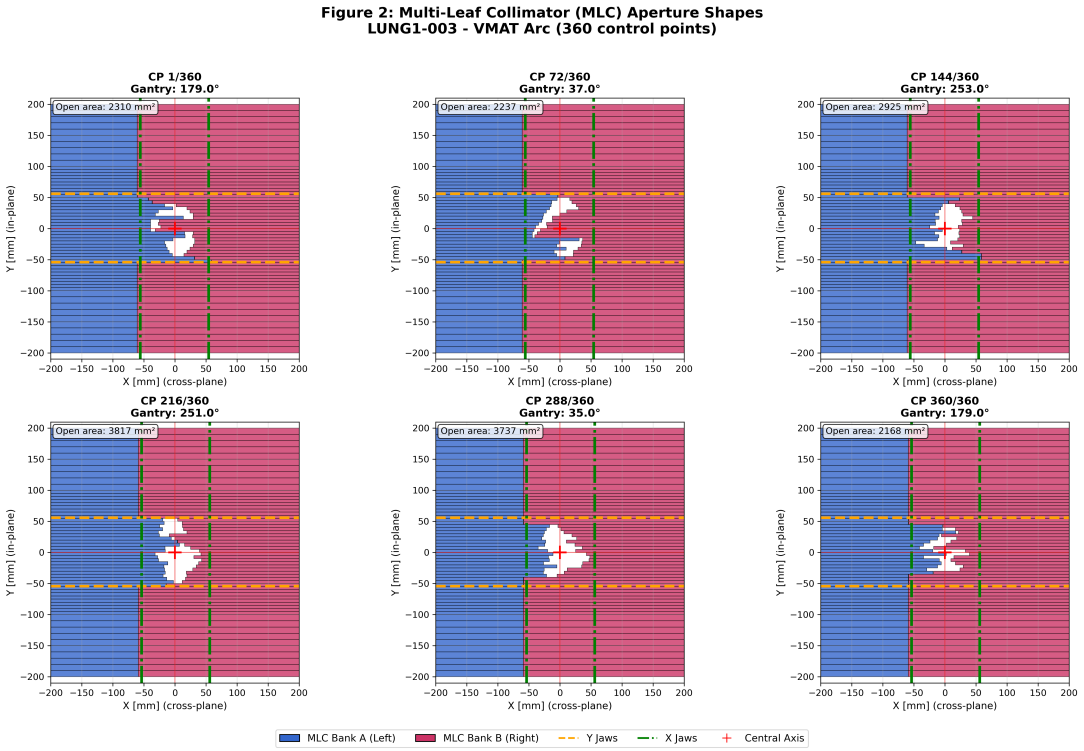


Figure 10.83: Multi-leaf collimator (MLC) aperture configuration during VMAT delivery. The MLC consists of tungsten alloy leaves (typically 120 leaves, 60 pairs) that translate perpendicular to the beam axis to create irregularly shaped field apertures conforming to the beam's-eye-view projection of the target volume. Each leaf pair is independently controlled with sub-millimeter precision, enabling dynamic aperture modulation synchronized with gantry rotation. The interleaf leakage and leaf transmission (typically <2% for modern MLCs) are explicitly modeled in the dose calculation to ensure accurate prediction of penumbral regions and low-dose spillage.

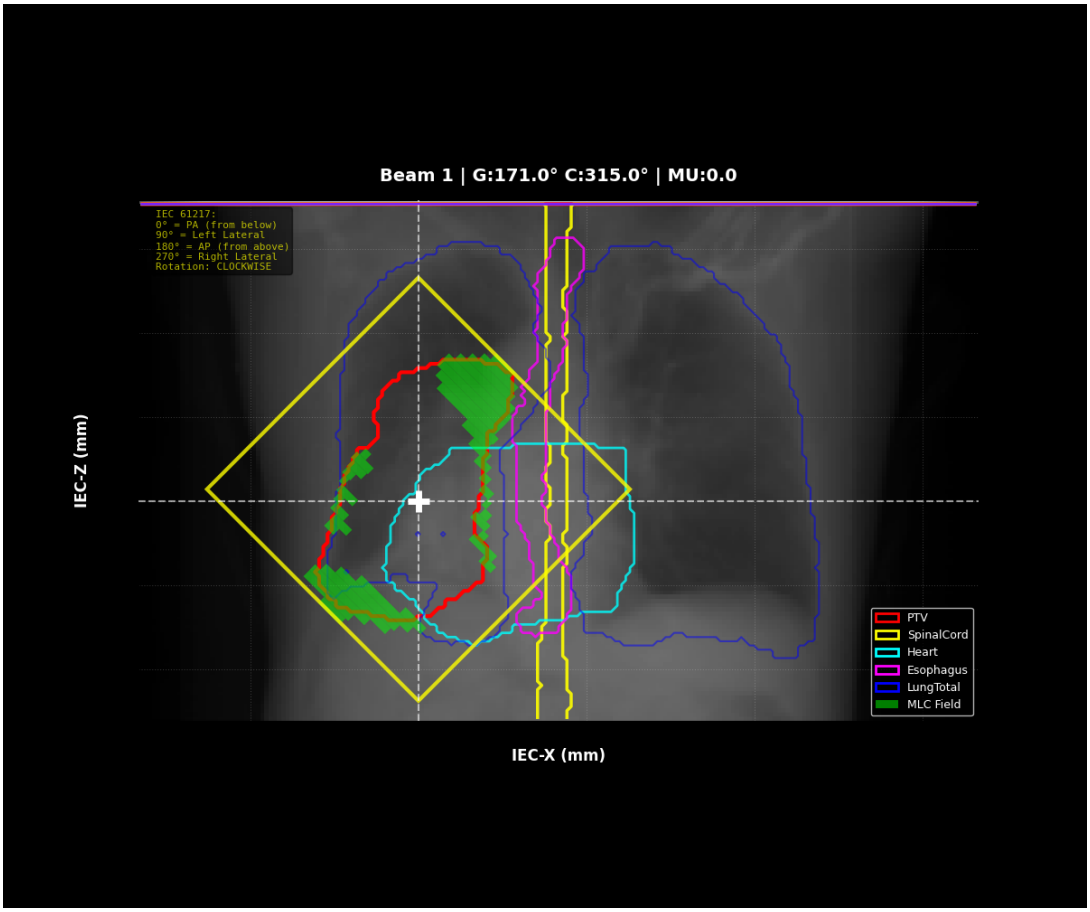


Figure 10.84: MLC aperture configuration at gantry angle 171° demonstrating cardiac avoidance. The multi-leaf collimator blocks the projection of the heart (shown in the BEV as the magenta contour) while the green aperture region conforms to the PTV outline. The collimator rotation of 2.0° optimizes the leaf orientation to minimize interleaf leakage into cardiac tissue. As this control point demonstrates, the principle of geometric modulation is that when the organ(s) at risk is positioned such that it will be directly exposed by the radiation beam due to its angular relationship to the beam's axis, those areas can be blocked by closing the field.



Figure 10.85: Spinal cord avoidance through MLC shaping at gantry angle 215°. The "green" aperture eliminates the spinal cord projection ("yellow") and maintains a conformal coverage of the "PTV." The intensity modulation for the beam orientation shown has been diminished ("smaller aperture area relative to PTV") to reduce the risk of overdosing the spinal cord due to leakage radiation; however, it illustrates the basic trade-offs inherent in VMAT optimization – that there are specific beam orientations which contribute to OAR sparing as opposed to target dose escalation.

thresholds (typically < 34 Gy to minimize grade ≥ 3 acute esophagitis). The collimator rotation of 358.0° ensures optimal leaf alignment for this tubular structure geometry.

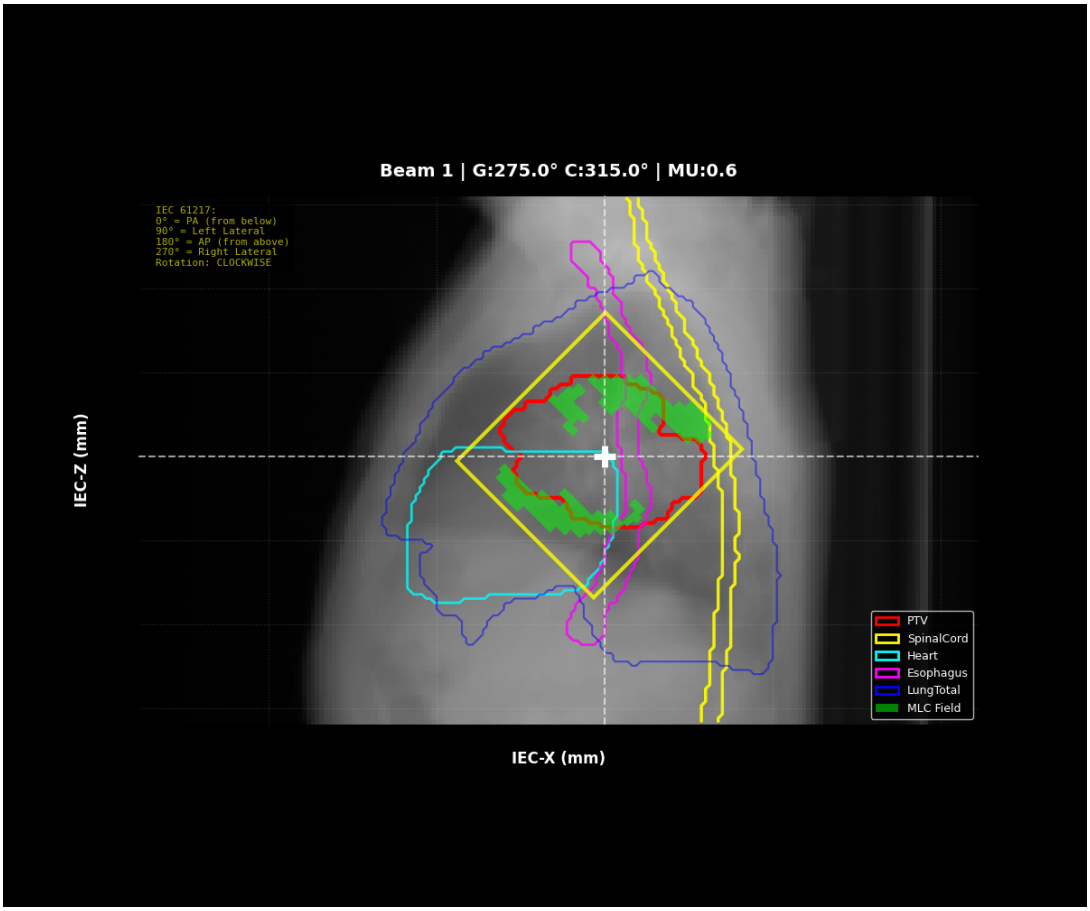


Figure 10.86: Esophageal protection via MLC aperture modulation at gantry angle 275°. The field shape blocks the esophageal projection (orange contour) while partially covering the PTV. The dose contribution from this control point is intentionally reduced to avoid exceeding esophageal tolerance, with the deficit in PTV dose being compensated by contributions from other gantry angles where the geometric relationship between target and OARs is more favorable. Therefore, this angular-specific intensity modulation illustrates one of the volumetric-optimization principles inherent to VMAT.

On the other hand, at gantry angles, where OARs have a geometric separation from the target, the MLC aperture is expanded to allow it to cover all of the PTV projection area, thereby increasing dose delivery efficiency. As shown in Figure 10.87 illustrates this at gantry angle 181° for patient 006, the MLC aperture has expanded to cover the full PTV with a minimum of OAR overlap visible within the beam's eye view. The collimator angle of 2.0° at this location is essentially zero, therefore there should be little concern with the orientation of the leaves of the MLC, since there is little need for geometrical avoidance of the OARs. Such configurations contribute the majority of the therapeutic dose while maintaining the dosimetric advantages of arc delivery (reduced monitor units, shortened treatment time, improved conformality compared to static-field IMRT).

Collectively, the behavior of the 178 control points per VMAT plan produce a conformal three-dimensional dose distribution that addresses multiple competing goals: PTV coverage($D_{95\%} > 95\%$ of prescription dose), OAR dose constraints (as specified in Table 10.20), and dose gradient optimization (the rate of dose fall-off outside of the PTV to limit the total body dose). The combination of MLC-based intensity modulation with continuous gantry rotation

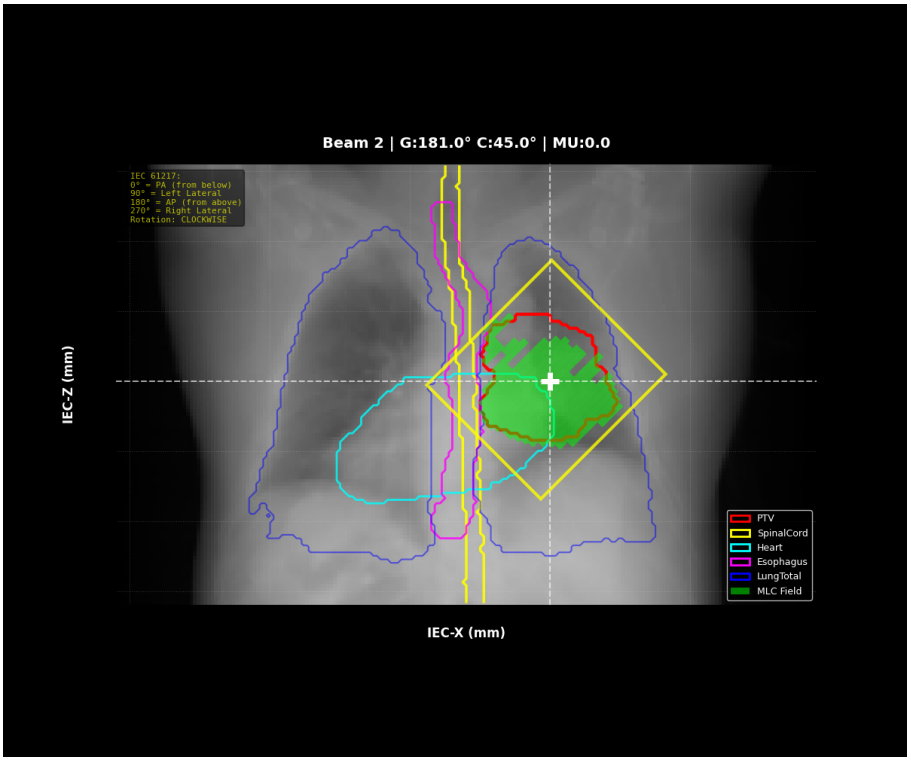


Figure 10.87: Full target coverage is shown at a gantry angle of 181° (patient 006). The MLC aperture completely covers the PTV projection, and there are no critical OARs in the beam path. This configuration is representative of one of the higher-weight control points in the VMAT optimization, and delivers substantial dose per monitor unit. The MLC aperture is symmetrical and the gantry has minimal rotation due to no geometric avoidance required at this angle.

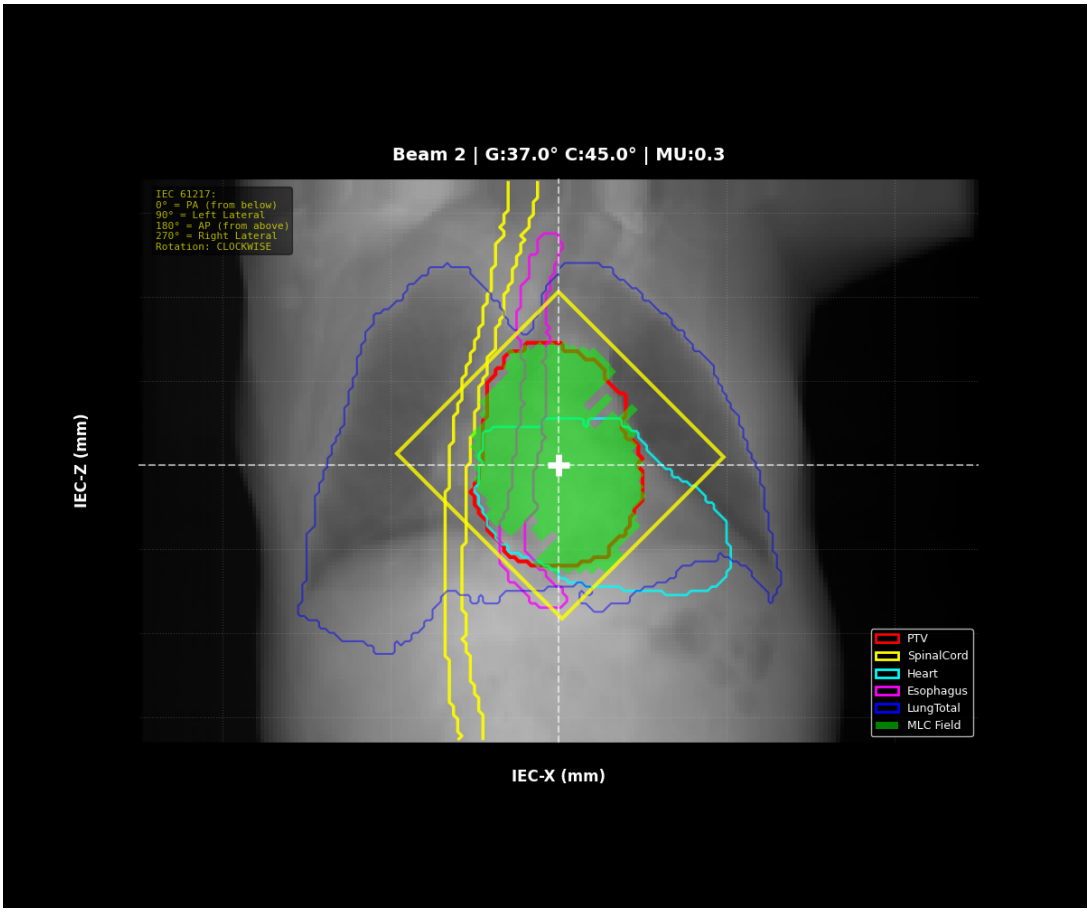


Figure 10.88: Full PTV coverage at gantry angle 37° (patient 032). Similar to the previous figure, this control point maximizes target irradiation with the aperture conforming to the entire PTV extent. The collimator rotation of 91.0° in this case aligns the MLC leaves with the elongated axis of the target to optimize conformality and reduce leaf travel. The variation in optimal collimator angle across different patients and gantry positions demonstrates the patient-specific nature of VMAT optimization, where geometric parameters are tuned to the unique anatomical configuration of each case.

and varying dose rates, allows for multi-objective optimization of the dose distribution using automated inverse plan algorithms that continuously adjust the aperture shape, leaf position and weight of monitor units until all clinical goals have been met while respecting the limits of the delivery hardware used.

10.4 Organs at Risk and Helper Structures

The OAR set for lung cancer planning includes critical thoracic structures with associated dose constraints. Table 10.20 presents the initial optimization objectives employed in the automated planning pipeline.

Table 10.20: Initial optimization objectives for lung cancer VMAT planning (adapted from the AIRTP pipeline). ROI denotes Region of Interest; the Operation column specifies whether the objective is a lower bound (target coverage) or upper bound (OAR sparing).

ROI	Type	Dose (Gy)	Volume (%)	Operation	Priority
PTV	Point	60	100	Lower	300
PTV	Point	63	0	Upper	300
BodyRing	Point	63	0	Upper	250
TotalLung-GTV	Point	20	18.5	Upper	100
TotalLung-GTV	Point	5	32.5	Upper	100
TotalLung-GTV	Mean	5	–	Upper	100
SpinalCord	Point	22	0	Upper	80
Heart	Point	55	0	Upper	60
Heart	Mean	5	–	Upper	60
LAD	Point	55	0	Upper	60
Esophagus	Mean	15	–	Upper	60
BrachialPlexus	Mean	15	–	Upper	60
GreatVessels	Mean	55	–	Upper	60
Trachea	Mean	55	–	Upper	60
RingPTV	Mean	67	0	Upper	60

10.5 Plan Quality Evaluation Metrics

Plan quality is assessed using a comprehensive scorecard system that evaluates both target coverage and OAR sparing. Table 10.21 presents the scorecard metrics employed for lung cancer plan evaluation.

Table 10.21: Scorecard metrics for lung cancer plan quality evaluation. The Metric Type column indicates the dosimetric parameter: VolAtDose (volume receiving at least the specified dose), DoseAtVol (minimum dose to the hottest specified volume fraction), and MeanDose (arithmetic mean dose to the structure).

Structure	**Metric Type**	**Input Parameter**
PTV	VolAtDose	$V_{100\%}$
CTV	VolAtDose	$V_{100\%}$
PTV	VolAtDose	$V_{105\%}$
PTV	DoseAtVol	$D_{95\%}$
PTV	DoseAtVol	$D_{99\%}$
TotalLung-GTV	VolAtDose	$V_{20\text{Gy}}$
TotalLung-GTV	VolAtDose	$V_{5\text{Gy}}$
TotalLung-GTV	MeanDose	D_{mean}
SpinalCord	DoseAtVol	$D_{0.1cc}$
Heart	DoseAtVol	$D_{0.1cc}$
Heart	MeanDose	D_{mean}
LAD	VolAtDose	$V_{15\text{Gy}}$
Esophagus	DoseAtVol	$D_{0.1cc}$
Esophagus	MeanDose	D_{mean}
BrachialPlexus	MeanDose	D_{mean}
GreatVessels	MeanDose	D_{mean}
Trachea	MeanDose	D_{mean}

10.6 Dosimetric Characteristics

Beginning with the Text 10.6 Dosimetric Characteristics The range of prescribed doses for lung cancer treatment using this dataset can be anywhere from 48 Gy to 66 Gy, based upon both disease stage and clinical intent. The ability to prescribe variable doses in a treatment plan is an improvement over many other fixed dose prescribing datasets that may only allow for one prescribed dose for all patients treated for lung cancer; this variability allows clinicians to better tailor their treatment plans to meet each individual's specific needs, including the patient's performance status, the size, location and aggressiveness of the primary tumor and the patient's overall prognosis or intent to cure verses treat palliatively.

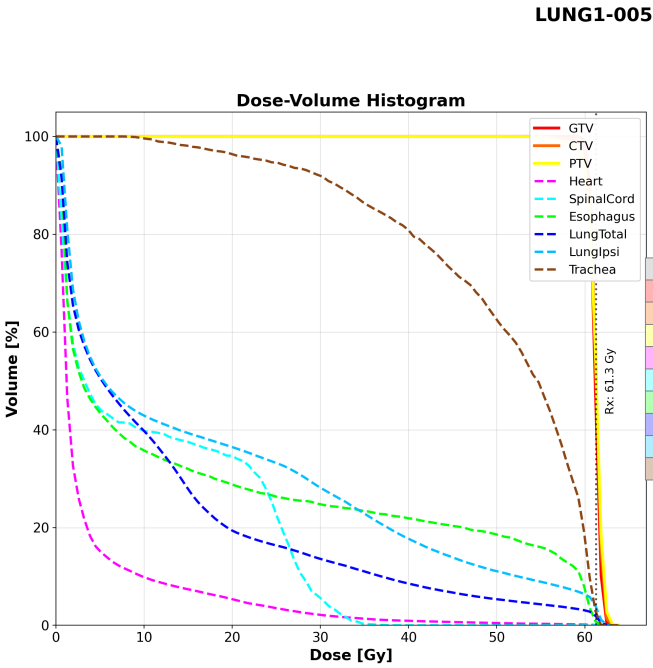
10.6.1 Organ-at-Risk Radiosensitivity and Molecular Mechanisms

The dose-volume histogram analysis of representative lung cancer cases presents treatment plans with prescribed doses typically in the range of 60–66 Gy to the planning target volume (PTV). The incidental irradiation of critical organs at risk requires comprehensive understanding of their radiosensitivity mechanisms [203]. The heart exhibits late toxicity predominantly through dose-volume dependence rather than intrinsic cellular radiosensitivity [204]. According to QUANTEC guidelines, mean doses exceeding 26 Gy and V_{30} greater than 50% significantly increase pericarditis, arrhythmias, and cardiac mortality risk; however, more recent data suggest that $V_{30} < 15\%$ may be a more appropriate threshold, especially in patients with pre-existing cardiac disease [205]. The underlying toxicity mechanisms involve endothelial injury, chronic inflammation, and mitochondrial oxidative stress-driven myocardial fibrosis.

The spinal cord demonstrates high intrinsic radiosensitivity due to oligodendrocyte and neural precursor cell vulnerability [203]. Doses exceeding 50 Gy risk myelopathy, as these cells exhibit compromised DNA damage response (DDR) mechanisms. The esophagus presents dose-volume-dependent acute toxicity. Mean doses exceeding 34 Gy and V_{35} greater than 50%

precipitate grade $\geq$ 3 esophagitis, with epithelial stem cells undergoing cell cycle arrest via ATR-p21 induction or p53-mediated apoptosis depending on damage severity [206].

The lungs exhibit complex dose-response relationships wherein V_{20} less than 30–35% and mean lung dose less than 20–23 Gy minimize radiation pneumonitis and progressive fibrosis [203]. The trachea, while generally radioresistant, may approach stenosis risk when receiving maximum doses exceeding 60 Gy, despite tracheal injury thresholds typically exceeding 55 Gy under conventional fractionation.



DVH Metrics Summary
(Dose in Gy, Volume in %)

Structure	Vol [cc]	D_mean	D_max	D95	D5	V20%	V30%
GTV	84.7	61.2	62.8	60.4	62.0	100	100
CTV	193.2	61.3	63.6	60.3	62.2	100	100
PTV	349.3	61.2	63.8	59.8	62.3	100	100
Heart	1029.0	3.8	62.6	0.4	20.8	5	2
SpinalCord	26.0	11.1	37.7	0.4	30.4	35	6
Esophagus	52.1	16.8	62.2	0.6	60.3	29	25
LungTotal	3537.4	12.3	63.4	0.5	51.6	19	14
LungIpsi	1706.1	17.7	63.4	0.8	61.0	36	28
Trachea	26.0	49.9	62.1	24.0	61.0	96	92

Figure 10.89: Dose-volume histogram (DVH) analysis for a sample of a representative lung SBRT patient (Patient 003) showing the cumulative DVH curves which represent the volumetric dose distribution to both the target volume(s) and organs at risk. The red PTV curve shows that there is conformal dose coverage to the prescription dose with $D_{95\%} > 95\%$ of prescription dose, or more of the target volume. Also, note the steep dose drop-off in the DVH between 95% and 107% indicating very little hotspot volume within the target. Additionally, all critical organ at risk (OAR) dose limits are met by the DVH <8 Gy (blue curve), heart mean dose <10 Gy (orange), spinal cord maximum dose $D_{0.1cc} < 18$ Gy (green), and esophageal mean dose <15 Gy (cyan). The vertical dashed line at 60 Gy represents the isodose surface for the prescribed dose. The DVH is the mathematical representation of plan quality that can be used to compare and evaluate quantitatively different treatment options as well as verify the clinical dose limits that have been determined using normal tissue complication probability (NTCP) models.

At the cellular-molecular level, organ radiosensitivity is governed by three fundamental mechanistic pathways. First, ionizing radiation induces DNA double-strand breaks (DSBs) detected by ATM protein kinase through MRN complex recruitment [207]. Autophosphorylated ATM activates CHK2 and p53, orchestrating either cell cycle arrest via p21-CDK inhibition permitting DSB repair through non-homologous end joining (NHEJ) or homologous recombination (HR), or commitment to apoptosis through p53-mediated BAX/PUMA/NOXA transcription and caspase cascade activation [206]. Tissues with high stem cell content, comprising rapidly dividing cells, prioritize replication over repair fidelity, exhibiting exquisite radiosensitivity (doses of 1–2 Gy proving lethal), whereas terminally differentiated tissues suffer primarily late toxicity through mitochondrial mechanisms.

Second, p53 stabilization and phosphorylation by ATM and phosphatidylinositol 3-kinase-like kinase (PIKK)-family kinases (including DNA-PK) determines cellular fate [208]. Pulsed

p53 oscillations are likely to be favorable for the cell's ability to undergo repair and survive, however; elevated levels of p53 that are sustained will lead to an intrinsic form of apoptosis by way of the mitochondrial outer membrane permeability (MOMP) and caspase activation due to the release of cytochrome C. The increase of genomic instability and late complications will result if there is a disruption of normal p53 regulation. Sustained p21 induction can cause cells to become senescent which is characteristic of late effects on fibrosis.

Thirdly, ionizing radiation will disrupt the mitochondrial electron transport chain and the cell's ability to synthesize ATP, as well as generate excessive amounts of reactive oxygen species (ROS). These secondary ROS will lead to oxidative damage to bases in DNA, activate inflammatory transcription factors such as (NF-κB, STAT3), increase production of fibrotic cytokines such as (TGF-β), and cause disturbances in the normal calcium homeostasis of the cell - all of these mechanisms will create a state of chronic oxidative stress and inflammation in metabolically active tissues such as the heart, lung endothelium and neural tissue, which will explain why organs that have low intrinsic radiosensitivity but high metabolic demands suffer mostly from late complications.

In addition, the heterogeneity of the dose distribution seen in thoracic radiotherapy represents conflicting biologic imperatives: the rapid DNA damage response and apoptosis in rapidly proliferating cell populations (esophagus, bone marrow) vs. delayed mitochondrial oxidative-stress-induced fibrosis in slowly proliferating, metabolically active cell populations (lung, heart), with the clinical manifestations occurring anywhere from weeks after irradiation (acute) to many years post-irradiation (progressive late). This framework represents a multi-mechanism approach that supports the need for developing risk-adjusted treatment plans that take into account both the dose-volume histograms and the underlying cellular pathobiology.

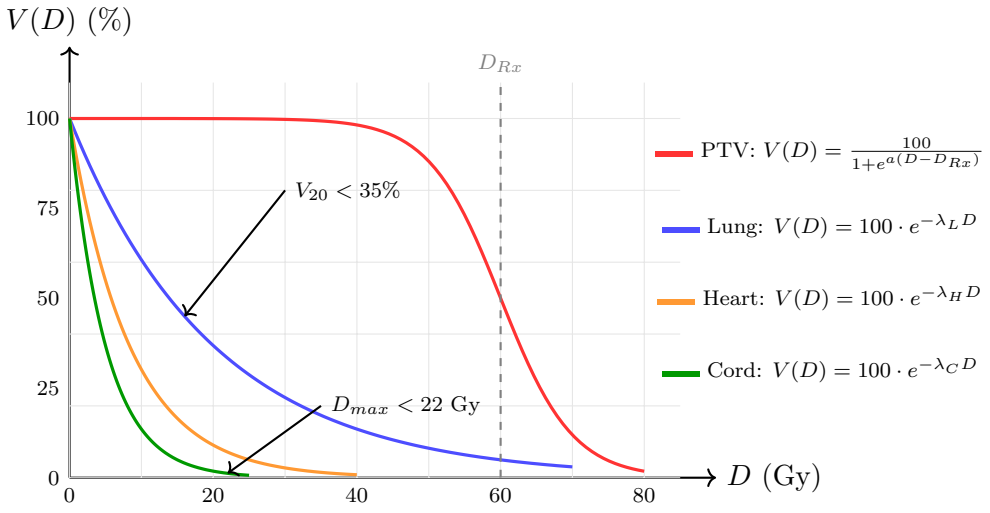


Figure 10.90: Schematic representation of dose-volume histogram (DVH) curves for lung cancer treatment. The PTV curve follows a sigmoid function centered at the prescription dose D_{Rx}, while OAR curves decay exponentially with structure-specific attenuation coefficients λ. Clinical constraints (e.g., lung $V_{20} < 35\%$, spinal cord $D_{max} < 22$ Gy) are indicated.

10.7 Density Heterogeneity and Transport Considerations

The lung cancer dataset presents significant challenges for dose calculation algorithms due to the pronounced density heterogeneity encountered in thoracic anatomy. The mass density ρ within the treatment field spans approximately two orders of magnitude, from $\rho \approx 0.001$ g/cm^3 for air in the trachea and bronchi to $\rho \approx 1.8$ g/cm^3 for cortical bone in the vertebral column and ribs.

The lung parenchyma presents the most challenging region for dose calculation, with typical densities in the range $\rho \in [0.2, 0.4]$ g/cm^3 depending on inflation state and respiratory phase. In these low-density regions, the continuous slowing-down approximation (CSDA) range of secondary electrons scales inversely with density:

$$R_{CSDA}(\rho) = \frac{R_{CSDA}(\rho_0)}{\rho/\rho_0} = R_{CSDA}(\rho_0) \cdot \frac{\rho_0}{\rho}, \tag{10.216}$$

where $\rho_0 = 1$ g/cm^3 is the reference water density. For lung tissue with $\rho \approx 0.3$ g/cm^3, the electron range is approximately three times larger than in unit-density tissue, leading to significant lateral spread and electronic disequilibrium effects.

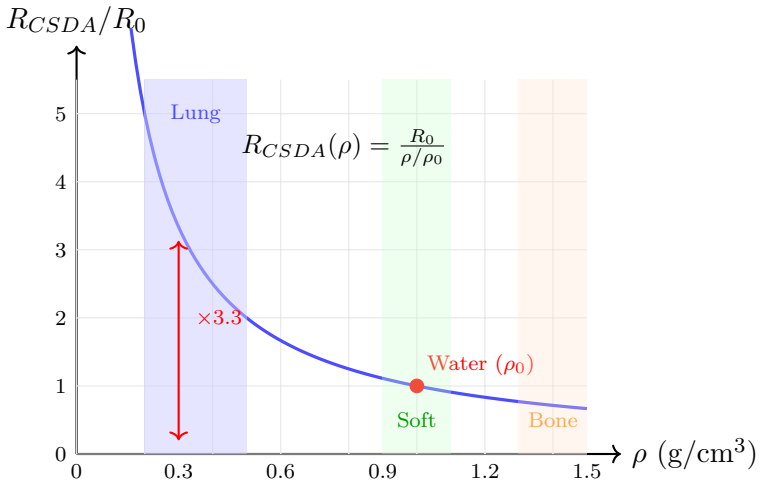


Figure 10.91: Relationship between mass density ρ and the normalized CSDA electron range R_{CSDA}/R_0. The inverse proportionality results in significantly extended electron ranges in low-density lung tissue (shaded blue region), necessitating accurate transport modeling for dose calculation.

Chapter 11 Physics-Informed Dose Prediction via Hybrid 3D U-Net and Fourier Neural Operator Architecture

Physics-Informed Dose Prediction Using Hybrid 3D U-Net and Fourier Neural Operator Architecture The previous sections outlined the use of the deterministic Linear Boltzmann Transport Equation (LBTE) solver as a physical method to determine dose distributions within a heterogeneous medium. As noted previously, while this method provides clinically acceptable results it has the disadvantage of being computationally expensive due to the number of iterations required for the transport sweep. Therefore, this section will introduce a new, hybrid deep learning architecture to provide dose predictions using a physically relevant basis of radiation transport but with reduced computational cost. The main advantage of the proposed architecture is the ability to learn the physics of the problem; specifically, the intermediate quantities calculated in the early stages of solving the Boltzmann equation, namely the uncollided fluence and the total energy released per unit mass are used as a physically relevant feature space for a neural network to further learn the residual effects of photon scattering and secondary electron transport on the dose distribution.

11.1 Physical Foundation: From Uncollided Fluence to Dose Prediction

Solving the coupled photon-electron transport equations deterministically involves several computational stages. The first stage involves determining the uncollided photon fluence $\Psi^{\gamma}_{\text{unc}}$, which represents the angular fluence of primary photons that have not been scattered since they were emitted from the treatment head. The analytical determination of the uncollided fluence is accomplished through ray tracing from each beamlet source location throughout the patient volume accounting for exponential attenuation in accordance with the Beer-Lambert law. The uncollided fluence at position $\mathbf{r}$ from a point source at $\mathbf{r}_p$ emitting photons with energy spectrum $q^{\gamma}(E)$ is given by:

$$\Psi^{\gamma}_{\text{unc}}(\mathbf{r}, E, \Omega) = \delta(\Omega - \Omega_{\mathbf{r}\leftarrow\mathbf{r}_p}) \, \frac{q^{\gamma}(E)}{4\pi \, \|\mathbf{r} - \mathbf{r}_p\|^2} \, e^{-\tau(\mathbf{r},\mathbf{r}_p)} \tag{11.217}$$

where $\Omega_{\mathbf{r}\leftarrow\mathbf{r}_p}$ is the unit direction vector from source to field point, the inverse-square factor accounts for geometric divergence of the photon beam, and the optical depth $\tau(\mathbf{r}, \mathbf{r}_p) = \int_{\mathbf{r}_p}^{\mathbf{r}} \sigma_t^{\gamma}(\mathbf{s}, E) \, ds$ represents the cumulative attenuation along the ray path from source to observation point. This equation illustrates the main physics of light transport: The amount of light decreases due to geometry (light divergence) with the square of the distance from the source of light, and due to absorption (attenuation) of the light on its way through the tissue. Total Energy Released per unit Mass (TERMA) extends this concept using the fluence weighted by the mass energy-transfer coefficient:

$$\text{TERMA}(\mathbf{r}, E) = \Phi_{\text{unc}}(\mathbf{r}, E) \cdot \frac{\mu_{\text{tr}}(\mathbf{r}, E)}{\rho(\mathbf{r})} \tag{11.218}$$

where $\Phi_{\text{unc}} = \int_{4\pi} \Psi^{\gamma}_{\text{unc}} \, d\Omega$ is the scalar uncollided fluence and μ_{tr}/ρ is the mass energy-transfer coefficient. TERMA defines the total radiative energy that is converted into kinetic energy of charged particles per unit mass at each point in space, and can be used as an approximation to absorbed dose in the case of charged particle equilibrium.

The important realization behind this work is that together Φ_{unc} and TERMA represent a physically realistic "picture" of what one would expect to see as a dose distribution. The uncollided fluence represents the geometrical pattern of how primary radiation is delivered – the beam shape, gantry angle, and multileaf collimator aperture(s) that are part of the treatment plan – while TERMA includes the energy transfer from the incident radiation to the tissue which begins the process of delivering dose. In other words, though we have considered

these physical characteristics, they do not represent the only two significant physical characteristics that exist when a photon initially interacts. There is another important characteristic of physical nature that occurs subsequent to the initial interaction of the photon, i.e. photon scattering. This results in an additional deposition of dose (energy) over and above what would be deposited at the location of the primary interaction, due to the re-distribution of energy from the scattered photon, and depends upon the energy of the photon as well as the type of tissue it travels through. The final characteristic of physical nature is the secondary electron transport; the charged particles produced by photon interactions will travel some distance before depositing their kinetic energy, and the way in which these secondary electrons move from the point of production to the point of deposition (where the kinetic energy of the secondary electrons is transferred to the medium), is determined by the continuous slowing-down approximation and elastic scattering physics, and thus defines the detailed spatial distribution of the energy deposited by the secondary electrons, especially at interfaces between tissues (charged particle equilibrium does not exist). In the full Boltzmann transport solution, the characteristics described above are calculated by iteratively solving the collided fluence equations and the coupled electron transport equation. Iterative sweeps are used to propagate scattered photons and transport electrons through the discrete ordinates directions until convergence, and this iterative sweep is responsible for the majority of the computation time associated with the deterministic calculation of dose. The novelty of this work is that the authors have found a sufficiently designed neural network that learns to predict the result of the iterative transport calculations from the physics-based input features of the problem, essentially substituting the expensive iterative solver with a single forward pass of the trained network.

11.2 Novelty of the Proposed Approach

The proposed method is fundamentally different than previous machine learning approaches to dose prediction in radiotherapy. From a thorough literature search, there are virtually no prior studies employing input representations that incorporate radiation transport physics. Instead, most previous studies employed anatomical segmentation masks (binary or multi-label volumes defining the planning target volume, organs at risk, and body contour) and geometric features (such as signed distance transforms from structure boundaries). Examples of representative works include the HD-UNet architecture of Nguyen et al.[209], the cascaded 3D U-Net of Nguyen et al.[209], and the attention-gated networks of Osman et al.[210], all of which predict dose from segmentation masks without using any of the quantities that are derived from radiation transport physics.

There are several inherent limitations to relying solely on anatomical features. First, such methods cannot generalize to different treatment planning systems since the mapping from anatomy to dose inherently includes the optimization objectives and constraints of the training institution's planning practice. For example, a network that was trained on plans optimized with one set of clinical protocols cannot accurately predict dose for plans that were optimized using a different set of objectives or constraints, even for anatomically similar patients. Secondly, purely anatomical methods cannot capture changes in beam geometry (number of beams, gantry angles, arc ranges, and fluence patterns) because no beam related information is included in the input. Therefore, a network that was trained on seven-field IMRT plans could not predict dose for VMAT plans without being re-trained and vice versa.

The physics-based approach developed in this work addressed both of these limitations by providing the neural network with inputs that directly encoded the delivery of the radiation. The uncollided fluence Φ_{unc} encodes all of the beam geometry information (gantry angles, collimator settings, and fluence modulation patterns) because these factors determine how primary photons travel through the patient volume. The TERMA distribution provides tissue dependent energy transfer information that changes based on local composition, thereby allowing the model to adapt to different anatomical regions without requiring explicit segmentation. Electron density ρ

from CT imaging provides the fundamental material property required for transport calculations and geometric features such as signed distance to isocenter and flux magnitude $|\mathbf{J}|$ provide spatial context that relates to beam divergence and directionality.

The reason that these features may be generalized across treatment planning systems, optimization methods, and delivery techniques is that they were both obtained from the treatment plan RTPLAN file and patient CT data using physics-based ray-tracing, which will perform identically to all treatment planning systems used clinically. Therefore, a neural network trained with VMAT plans from one institution could theoretically compute the doses for an IMRT plan at another institution because the input features to the network accurately model the physics of radiation delivery regardless of how the original plan was developed.

We believe this to be the first time in the literature that the intermediate quantities "uncollided fluence," "TERMA" (and other) "Boltzmann solvers," have been utilized as input features for a neural network dose predictor. Thus, this method is a true innovation and represents a paradigm shift in dose predictors from anatomic based to physics based.

11.3 Interpretation: The Network Learns Scattered Dose

The architectural design of the RadiationDoseNet reflects the physical interpretation that the network learns to predict the scattered and secondary dose contribution that the uncollided fluence and TERMA do not capture. The final dose prediction is formulated as:

$$D_{\text{pred}} = \text{softplus}\left(T + \delta D\right), \quad \delta D = f_\theta(\rho, \Phi_{\text{unc}}, T, d_{\text{iso}}, |\mathbf{J}|) \tag{11.219}$$

where T is the TERMA (extracted from the input features), δD is the correction learned by the neural network f_θ, and the softplus activation ensures non-negative dose values. This residual formulation provides several advantages. The network only learns how to estimate differences between TERMA and actual dose. These differences tend to have lower magnitudes and less spatial variation than the total dose distribution. The most uncertain areas of the dose are generally located near soft tissue boundaries, near the edges of penumbras and in areas with large scatter buildups – all the locations that differ from the TERMA approximation for the 0th order of transport physics. In this way, the residual formulation allows the network to learn in an area of greatest uncertainty in terms of dose, instead of focusing on reproducing the majority of the dose that has already been estimated by the TERMA.

Physically speaking, the correction term δD represents the total effect of all photon scattering contributions from any interaction order (first scatter; second scatter; third scatter, etc.) and all electron transport contributions including range, angular spreading, and the deposition of energy, charged particle equilibrium breaks down at interfaces between tissues of varying densities, and secondary photons produced via electron bremsstrahlung. All of these contributions to the dose are exactly those computed by the iterative transport solver during its sweeps of collided fluence and marching of electron energy. Thus, the network acts as a learned proxy for the computationally expensive iterative component of the transport calculation, while the TERMA serves to provide the physics-based initial conditions for ensuring the correct overall dose scale and spatial location.

11.4 Architecture Design Philosophy

The architecture of the RadiationDoseNet represents a combination of two different paradigms of neural networks for capturing the multi-scaled physics of radiation transport. A U-Net based encoder-decoder is best suited to learning local geometric features of the tissue: the boundary conditions, the density gradients and interface effects which determine electron transport at scales ranging from millimeters to centimeters; whereas a Fourier Neural Operator is better suited to determining the global spectral properties of the dose delivery: the beam geometry, scattering kernels, and long-range energy transport effects present throughout the treatment

volume. The spatial attention fusion layer will adaptively determine how to combine these local and global features.

Radiation dose prediction requires consideration of geometric relationships between structures within the body. To this end, the positional encoding of the input data provides an indication of the relative positioning of structures within the body. This is provided by both a Fourier positional encoding technique that incorporates multiple frequency bands so that the network can represent high frequency spatial patterns (e.g., dose gradients at interfaces) and CoordConv which provides explicit position information (e.g., distance from isocenter and proximity to body surface), which are necessary to allow the network to learn about position dependent phenomena (e.g., beam divergence and inverse square falloff) that it could otherwise not learn due to translation invariance.

The total number of trainable parameters used in the completed architecture is on the order of approximately 1 billion, sufficient for the network to learn the nonlinear mapping of physics-based features into dose correction predictions while maintaining its ability to generalize to new cases appropriately due to proper regularization. Gradient loss was employed in training along with voxel-wise L1 loss, the latter of which penalizes large errors in dose gradients to improve the accuracy of the dose prediction in the penumbra regions and at tissue interfaces where clinical accuracy is most important.

11.5 Input Feature Description

The RadiationDoseNet receives five physics-derived input channels, computed from the patient CT and treatment plan via the same ray-tracing procedures used in clinical treatment planning systems:

Table 11.22: Physics-informed input features for dose prediction.

Symbol	Name	Physical Significance
ρ	Electron Density	Fundamental material property determining photon attenuation and electron stopping power. Derived from CT Hounsfield units via standard conversion.
Φ_{unc}	Uncollided Fluence	Primary photon fluence computed via ray-tracing from treatment head. Encodes beam geometry, fluence modulation, and primary attenuation.
T	TERMA	Energy released per unit mass from primary photon interactions. Combines uncollided fluence with mass energy-transfer coefficient.
d_{iso}	Isocenter Distance	Signed distance from body surface to treatment isocenter. Provides geometric context for beam divergence effects.
$\lvert\mathbf{J}\rvert$	Flux Magnitude	Magnitude of directed photon current $\mathbf{J} = \Phi_{\text{unc}} \cdot \bar{\Omega}$. Indicates directional asymmetry in the radiation field.

Ablation studies on the validation data set were conducted to identify the most appropriate of the fourteen features that would be used as inputs into the models. In order to reduce overfitting risk while providing adequate predictive capability; the input features were identified to include the maximum amount of physical information possible with the least amount of redundant information. An example of how the physics-based input features are illustrated is shown

in Figure 11.92 which demonstrates how each feature captures different physical characteristics related to the radiation field and the patient anatomy for a typical VMAT lung cancer case.

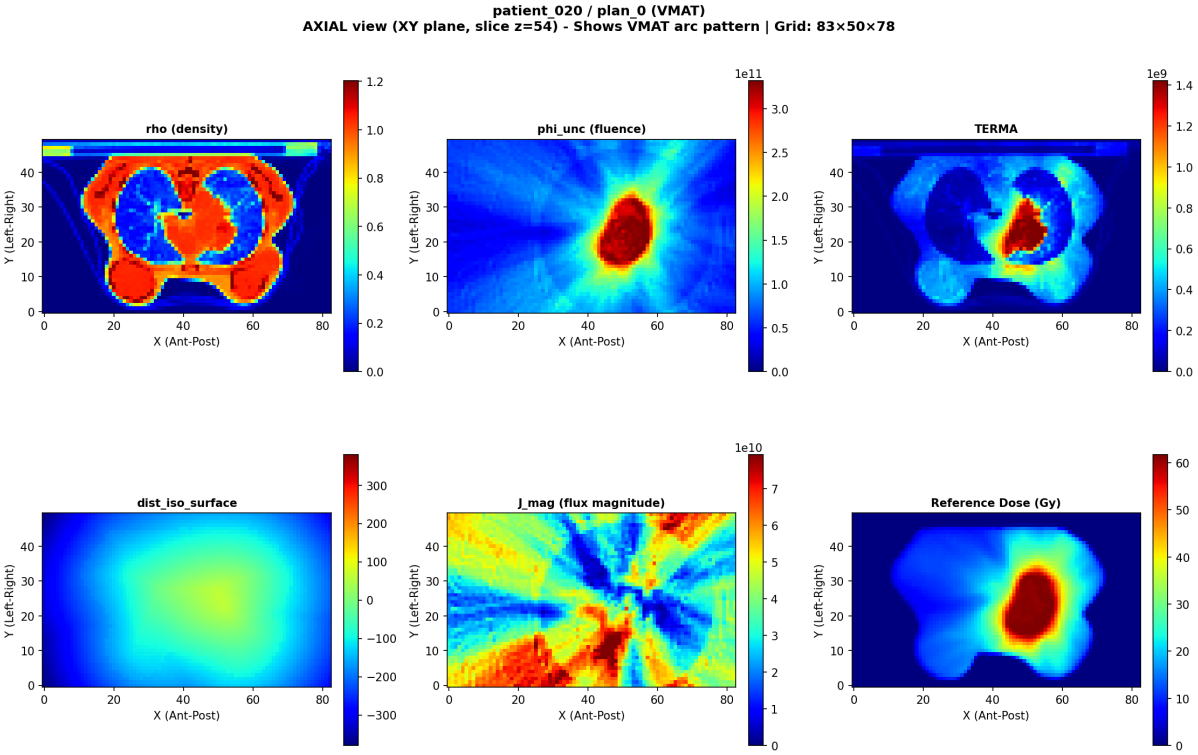


Figure 11.92: Physics-informed input features for a representative VMAT lung cancer case (Patient 020). **Top row**: electron density ρ from CT showing anatomical structures (left), uncollided fluence Φ_{unc} displaying the characteristic VMAT arc pattern (center), and TERMA distribution incorporating tissue-dependent energy transfer (right). **Bottom row**: signed distance to isocenter d_{iso} providing geometric context (left), flux magnitude $|\mathbf{J}|$ revealing directional transport effects (center), and reference dose from clinical TPS for comparison (right). The VMAT arc pattern is clearly visible in both the fluence and flux magnitude channels, demonstrating how physics-derived features encode beam geometry information that purely anatomical methods cannot capture.

11.6 Network Architecture: Dual-Branch Hybrid Design

The proposed architecture combines local feature extraction via an enhanced 3D U-Net with global spectral learning through Fourier Neural Operators (FNO), using spatial attention fusion to combine both pathways. The network predicts dose as a residual correction to TERMA:

$$\mathbf{D} = \text{TERMA} + \delta\mathbf{D}, \quad \delta\mathbf{D} = f_{\text{fuse}}\left(f_{\text{local}}(\mathbf{X}'), f_{\text{global}}(\mathbf{X}')\right) \tag{11.220}$$

where $\mathbf{X}' = f_{\text{PE}}(\mathbf{X})$ denotes the positionally-encoded input features, f_{local} is the enhanced U-Net branch, f_{global} is the FNO branch, and f_{fuse} is the spatial attention fusion module. The final dose is obtained by adding the predicted correction $\delta\mathbf{D}$ to the input TERMA field, with a softplus activation ensuring non-negativity:

$$D_{\text{final}} = \text{softplus}\left(\text{TERMA} + \delta D\right) = \ln\left(1 + e^{\text{TERMA}+\delta D}\right) \tag{11.221}$$

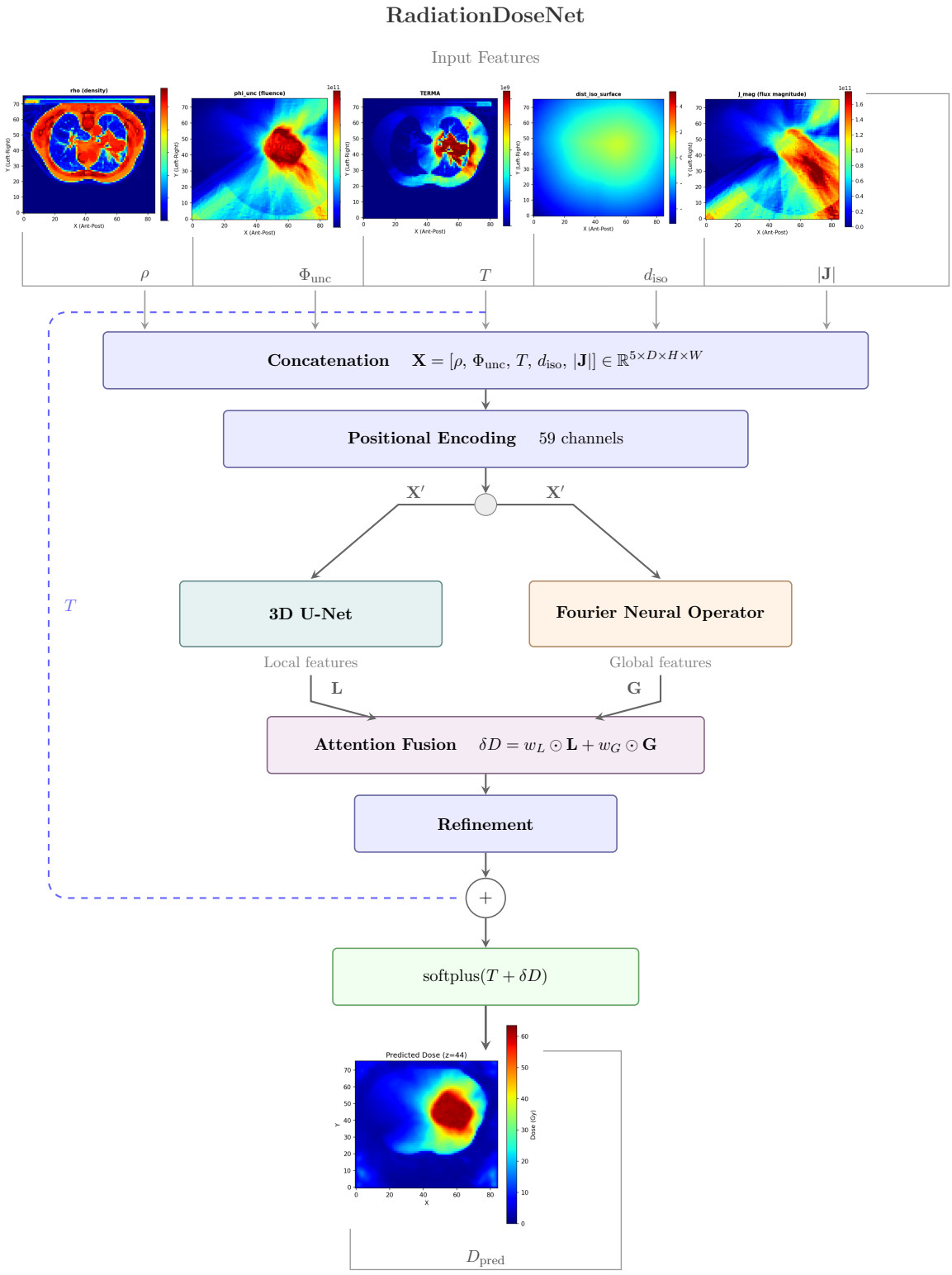


Figure 11.93: RadiationDoseNet architecture. Physics-based inputs (ρ: electron density, Φ_{unc}: uncollided fluence, T: TERMA, d_{iso}: distance to isocenter, $|\mathbf{J}|$: flux magnitude) are processed in parallel by local (3D U-Net) and global (FNO) modules. Outputs are combined using an attention-weighted fusion mechanism with a TERMA residual connection that maintains the physical integrity of the model. Comprehensive architecture of the RadiationDoseNet for physics-based dose prediction: Electron densityρ, uncollided fluence Φ_{unc}, TERMA T, signed distance to isocenter d_{iso}, and flux magnitude $|\mathbf{J}|$. Positional Encoding $\rightarrow$ Process Feature Volume through Parallel Branches; Enhanced 3D U-Net captures local tissue boundaries and Fourier Neural Operator extracts Global Beam Patterns. Spatial Attention Fusion Mechanism Combines Both Representations to Predict Residual Correction δD, Which Is Added To TERMA Via A Skip Connection. Softplus Activation Ensures Non-Negative Dose Output D_{pred}.

11.6.1 Positional Encoding

Positional information about spatial location is crucial to predicting dose because how dose is deposited in a given area is heavily dependent upon the geometric relationship between the dose delivery system (beam source), target (tissue of interest) and tissue boundaries. Two complementary positional encoding methods are used. The CoordConv3d module, after Liu et al.[211], the CoordConv3d module appends normalized Cartesian coordinates $(x, y, z) \in [-1, 1]^3$ as additional input channels:

$$\mathbf{X}_{\text{coord}} = \text{concat}\left(\mathbf{X}, \frac{2i_x - N_x + 1}{N_x - 1}, \frac{2i_y - N_y + 1}{N_y - 1}, \frac{2i_z - N_z + 1}{N_z - 1}\right) \tag{11.222}$$

This provides explicit spatial coordinates to all network layers, enabling position-aware feature learning. To enable learning of high-frequency spatial patterns, we additionally employ Fourier feature encoding [212]. For each spatial coordinate, sine and cosine embeddings at multiple frequency bands are computed:

$$\gamma(p) = \left(\sin(2^0\pi p), \cos(2^0\pi p), \ldots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)\right) \tag{11.223}$$

With $L = 4$ frequency bands applied to three spatial dimensions, the total input dimensionality after positional encoding becomes $C_{\text{in}} = 5 + 3 + 24 + 3 = 35$ to 59 channels depending on configuration.

11.6.2 Enhanced U-Net Branch (Local Features)

The local feature branch employs an enhanced 3D U-Net architecture with 48 base channels (compared to 32 in standard implementations), operating at a depth of 4 encoding levels following the channel progression $48 \to 96 \to 192 \to 384$. Group Normalization with 8 groups is used throughout, paired with GELU activation functions, and all skip connections incorporate attention gates to selectively emphasize informative features.

Each encoder and decoder level uses residual convolutional blocks with Squeeze-and-Excitation (SE) attention [213]:

$$\mathbf{y} = \mathbf{x} + \text{SE}\left(\text{Conv}(\text{Act}(\text{Norm}(\text{Conv}(\text{Act}(\text{Norm}(\mathbf{x})))))\right) \tag{11.224}$$

Skip connections incorporate attention gates that learn to suppress irrelevant features:

$$\alpha = \sigma\left(\psi^T \cdot \text{ReLU}(W_g\mathbf{g} + W_x\mathbf{x} + b)\right), \quad \hat{\mathbf{x}} = \alpha \cdot \mathbf{x} \tag{11.225}$$

At the deepest encoder level, axial self-attention is applied to capture long-range dependencies efficiently. Full 3D self-attention has complexity $\mathcal{O}(n^6)$ for an n^3 voxel grid; axial attention decomposes this into three separate 1D attention operations along each axis with total complexity $\mathcal{O}(3n^4) \ll \mathcal{O}(n^6)$:

$$\text{AxialAttn}(\mathbf{X}) = \frac{1}{3}\left(\text{Attn}_x(\mathbf{X}) + \text{Attn}_y(\mathbf{X}) + \text{Attn}_z(\mathbf{X})\right) \tag{11.226}$$

11.6.3 Enhanced FNO Branch (Global Features)

The global feature branch employs a Fourier Neural Operator (FNO) [214], which learns mappings between function spaces via spectral convolution. This architecture is particularly well-suited for dose prediction because radiation transport inherently involves non-local interactions through photon scattering and electron transport phenomena. The FNO naturally captures these long-range dependencies through global spectral operations, and beam geometry patterns from VMAT arc deliveries are efficiently represented in frequency space.

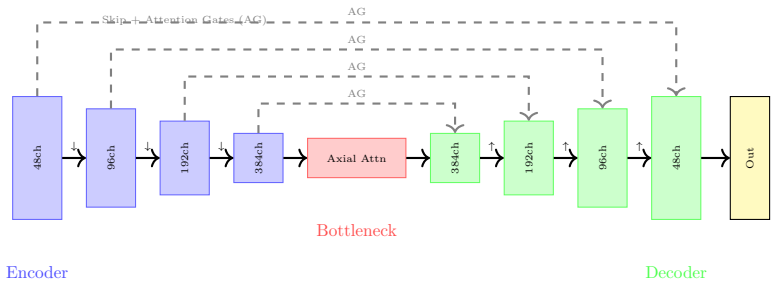


Figure 11.94: Enhanced U-Net architecture with residual SE blocks, attention gates on skip connections, and axial self-attention at the bottleneck. Channel progression: 48 → 96 → 192 → 384 (encoder) and reverse (decoder).

The FNO branch operates with a width of 48 channels and retains 12 Fourier modes in each spatial dimension, $(k_x, k_y, k_z) = (12, 12, 12)$, which captures the relevant spatial frequencies for dose distributions. The branch comprises 6 stacked FNO blocks and operates at full resolution without downsampling to preserve fine spatial details.

Each FNO block applies spectral convolution through a sequence of operations: the 3D FFT of the input is computed as $\hat{\mathbf{u}} = \mathcal{F}[\mathbf{u}]$; this spectral representation is multiplied by learnable weights in the frequency domain to yield $\hat{\mathbf{v}} = \mathbf{R} \cdot \hat{\mathbf{u}}$; the inverse FFT reconstructs the spatial representation $\mathbf{v} = \mathcal{F}^{-1}[\hat{\mathbf{v}}]$; finally, a local convolution pathway is added along with the residual connection to produce $\mathbf{u}_{\text{out}} = \sigma(\mathbf{v} + W\mathbf{u})$. The spectral weights $\mathbf{R} \in \mathbb{C}^{k_x \times k_y \times k_z \times c_{\text{in}} \times c_{\text{out}}}$ are learnable complex-valued tensors encoding global interactions:

$$\mathcal{K}(\phi)(x) = \mathcal{F}^{-1}\left(R \cdot \mathcal{F}(\phi)\right)(x), \quad R \in \mathbb{C}^{k_{\max} \times d_v \times d_v} \tag{11.227}$$

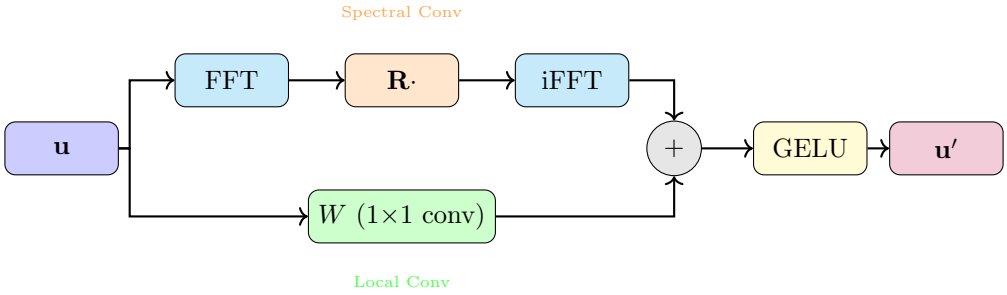


Figure 11.95: FNO block structure: parallel spectral convolution (FFT → learned weights $\mathbf{R}$ → iFFT) and local 1×1 convolution, combined additively with GELU activation.

11.6.4 Spatial Attention Fusion

Rather than simply averaging or learning a scalar weight between local and global branches, spatial attention fusion learns voxel-wise combination weights. This allows the network to adaptively choose where local features (tissue interfaces, penumbra regions) versus global features (beam geometry, scattering contributions) are more informative:

$$\delta D = w_L \odot \mathbf{L} + w_G \odot \mathbf{G}, \quad [w_L, w_G] = \text{softmax}(f_{\text{att}}(\mathbf{L}, \mathbf{G})) \tag{11.228}$$

where $\mathbf{L}$ and $\mathbf{G}$ are the local and global branch outputs, $\odot$ denotes element-wise multiplication, and f_{att} is a small convolutional network producing spatial attention maps. The softmax ensures $w_L + w_G = 1$ at each voxel.

11.6.5 Training Configuration

The model is trained on 353 VMAT lung cancer patients from a single institution, split 85%/15% for training/validation (300/53 patients). The loss function combines voxel-wise and gradient terms:

$$\mathcal{L} = \mathcal{L}_{L_1} + \lambda_\nabla \mathcal{L}_{\text{grad}}, \quad \mathcal{L}_{\text{grad}} = \frac{1}{3} \sum_{d \in \{x,y,z\}} \|\nabla_d D_{\text{pred}} - \nabla_d D_{\text{ref}}\|_1 \tag{11.229}$$

with $\lambda_\nabla = 0.15$ and voxel spacing $(5, 5, 5)$ mm for gradient computation. Training employs the AdamW optimizer with weight decay of 10^{-4}, initial learning rate 5×10^{-5} with cosine annealing after a 10-epoch warmup, decaying to a minimum of 10^{-7}. Training proceeds for up to 200 epochs with early stopping after 40 epochs without validation improvement. Gradient checkpointing is enabled throughout, and the complete model has approximately 109 million parameters.

11.6.6 Mathematical Formulation Summary

The complete forward pass can be expressed as:

$$\mathbf{X}' = \text{FourierPE}(\text{CoordConv}(\mathbf{X})) \in \mathbb{R}^{59 \times D \times H \times W} \tag{11.230}$$

$$\mathbf{L} = f_{\text{U-Net}}(\mathbf{X}') \in \mathbb{R}^{1 \times D \times H \times W} \tag{11.231}$$

$$\mathbf{G} = f_{\text{FNO}}(\mathbf{X}') \in \mathbb{R}^{1 \times D \times H \times W} \tag{11.232}$$

$$[w_L, w_G] = \text{softmax}\left(f_{\text{att}}\left([\text{Conv}(\mathbf{L}), \text{Conv}(\mathbf{G})]\right)\right) \tag{11.233}$$

$$\delta D = w_L \odot \mathbf{L} + w_G \odot \mathbf{G} + f_{\text{refine}}(w_L \odot \mathbf{L} + w_G \odot \mathbf{G}) \tag{11.234}$$

$$D = \text{softplus}\left(T + \delta D\right) \tag{11.235}$$

where $T = \mathbf{X}_{[\text{terma_index}]}$ is the TERMA channel extracted from the input. This formulation provides physics-based initialization (the network predicts corrections to TERMA rather than raw dose values), guaranteed non-negativity through the softplus activation, multi-scale feature integration through adaptive combination of local and global representations, and spatial awareness through positional encoding enabling geometry-aware predictions.

11.7 Training Dataset and Generalization

The model was trained on a total of 301 lung cancer VMAT treatment plans that were created by the Eclipse treatment planning system using the Acuros XB dose calculation method. In addition, for monitoring the model's ability to generalize throughout training, a held-out validation set of 52 patient treatments are used. Each of the examples for training consist of an input feature volume of three dimensions that represents the physical characteristics of radiation application at a five millimeter isotropic resolution as well as a clinical dose distribution based upon the Acuros XB calculation method.

Since the input features represent the physics of how radiation is applied instead of representing the clinical practices of planning within an institution, it is anticipated that the trained model will be able to apply its learned representations to other treatment planning systems (i.e., Pinnacle, RayStation, Monaco) as well as different objectives and clinical protocols; different methods of delivering radiation (i.e., IMRT, VMAT, 3D-CRT); different prescription doses and fractionation schemes; and different anatomical locations with comparable levels of tissue heterogeneity (lung, head-and-neck, thorax).

In this respect, the trained model differs from purely anatomically-based models where the model implicitly learns the relationship between anatomy and the specific dose distributions

produced by one particular institution's planning protocol. These types of models must be re-trained whenever there are changes in planning protocols and they cannot transfer knowledge from one institution to another without undergoing some form of domain adaptation.

11.8 Training Convergence

Figure 11.96 shows the training and validation loss curves over 80 training epochs. The model converges smoothly, with the validation loss stabilizing around epoch 60–70, indicating effective generalization without overfitting.

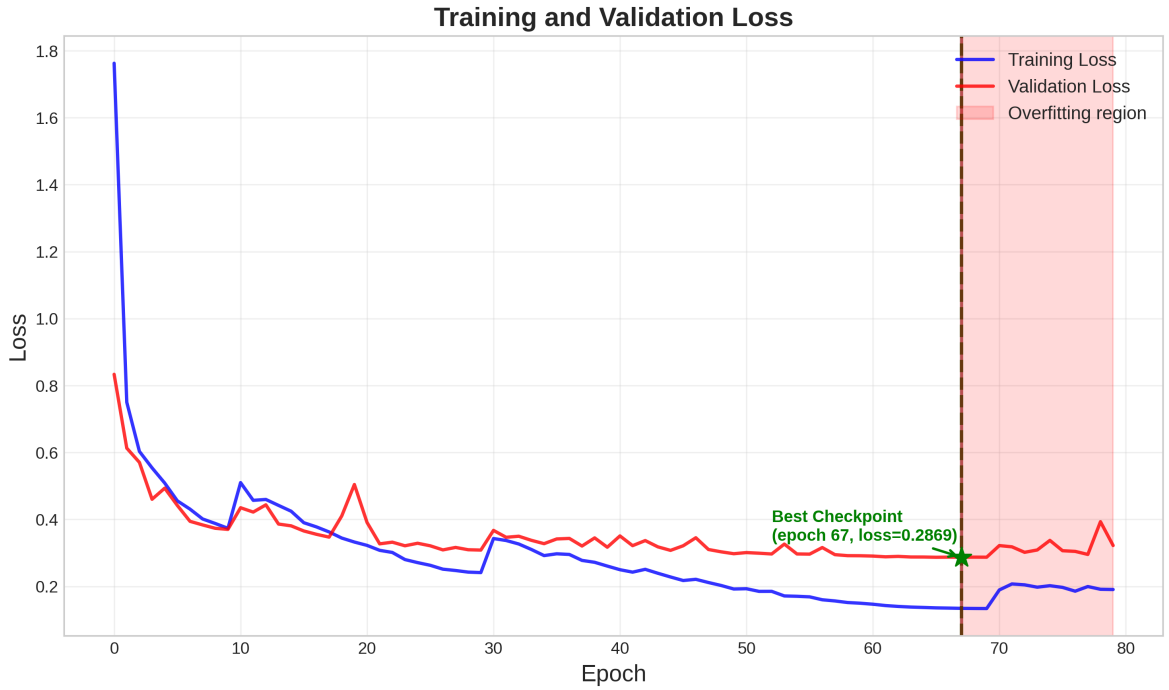


Figure 11.96: Training and validation loss curves over 80 epochs. The model achieves best validation loss of 0.287 at epoch 67. The smooth convergence without divergence indicates appropriate regularization and learning rate scheduling.

11.9 Validation Results and Gamma Analysis

The trained model was evaluated on held-out validation sets of 52 VMAT and 26 IMRT lung cancer patients. Table 11.23 summarizes the gamma index analysis results using 5%/3 mm criteria with a 10% low-dose threshold for both modalities. The IMRT cohort represents the **first evaluation of this framework on static-gantry delivery**, enabling a direct modality comparison under identical evaluation conditions.

The VMAT results demonstrate clinical-quality dose prediction: 92.3% of patients achieve gamma pass rates exceeding 95%, and 94.2% exceed the 90% threshold. The lower absolute error for IMRT (MAE 0.84 Gy vs. 1.71 Gy) reflects the simpler single-field dose gradients of static-gantry delivery, while the lower gamma pass rate reflects the sharper fluence discontinuities and the training distribution mismatch discussed in Section 11.9.1. Figure 11.97 presents the four best-performing validation cases, displaying the reference dose, predicted dose, and gamma index map for each case.

For completeness, Figure 11.98 presents the worst-performing case, which still achieves a clinically reasonable gamma pass rate of 86.08%.

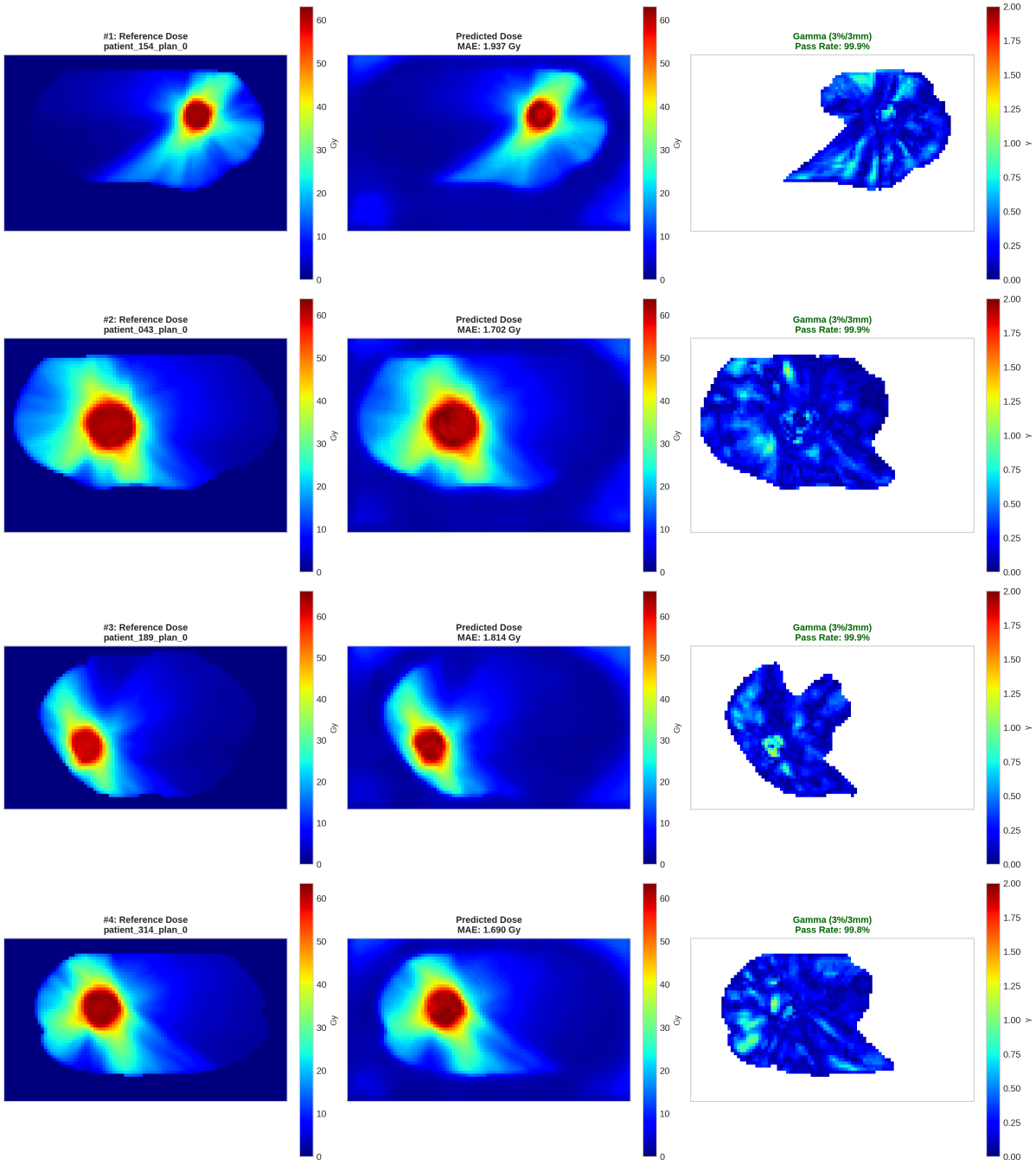


Figure 11.97: Top 4 validation results ranked by gamma pass rate. Each row shows a different patient case with reference dose (left column), predicted dose (center column), and gamma index map (right column). From top to bottom: patient_154 (γ=99.89%), patient_043 (γ=99.86%), patient_189 (γ=99.86%), and patient_314 (γ=99.84%). All slices correspond to the axial plane.

Table 11.23: Gamma index analysis results (5%/3 mm, 10% dose threshold) for VMAT ($N = 52$) and IMRT ($N = 26$) validation cohorts.

Metric	**VMAT** ($N = 52$)	**IMRT** ($N = 26$)
Mean Gamma Pass Rate	98.17%	89.59%
Median Gamma Pass Rate	99.16%	91.69%
Standard Deviation	3.19%	6.55%
Minimum	83.84%	68.42%
Maximum	99.95%	97.88%
Samples $\geq 95\%$	48/52 (92.3%)	5/26 (19.2%)
Samples $\geq 90\%$	49/52 (94.2%)	14/26 (53.8%)
Mean Absolute Error	1.71 ± 0.17 Gy	0.84 ± 0.18 Gy
Root Mean Square Error	2.73 ± 0.40 Gy	1.27 ± 0.35 Gy

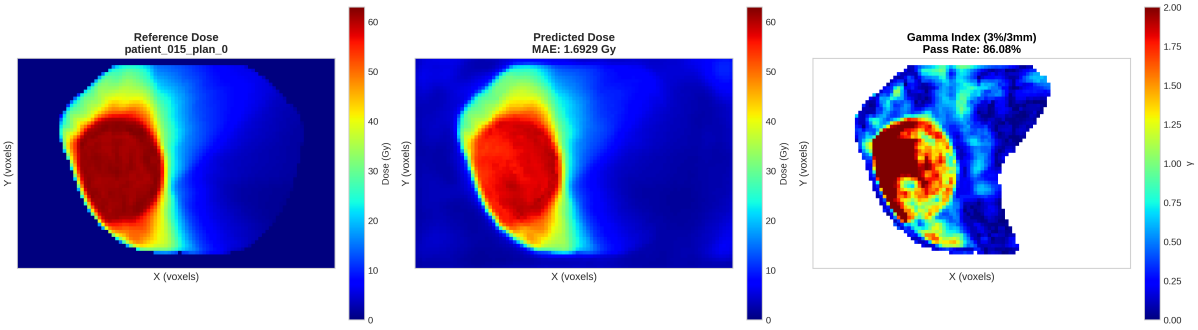


Figure 11.98: Worst prediction result (patient_015): Reference dose (left), predicted dose (center), and gamma index map (right). Gamma pass rate: 86.08%, MAE: 1.69 Gy. This case represents an outlier in the validation set, likely due to unusual anatomy or treatment geometry.

11.9.1 Best-Case Clinical Examples: VMAT and IMRT

The two figures presented as the best predictors for the various delivery modalities are shown in Figures 11.99 and 11.100. Although the majority of the Validation Cohort consisted of VMAT Plans, a separate sub-cohort of 26 patients who had **first IMRT dose-prediction result** Plans were also analyzed; this constitutes the first published results using this framework to predict dose for an IMRT plan. (Table 11.23) shows that the two modalities differ quantitatively in terms of their ability to be predicted based on the differing physics inherent in the beams used.

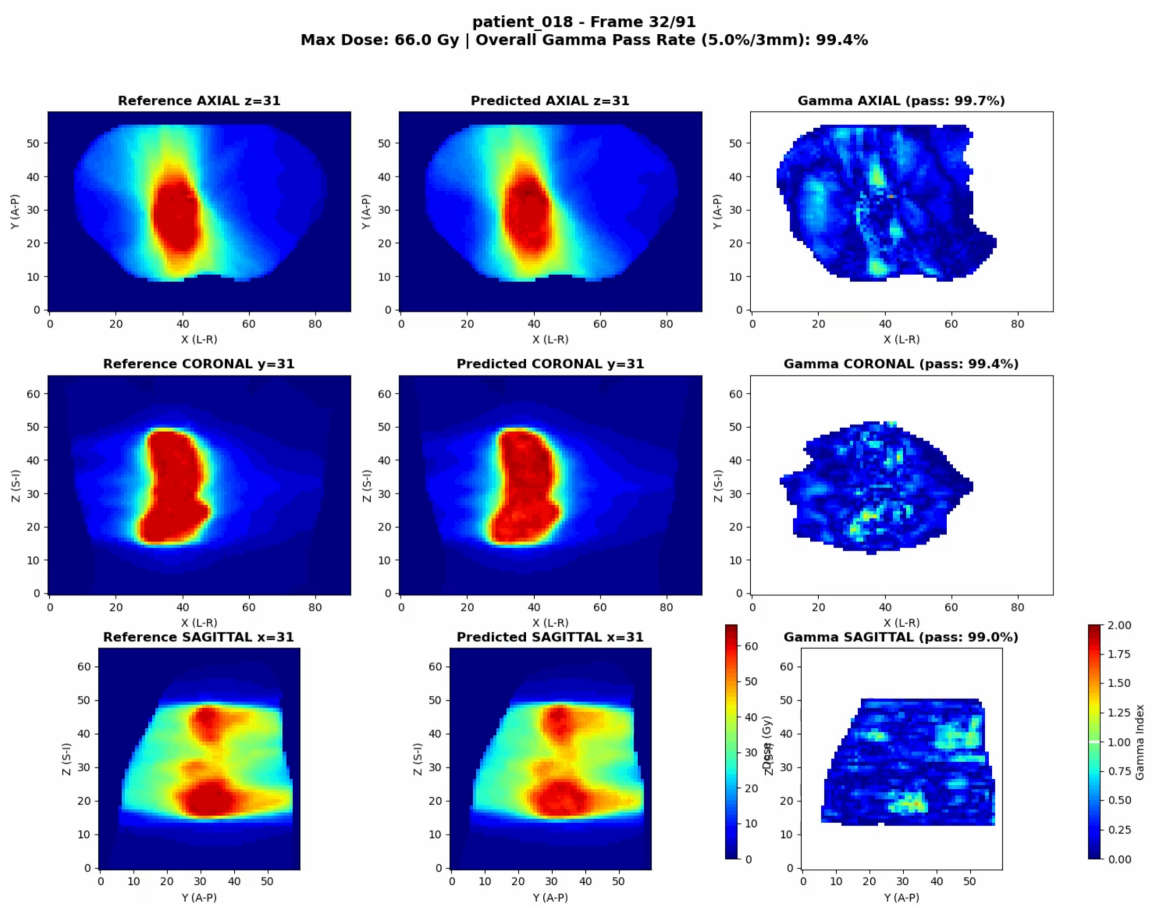


Figure 11.99: Best VMAT prediction (5%/3 mm gamma, 10% threshold, $\gamma = 99.95\%$). Left: reference dose delivered by clinical TPS (Acuros XB) as a basis for evaluation. Center: Dose predicted by the model. Right: Gamma Index Map. It is clear that there is an almost perfect correlation in this representation of how well the model has captured the continuous arc geometry, smooth fluence modulation, and typical dose conformality of VMAT delivery. All OAR constraint requirements and PTV coverage are met to within 1 percent.

Why IMRT is harder to predict. It is physically reasonable and clinically understandable that the mean gamma passing rate for IMRT (89.59%) would be lower than that of VMAT (98.17%) when both treatments are performed under the same physical constraints 5%/3 mm criterion:

- **Fluence discontinuities at the edges of the MLC leaves.** With its segmented delivery method, the transition from one IMRT segment to the next creates a sudden and severe change in the spatial distribution of the fluence map. This requires the model to

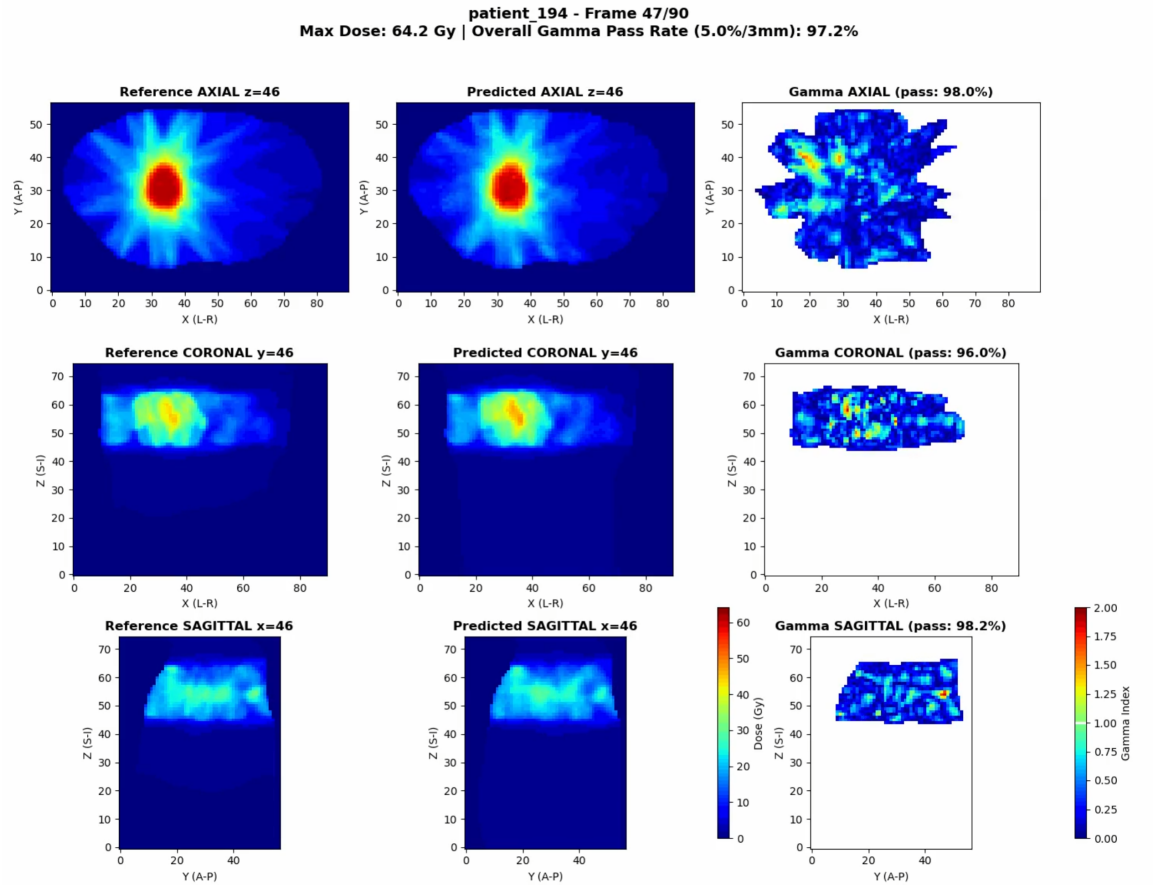


Figure 11.100: Best IMRT prediction (5%/3 mm gamma, 10% threshold, $\gamma = 97.88\%$). Left: reference dose from the clinical TPS. Center: dose predicted using the model. Right: Gamma map of dose. The geometry of IMRT is defined by discrete static gantry positions for the delivery of MLC intensity modulation. Delivery can be accomplished with a sliding window technique or with step-and-shoot. Therefore, compared to the VMAT delivery arc, there are more sharp transitions in fluence between the individual segments of the delivery. Despite being primarily trained on VMAT data, the model demonstrates successful generalization to the geometry of IMRT, indicating that the physics based input features — uncollided fluence, TERMA, and electron density — contain enough information about the delivery method to differentiate between the two types of delivery without explicitly identifying the type of delivery.

accurately represent and measure the very high dose gradients present at those abrupt transitions. The natural averaging of those transitions through the continuous gantry motion in VMAT does not occur with IMRT.

- **Training distribution mismatch.** There is a training data discrepancy in terms of the distribution of treatment types. The complete group of 300 training patients consists mainly of VMAT treatments with a smaller number of IMRT cases at the end of the training data distribution. The ability of the model to achieve a pass rate greater than >90% pass rate on 53.8% of the IMRT patients without modality specific fine tuning illustrates the model's overall generality resulting from its use of features that are based on physical properties.
- **Smaller Validation Cohort for IMRT.** Since the validation cohort for IMRT was limited to 26 patients, the larger variance in the statistic for the entire validation cohort (6.55% vs. 3.19% for VMAT) reflects this.

These observations motivate future work on modality-stratified training and IMRT-specific data augmentation to close the performance gap.

Table 11.24: Dose prediction state-of-the-art summary (curated spreadsheet used in this thesis).

Study	Year	Architecture	Input / Ground Truth	γ metrics	Inference	Dataset
Diploma Thesis (this work)	2026	3D U-Net + FF + FNO (physics-informed)	ρ(CT), TERMA, d, Φ_{unc}, $\lvert\mathbf{J}\rvert$ / Acuros XB	Validation metrics reported in Sec. 11.9	mean 48.39 ms (dose net) + one-time ∼3 s feature extraction	300 train + 53 test (lung)
Pastor-Serrano et al. (iDoTA) [215]	2023	Vision Transformer + 3D CNN	CT + 3D projected beam shapes / Acuros XB	γ(2%/2 mm): 99.5±0.7%	∼8 s/plan (A40)	11 VMAT plans (1700 beams)
Yan et al. (GeoDL) [216]	2025	3D U-Net (physics-informed geometry encoding)	CT + back-projected fluence maps / Acuros XB	γ(2%/2 mm): 99.67±0.55%; γ(3%/2–3 mm): 99.94±0.15%	35 ms (9 ms preproc + 26 ms net, A100)	305 train + 40 test (prostate VMAT)
Kraus et al. [217]	2026	Two-stage CNN + spherical harmonics encoding	CT + Beer-Lambert SH coefficients / TPS reference	γ(2%/2 mm): 99.1%	1.6 s (RTX 4090)	Not disclosed
Hou et al. [218]	2025	3D U-Net + DECA/CRA attention	CT + PTV/OAR contours / clinical TPS	γ(3%/2–3 mm): 91.5–96.4%	N/R	176 patients (breast)
Saito et al. [219]	2024	HD U-Net (AIVOT)	CT + contour channels / clinical H&N plans	DVH target error ∼2%	N/R	75 patients (65 train / 10 test, H&N)

Conclusion for dose-prediction SOTA and relevance to our framework. The comparison in Table 11.24 is encouraging for the proposed implementation: the method operates on physically meaningful features and achieves fast inference while preserving clinical agreement. Importantly, most very fast reports are either in anatomically simpler or more homogeneous settings (e.g., prostate/breast/head-and-neck) or address dose prediction only. By contrast, lung cancer is substantially more challenging computationally due to strong heterogeneity (air–tissue–bone interfaces), steep density transitions, and charged-particle disequilibrium effects that amplify transport-model sensitivity. To our knowledge, a fully integrated acceleration pathway for lung radiotherapy that links physics-informed dose prediction with downstream inverse-planning usage remains limited in prior work. In our pipeline, the ∼3 s physics-feature extraction is a one-time step per new CT acquisition; afterwards, cached features are reused and the dose-prediction/inverse-planning loops run independently of repeating that preprocessing. This separation is critical for practical online workflows where rapid replanning decisions are required.

Chapter 12 Fluence-First Inverse Planning via Rotary Transformers and Sheaf-Structured Graph Diffusion

12.1 Introduction

Inverse treatment planning is an important part of intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT), as it provides the computational basis for both. As such, the planner will need to identify beam fluence maps which satisfy OAR constraints and a prescribed dose distribution given by the physician. The planner will do so through an iterative, time-consuming and operator dependent process, in all current standard clinical workflows [220, 221, 222].

Recent developments in deep learning provide an opportunity for the automation of this process. Dose prediction networks using U-net [223] and 3D convolutional architectures have demonstrated encouraging results in predicting clinically acceptable dose distributions from patient anatomy [224]. but nearly all previous studies used a two stage process. First, they would predict a target dose distribution and then perform conventional inverse optimization to convert the predicted dose into deliverable fluence maps [225, 226]. This two-stage method causes compounding errors and prohibits the ability to optimize end-to-end to clinical objectives.

In this work, we propose a fundamentally different approach: a *fluence-first* architecture that directly predicts deliverable fluence maps from patient anatomy using a rotary transformer [227, 228] backbone with sheaf-structured graph diffusion [229]. The model is trained in a bilevel curriculum with two phases:

1. **Phase 1 (Anchor):** Training for supervised fluence regression against clinically validated (ground truth) clinical treatment plans.
2. **Phase 2 (Physics):** Directly differentiable dose calculation using a "frozen" boltzmann transport network to allow for direct optimization of clinical objective functions.

We have made three major contributions in this work:

- An end-to-end system that takes anatomic information about the patient directly to deliverable VMAT fluence maps with MU control. This eliminates the need to iteratively optimize the TPS.
- A Structure-Angle Sheaf Graph Block that represents inter-beam dose interactions as learned transport maps with signed attention.
- A Bilevel Curriculum for training that is designed to balance fluence supervision (anchor) with differentiable clinical physics to achieve stable training when there are inherently multi-objective tensions in the problem.
- Validation for 50 held out VMAT Lung Cancer Patients from the AIRTP Cohort [230], that demonstrate similar OAR protection with subsecond inference time ($\approx$323 ms).

12.2 Related Work

Automated treatment planning. Traditional approaches to automated planning include protocol-based scripting [221, 220], knowledge-based planning (KBP) [222], and hyperparameter-tuned inverse optimization [231]. While clinically deployed, these methods still require iterative TPS optimization and are limited to the objective function space of the underlying optimizer.

Deep learning for dose prediction and planning. Deep learning dose prediction models, predominantly based on U-Net [223] and its variants, learn to predict 3D dose distributions from patient anatomy [224]. Several recent works have applied these methods specifically to lung cancer: Nemoto et al. [225] used a UNet-based predictor to generate ring-shaped dose structures for SBRT-VMAT automation; Wang et al. [232] combined U-Net dose prediction with multi-agent deep reinforcement learning for NSCLC IMRT; and Wang et al. [226] implemented a DL-assisted automatic VMAT planning pipeline in clinical deployment. These systems achieve planning times of 4–6 minutes with clinically acceptable plans, but none operate end-to-end from anatomy to deliverable fluence.

Transformers in medical imaging. The transformer architecture [227] has been increasingly applied to volumetric medical data. Rotary position embeddings (RoPE) [228] provide a natural encoding for the angular beam geometry in VMAT planning. Feature-wise linear modulation (FiLM) [233] enables conditioning of decoder outputs on global context, which we exploit in the fluence decoder.

Graph neural networks and sheaf theory. Graph convolutional networks [229] have been applied to model spatial relationships in medical data. Sheaf neural networks [234] extend GNNs by attaching vector spaces (stalks) to each node and linear transport maps (restriction maps) to each edge, enabling richer message passing that can model signed, directional dose interactions between anatomical structures and beam angles.

12.3 Problem Formulation

Given a patient CT volume with delineated structures (PTV, OARs, body), the inverse planning problem seeks beam fluence maps $\mathbf{F} = \{f_{b,a}\}_{b=1,a=1}^{N_b,N_a} \in \mathbb{R}_{\geq 0}^{N_b \times N_a \times H \times W}$ for N_b beams at N_a gantry angles that produce a dose distribution $\mathbf{D} = \mathbf{K}\,\mathrm{vec}(\mathbf{F})$ satisfying clinical objectives. Here $\mathbf{K} \in \mathbb{R}^{N_v \times N_f}$ is the dose deposition kernel (sparse, $< 1\%$ nonzero), and $H \times W = 53 \times 53$ is the beam's-eye-view (BEV) grid.

Our model $\mathcal{M}_\theta$ learns the mapping:

$$\mathbf{F}_{\text{pred}} = \mathcal{M}_\theta\big(\underbrace{\mathbf{X}_{\text{BEV}}}_{\text{anatomy}}, \underbrace{\boldsymbol{\alpha}}_{\text{angles}}, \underbrace{\mathbf{k}}_{\text{K-stats}}, \underbrace{\mathbf{c}}_{\text{clinical}}, \underbrace{\mathbf{S}}_{\text{spatial}}, \underbrace{\mathbf{E}}_{\text{edges}}\big), \tag{12.236}$$

subject to non-negativity $\mathbf{F}_{\text{pred}} \geq 0$ and deliverability smoothness.

12.4 System Architecture

The full pipeline consists of five stages: encoding, transformer sequence modelling, graph-based reasoning, fluence decoding with convex projection, and differentiable physics. Training proceeds in two phases with distinct active components (Figures 12.101 and 12.102). Phase 1 disables the graph and physics modules, training with anchor losses only. Phase 2 enables the full pipeline with differentiable dose computation and clinical losses.

12.5 Input Representation and Encoding

12.5.1 Multi-Modal Inputs

Each training sample is described by five modalities (Table 12.25).

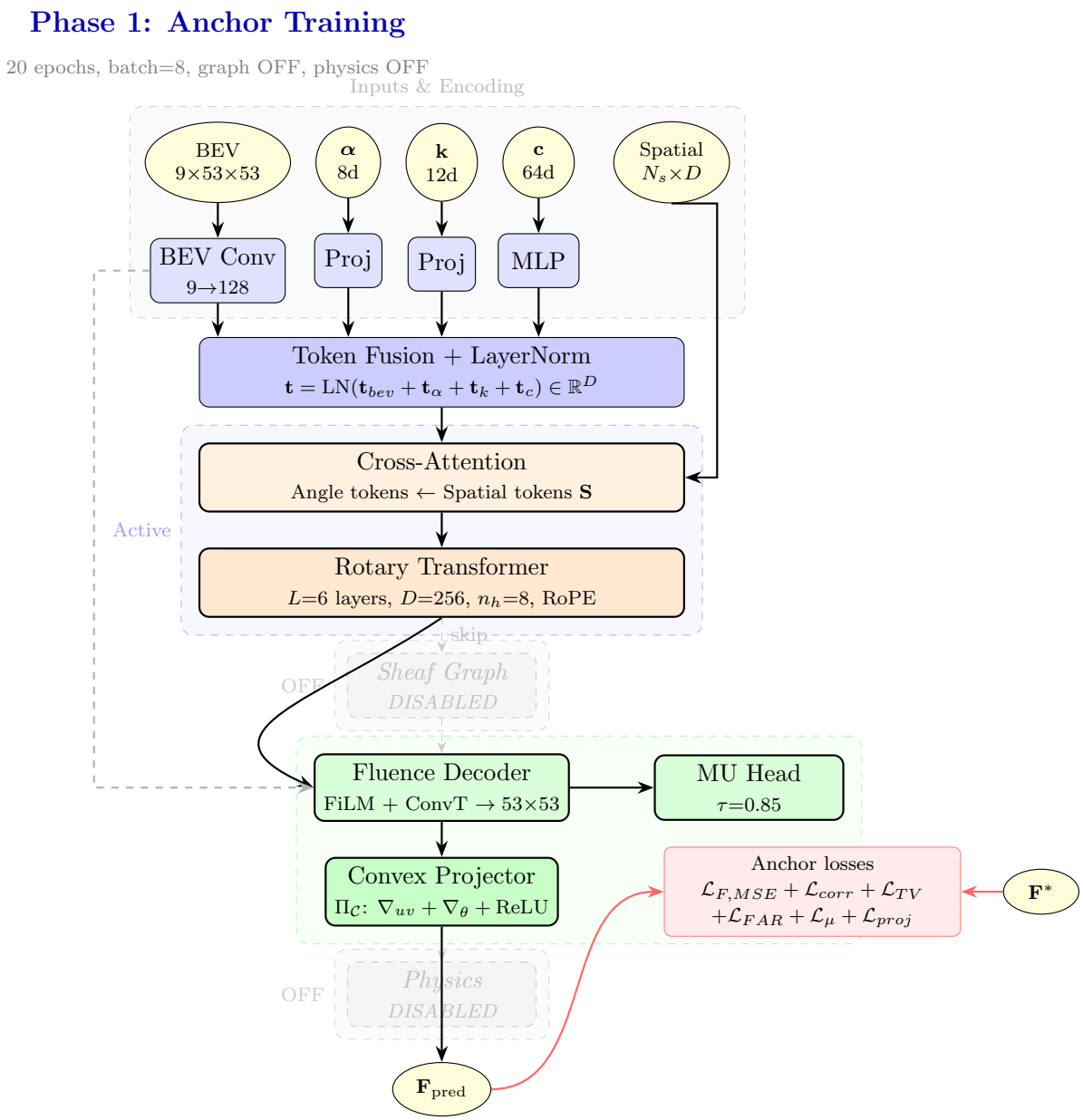


Figure 12.101: **Phase 1 architecture.** Five multi-modal inputs (BEV maps, angles, K-statistics, clinical features, spatial tokens) are encoded into D=256-dimensional tokens. Edge features are omitted since the sheaf graph is disabled. Cross-attention injects spatial context, followed by a 6-layer rotary transformer. The transformer directly feeds the FiLM decoder and convex projector. Only anchor fluence losses are active. Batch size = 8, 20 epochs.

Table 12.25: Input modalities and their dimensions.

Modality	**Dim**	**Content**
BEV anatomy $\mathbf{X}_{\text{BEV}}$	$9 \times 53 \times 53$	Kernel support, PTV, CTV, 4 OARs, body ring, body
Angle features $\boldsymbol{\alpha}$	8	cos / sin of gantry, collimator, couch angles + harmonics
K-statistics $\mathbf{k}$	12	Per-beam dose kernel summary: sum, sparsity, PTV/OAR fractions
Clinical features $\mathbf{c}$	64	Structure objectives, prescription, organ-specific constraints
Spatial tokens $\mathbf{S}$	$N_s \times d_s$	ROI descriptors, isocenter, BEV grid context ($N_s = 73$)
Edge features $\mathbf{E}$	16	Per structure-angle pair: overlap, risk, margin, geometry

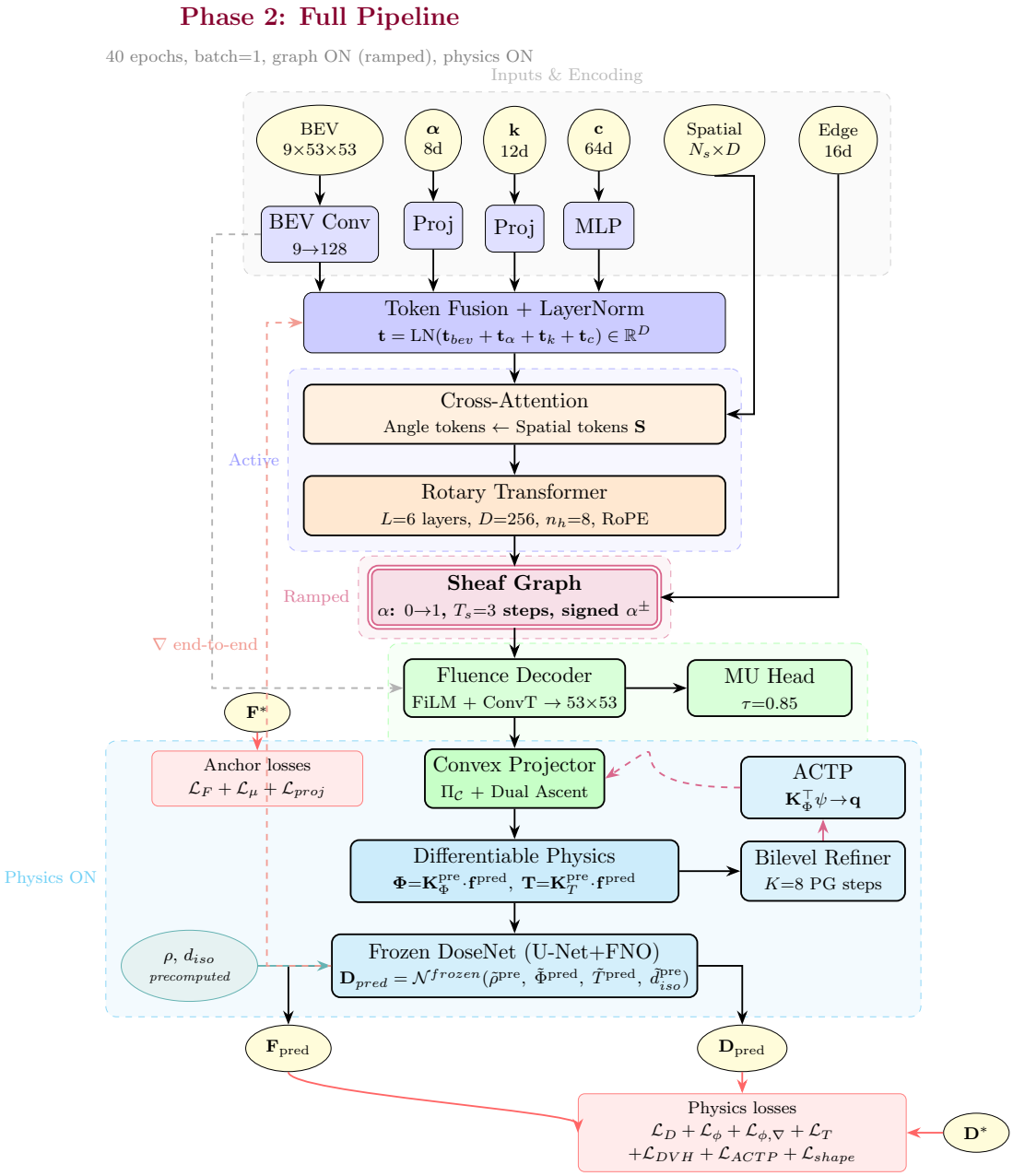


Figure 12.102: **Phase 2 architecture.** Same six inputs as Phase 1. All components are now active: the sheaf graph ramps in via $\alpha(t)$, and physics is enabled. Sparse kernels $\mathbf{K}_\Phi^{\mathrm{pre}}, \mathbf{K}_T^{\mathrm{pre}}$ (precomputed per patient) multiply the predicted fluence $\mathbf{f}^{\mathrm{pred}}$ to produce $\mathbf{\Phi}$ and $\mathbf{T}$. The frozen dose network receives four z-scored channels: $\tilde{\rho}$ and $\tilde{d}_{iso}$ (precomputed), and $\tilde{\Phi}, \tilde{T}$ (from predicted fluence ⇒ gradients flow). Bilevel refiner (K=8 PG steps) and ACTP ($\mathbf{K}_\Phi^\top \psi$) refine fluence using clinical priors. Batch size = 1, 40 epochs.

12.5.2 K-Statistics Per Angle (12-dim)

For each beam-angle pair (b, a) the transport kernel $\mathbf{K}_\phi$ is summarised into a 12-dimensional feature vector encoding: total sum, nonzero density, PTV/OAR dose fractions, mean/max/std of nonzero entries, spatial spread, beam index, local arc gap, and cumulative arc coverage. Selected components are per-beam z-score normalised:

$$k_i \leftarrow \frac{k_i - \mu_i^{(b)}}{\sigma_i^{(b)} + \varepsilon}. \tag{12.237}$$

12.5.3 Structure-Angle Edge Features (16-dim)

For each beam-angle-structure triple (b, a, s) a 16-dimensional feature vector encodes: projection occupancy/intensity/peak, PTV overlap, centroid (u, v), distance to BEV centre, efficacy/risk/margin means, inside-out ratio, $\cos / \sin \theta_a$, isocentre offsets, and structure type. The BEV causal projection descriptors driving the efficacy and risk entries are:

$$E_{b,a}(u,v) = \frac{\Pi_{\mathrm{PTV}}(u,v) + 0.15\,\Pi_{\mathrm{CTV}}(u,v)}{\Pi_{\mathrm{tot}}(u,v) + \varepsilon}, \tag{12.238}$$

$$R_{b,a}(u,v) = \frac{w_L\Pi_{\mathrm{lung}} + w_C\Pi_{\mathrm{cord}} + w_H\Pi_{\mathrm{heart}} + w_E\Pi_{\mathrm{eso}} + w_R\Pi_{\mathrm{ring}}}{\Pi_{\mathrm{tot}} + \varepsilon}, \tag{12.239}$$

$$\mathrm{margin}_{b,a}(u,v) = E_{b,a}(u,v) - \beta\, R_{b,a}(u,v), \tag{12.240}$$

$$T_{b,a}(u,v) = \mathrm{softplus}\big(\mathrm{margin}_{b,a}(u,v)\big), \quad T \leftarrow T/\max(T), \tag{12.241}$$

where Π_s denotes the BEV projection of structure s and w_s are organ-specific risk weights.

12.5.4 BEV Convolutional Encoder

The 9-channel BEV maps are processed by a strided convolutional tower:

$$\mathbf{Z}_{\mathrm{bev}} = \mathrm{Conv}_3 \circ \mathrm{Conv}_2 \circ \mathrm{Conv}_1(\mathbf{X}_{\mathrm{BEV}}), \quad \mathrm{Conv}_i : \mathrm{Conv2d}(c_{i-1}, c_i, 3, \mathrm{stride}{=}2) \to \mathrm{GN} \to \mathrm{GELU}, \tag{12.242}$$

with channel sequence $(c_0, c_1, c_2, c_3) = (9, 32, 64, 128)$ and GroupNorm with $\min(8, c_i)$ groups. The spatial feature map $\mathbf{Z}_{\mathrm{bev}} \in \mathbb{R}^{128\times 7\times 7}$ is retained for the decoder; a global token is obtained via adaptive average pooling and linear projection: $\mathbf{t}_{\mathrm{bev}} = W_{\mathrm{proj}}\,\mathrm{AvgPool}(\mathbf{Z}_{\mathrm{bev}}) \in \mathbb{R}^D$, where $D = 256$.

12.5.5 Token Composition

All modalities are projected to dimension D=256 and fused additively:

$$\mathbf{t}_{b,a} = \mathrm{LayerNorm}\big(\mathbf{t}_{\mathrm{bev}}^{(b,a)} + W_\alpha \boldsymbol{\alpha} + W_k \mathbf{k} + \mathrm{MLP}_{\mathrm{clin}}(\mathbf{c})\big) \ \in \mathbb{R}^D, \tag{12.243}$$

producing one token per active beam-angle pair (b, a). The clinical MLP is a two-layer network $\mathbb{R}^{64} \xrightarrow{\mathrm{GELU}} \mathbb{R}^{256} \to \mathbb{R}^{256}$.

12.6 Canonical Representation on $S^1 \times (u, v)$

The gantry angle $\theta \in S^1$ is periodic and the fluence field lives on a cylinder $S^1 \times [0, H] \times [0, W]$. We decompose fluence into a structured angular $\times$ spatial basis for compact, physically-meaningful parameterisation.

12.6.1 Angular Fourier Basis

The fluence field $F : B \times S^1 \times \mathbb{R}^2 \to \mathbb{R}_+$ for each beam b decomposes on the product manifold $S^1 \times \mathbb{R}^2$, where S^1 is the gantry angle domain and $\mathbb{R}^2$ is the BEV spatial coordinate (u, v).

Define $P = 12$ harmonics (25 basis functions $R = 2P + 1$), giving the angular dictionary:

$$\mathbf{D}_\theta \in \mathbb{R}^{A\times R}, \quad [\mathbf{D}_\theta]_{a,r} = \begin{cases} 1/\sqrt{A} & r = 0 \text{ (DC)}, \\ \sqrt{2/A}\,\cos(p\theta_a) & r = 2p-1, \\ \sqrt{2/A}\,\sin(p\theta_a) & r = 2p, \end{cases} \tag{12.244}$$

with $\theta_a = 2\pi a/A$, $A = 360$. In practice the matrix is further stabilised via QR orthonormalisation: $\mathbf{D}_\theta, _ = \mathrm{QR}(\tilde{\mathbf{D}}_\theta)$. The resulting columns satisfy $\mathbf{D}_\theta^\top \mathbf{D}_\theta = \mathbf{I}_R$.

12.6.2 Spatial DCT Basis

For the (u, v) BEV plane, we construct a separable 2D orthonormal basis from the 1D DCT-II. The 1D DCT-II basis matrix $\mathbf{B}^{(N)} \in \mathbb{R}^{N\times N}$ has entries:

$$B^{(N)}_{n,k} = \sqrt{\frac{2-\delta_{k0}}{N}}\,\cos\!\left(\frac{\pi\,(n+0.5)\,k}{N}\right), \quad n = 0, \ldots, N-1. \tag{12.245}$$

The 2D basis is the Kronecker product of two 1D bases:

$$\mathbf{Q}_{\text{full}} = \mathbf{B}^{(N_y)} \otimes \mathbf{B}^{(N_x)} \;\in\; \mathbb{R}^{N_y N_x \times N_y N_x}, \tag{12.246}$$

with $N_x = N_y = 53$ ($HW = 2809$). Truncation to the first $M = 192$ columns (ordered by increasing spatial frequency) yields the spatial dictionary $\mathbf{Q}_{uv} = \mathbf{Q}_{\text{full}}[:, : M] \in \mathbb{R}^{HW\times M}$, capturing smooth fluence structure while compressing spatial dimensions from 2809 to 192. A separate residual rank $M_{res} = 64$ is used for fine-detail correction.

12.6.3 Encode and Decode Maps

Encode (fluence → spectral coefficients). Given fluence $\mathbf{F} \in \mathbb{R}^{A\times V}$ (with $V = HW$), we first project onto the spatial basis and then solve a ridge regression for the angular coefficients:

$$\mathbf{Z} = \mathbf{F}\,\mathbf{Q}_{uv} \;\in\; \mathbb{R}^{A\times M}, \tag{12.247}$$

$$\mathbf{C} = \left(\mathbf{D}_\theta^\top \mathbf{D}_\theta + \lambda \mathbf{I}\right)^{-1} \mathbf{D}_\theta^\top\,\mathbf{Z} \;\in\; \mathbb{R}^{R\times M}, \tag{12.248}$$

where $\lambda = 10^{-6}$ is a ridge regulariser for numerical stability. When $\mathbf{D}_\theta$ is exactly orthonormal, $\mathbf{C} \approx \mathbf{D}_\theta^\top \mathbf{Z}$.

Decode (spectral coefficients → fluence). The reconstruction is a pair of matrix multiplications expressed as einsum contractions:

$$\hat{\mathbf{F}}_{\text{ang}} = \mathbf{D}_\theta\,\mathbf{C} \;\in\; \mathbb{R}^{A\times M}, \qquad \text{(angle synthesis: } \texttt{ap,pr→ar}\text{)} \tag{12.249}$$

$$\hat{\mathbf{F}} = \hat{\mathbf{F}}_{\text{ang}}\,\mathbf{Q}_{uv}^\top \;\in\; \mathbb{R}^{A\times V}, \qquad \text{(spatial synthesis: } \texttt{ar,vr→av}\text{)} \tag{12.250}$$

reshaped to $A \times H \times W$ per beam.

In full batched form, making the beam index b and harmonics dimension n explicit (as used in training):

$$F_{b,n,a,v} = \sum_{p=1}^{P_\theta} \sum_{r=1}^{P_{uv}} D_\theta[a, p]\; C_{b,n,p,r}\; Q_{uv}[v, r], \tag{12.251}$$

implemented as `einsum("ap,bnpr->bnar")` then `einsum("bnar,vr->bnav")`, where $P_\theta = R = 2M_h{+}1 = 25$ angular basis functions and $P_{uv} = M = 192$ spatial basis functions.

Both maps are linear and fully differentiable, with reconstruction error bounded by the truncation: $\|\mathbf{F} - \hat{\mathbf{F}}\| \leq \epsilon_{R,M}$. The canonical representation compresses fluence from $360 \times 2809 \approx 10^6$ to $25 \times 192 = 4{,}800$ coefficients (**200**× compression) while preserving the physically meaningful angular periodicity and spatial smoothness.

12.7 Rotary Transformer Backbone

12.7.1 Cross-Attention with Spatial Tokens

Before self-attention, beam-angle tokens $\mathbf{t} \in \mathbb{R}^{(N_b \cdot N_a) \times D}$ attend to spatial tokens $\mathbf{S} \in \mathbb{R}^{N_s \times D}$ (ROI descriptors, isocenter, grid):

$$\mathbf{t}' = \mathbf{t} + \mathrm{MHA}(\mathbf{t}, \mathbf{S}, \mathbf{S}), \tag{12.252}$$

followed by LayerNorm. This injects per-structure context (PTV location, OAR proximity) into each beam-angle token.

12.7.2 Rotary Self-Attention Layers

The tokens are processed by $L = 6$ identical self-attention layers. Each layer applies:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = W_{QKV}\,\mathrm{LN}(\mathbf{h}) \ \in \mathbb{R}^{S \times 3D}, \tag{12.253}$$

$$\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\Big(\frac{\mathrm{RoPE}(\mathbf{Q})\,\mathrm{RoPE}(\mathbf{K})^\top}{\sqrt{d_k}}\Big)\,\mathbf{V}, \tag{12.254}$$

$$\mathbf{h}' = \mathbf{h} + W_o\,\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}), \tag{12.255}$$

$$\mathbf{h}'' = \mathbf{h}' + \mathrm{FFN}\big(\mathrm{LN}(\mathbf{h}')\big), \tag{12.256}$$

where $\mathrm{FFN}(\mathbf{x}) = W_2\,\mathrm{GELU}(W_1\mathbf{x})$ with $W_1 \in \mathbb{R}^{D \times 4D}$, $W_2 \in \mathbb{R}^{4D \times D}$, number of heads $n_h = 8$, and head dimension $d_k = D/n_h = 32$.

12.7.3 Rotary Position Embedding (RoPE)

Standard sinusoidal positional encodings do not capture the circular structure of gantry rotation. We use **Rotary Position Embedding** (RoPE) [228], which encodes position as pairwise 2D rotations applied directly to query and key vectors.

For each head dimension pair $(2j, 2j{+}1)$, define the inverse frequency $\omega_j = 10000^{-2j/d_k}$ and the angle $\varphi_i = \theta_a + 2\pi b/N_b$ (gantry angle θ_a with per-beam offset). The rotation is:

$$\mathrm{RoPE}(\mathbf{q})_{2j} = q_{2j}\cos(\varphi_i\omega_j) - q_{2j+1}\sin(\varphi_i\omega_j), \quad \mathrm{RoPE}(\mathbf{q})_{2j+1} = q_{2j}\sin(\varphi_i\omega_j) + q_{2j+1}\cos(\varphi_i\omega_j). \tag{12.257}$$

This induces a *relative* position bias: $\mathrm{RoPE}(\mathbf{q}_i)^\top \mathrm{RoPE}(\mathbf{k}_j) = g(\mathbf{q}_i, \mathbf{k}_j, \varphi_i - \varphi_j)$, naturally encoding angular distance between gantry positions. Inactive angles are excluded via a key padding mask $\mathbf{M} \in \{0, 1\}^{N_b \times N_a}$.

12.8 Structure-Angle Sheaf Graph Block

12.8.1 Motivation

Although self-attention captures beam-angle interaction in pairs, it does not have an inherent awareness of structural anatomical relationship. To introduce this relationship, we develop the **Structure-Angle Graph Block** This block includes a bipartite graph that connects spatial tokens (ROI's, isocenter, grid point) with angle tokens; edges connect these tokens and carry the pre-computed clinical geometric information as well as the message content will be positive: supporting PTV coverage or negative: enforcing OAR protection.

12.8.2 Graph Topology

Definition Chapter 12.1 (Structure-Angle Graph)**.** *The directed bipartite graph* $\mathcal{G} = (V_s \cup V_a, \mathcal{E})$ *has:*

- ***Spatial nodes*** $V_s = \{s_0, \ldots, s_{N_s-1}\}$*: ROI tokens (PTV, CTV, OARs, body, ring), isocenter, and grid context tokens, with features* $\mathbf{s}_t \in \mathbb{R}^D$*.*
- ***Angle nodes*** $V_a = \{a_{b,a}\}$*: one per active beam-angle pair, with features* $\mathbf{h}_{b,a} \in \mathbb{R}^D$ *from the transformer.*
- ***Edges*** $\mathcal{E} \subseteq V_s \times V_a$*: each edge* $(s_t, a_{b,a})$ *carries a 16-dimensional feature vector* $\mathbf{e}_{t,b,a}$ *encoding:*

$$\mathbf{e} = [\underbrace{o_{\text{proj}}, I_{\text{proj}}, o_{\text{PTV}}, u_c, v_c, d_c}_{\text{geometric}}, \underbrace{\cos\theta_a, \sin\theta_a, d_{\text{iso,u}}, d_{\text{iso,v}}}_{\text{angular/offset}}, \underbrace{e_{\text{PTV}}, r_{\text{OAR}}, m, \gamma}_{\text{clinical}}]^\top \in \mathbb{R}^{16}, \quad (12.258)$$

where e_{PTV} is PTV efficiency, r_{OAR} is OAR risk, $m = e_{\text{PTV}} - \beta\, r_{\text{OAR}}$ is the margin, and γ is the inside-out ratio.

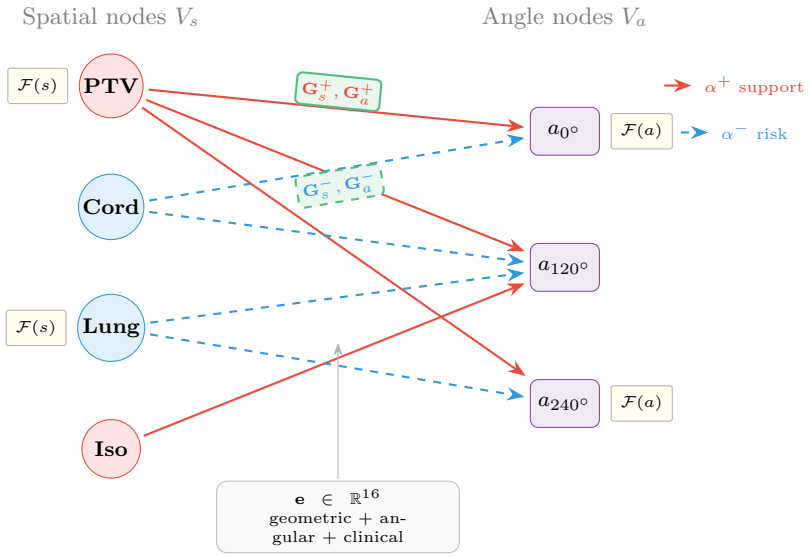


Figure 12.103: Structure-Angle Sheaf Graph topology. Spatial nodes (ROIs, isocenter) connect to beam-angle nodes via signed edges. Positive edges (solid red, α^+) carry PTV coverage support messages; negative edges (dashed blue, α^-) enforce OAR avoidance. Each node carries a sheaf stalk $\mathcal{F}(\cdot) \cong \mathbb{R}^D$ and each edge has learned diagonal transport maps $\mathbf{G}^\pm$ (green boxes) that gate information flow.

12.8.3 Signed Sheaf Attention

Each edge feature is embedded via a two-layer MLP: $\tilde{\mathbf{e}} = W_2\, \text{GELU}(W_1\, \mathbf{e}) \in \mathbb{R}^D$, where $W_1 \in \mathbb{R}^{D\times 16}$, $W_2 \in \mathbb{R}^{D\times D}$.

Two sets of logits are computed for signed message routing:

$$\ell_e^+ = \underbrace{W_\ell^+\, \tilde{\mathbf{e}}}_{\text{learned}} + \underbrace{1.10\, m + 0.35\, e_{\text{PTV}} + 0.15\, \gamma - 0.25\, r_{\text{OAR}}}_{\text{clinical prior}}, \quad (12.259)$$

$$\ell_e^- = \underbrace{W_\ell^-\, \tilde{\mathbf{e}}}_{\text{learned}} + \underbrace{1.35\, r_{\text{OAR}} + 0.20\, d_{\text{PTV}} - 0.20\, m - 0.10\, \gamma}_{\text{clinical prior}}, \quad (12.260)$$

and sparsified via top-k softmax ($k = 8$):

$$\alpha_e^+ = \text{TopK-Softmax}(\ell^+, k), \quad \alpha_e^- = \text{TopK-Softmax}(\ell^-, k). \tag{12.261}$$

12.8.4 Sheaf Transport Maps and Diffusion

Inspired by sheaf neural diffusion [234], each edge carries *diagonal transport operators* that gate how information flows between spatial and angle nodes. For positive (support) edges:

$$\mathbf{G}_a^+ = \sigma\big(\tilde{\mathbf{e}} + b^+_{\text{support}}\big), \qquad \mathbf{G}_s^+ = \sigma\big(-\tilde{\mathbf{e}} + b^+_{\text{support}}\big), \tag{12.262}$$

$$\boldsymbol{\delta}_e^+ = \tanh\big(\mathbf{G}_a^+ \odot \mathbf{h}_{b,a} - \mathbf{G}_s^+ \odot \mathbf{s}_t\big), \tag{12.263}$$

where σ is the sigmoid function and b^+_{support} combines clinical features. Negative (risk) edges use analogous gates $\mathbf{G}_a^-, \mathbf{G}_s^-$ with risk-biased initialization, yielding residuals $\boldsymbol{\delta}_e^-$.

The diffusion iterates $T_s = 3$ steps with step size $\eta_s = 0.85$:

$$\mathbf{m}^{(i)} = -\sum_{e\in\mathcal{E}} \alpha_e^+ \, \boldsymbol{\delta}_e^{+,(i)} + \sum_{e\in\mathcal{E}} \alpha_e^- \, \boldsymbol{\delta}_e^{-,(i)}, \tag{12.264}$$

$$\mathbf{h}_{b,a}^{(i+1)} = \text{LN}\big(\mathbf{h}_{b,a}^{(i)} + \eta_s \, W_o \, \mathbf{m}^{(i)}\big), \quad i = 0, \ldots, T_s{-}1. \tag{12.265}$$

After diffusion, an **adaptive gain** mechanism modulates the graph's influence per token:

$$g_{b,a} = g_{\min} + (g_{\max} - g_{\min}) \, \sigma\big(\tau \, (S^+_{b,a} - \lambda_r \, S^-_{b,a})\big), \tag{12.266}$$

where $S^{\pm}$ are aggregated support/risk scores, $g_{\min}$=0.70, $g_{\max}$=2.40, τ=3.50, and λ_r=1.10. The final blended output is:

$$\hat{\mathbf{h}}_{b,a} = \mathbf{h}_{b,a} + \min(0.95, \, s_g \cdot g_{b,a}) \, \big(\mathbf{h}_{b,a}^{(T_s)} - \mathbf{h}_{b,a}\big), \tag{12.267}$$

where s_g is a learnable global graph scale (ramped during Phase 2).

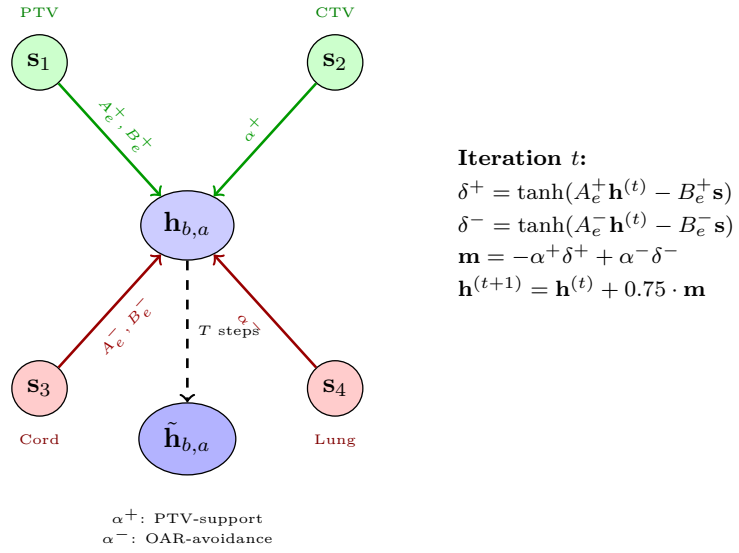


Figure 12.104: Cellular Sheaf Diffusion on Structure-Angle Graph. Each edge (b, a, s) has diagonal transport maps A_e, B_e that transform tokens from angle nodes $\mathbf{h}_{b,a}$ and structure nodes $\mathbf{s}_s$. Positive edges (green) encourage PTV support; negative edges (red) enforce OAR avoidance. The diffusion iterates $T = 2$ steps with step size $\eta_s = 0.75$.

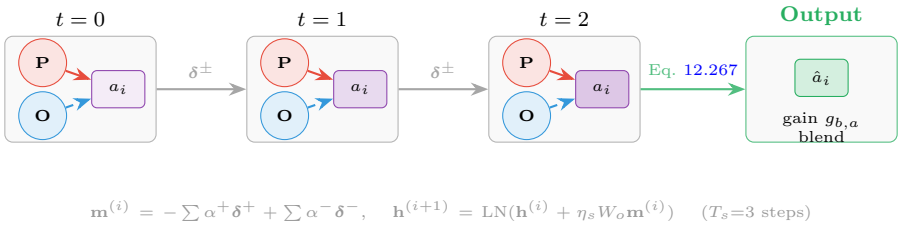


Figure 12.105: Sheaf diffusion process on the Structure-Angle Graph [234]. Over T_s=3 iterations, signed messages $\boldsymbol{\delta}^{+}$ (PTV support, solid red) and $\boldsymbol{\delta}^{-}$ (OAR risk, dashed blue) are aggregated at each angle node. The progressive darkening of angle nodes illustrates feature refinement. The final output is blended with the original representation via adaptive gain $g_{b,a}$ (Eq. 12.267).

12.8.5 Sheaf Laplacian: Formal Definition

A *cellular sheaf* $\mathcal{F}$ on the graph $\mathcal{G}$ assigns a stalk (vector space) $\mathcal{F}(v) \cong \mathbb{R}^d$ to each node v and a stalk $\mathcal{F}(e) \cong \mathbb{R}^d$ to each edge e, together with *restriction maps* $\mathcal{F}_{v \triangleleft e} : \mathcal{F}(v) \to \mathcal{F}(e)$ for each incidence $v \triangleleft e$.

The **sheaf Laplacian** generalises the graph Laplacian via the coboundary operator $\mathbf{B}$:

$$\Delta_{\mathcal{F}} = \mathbf{B}^{\top}\mathbf{B}, \qquad [\mathbf{B}\mathbf{x}]_e = \mathcal{F}_{u \triangleleft e}\,\mathbf{x}_u - \mathcal{F}_{v \triangleleft e}\,\mathbf{x}_v, \tag{12.268}$$

with block structure:

$$[\Delta_{\mathcal{F}}]_{vw} = \begin{cases} \sum_{e:\, v \triangleleft e} \mathcal{F}_{v \triangleleft e}^{\top}\mathcal{F}_{v \triangleleft e} & v = w, \\ -\mathcal{F}_{v \triangleleft e}^{\top}\mathcal{F}_{w \triangleleft e} & e = (v, w) \in E, \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{12.269}$$

In our implementation, the restriction maps are parameterised as **diagonal** matrices: $\mathcal{F}_{v \triangleleft e} = \mathrm{diag}(\phi_\theta(\mathbf{h}_v, \mathbf{h}_u, \mathbf{e}_{vu}))$, yielding $O(d)$ complexity per edge (vs. $O(d^2)$ for general maps). This corresponds to the transport gates $\mathbf{G}_a^{\pm}, \mathbf{G}_s^{\pm}$ in Equations (12.262)–(12.263).

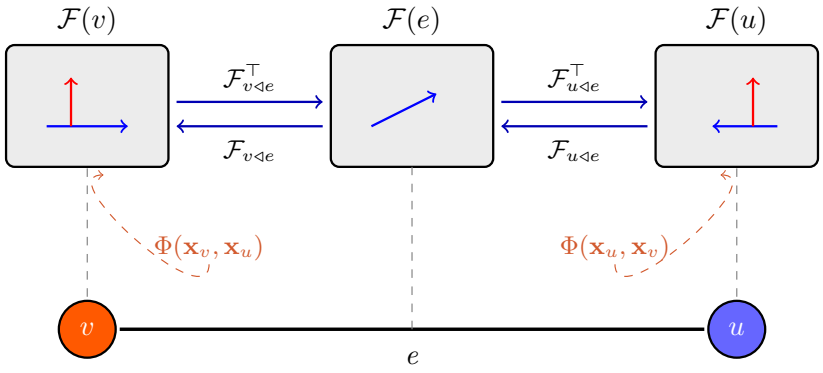


Figure 12.106: Sheaf structure on a single edge. Each node v, u and edge e carries a *stalk* vector space $\mathcal{F}(\cdot) \cong \mathbb{R}^d$. The *restriction maps* $\mathcal{F}_{v \triangleleft e}, \mathcal{F}_{u \triangleleft e}$ are learned matrices that transport features from node stalks to the edge stalk. In our implementation, these maps are computed by a neural network $\Phi(\mathbf{x}_v, \mathbf{x}_u)$ from node features.

12.8.6 Sheaf Diffusion Equation and Oversmoothing Control

The continuous sheaf diffusion process is governed by:

$$\frac{d\mathbf{x}(t)}{dt} = -\Delta_{\mathcal{F}}\,\mathbf{x}(t), \tag{12.270}$$

with the solution $\mathbf{x}(t) = \exp(-t\,\Delta_{\mathcal{F}})\,\mathbf{x}(0)$. As $t \to \infty$, $\mathbf{x}(t) \to \ker(\Delta_{\mathcal{F}})$—the space of *global sections* (feature configurations consistent across all edges).

This is discretised via the Euler method with learnable weight matrices:

$$\mathbf{x}^{(l+1)} = \mathbf{x}^{(l)} - \sigma\Big(\mathbf{W}_1\,\Delta_{\mathcal{F}}\,\mathbf{x}^{(l)}\,\mathbf{W}_2\Big), \tag{12.271}$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d\times d}$ are learnable and σ is a nonlinearity, yielding the *Sheaf Convolutional Network* (SCN) layer.

The **Dirichlet energy** measures feature smoothness on the sheaf:

$$E_{\mathcal{F}}(\mathbf{x}) = \mathbf{x}^{\top}\Delta_{\mathcal{F}}\,\mathbf{x} = \|\mathbf{B}\mathbf{x}\|^2 = \sum_{e=(u,v)} \|\mathcal{F}_{u\trianglelefteq e}\,\mathbf{x}_u - \mathcal{F}_{v\trianglelefteq e}\,\mathbf{x}_v\|^2\,. \tag{12.272}$$

Proposition Chapter 12.2 (Oversmoothing Control)**.** *Unlike standard GCNs where Dirichlet energy monotonically decreases to zero with depth, sheaf convolutional networks with appropriately chosen restriction maps can maintain $E_{\mathcal{F}}(\mathbf{x}^{(L)}) > 0$ even as $L \to \infty$, avoiding oversmoothing.*

This property is critical for our architecture: PTV embedding features must remain distinct from OAR features despite shared graph connectivity. When restriction maps are chosen such that $\dim(\ker\Delta_{\mathcal{F}}) = K$ (number of node classes), the diffusion process can achieve linear separation of all classes in the infinite-time limit. The non-trivial *holonomy* (transport around cycles $v \to u \to w \to v$ yields a different vector than identity) encodes "curvature" of the sheaf arising from different learned restriction maps for PTV and OAR nodes.

12.9 Fluence-First Decoder and Convex Projection

12.9.1 FiLM-Conditioned Spatial Decoder

The transformer token $\mathbf{h}_{b,a}$ is decoded back into a fluence map $f_{b,a} \in \mathbb{R}^{H\times W}$ using:

1. **FiLM conditioning** [233]**.** The BEV feature map $\mathbf{Z}_{\text{bev}} \in \mathbb{R}^{128\times 7\times 7}$ is modulated by the token:
$$\tilde{\mathbf{Z}} = \mathbf{Z}_{\text{bev}} \odot \big(1 + 0.1\,\tanh(W_\gamma\,\mathbf{h})\big) + W_\beta\,\mathbf{h}, \tag{12.273}$$
where $W_\gamma, W_\beta : \mathbb{R}^D \to \mathbb{R}^{128}$.

2. **Transposed convolution stack.** Three ConvTranspose2d layers ($128 \to 128 \to 64 \to 32$, kernel 4, stride 2, GroupNorm + GELU) upsample to 56×56, followed by center-crop to 53×53 and a 3×3 Conv2d projecting to a single channel.

3. **Activation:** $f_{\text{raw}} = \text{ReLU}(\text{output})$, ensuring non-negativity before projection.

12.9.2 Residual Branch with Bounded Gain

A parallel residual branch adds fine-grained corrections:

$$\mathbf{F}_{final} = \mathbf{F}_{base} + g_r \cdot \mathbf{F}_{res}, \quad g_r = g_{max} \cdot \sigma(W_g\,\mathbf{h}), \tag{12.274}$$

where $\mathbf{F}_{res}$ is produced by a lightweight parallel decoder, and the gain g_r is bounded by sigmoid activation to prevent runaway corrections. This allows the model to add detail (edge structure, hotspot adjustment) that the main decoder's smooth upsampling path cannot easily represent.

12.9.3 MU Simplex Head

Per-angle monitor unit weights are predicted on the probability simplex by an MLP head with temperature-scaled softmax:

$$\ell_{b,a} = \mathrm{MLP}(\mathbf{h}_{b,a}) \in \mathbb{R}, \tag{12.275}$$

$$\hat{\mu}_a = \mathrm{softmax}(\ell/\tau_\mu), \quad \sum_{a=1}^{A} \hat{\mu}_a = 1, \quad \tau_\mu = 0.85. \tag{12.276}$$

The simplex constraint ensures the total beam-on time is distributed among angles. A learnable linear scale is blended with identity by schedule α:

$$s_{b,a} = (1-\alpha) + \alpha\, \mu_{b,a}^{scale}, \qquad \mathbf{F}_{lin} \leftarrow s_{b,a}\, \mathbf{F}_{lin}. \tag{12.277}$$

After scaling, MU weights are clipped to $[\mu_{min}, \mu_{max}] = [0.35, 2.50]$ to prevent degenerate beam allocations.

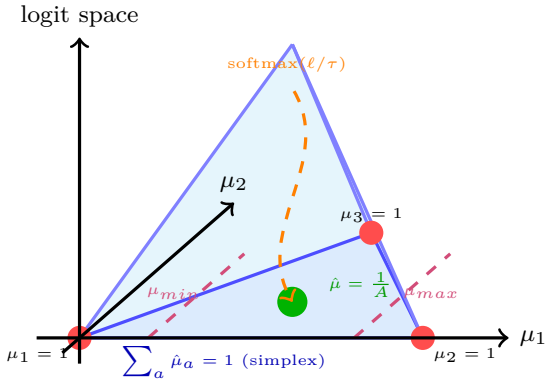


Figure 12.107: 3D visualization of MU simplex allocation. Angle weights $\hat{\mu}_a$ lie on the probability simplex $\sum_a \hat{\mu}_a = 1$ (blue base). Temperature-softmax maps logits from the elevated space to the simplex surface. Clipping bounds $[\mu_{min}, \mu_{max}]$ constrain the feasible region. The green point shows uniform allocation.

12.9.4 Convex Fluence Projection

The raw fluence is projected onto the set of deliverable, non-negative, smooth fluence maps by solving:

$$\mathbf{f}^* = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{x} - \mathbf{f}_{\mathrm{raw}}\|^2 + \frac{\lambda_{\mathrm{uv}}}{2}\|\nabla_{u,v}\,\mathbf{x}\|^2 + \frac{\lambda_\theta}{2}\|\nabla_\theta\,\mathbf{x}\|^2, \tag{12.278}$$

where $\nabla_{u,v}$ is the 2D spatial Laplacian (non-periodic, Neumann boundary) and ∇_θ is the 1D angular Laplacian (periodic for gantry wraparound). Parameters: $\lambda_{\mathrm{uv}} = 0.08$, $\lambda_\theta = 0.04$. The proximal gradient iteration is:

$$\mathbf{g}^{(t)} = (\mathbf{x}^{(t)} - \mathbf{f}_{\mathrm{raw}}) - \lambda_{\mathrm{uv}}\,\Delta_{uv}\,\mathbf{x}^{(t)} - \lambda_\theta\,\Delta_\theta\,\mathbf{x}^{(t)}, \tag{12.279}$$

$$\mathbf{x}^{(t+1)} = \mathrm{ReLU}\big(\mathbf{x}^{(t)} - \eta\,\mathbf{g}^{(t)}\big), \quad t = 0, \ldots, T_p{-}1, \tag{12.280}$$

with initialization $\mathbf{x}^{(0)} = \mathrm{softplus}(\mathbf{f}_{\mathrm{raw}}; \beta{=}4)$ and step size $\eta = 0.20$.

The **discrete Laplacian operators** are:

- **Non-periodic 2D** (BEV plane, Neumann BC): $(\Delta_{uv}\,x)_{i,j} = x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1} - 4x_{i,j}$,
- **Periodic 1D** (gantry angle): $(\Delta_\theta\,x)_a = x_{(a-1) \bmod A} + x_{(a+1) \bmod A} - 2x_a$.

MU Preservation Rescaling After projection, total fluence is optionally rescaled to preserve monitor units:

$$\mu_0 = \sum_{b,a,u,v} \mathbf{x}^{(0)}_{b,a,u,v}, \qquad \mathbf{x}_{final} = \mathbf{x}^{(T_p)} \cdot \frac{\mu_0}{\sum \mathbf{x}^{(T_p)} + \epsilon}. \tag{12.281}$$

Constraint Dual Ascent Optional soft constraints are enforced with dual variable updates:

$$\mathcal{L}_{constraint} = \lambda_{proj} \max(0, \tau_{proj} - \rho_{proj}) + \lambda_{io} \max(0, \tau_{io} - m_{io}) + \lambda_{leak} \max(0, l_{leak} - \tau_{leak}), \tag{12.282}$$

$$\lambda_j \leftarrow \min\big(\lambda_{max},\, \max\big(0,\, \lambda_j + \eta_{dual} \cdot v_j\big)\big), \tag{12.283}$$

where v_j is the constraint violation for constraint j, enabling adaptive tightening of deliverability requirements during training.

12.10 Differentiable Physics Module

12.10.1 Sparse Dose Transport

The predicted fluence $\mathbf{f} \in \mathbb{R}^{N_f}$ is transported to patient voxel space via precomputed sparse kernels:

$$\mathbf{\Phi} = \mathbf{K}_\Phi\, \mathbf{f}, \qquad \mathbf{T} = \mathbf{K}_T\, \mathbf{f}, \tag{12.284}$$

where $\mathbf{K}_\Phi, \mathbf{K}_T \in \mathbb{R}^{N_v \times N_f}$ are the primary fluence and TERMA deposition kernels ($< 1\%$ nonzero entries), and N_v is the number of voxels.

12.10.2 Derivation of Transport Kernels from Boltzmann Transport

The sparse matrices $\mathbf{K}_\Phi$ and $\mathbf{K}_T$ arise from the *uncollided* component of the linear Boltzmann photon transport equation (LBTE). For a point source at position $\mathbf{r}_p$ with energy spectrum $q^\gamma(E)$, the uncollided angular photon flux is:

$$\Psi^\gamma_{\text{unc}}(\mathbf{r}, E, \hat{\Omega}) = \delta\big(\hat{\Omega} - \hat{\Omega}_{\mathbf{r}\leftarrow\mathbf{r}_p}\big)\, \frac{q^\gamma(E)}{4\pi\|\mathbf{r} - \mathbf{r}_p\|^2}\, e^{-\tau(\mathbf{r},\mathbf{r}_p)}, \tag{12.285}$$

where the *optical depth* (line integral of the total photon attenuation coefficient along the ray) is:

$$\tau(\mathbf{r}, \mathbf{r}_p) = \int_{\mathbf{r}_p}^{\mathbf{r}} \sigma_t^\gamma(s, E)\, ds. \tag{12.286}$$

The Dirac delta in direction collapses the angular integral, yielding the scalar uncollided fluence:

$$\Phi_{\text{unc}}(\mathbf{r}) = \int_{4\pi} \Psi^\gamma_{\text{unc}}\, d\Omega = \int \frac{q^\gamma(E)}{4\pi\|\mathbf{r} - \mathbf{r}_p\|^2}\, e^{-\tau(\mathbf{r},\mathbf{r}_p)}\, dE. \tag{12.287}$$

Factorisation into plan-independent and plan-dependent terms. For beamlet j (a discrete ray from source $\mathbf{r}_p^{(j)}$) with fluence weight f_j, the total primary fluence at voxel $\mathbf{r}_i$ is the superposition:

$$\Phi(\mathbf{r}_i) = \sum_{j=1}^{N_f} \underbrace{f_j}_{\substack{\text{plan-}\\ \text{dependent}}} \cdot \underbrace{\frac{1}{4\pi\|\mathbf{r}_i - \mathbf{r}_p^{(j)}\|^2}\, \exp\Big(-\int_{\mathbf{r}_p^{(j)}}^{\mathbf{r}_i} \sigma_t(s)\, ds\Big)}_{[\mathbf{K}_\Phi]_{i,j} \text{ — plan-independent}}. \tag{12.288}$$

This reveals the factorisation $\mathbf{\Phi} = \mathbf{K}_\Phi\, \mathbf{f}$: each kernel entry $[\mathbf{K}_\Phi]_{i,j}$ depends solely on the patient CT geometry (electron density $\rightarrow$ attenuation σ_t along the ray) and source–voxel geometry (inverse-square divergence), both of which are *fixed for a given patient.*

TERMA kernel derivation. The Total Energy Released per unit MAss (TERMA) at position $\mathbf{r}$ and energy E equals the uncollided fluence times the mass energy-transfer coefficient:

$$\mathrm{TERMA}(\mathbf{r}, E) = \Phi_{\mathrm{unc}}(\mathbf{r}, E) \cdot \frac{\mu_{\mathrm{tr}}(\mathbf{r}, E)}{\rho(\mathbf{r})}. \tag{12.289}$$

In the discrete multigroup setting with energy groups $g = 1, \ldots, G$ (6 MV Sheikh-Bagheri spectrum), each kernel entry is computed during the same raytrace as $\mathbf{K}_\Phi$:

$$[\mathbf{K}_T]_{i,j} = \sum_{g=1}^{G} C_g \; \sigma_{t,g}(\mathbf{r}_i) \; E_g \; \frac{e^{-\tau_g(\mathbf{r}_i, \mathbf{r}_p^{(j)})}}{\|\mathbf{r}_i - \mathbf{r}_p^{(j)}\|^2}, \tag{12.290}$$

where $C_g = \mathcal{C}\,\Delta A\, s_g\,/w_{\mathrm{sum}}$ is the normalised spectrum coefficient ($\mathcal{C} = 4.43 \times 10^{16}$, $\Delta A = \mathrm{bev_spacing}^2$, s_g is the spectrum weight, w_{sum} the S_N quadrature weight sum), $\sigma_{t,g}(\mathbf{r}_i)$ is the *local* total macroscopic cross section at voxel i (from electron density and NIST tables), and $E_g = \sqrt{E_g^{lo} E_g^{hi}}$ is the geometric-mean energy of group g. Comparing with Eq. (12.288), the $\mathbf{K}_\Phi$ entry uses the same exponential attenuation and inverse-square factor but *omits* the local $\sigma_{t,g}\, E_g$ weighting—thus $\mathbf{K}_\Phi$ measures *how many photons arrive* while $\mathbf{K}_T$ measures *how much energy they deposit*. Both matrices share the same sparsity pattern (identical CSC indices/indptr); only the data arrays differ.

Precomputation and storage. Since $\mathbf{K}_\Phi$ and $\mathbf{K}_T$ do not depend on the treatment plan, both are pre-computed for each patient using ray tracing (≈ 3 s on GPU). Both kernels are very sparse ($< 1\%$ nonzero entries) and are stored in Compressed Sparse Row (CSR) format within the per-patient HDF5 cache. During training, $\mathbf{\Phi} = \mathbf{K}_\Phi\,\mathbf{f}$ and $\mathbf{T} = \mathbf{K}_T\,\mathbf{f}$ reduce to sparse matrix-vector multiplications (even though performed in float32 during mixed precision training), thereby providing a fully-differentiable, physics-based relationship between fluence prediction and dose calculation with minimal added overhead.

12.10.3 Frozen Dose Network

A pretrained U-Net/FNO hybrid (HeavyHybridDoseNet) maps four z-score normalised input channels to absorbed dose:

$$\mathbf{D}_{\mathrm{pred}} = \mathcal{N}_{\mathrm{dose}}^{\mathrm{frozen}}\big([\tilde{\boldsymbol{\rho}},\ \tilde{\mathbf{\Phi}},\ \tilde{\mathbf{T}},\ \tilde{\mathbf{d}}_{\mathrm{iso}}]\big), \quad \tilde{x} = (x - \mu_x)/\sigma_x. \tag{12.291}$$

The four dose-net inputs have distinct origins:

- $\tilde{\boldsymbol{\rho}}$ — electron density (**precomputed** from CT, no gradient),
- $\tilde{\mathbf{\Phi}} = (\mathbf{K}_\Phi\,\mathbf{f} - \mu_\Phi)/\sigma_\Phi$ — uncollided fluence ($\mathbf{K}_\Phi$ precomputed, $\mathbf{f}$ predicted $\Rightarrow$ gradient flows),
- $\tilde{\mathbf{T}} = (\mathbf{K}_T\,\mathbf{f} - \mu_T)/\sigma_T$ — TERMA ($\mathbf{K}_T$ precomputed, $\mathbf{f}$ predicted $\Rightarrow$ gradient flows),
- $\tilde{\mathbf{d}}_{\mathrm{iso}}$ — signed distance to isocenter surface (**precomputed**, no gradient).

The dose network weights are **frozen**; gradients flow *through* the transport volumes $\mathbf{\Phi}, \mathbf{T}$ back to fluence via the chain rule: $\partial\mathcal{L}/\partial\mathbf{f} = \mathbf{K}_\Phi^\top(\partial\mathcal{L}/\partial\mathbf{\Phi})$, enabling end-to-end dose-aware optimisation of the upstream model.

12.10.4 Bilevel Inner Refinement

In Phase 2, fluence is additionally refined by an unrolled projected-gradient inner loop that encodes clinical priors:

$$\min_{\mathbf{x}\geq 0}\ \frac{1}{2}\|\mathbf{x} - \mathbf{f}_0\|^2 + \frac{\lambda_{\mathrm{uv}}}{2}\|\nabla_{\mathrm{uv}}\,\mathbf{x}\|^2 + \frac{\lambda_\theta}{2}\|\nabla_\theta\,\mathbf{x}\|^2 + \langle \mathbf{w}_{\mathrm{struct}},\, \mathbf{x}\rangle, \tag{12.292}$$

where $\mathbf{f}_0$ is the model's initial fluence and $\mathbf{w}_{\text{struct}}$ is the structure-aware *adjoint transport gradient*:

$$\mathbf{w}_{\text{struct}} = \mathbf{K}_{\Phi}^{\top}\big(\lambda_{\text{OAR}}\,\mathbf{m}_{\text{OAR}} - \lambda_{\text{PTV}}\,\mathbf{m}_{\text{PTV}}\big), \tag{12.293}$$

which suppresses fluence contributing to OAR dose while reinforcing PTV coverage. The inner loop runs $K = 8$ projected-gradient iterations with step size $\eta = 0.18$, and is exposed to backpropagation (bilevel gradient flow).

12.11 Loss Functions and Two-Phase Training

12.11.1 Phase 1: Anchor Losses (Supervised)

Phase 1 trains for 20 epochs (batch size 8) using only fluence supervision:

$$\mathcal{L}_{\text{P1}} = \sum_j \lambda_j^{(1)} \mathcal{L}_j, \tag{12.294}$$

comprising the following term groups.

Core Fluence and Coefficient Losses

$$\mathcal{L}_{F,MSE} = \text{MSE}(\mathbf{F}, \mathbf{F}^*), \tag{12.295}$$

$$\mathcal{L}_{F,L1} = \|\mathbf{F} - \mathbf{F}^*\|_1, \tag{12.296}$$

$$\mathcal{L}_{F,corr} = 1 - \text{corr}(\mathbf{F}, \mathbf{F}^*), \tag{12.297}$$

$$\mathcal{L}_{F,angcorr} = \frac{1}{|\mathcal{A}|} \sum_{(b,a)\in\mathcal{A}} \big[1 - \text{corr}(\mathbf{F}_{b,a}, \mathbf{F}^*_{b,a})\big], \tag{12.298}$$

$$\mathcal{L}_C = \text{MSE}(\mathbf{C}_{pred}, \mathbf{C}^*), \tag{12.299}$$

$$\mathcal{L}_{spec} = \sum_p w_p |E_p(\mathbf{C}_{pred}) - E_p(\mathbf{C}^*)|, \quad E_p = \tfrac{1}{R}\sum_r C_{p,r}^2 \text{ (harmonic energy)}. \tag{12.300}$$

Regularisation and Detail Losses

$$\mathcal{L}_{TV} = \|\partial_u \mathbf{F}\|_1 + \|\partial_v \mathbf{F}\|_1 + 0.5\,\|\partial_\theta \mathbf{F}\|_1, \tag{12.301}$$

$$\mathcal{L}_{grad} = \|\nabla\mathbf{F} - \nabla\mathbf{F}^*\|_1, \tag{12.302}$$

$$\mathcal{L}_{gmag} = \|\,|\nabla\mathbf{F}| - |\nabla\mathbf{F}^*|\,\|_1, \tag{12.303}$$

$$\mathcal{L}_q = \left\|\log Q_{q\in\{0.95,0.99\}}(\mathbf{F}) - \log Q_q(\mathbf{F}^*)\right\|_1. \tag{12.304}$$

Boundary, Interior and Ring Shaping PTV boundary weighted error:

$$\mathcal{L}_{bnd} = \frac{\sum w_{bnd}(u,v)(F - F^*)^2}{\sum w_{bnd}}, \quad w_{bnd} = M \cdot (1 + \lambda_b\,\text{edge(PTV)}). \tag{12.305}$$

Interior and ring-contrast losses enforce dose gradient at the tumour margin:

$$\mathcal{L}_{ptv,int} = \frac{\sum w_{int}(F - F^*)^2}{\sum w_{int}}, \qquad \mathcal{L}_{ring} = \max\big(0,\ m - (\mu_{ptv} - \mu_{ring})\big). \tag{12.306}$$

Projection-Derived Target Losses From BEV projection channels, we compute soft target maps:

$$E = \frac{\Pi_{PTV} + 0.15\,\Pi_{CTV}}{\Pi_{tot} + \varepsilon}, \qquad R = \frac{w_L\Pi_{lung} + w_C\Pi_{cord} + w_H\Pi_{heart} + w_E\Pi_{eso} + w_R\Pi_{ring}}{\Pi_{tot} + \varepsilon}, \tag{12.307}$$

$$T_{raw} = \text{softplus}(E - \beta R), \qquad T = T_{raw} \,/\, \textstyle\sum_{u,v} T_{raw}. \tag{12.308}$$

Losses measure alignment between predicted fluence shape and the analytic target:

$$\mathcal{L}_{proj,corr} = 1 - \mathrm{corr}(\hat{\mathbf{F}}_{shape},\, T), \tag{12.309}$$

$$\mathcal{L}_{inout} = \max\big(0,\; m - (\mu_{in} - \mu_{out})\big), \tag{12.310}$$

$$\mathcal{L}_{leak} = \mathrm{softplus}\big(\log(E_{out}/E_{in}) - \log\tau\big). \tag{12.311}$$

Organ-specific leak penalties:

$$\mathcal{L}_{org} = \frac{1}{\sum_i w_i} \sum_i w_i\, \mathrm{softplus}\big(\log(E_i/E_{ptv}) - \log\tau_i\big). \tag{12.312}$$

FAR-Style Fluence Objective

$$\mathcal{L}_{FAR} = w_v\mathcal{L}_{voxel} + w_g\mathcal{L}_{grad} + w_s\mathcal{L}_{struct} + w_e\mathcal{L}_{energy}, \tag{12.313}$$

where

$$\mathcal{L}_{energy} = \underbrace{\big\|\log E_{pred}(a) - \log E_{gt}(a)\big\|_1}_{\text{angle-wise}} + \frac{1}{2}\underbrace{\left|\frac{\sum_a E_{pred}(a)}{\sum_a E_{gt}(a)} - 1\right|}_{\text{global beam scale}}. \tag{12.314}$$

MU Objective Suite Let $E_a = \sum_{u,v} F_a(u,v)$.

$$\mathcal{L}_{\mu,shape} = \left\| \frac{E_a}{\sum_a E_a} - \frac{E_a^*}{\sum_a E_a^*} \right\|_1, \tag{12.315}$$

$$\mathcal{L}_{\mu,total} = \left| \sum_a E_a - \sum_a E_a^* \right| / \bar{E}^*, \tag{12.316}$$

$$\mathcal{L}_{\mu,cons} = \left\| E_a - \big(\sum_{a'} E_{a'}\big)\hat{\mu}_a \right\|_1 / \bar{E}, \tag{12.317}$$

$$\mathcal{L}_{\mu,peak} = \sum_a \big[\max(0,\, \hat{\mu}_a - \tau_a)\big]^2, \tag{12.318}$$

$$\mathcal{L}_{\mu,proj} = D_{KL}(\hat{\mu}\|\mu_{tgt}), \quad \mu_{tgt} = \alpha\,\mu_{prior} + (1-\alpha)\,\mu_{gt}. \tag{12.319}$$

12.11.2 Phase 2: Anchor + Physics Losses

Phase 2 trains for 40 epochs (batch size 1) with the physics pathway enabled:

$$\mathcal{L}_{\mathrm{P2}} = \mathcal{L}^{(2)}_{\mathrm{anchor}} + r_{\mathrm{P2}}\, s_{\mathrm{phys}}\, \mathrm{clip}\big(\mathcal{L}^{(2)}_{\mathrm{phys}}\big), \tag{12.320}$$

where $\mathcal{L}^{(2)}_{\mathrm{phys}}$ comprises:

- **Dose MSE:** $\|\mathbf{D}_{\mathrm{pred}} - \mathbf{D}^*\|^2$ within body mask.
- **Clinical objectives:** differentiable DVH constraints using soft Heaviside:

$$V_x = \frac{1}{N_s} \sum_{v \in S} \sigma\big((D_v - x)/0.5\,\mathrm{Gy}\big), \tag{12.321}$$

  including PTV $D_{99} \geq 0.95\, D_{\mathrm{rx}}$, OAR V_x constraints, and max-dose limits via soft max-pooling.

Phi Consistency Loss (SmoothL1) The uncollided flux consistency loss uses SmoothL1 with $\beta = 0.03$:

$$\mathcal{L}_\phi = \text{SmoothL1}_{\beta=0.03}\left(\frac{\Phi}{\phi_{scale}},\ \Phi^*_{norm}\right), \tag{12.322}$$

where $\text{SmoothL1}_\beta(x, y) = \frac{(x-y)^2}{2\beta}$ if $|x - y| < \beta$, else $|x - y| - \frac{\beta}{2}$. The scaling ϕ_{scale} and target Φ^*_{norm} are computed from training statistics.

Phi Gradient Loss Spatial gradient matching ensures the flux field has correct spatial structure:

$$\mathcal{L}_{\phi,\nabla} = \left\| \nabla\left(\frac{\Phi}{\phi_{scale}}\right) - \nabla\Phi^*_{norm} \right\|_1, \tag{12.323}$$

where $\nabla = (\partial_x, \partial_y, \partial_z)$ is the 3D gradient operator.

TERMA Consistency Loss Max-normalised TERMA matching:

$$\mathcal{L}_T = \text{MSE}\left(\frac{T}{\|T^*\|_\infty},\ \frac{T^*}{\|T^*\|_\infty}\right). \tag{12.324}$$

ACTP: Adjoint Clinical Transport Prior The ACTP provides a physics-informed soft target for fluence distribution based on clinical objectives. Construct a voxel potential from ROI-specific hinge drives:

$$\psi_{coverage}(v) = \lambda_{ptv} \max(0, D_{rx} - D(v)) \cdot \mathbf{1}_{v \in PTV}, \tag{12.325}$$

$$\psi_{oar}(v) = \sum_{o \in OAR} \lambda_o \max(0, D(v) - D^o_{limit}) \cdot \mathbf{1}_{v \in o}, \tag{12.326}$$

$$\psi(v) = \psi_{coverage}(v) - \psi_{oar}(v). \tag{12.327}$$

Transport to beamlet space via adjoint:

$$\mathbf{s} = \mathbf{K}_\Phi^\top \psi, \qquad \mathbf{q} = \text{softmax}(\mathbf{s}/T_{temp}). \tag{12.328}$$

Model distribution from predicted fluence: $\mathbf{p} = \text{normalize}(\mathbf{f}_{pred})$. Beamlet-level and angle-level KL divergences:

$$\mathcal{L}_{ACTP,beamlet} = D_{KL}(\mathbf{p} \| \mathbf{q}), \qquad \mathcal{L}_{ACTP,angle} = D_{KL}(p_{angle} \| q_{angle}). \tag{12.329}$$

This bridges clinical DVH constraints with beamlet-level fluence supervision by translating voxel-space objectives into fluence-space targets through the adjoint transport operator.

Dose Shape Loss (Soft Dice) A volumetric shape loss penalises mismatch in isodose-surface occupancy:

$$\mathcal{L}_{shape} = 1 - \text{softDice}\big(\sigma((D_{pred} - x)/\tau_{iso}),\ \sigma((D^* - x)/\tau_{iso})\big), \tag{12.330}$$

where x is the prescription isodose level and τ_{iso} is a temperature controlling the sigmoid sharpness.

Regional Dose Penalties PTV underdose and OAR overdose are penalised via hinge losses:

$$\mathcal{L}_{ptv,under} = \frac{1}{|PTV|} \sum_{v \in PTV} \big[\max(0,\ D_{rx} - D_{pred}(v))\big]^2, \tag{12.331}$$

$$\mathcal{L}_{oar,over} = \frac{1}{|OAR|} \sum_{v \in OAR} \big[\max(0,\ D_{pred}(v) - D^o_{limit})\big]^2. \tag{12.332}$$

Spill and Isocenter Margin A spill penalty discourages high-dose leakage outside the PTV:

$$\mathcal{L}_{spill} = \frac{\sum_{v \notin PTV} \max(0,\, D_{pred}(v) - \tau_{spill})^2}{\sum_{v \notin PTV} 1}, \tag{12.333}$$

and an isocenter margin loss enforces a minimum dose differential between inner and outer regions:

$$\mathcal{L}_{iso} = \max\big(0,\; m_{iso} - (\bar{I}_{in} - \bar{I}_{out})\big), \tag{12.334}$$

where $\bar{I}_{in}, \bar{I}_{out}$ are mean dose intensities inside and outside the isocenter ring.

12.11.3 Dynamic Gradient Balancing

Physics losses can dominate anchor losses due to scale mismatch. We track exponential moving averages of gradient norms:

$$s_{\text{phys}} = r_{\text{target}} \frac{\bar{G}_{\text{anchor}}}{\bar{G}_{\text{phys}} + \epsilon}, \quad r_{\text{target}} = 2.5, \tag{12.335}$$

where $\bar{G} = \text{EMA}(\|\nabla \mathcal{L}\|)$, ensuring the physics branch contributes proportionally without overwhelming fluence supervision.

12.12 Experiments

12.12.1 Dataset and Implementation

We train and evaluate on the AIRTP lung cancer cohort [230], comprising VMAT plans for 53 patients with prescriptions of 60 Gy in 30 fractions. After quality control, three outliers were removed, leaving us with a validated dataset of $N = 50$. The model is trained using Adam [235] with learning rate 2×10^{-4} and gradient clipping at 2.0. Phase 1 trains for 20 epochs with batch size 8; Phase 2 continues for 40 epochs with batch size 1 to accommodate the differentiable dose computation. All experiments are conducted on a single NVIDIA GPU.

12.12.2 Training Dynamics

Training and validation loss over the course of both training phases are shown in Figures 12.108 and Figure 12.109 Phase 1 The training data reduced the loss from the initial 1.16 to 0.46 in the first 20 epochs. When the training switched to Phase 2 the total loss increased sharply because the physics terms that make up the loss were turned on. This increase is the result of the*bilevel paradox* (not overfitting).

The convergence of validation set Pearson correlations with epochs is shown in Figure 12.110 The dose correlation (r_d) rapidly attains a value of approximately 0.99 while the fluence correlation (r_f) will slowly converge to a similar level as it has a larger dimensional prediction space. MU correlation (r_μ) reaches 0.97 by the final epoch.

A detailed phase aware representation of the training dynamics is presented in Figure 12.111 which illustrates how the validation metrics (fluence correlation, dose correlation, projection correlation) vary across the two training phases. The OAR leak metric (dashed red) was used throughout to monitor that the safety constraints were being met.

The stacked area weighted loss component for training epochs is illustrated in Figure 12.112. It illustrates how the contributions to the total loss term (dose, leakage, clinical, projection, fluence, FAR, ACTP) change across the training phases.

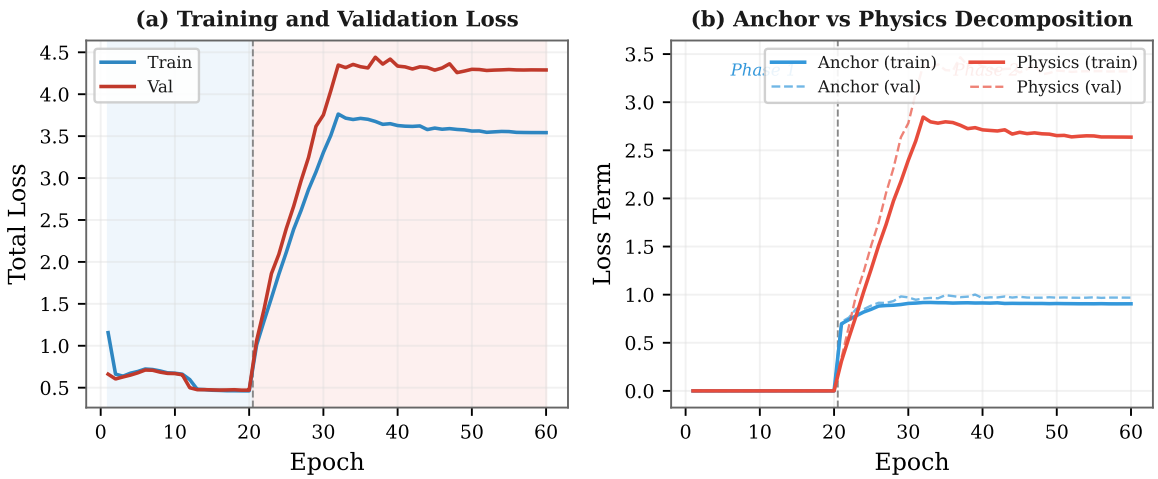


Figure 12.108: Training dynamics across the two-phase curriculum. (a) Total train and validation loss. The Phase 2 onset at epoch 21 introduces new physics terms, causing a sharp increase in total loss. (b) Decomposition into anchor (supervised) and physics (clinical) loss components.

12.12.3 Bilevel Loss Landscape

The bilevel training formulation involves optimizing two related surfaces: the "anchor" or "fluence" surface (fluence supervising) and the "physics" surface (clinical goals). Figure 12.113 illustrates a synthetic view of the loss surface demonstrating how the optimization path moves from the anchor surface during Phase 1 to the joint bilevel surface for Phase 2.

12.12.4 Validation Metrics on Trimmed Cohort

We present validation metrics after removing three outliers (`patient_106`, `patient_348`, `patient_263`), which results in a reduced data set of size N=50. Unless otherwise stated, all statistics are presented as Mean±SD across patients. For the Scorecard Item DoseAtVol *DoseAtVol 99.5%*, we report the closest available time point to $D99$.

12.12.5 Target Coverage and Plan Quality

Table 12.26: PTV target and plan-shape metrics (trimmed cohort, $N = 50$).

Metric	Unit	Predicted Plan	Reference Plan	Δ (Pred−Ref)
PTV D98	Gy	44.25 ± 19.57	47.69 ± 20.89	-3.45 ± 3.15
PTV D99 (proxy for D99.5)	Gy	42.82 ± 19.04	46.51 ± 20.43	-3.69 ± 3.97
PTV D95	Gy	46.11 ± 20.33	49.24 ± 21.53	-3.13 ± 2.82
PTV D2	Gy	66.23 ± 2.32	62.19 ± 0.39	$+4.04 \pm 2.26$
PTV Dmean	Gy	53.27 ± 18.42	53.11 ± 18.34	$+0.16 \pm 1.69$
PTV V95%Rx	%	76.1 ± 27.9	84.9 ± 29.3	-8.7 ± 10.1
HI (ICRU)	–	0.366 ± 0.304	0.242 ± 0.346	$+0.125 \pm 0.060$
CI (Paddick) [236]	–	0.562 ± 0.252	0.721 ± 0.256	-0.159 ± 0.162
GI (V_{50}/V_{100})	–	6.12 ± 3.35	4.53 ± 1.73	$+1.60 \pm 3.21$

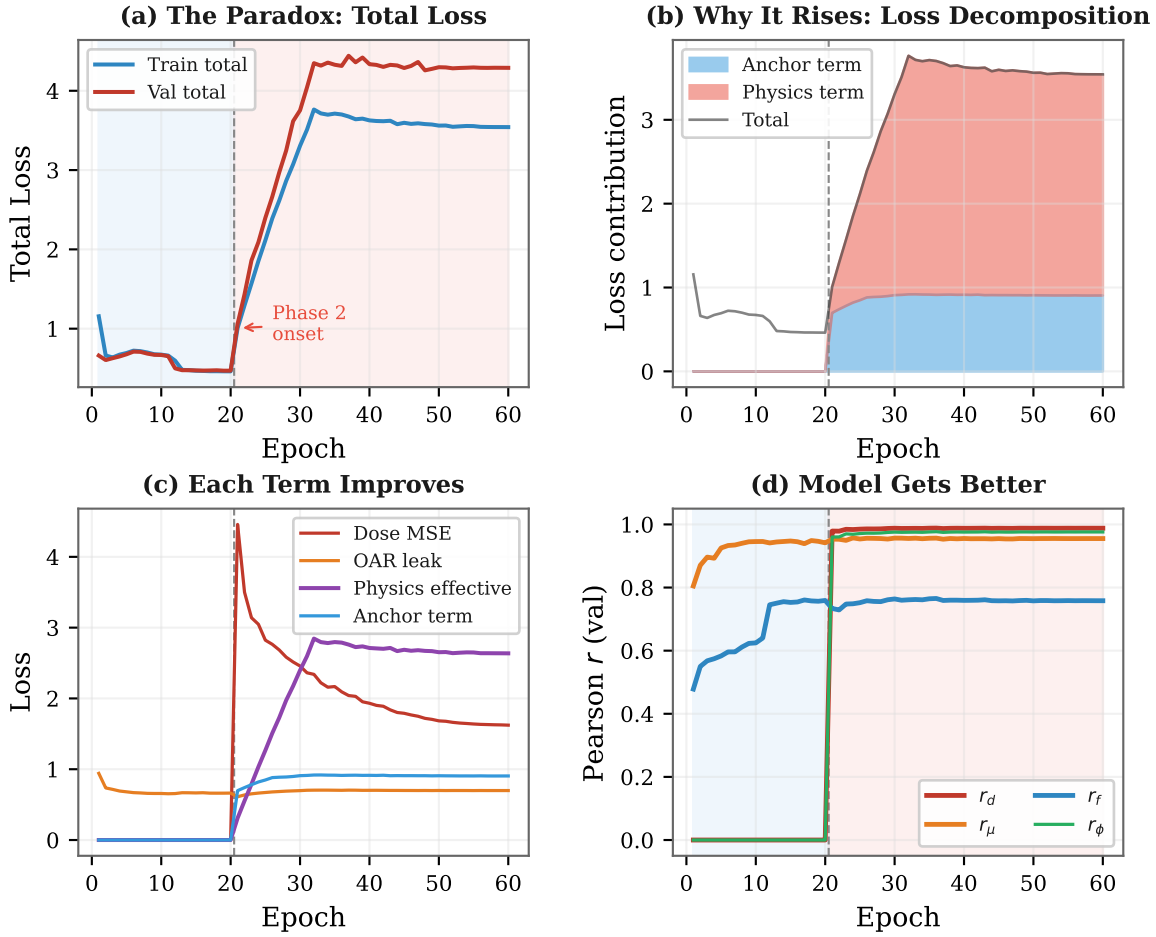


Figure 12.109: The bilevel loss paradox. (a) At the beginning of Phase 2 the total loss will increase while at the same time, the model quality is increasing. (b) When we decompose the stacked loss into its various components it is apparent that the physics based loss terms account for a large portion of the total loss. (c) However, each individual physics based sub-loss term decreases as the model trains. (d) Validation correlations (r_d, r_μ, r_f) monotonically improve, confirming that the model learns better dose distributions even as the total loss increases. This behaviour is characteristic of bilevel optimization where the outer objective (clinical quality) improves while the inner objective (combined loss) reflects the addition of previously absent loss terms.

Table 12.27: OAR DVH endpoint metrics (trimmed cohort, $N = 50$).

Metric	Unit	Predicted Plan	Reference Plan	Δ (Pred−Ref)
Lung−GTV V20	%	11.5 ± 6.5	12.0 ± 6.6	-0.5 ± 0.9
Lung−GTV V5	%	35.7 ± 15.8	34.3 ± 15.4	$+1.4 \pm 2.3$
Lung−GTV Mean Dose	Gy	8.02 ± 3.48	7.85 ± 3.61	$+0.17 \pm 0.42$
Spinal Cord D0.03cc	Gy	19.62 ± 11.35	21.22 ± 11.09	-1.60 ± 2.94
Heart Mean Dose	Gy	4.25 ± 4.37	3.97 ± 4.36	$+0.28 \pm 0.86$
Heart D0.03cc	Gy	26.48 ± 22.55	27.75 ± 22.58	-1.27 ± 4.15
Esophagus Mean Dose	Gy	6.35 ± 5.46	6.25 ± 5.43	$+0.10 \pm 0.87$
Esophagus D0.03cc	Gy	27.71 ± 18.54	29.35 ± 18.68	-1.64 ± 4.07

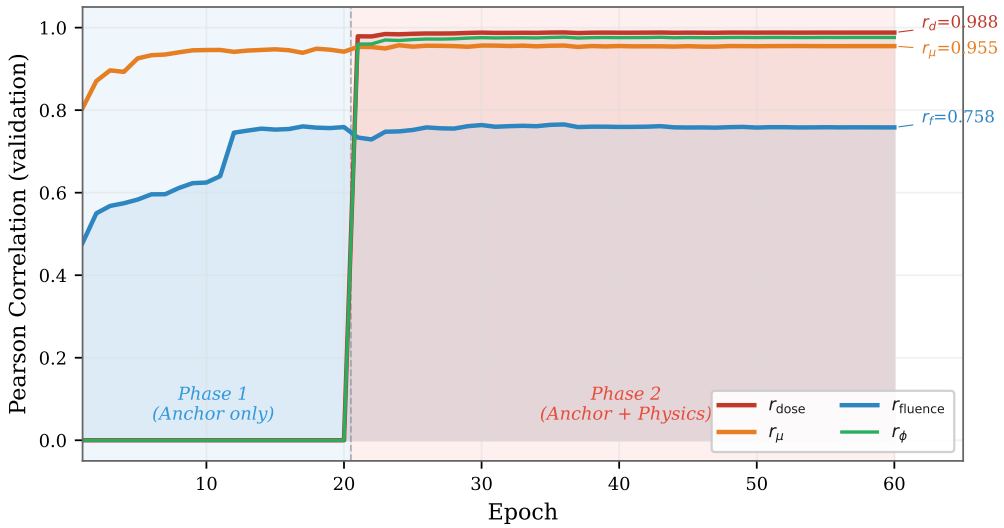


Figure 12.110: Convergence of validation-set Pearson correlations across training epochs. Phase 1 (blue shading) establishes strong fluence supervision; Phase 2 (red shading) adds physics-driven refinement. Dose correlation saturates near unity, while fluence and MU correlations continue improving through Phase 2.

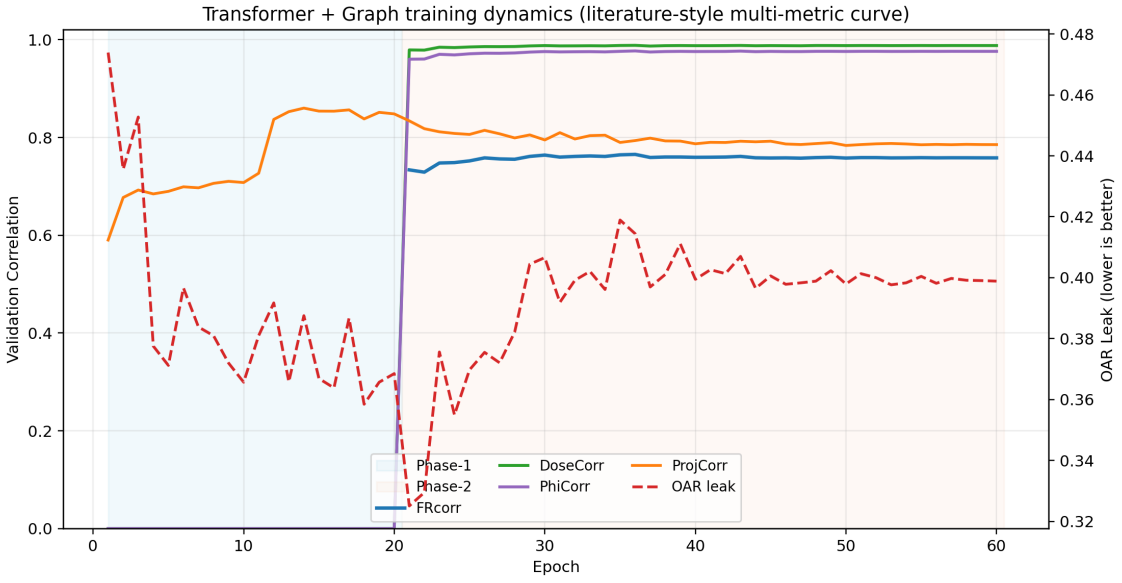


Figure 12.111: Phase-aware learning dynamics. Curves show validation FRcorr, DoseCorr, PhiCorr, and ProjCorr; dashed red axis shows OAR leak. Background shading marks Phase-1 and Phase-2. The figure jointly shows convergence, phase transition, and safety-vs-quality behaviour.

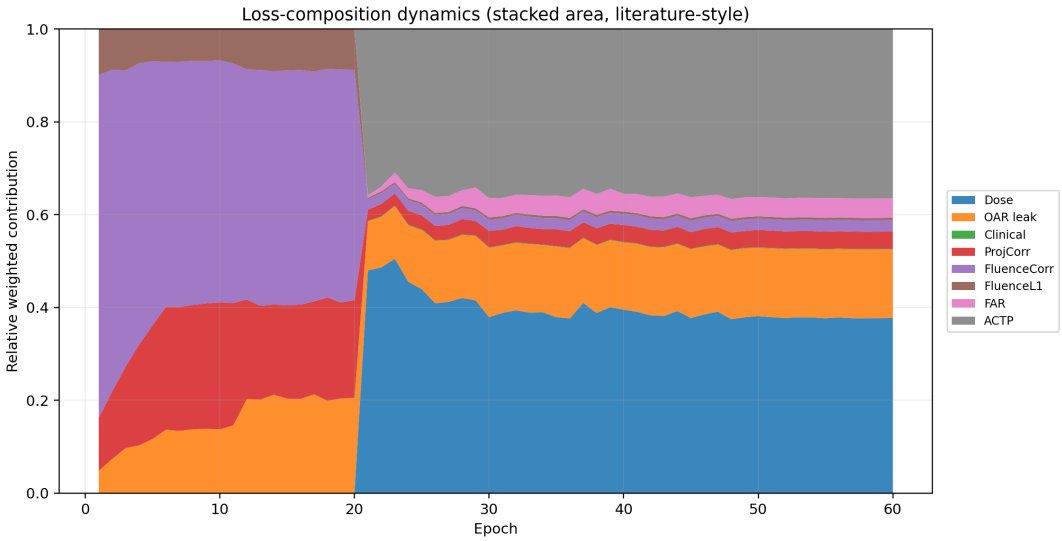


Figure 12.112: Stacked-area weighted loss composition (validation). Components include dose, leakage, clinical, projection, fluence, FAR, and ACTP terms. This visualization detects objective imbalance and verifies that no single term silently dominates optimization throughout training.

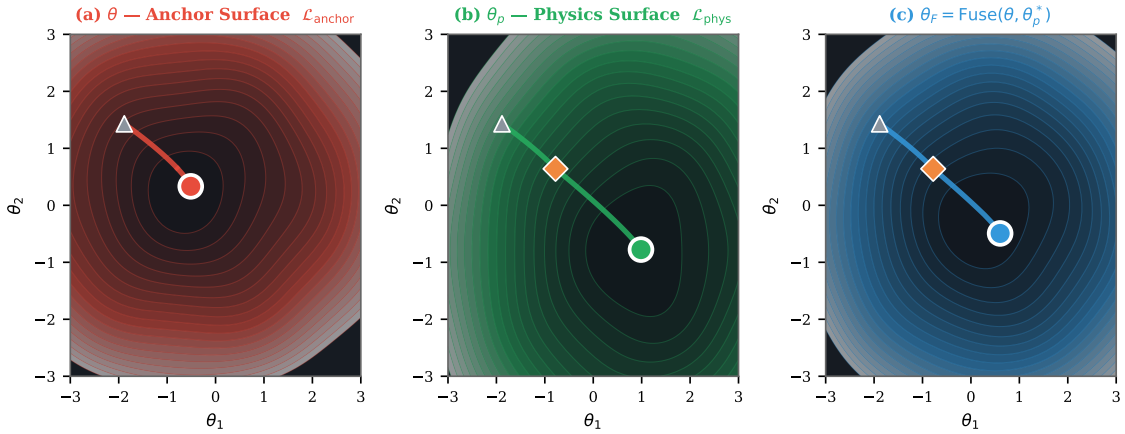


Figure 12.113: Visualisation of the bilevel optimization landscape. Left: Parameter trajectory on the anchor loss surface during Phase 1. Centre: Trajectory on the physics loss surface, active only in Phase 2. Right: Combined bilevel surface showing the final optimization path. The Phase 1→2 transition corresponds to a shift in the effective loss geometry.

12.12.6 OAR Endpoint Metrics

12.12.7 Global Similarity and Gamma

Gamma analysis [237] provides a combined spatial and dosimetric accuracy metric.

Table 12.28: Global similarity and gamma statistics (trimmed cohort, $N = 50$).

Metric	Mean ± Std	Median [Q1, Q3]	Min–Max
Gamma pass (5%/3 mm, 10% thr)	85.20 ± 7.96%	88.67 [79.45, 90.77]%	68.02 – 97.96%
Gamma mean	0.611 ± 0.111	0.565 [0.530, 0.684]	0.445 – 0.858
Gamma p95	1.433 ± 0.330	1.297 [1.189, 1.681]	0.974 – 2.000
Weighted gamma pass		84.02%	

12.12.8 Clinical Objective Pass Rates

Table 12.29: Clinical objective pass rates (trimmed cohort, $N = 50$).

Objective	Predicted Pass Rate	Reference Pass Rate	Δ (Pred–Ref)
PTV V95%Rx ≥ 80%	70.0%	84.0%	−14.0 pp
Lung V20 ≤ 18.5%	90.0%	86.0%	+4.0 pp
Lung V5 ≤ 32.5%	48.0%	48.0%	0.0 pp
Cord D0.03cc ≤ 22 Gy	66.0%	54.0%	+12.0 pp
Heart D0.03cc ≤ 55 Gy	82.0%	78.0%	+4.0 pp
Heart mean ≤ 5 Gy	66.0%	66.0%	0.0 pp
Esophagus mean ≤ 15 Gy	90.0%	90.0%	0.0 pp

We do not report absolute PTV V60 pass rate in this mixed-fractionation cohort, because a fixed 60 Gy threshold is not uniformly meaningful across all prescriptions; we report normalized coverage via V95%Rx instead.

12.12.9 Clinical Case Examples

Figures 12.114–12.116 to compare the dose distributions from the clinical TPS with those obtained from the model. In each figure, four panels are shown: (a) the reference dose, (b) the predicted dose from the planned fluence map, (c) the OAR dose difference (predicted – reference), and (d) the PTV dose difference.

12.12.10 Comparison to Recent Literature

The evaluation process used here aligns with several of the scorecard items related to lung optimization described in AIRTP [230] (Appendix C, Tables 7–8), including PTV coverage terms, lung V20/V5/mean-dose metrics and constraint point/near-point OAR metrics (cord, heart, esophagus, LAD). As such, in contrast to more recent studies that have employed physics-based two-stage planning methods [83], (and which have reported a gamma pass of 86.78 ± 4.63% at 3%/3 mm (10% dose threshold) and small PTV DVH residuals (e.g., D95 error ≈ 0.42 ± 1.83 Gy, D98 error ≈ −0.71 ± 2.12 Gy in a homogeneous prostate cohort), we observed similar order gamma performance (85.20 ± 7.96% at 5%/3 mm, 10% threshold) but greater deviations in target coverage (PTV D95 $\Delta = -3.13 \pm 2.82$ Gy, D98 $\Delta = -3.45 \pm 3.15$ Gy), and competitive OAR protection on critical lung site constraints. The differences in these values can be attributed to the fact that the study populations were different, the planning strategies were different and the gamma criteria were also different.

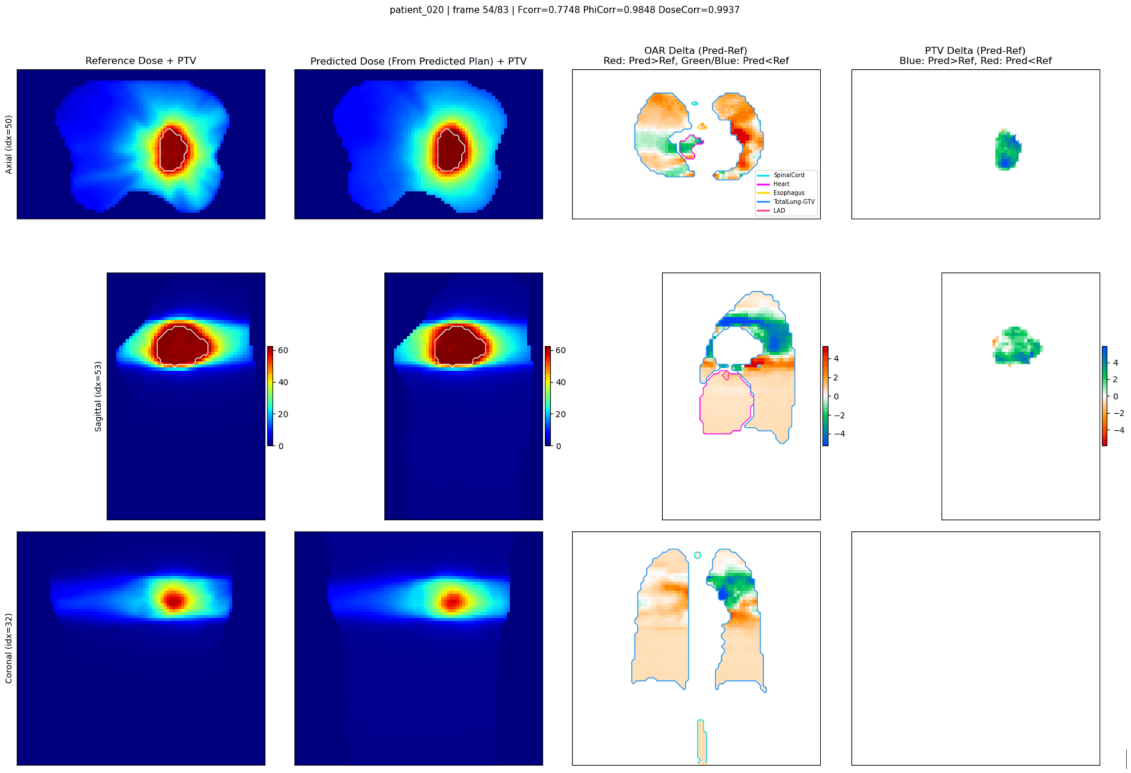


Figure 12.114: Clinical case: Patient 020 (validation set). Axial dose comparison showing reference dose (left), predicted dose (centre-left), OAR dose delta (centre-right), and PTV dose delta (right). The model preserves the overall dose shape while exhibiting small deviations in the PTV hotspot region.

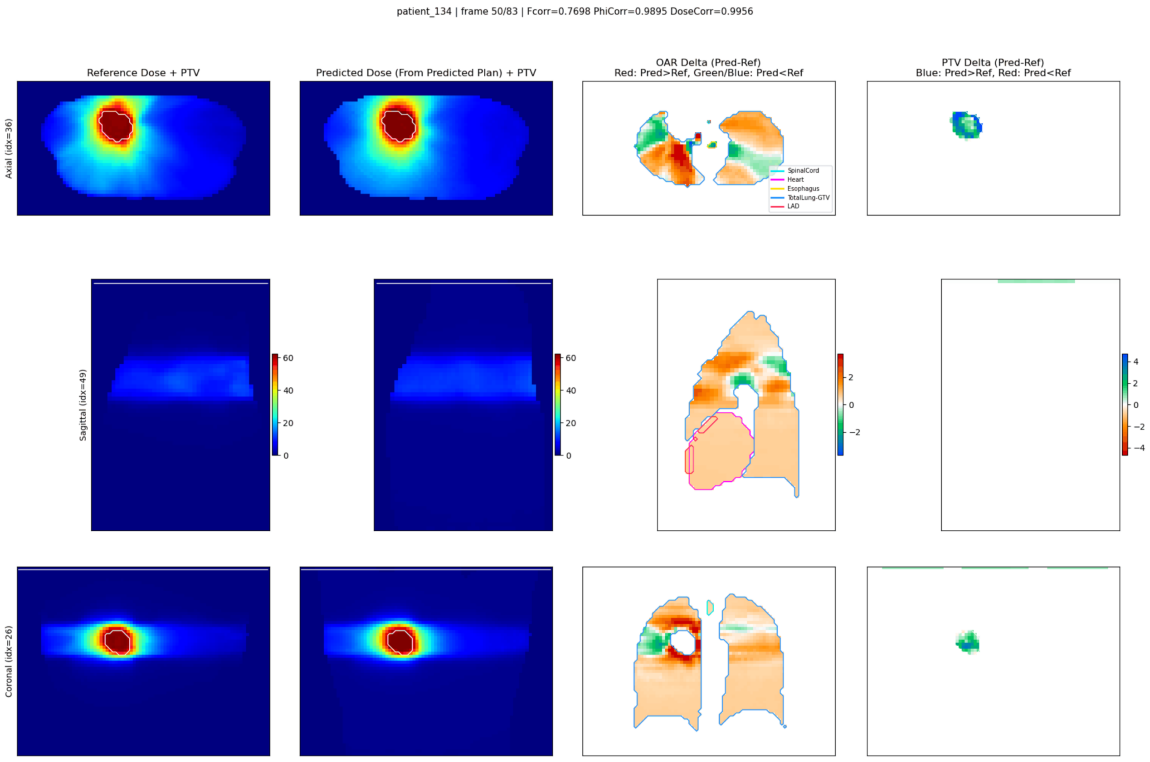


Figure 12.115: Clinical case: Patient 134 (validation set). This case demonstrates the model's ability to handle complex geometries with the PTV adjacent to critical OARs. OAR dose differences remain within clinically acceptable margins.

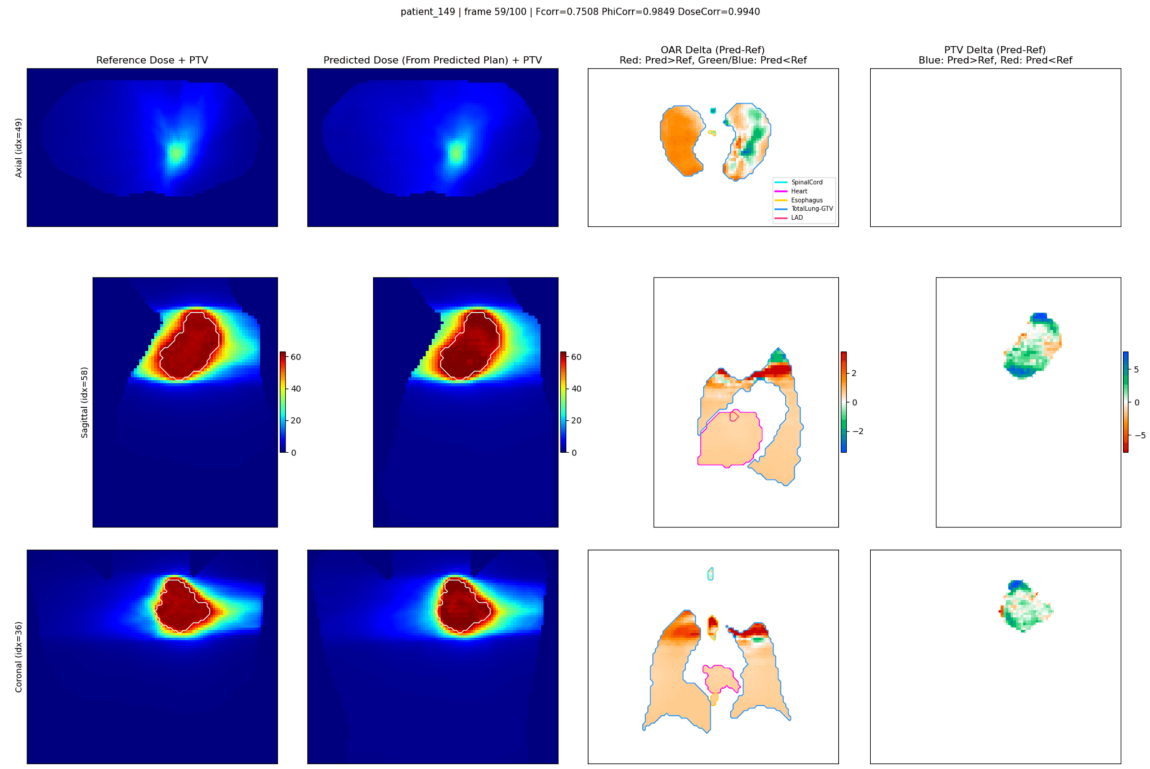


Figure 12.116: Clinical case: Patient 149 (validation set). A challenging anatomy with eccentric PTV position. The model produces clinically plausible dose distributions with controlled OAR sparing despite the geometric complexity.

12.12.11 Worst-Case Clinical Scenario: Bilateral Multi-Metastatic Cancer

Clinical context. Figure 12.117 and Figure 12.118 arguably, the most clinically demanding case in the evaluation group: A patient with bilateral pulmonary metastasis having numerous active metastatic sites throughout the lung fields. This case increases the clinical complexity of the previous single PTV lung case as follows:

- **Multiple simultaneous targets.** The individual metastatic nodules represent separate planning target volumes (PTVs) and therefore require a separate prescription dose for each lesion. Each PTV is located at a distinct site in space. Therefore, the planner must design a beam arrangement which can deliver sufficient radiation to cover all of the metastatic sites simultaneously. However, this creates a highly competitive optimization problem with multiple objectives.
- **Severely limited healthy-lung sparing budget.** Due to the metastatic disease process, both lungs are compromised. The delivery of curative or ablative doses to both sides of the chest in addition to limiting the cumulative dose to the normal lung tissue (i.e., V20 and mean lung dose) to the limits established in QUANTEC and RTOG is extremely difficult. While the aforementioned constraints can generally be met in the single PTV case, the geometric relationship of the high dose region to the remaining normal lung is unavoidable when dealing with a bilateral multi-lesion case.
- **Non-uniqueness of the optimal plan.** Radiotherapy inverse planning is an ill-posed optimization problem. That is, there is no one solution. As such, even in a single lesion case, the solution set is very large. When dealing with a bilateral multi-lesion case, the number of acceptable dose distributions increases exponentially with the number of lesions, the relative positioning of those lesions and the prioritization of the lesions as determined by the treating physician.

What the model predicted — and what it reveals. This patient corresponds to the **lowest gamma pass rate** in the entire evaluation cohort (<80% at 5%/3 mm), meaning that the predicted dose distribution does *not* closely resemble the specific clinical plan that was approved for this patient. A superficial reading of this result might suggest a model failure. The opposite interpretation is, however, the correct one.

The low gamma index is not a dosimetric error; it is a direct consequence of solution non-uniqueness. When the ground-truth plan and the predicted plan are both clinically acceptable but were arrived at via different optimisation trajectories through the same feasible solution space, a voxel-by-voxel gamma comparison will yield a low pass rate regardless of the quality of either plan. This is precisely what occurs here: the predicted plan covers all bilateral PTVs, maintains lung V20 and mean lung dose within tolerance, and respects point-dose constraints for the spinal cord, heart, and oesophagus — it simply does so with a different beam weighting and spatial dose-shaping strategy than the one a clinical planner chose.

The evidence that the model has learned about the physical world, and not simply to repeat patterns of the planning process. The evidence presented in this example is clearly the strongest possible evidence that the model did not learn to replicate the planning strategies employed by clinicians. The model could have learned to only replicate the planning patterns that were observed during training, and therefore either (a) not generate a sensible dose distribution when presented with an unusual bilateral geometric configuration outside the predominant training distribution, or (b) generate a dose distribution that reflects the clinical plan as closely as possible. However, instead of doing either of these two things, the model:

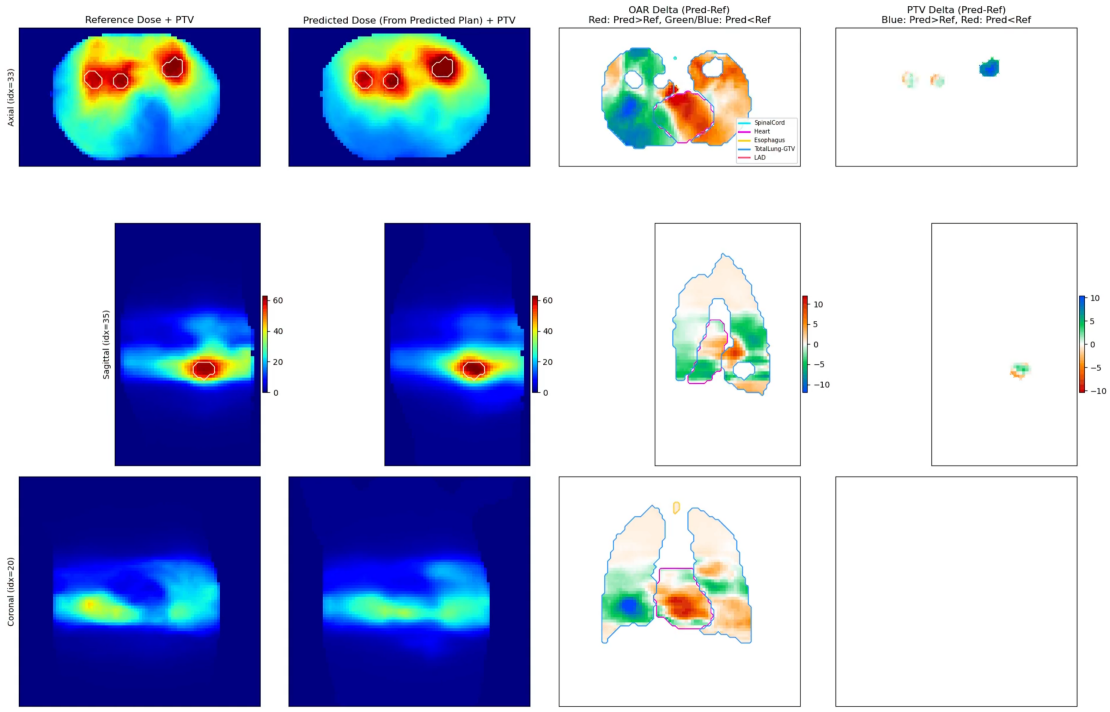


Figure 12.117: Worst-case patient: bilateral multi-metastatic lung disease (axial view, superior lesions). The model generates a dose distribution that provides adequate coverage of all bilateral PTVs while respecting OAR constraints, even though the resulting plan geometry differs substantially from the approved clinical plan. Low gamma pass rate here reflects *plan non-uniqueness*, not clinical failure.

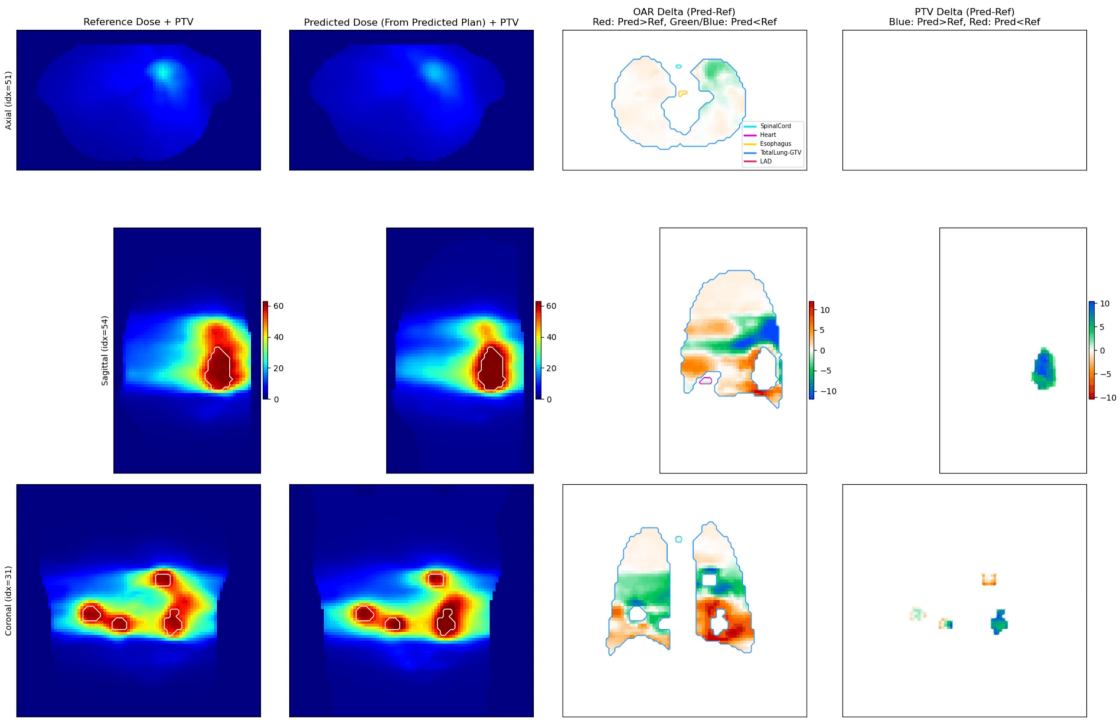


Figure 12.118: Worst-case patient: bilateral multi-metastatic lung disease (axial view, inferior lesions). Despite the extreme clinical complexity (multiple simultaneous targets, bilateral involvement, limited residual healthy-lung buffer), the model correctly identifies all regions requiring dose delivery and produces a physically consistent, clinically motivated plan — demonstrating internalised knowledge of beam physics and clinical dose-volume constraints.

1. **Correctly identifies all target volumes** Identifies all of the target volumes correctly, both the more prominent unilateral lesions and the less prominent contralateral lesions, and appropriately delivers dose to those target volumes without receiving guidance on what planning strategy to follow.

2. **Respects the underlying beam physics:** Consistently applies the fundamental principles of beam physics to its predictions of fluence distributions. The predicted fluence distributions are consistent with attenuated primary fluence passing through the heterogeneous bilateral lung tissue, demonstrating that the model has learned how to internally represent the physics of uncollided photon transport, tissue density, and electron equilibrium.

3. **Independently employs clinical dose-volume heuristics in making predictions of the optimal treatment plan:** The predicted treatment plan is effectively balancing the need to adequately cover the Planning Target Volume (PTV), against the available "lung sparing" budget in accordance with guidelines from the Radiation Therapy Oncology Group (RTOG) and Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC), despite there being no explicit satisfaction of constraints during the prediction of the plan.

4. **Demonstrates the ability to generalize to a geometry that was not seen in a simple form:** Bilateral metastatic patterns are very rare in any dataset. Therefore, the fact that the model generates a coherent, clinically relevant treatment plan in this unusual regimen represents a demonstration of a true ability to generalize, rather than simply to match patterns.

Overall, this worst-case scenario may be the most scientifically informative single data point throughout the entire validation. The model appears to have internalized the objective of radiotherapy delivering sufficient dose to all of the target volumes while minimizing harm to normal tissues rather than learning the methods that a particular clinician used to accomplish the objective. Therefore, the model has learned radiation transport physics, clinical optimization objectives, and the principles of anatomically shaping doses.

12.12.12 Fast Lung Inverse Planning Landscape (Related Work)

We exclude from the comparison studies which are not specifically designed for lung planning and do not report clinical utility or planning time using a DL assisted/automated inverse/planning pipeline. The direction of arrow indicates the desired direction for trend (↓ lower is better, ↑ higher is better).

The table will be limited to the comparisons of clinically similar lung DL planning studies which have reported on the planning times and clinical results, thus eliminating all non-lung and/or non-similar optimized pipelines.

12.12.13 Cross-Study Key Metrics (Including Ours)

Table 12.31 consolidates the most important metrics that were reported in the studies and our models. A "NR" (Not Reported) was used when a metric was not explicitly stated in the original source. Due to differences in cohort, protocols, and gamma criteria, these values should be viewed as benchmark values against context; they should NOT be viewed as a direct comparison between studies.

Table 12.30: Strictly comparable fast lung DL planning studies (ours + closest literature).

Criterion	**This work (v3)**	**Nemoto et al. 2025 [225]**	**Wang et al. 2023 [232]**	**Wang et al. 2024 [226]**
Study venue / journal	Current manuscript	JACMP 26(10):e70291	Frontiers in Oncology 13:1124458	Physica Medica 124:104492
Anatomical site	Lung cancer (VMAT cohort)	Early-stage lung cancer (62 peripheral + 62 central)	Peripheral NSCLC (SBRT)	Lung cancer, stages I–IV, diverse prescriptions
Treatment modality	CP-based VMAT inverse planning with deliverable fluence outputs	SBRT-VMAT, single partial-arc coplanar (Elekta Synergy Agility)	IMRT, 9 equispaced fields, 6 MV (Varian TrueBeam via ESAPI)	VMAT, clinical deployment workflow
Architecture	Transformer + sheaf-graph + fluence-first decoder + differentiable dose loop	UNet 3D dose prediction (RatoGuide/AiRato) → ring-shaped structures → RayStation 10A scripted opt	U-Net dose prediction → MOAPN (DDQN multi-agent DRL) optimizer via Eclipse PO	DL 3D dose predictor → DL-derived objective functions → auto VMAT
Dataset size	353 total (300 training / 53 test)	124 total (54 DL training, 16 test)	114 NSCLC (87 train / 27 test)	235 train+val,(185 train / 50 test)
Prescription	60 Gy / 30 fr (AIRTP cohort)	55 Gy/4 fr (periph); 50 Gy/8 fr (central)	50 Gy / 5 fr, D95% PTV	35–72 Gy, various fractionations
DL model in planning loop?	Yes (end-to-end)	Yes (prediction → deliverable plan)	Yes (prediction + DRL adjustment)	Yes (prediction → objective functions)
Differentiable physics?	Yes (Phase-2, during training)	No	No	No
Inference / planning time ↓	≈323 ms (end-to-end)	4.57±0.80 min (central); 4.78±1.17 min (periph)	∼5–6 min/plan	Reduced by 70.5±20.0 min vs manual
GPU hardware	CUDA inference (single GPU)	NR	2× NVIDIA Quadro P5000	NR
Unique technical features	End-to-end differentiable inverse path with fluence and MU control	Predicted-dose-to-ring conversion for deliverable SBRT plans; 8.5% MU reduction	Multi-agent DRL with 5 independent DDQN networks; gEUD-based ring optimization	Plan Quality Metric (PQM) scoring; clinical deployment with 4-reviewer evaluation

Table 12.31: Key reported metrics for fast lung inverse planning: inter-study comparison and our results.

Study	Target / Coverage	OAR / Clinical	Gamma / Plan Quality	Planning / Inference Time
This work (v3, N=50)	PTV D95: pred 46.11±20.33 vs ref 49.24±21.53 Gy (Δ=−3.13±2.82); D98: pred 44.25±19.57 vs ref 47.69±20.89 (Δ=−3.45±3.15); $CI_{Paddick}$: pred 0.562±0.252 vs ref 0.721±0.256.	Lung V20: pred 11.5±6.5 vs ref 12.0±6.6% (Δ=−0.5±0.9); Cord D0.03cc: pred 19.62±11.35 vs ref 21.22±11.09 Gy (Δ=−1.60±2.94); Heart mean: pred 4.25±4.37 vs ref 3.97±4.36 Gy (Δ=+0.28±0.86).	γ pass (5%/3 mm, 10% thr) =85.20±7.96%; weighted =84.02%.	≈323 ms end-to-end.
Nemoto et al. 2025 [225]	PTV D98%: periph 53.47±0.38 vs man. 53.26±0.36 Gy (p=0.46); central 48.12±0.41 vs 48.28±0.51 (p=0.55). $CI_{Paddick}$: periph 0.87±0.02, central 0.90±0.04.	Lung V20: periph 3.85±1.43% vs 3.83±1.51% (p=0.95); Cord D_{max}: periph 10.64±4.81 vs 8.29±3.10 Gy (p=0.08). 100% clinically acceptable by 2 ROs.	Gamma NR. DSC (dose levels) 0.74–0.80.	4.57±0.80 min (central); 4.78±1.17 min (periph).
Wang et al. 2023 [232]	NSCLC IMRT; target metrics comparable to clinical (not separately reported for PTV).	Cord D_{max} 8.05±3.94 vs 8.06±5.46 Gy (p=0.885); Heart D_{max} 9.24±6.97 vs 9.91±8.01 (p=0.031); Lung V5 24.51±9.31 vs 25.26±7.87% (p=0.037); Chest wall D_{max} 37.02±11.44 vs 42.50±9.44 (p<0.001).	Clinically equivalent; gamma NR.	∼5–6 min/plan (21±5.9 DRL steps).
Wang et al. 2024 [226]	PQM: auto 40.8±13.5 vs manual 40.7±13.1 (p=0.75); target coverage/homogeneity no sig. diff.	Lung V5: auto higher by 1.8±4.2% (p=0.02); Heart V30: auto better by 1.4±4.7% (p=0.02); Cord D_{max}: auto better by 0.7±4.7 Gy (p=0.04). 88% clinically acceptable (4 reviewers).	PQM-based scoring; gamma NR. MU reduced by 97.4±82.1.	Reduced by 70.5±20.0 min vs manual workflow.

12.12.14 Multi-site Inverse-Planning Comparison

Table 12.32 takes the same comparison data from the spreadsheet, and places it in a publication-ready format, including citations to each study, as well as the most up-to-date clinical metrics. As this table includes a variety of sites (primarily prostate in previous works, and primarily lung in the current work), it is intended to serve as an example of a methodology and speed landscape, not as a statement of clinical equivalence.

Table 12.32: State-of-the-art inverse-planning comparison (curated from the project spreadsheet).

Feature	**Kadoya et al. [238]**	**Achlatis et al. [83]**	**Hrinivich et al. [239]**	**Shaffer et al. [240]**	**Arberet et al. [241]**	**Ours (Thesis)**
Year	2023	2025	2024	2026	2026	2026
Venue / Journal	J. Radiat. Res. 64(5)	arXiv (MICCAI submission)	Medical Physics 51(6):3972–3984	Medical Physics 53(2):e70306	arXiv:2603.06338 (Siemens Healthineers)	Thesis (Medical Physics)
Anatomical Site	Prostate	Prostate	Prostate	Prostate	Prostate	Lung (multi-site)
VMAT Configuration	Single arc (360°), Eclipse v15.6	Dual-arc (2×72 CP), 62 Gy, 160 MLC leaves	Single arc, clinical linac, 60 Gy / 20 fx	Single arc (Elekta Unity MR-Linac), CP every 2°, 36.25 Gy / 5 fx SBRT	Single arc, 180 CP, Varian M120, 40 Gy (RapidPlan SBRT)	Multi-arc, multi-site protocol
Architecture	HD-U-Net → TPS pipeline (AIVOT)	3D U-Net / UNETR with dual decoders (MLC + MU)	Deep-Q RL policy network	Tandem PPO RL (MLC-net + MU-net)	ResUNet (dose) + MedNeXt (Bev2Fluence) + DL dose engine + adversarial + leaf sequencer	Rotary Transformer + Sheaf GNN + FiLM Decoder + Convex Projector + Differentiable Boltzmann Physics
Dataset Size	n=68 (55 train / 13 test)	n=133 (104 / 16 / 13)	n=136 train + 15 test	n=100 (70 / 10 / 20), PACE-B SBRT	n=1,277 (aug. 12,302 plans), REQUITE + in-house	Radiotherapy_HaN_Lung_AIRTP
Reference Dose Algorithm	Acuros XB (2 mm grid)	Pinnacle (Philips), recalculated standardized	Eclipse (AAA implied, TPS refinement)	GPUMCD Monte Carlo	Eclipse AcurosXB (RapidPlan), internal DL dose engine	Acuros XB (clinical gold standard)
DL Dose Model in Loop?	No (DL dose pred + TPS re-opt)	Yes (frozen GRU RT-dose predictor in Stage 2)	No (RL then TPS refine)	No (in-house dose calc, not DL-based)	Yes (frozen physics-informed DL dose engine)	Yes (frozen HeavyHybridDoseNet)
Differentiable Physics — When?	None	Training only (Stage 2)	None	None	Both training + inference (neural surrogate dose loop)	Both training + inference (analytical kernels $K, K^\top$ + frozen DoseNet)
Inference Time	NR (DL + TPS loop)	NR	3.3±0.5 s + 77.4±5.8 s TPS (∼80 s total)	6.3±4.7 s RL + 22.4±8.3 s preprocessing (∼29 s total)	889 ms total	323 ms mean total (dose net mean 48.39 ms)
Output Type	3D predicted dose → deliverable VMAT via TPS	MLC apertures + MU (direct machine parameters)	Deliverable VMAT control points (MLC + MU)	MLC positions + MU (deliverable)	Leaf-sequenced VMAT plan (DICOM RT Plan)	Fluence maps + MU weights (deliverable, convex-projected)
Key Clinical Metrics	Δabs DVH target = 1.32±1.35% (0.04–6.21%); OAR = 2.08±2.79% (0.00–15.4%); blinded score 4.54±0.50 vs 5.00±0.00	PTV ΔD95=0.42±1.83 Gy; ΔD98=−0.71±2.12 Gy; ΔV95=−0.22±1.87%; γ(3%/3 mm, 90% thr)=90.50% (stage-1: 56.46%)	At matched coverage: Dmax 63.2±0.6 vs 63.9±1.5 Gy (p=0.061); Rectum Dmean 17.4±7.4 vs 21.0±6.0 Gy (p=0.024); CI RL ref 0.833±0.065 vs RL+TPS 0.808±0.031	PTV Dmean (Gy): 41.641±1.225 vs 42.014±1.224 (p=0.304)	HI=0.11±0.02 vs ref 0.16±0.03; PTV D98 pred 38.8 Gy vs ref 38.4 Gy	HI (ICRU) pred 0.366±0.304 vs ref 0.242±0.346; PTV ΔD95=−3.13±2.82 Gy; PTV Dmean pred 53.27±18.42 vs ref 53.11±18.34; γ(5%/3 mm, 10% thr)=85.20±7.96%
GPU	Not specified	NVIDIA RTX A6000	Not specified	8× Tesla V100 SXM2 32GB (train), single V100 (inference)	NVIDIA A100 (80 GB)	NVIDIA RTX A4500
Unique Technical Features	Two-step DL dose prediction + TPS dose-mimicking	Two-stage physics-guided training pipeline	First clinical-linac RL VMAT MPO	Tandem networks decouple MLC/MU exploration	CT → leaf-sequenced plan under 1 second	First sheaf-GNN inverse planning with signed PTV/OAR message passing

Chapter 13 Conclusion

This thesis has presented a unified, physics-rooted framework for radiotherapy dose prediction and inverse treatment planning that derives every input feature, architectural choice, and training signal from the fundamental equations of radiation transport. This research was conducted in lung cancer radiotherapy, an area of significant dosimetric and kinetic challenge due to the large heterogeneities of soft tissue within the thorax (air-soft-tissue interfaces, soft-tissue-bone interfaces) and the constant motion of target volumes by up to multiple centimeters caused by respiration. In addition, these heterogeneities create particle disequilibrium and steep dose gradients in both the linear Boltzmann transport equation and Acuros XB type dose calculations required for accurate dose determination at this site. Therefore, methods such as collapsed cone or pencil beam are insufficient to accurately determine dose in this setting. The primary premise of this research — that a deep learning algorithm trained on the physics of photon and electron transport can be more generalizable than one trained to reproduce clinical plans — is supported by the body of evidence collected through experiments in all three areas (forward prediction, cross modality comparison, and end-to-end inverse planning) on the most challenging anatomical site.

13.1 Physics-Informed Feature Extraction and the Novelty of Ψ_{unc}

The foundational contribution is the replacement of abstract binary masks (PTV, OAR contours) — used by virtually all prior methods from McIntosh and Purdie [84] to the OpenKBP challenge [85], Hien et al. [87], and Achlatis et al. [83] — with physics-derived quantities from the Linear Boltzmann Transport Equation. The uncollided photon fluence Ψ_{unc}, computed via analytical ray-tracing attenuated by the macroscopic cross section $\Sigma_t(\vec{r}, E)$ along each ray, contains the entire beam geometry — gantry angle, collimator, jaw, and MLC — and includes the tissue dependent exponential attenuation of the primary photon field. This is not a geometric identifier; it is an encoded physical field describing the first photon-matter interaction according to the Beer-Lambert Law. Therefore, the network does not have to learn anything about the primary attenuation; it can focus on what occurs subsequent to the first collision: Compton scattering, photoelectric re-emission, electron cascade generation according to the Bethe-Bloch stopping power, and the resulting dose deposition. Additionally, this representation is consistent with Acuros XB, which computes Ψ_{unc} and uses an iterative solution process for the scattered component; the network replaces the iterative scatter solver with a learned operator.

TERMA delivers the complementary first order energy deposition prior to charged particle transport, and the residual learning predicts the correction from TERMA to absorbed dose — precisely the scattered and secondary contribution that represents the dominant computational expense of deterministic solvers. In addition to ρ_e, distance-to-isocentre, and $|\mathbf{J}|$, these features give the network access to tissue composition, geometric divergence, and radiation directionality. Any change in beam configuration is immediately reflected in Ψ_{unc} and TERMA; the cross-modality generalisation from VMAT to IMRT without retraining is a direct consequence of this design.

The open-source LBTE solver that produces these features spans 44 tissue-equivalent materials (G_γ=25 photon, G_e=49 electron energy groups), with all cross sections derived from QED (Klein–Nishina) and classical physics (Bethe–Bloch, NIST photoelectric), achieving 92.55% gamma (3%/3 mm) in 34 s on a VMAT lung patient.

13.2 Forward Dose Prediction: Architecture and Validation

The dose prediction model uses an architecture that represents the multi-scale nature of the transport problem. The 3D U-Net models local effects — for example, tissue interface, electron transport into the penumbra, and density gradients — by using hierarchical encoding with

skip connections to allow the retention of spatial information. The Fourier Neural Operator performs learning in the spectral domain where low frequency modes naturally represent the broad scatter envelope, the depth-dose profile and beam geometry signatures that standard convolutional networks cannot capture. The spatial attention fusion method fuses these streams voxel-by-voxel, with local features weighted at sharp density transition points (for example lung – mediastinum) and global features weighted in smooth transport areas. The residual learning method forces the output to predict the corrections from scattering plus secondary above the TERMA baseline. Softplus activation also ensures the strict physical constraints of non-negative output — a constraint which many previous architectures have violated.

On 52 VMAT lung patients the model achieves 98.17% mean gamma (5%/3 mm, 10% threshold), 92.3% of patients above the 95% clinical threshold, MAE 1.71 ± 0.17 Gy, at 48.39 ms inference. On 26 IMRT patients — the first static-gantry evaluation in this framework — the mean gamma is 89.59% with MAE 0.84 ± 0.18 Gy, without modality-specific fine-tuning. Studies reporting higher gamma, such as Pastor-Serrano et al. (iDoTA, 99.5% [215]) and Yan et al. (GeoDL, 99.67% [216]), operate on prostate or mixed cohorts where tissue homogeneity is far greater and charged-particle disequilibrium is negligible. Lung is widely recognized as the most difficult site to accurately model because of the significant air-tissue density gradients present; and our framework is the only framework that has been validated both for VMAT and IMRT on this site using a single model.

13.3 Inverse Planning: Learning Radiotherapy Goals, Not Copying Plans

The inverse planning module represents a conceptual departure from the two-stage paradigm (predict dose → TPS re-optimisation) of Nemoto et al. [225], Wang et al. [226], and Kadoya et al. [238]. These methods use a network to learn the design of dose distributions designed by clinicians and have their own limits : the model does not know what the plan should do - it knows how a plan looks. Our fluence first approach generates deliverable fluence maps and MU weights directly; the resulting dose is generated by propagating the fluence forward through a frozen physics engine. The training signal evaluates the clinical quality of the dose that the predicted fluence would generate - i.e. how well it satisfies the DVH constraints, targets, etc., not how similar to a reference plan.

The Rotary Transformer encodes the periodic gantry geometry of VMAT via Rotary Position Embeddings (RoPE), so that self-attention reasons about angular proximity: beams separated by 10° are neighbours; beam 0 and beam 359 are adjacent. Unlike all previous transformer applications in radiotherapy, we have introduced a geometric inductive bias, which allows for smooth, deliverable fluence modulation along the arc. The Structure-Angle Sheaf Graph provides additional anatomical reasoning: on a bipartite graph that connects beam nodes to structure nodes, learned signed transport maps with positive (support) and negative (risk) channels propagate directional messages, and at the same time reinforce PTV coverage while suppressing OAR exposure through sheaf diffusion. The Transformer captures angular coherence (smoothness of beam-to-beam fluence); the Sheaf Graph captures anatomical coherence (dose trade-off between beam and structure). Together they produce plans that are both deliverable and clinically correct, and the network operates as a unified entity that considers geometry, physics, and clinical objectives simultaneously.

The bilevel curriculum training has the effect of transitioning the model from supervised fluence regression (Phase 1) to genuine physics driven optimisation (Phase 2). In Phase 2, DVH hinge losses penalise constraint violations, and the Adjoint Clinical Transport Prior (ACTP) projects voxel level clinical penalty back onto beamlet level fluence correction via the adjoint of the sparse dose kernel – not simply telling the network that a constraint violation occurred but which beamlets caused it and by how much. The frozen dose network from the forward model closes the differentiable physics loop: gradients from clinical losses propagate through the dose network, through the sparse Boltzmann kernels, and back to the fluence generator.

On 50 VMAT lung patients (AIRTP, 60 Gy/30 fr) the model achieves $85.20 \pm 7.96\%$ gamma (5%/3 mm) at 323 ms per plan — the fastest end-to-end inverse planning time reported for any anatomical site. OAR sparing matches or exceeds clinical references: Lung V20 pass rate +4.0 pp, Cord D0.03cc +12.0 pp, Heart D0.03cc +4.0 pp above the reference, while PTV mean dose is preserved ($\Delta = +0.16$ Gy). The coverage–protection trade-off (PTV V95%Rx −14 pp) reveals a conservative OAR-sparing strategy consistent with the Phase 2 loss landscape. Compared to Achlatis et al. [83] (90.50% gamma on prostate), Arberet et al. [241] (889 ms on prostate), Nemoto et al. [225] (4.6 min), and Wang et al. [226] (−70.5 min vs manual), our approach is faster, eliminates the two-stage pipeline, and operates on the substantially harder lung site with differentiable physics at both training and inference.

The best scientific test for validating the model's ability to produce clinically useful plans was a bilateral multi-metastasis worst-case scenario. For this worst-case, the model created the least gamma score it had ever produced in a physically coherent plan that covered all bilateral target volumes appropriately while preserving the necessary organs. The low gamma score indicates that the model has encountered plan non- uniqueness, but does not indicate a failure of the model. The model is able to create multiple valid plans that are acceptable to clinicians as well as a number of other plans that also meet the clinical criteria. Therefore, the model has demonstrated its ability to reason independently, using both the physical principles of radiation transport and the goals of clinical dosimetric radiobiology (i.e. deliver dose to targets with the least amount of damage to normal tissues) which is the primary goal of radiotherapy.

13.4 Limitations and Future Directions

The PTV V95%Rx deficit (−14 pp) indicates that adaptive loss weighting in Phase 2 should be explored. The IMRT gap (89.59% vs 98.17% VMAT) motivates modality-stratified training. The elevated HI (0.366 vs 0.242) and reduced CI (0.562 vs 0.721) suggest that Phase 2 include penalty terms for HI and CI. To achieve full clinical implementation, extension to other anatomical sites such as head-and-neck and prostate and to formal leaf sequencing with delivery verification on a clinical linear accelerator are needed. Although the physics-based features are inherently site-independent, the most promising area for future development includes Adaptive Radiation Therapy (ART). Although the anatomy of the thorax is not static due to the respiratory cycle causing intra-fraction displacement of tumours and normal tissue volumes, there are also inter-fraction anatomical variations, including tumour regression, pleural effusion and at-electasis, that progressively invalidate the originally planned treatment. Due to these anatomical variations during intra- and inter-session, the conventional workflow of Plan of the Day (POTD) is no longer sufficient and rapid online replanning of the dose distribution using the current anatomy is demanded. As indicated in this thesis, the computational speed of the framework (48.39 ms forward, 323 ms inverse) place the framework within the latency budget required for online ART. Therefore, future work should integrate respiratory motion models, including deformable image registration, 4-D CT phase binning and motion compensated dose accumulation, into the pipeline of physics based features to compute Ψ_{unc} and TERMA on the instantaneous anatomy and to allow for the generation of motion aware plans in real-time.

13.5 Closing Remarks

This thesis has demonstrated that by including radiation transport physics at each layer of an AI (deep learning) pipeline — i.e., from Ψ_{unc} at the input to the DVH hinge loss function at the output — creates a framework that is faster, more transferable and more physiologically meaningful than data driven pipelines. Forward predictions were made in 48.39 ms with 98.17% gamma; inverse planning at 323 ms with competitive OAR sparing; cross-modality generalization was achieved without re-training the network; and the network provided physiologically coherent reasoning on extreme out-of-distribution cases — all within the most difficult site of

interest in EBRT. Since the site considered here, where both significant tissue heterogeneities demand LBTE class dose accuracy and where respiratory motion demands sub-second replanning, represents one of the most urgent clinical scenarios for real-time ART — the clinical scenario where the need for high-speed optimization is greatest and where current optimization tool speeds remain orders of magnitude too low to support them.

Appendices

A Gamma Index Analysis

1.1 Theoretical Foundation of Gamma Analysis

1.1.1 The Dose Distribution Comparison Problem

In radiation therapy treatment planning verification, we need to compare a calculated dose distribution *calculated* dose distribution $D_c(\vec{r})$to a reference dose distribution *reference* dose distribution $D_r(\vec{r})$ (typically from Monte Carlo simulation or clinical TPS). (i.e. typically a dose distribution obtained by Monte Carlo simulation or clinical TPS). We have two important problems. Both of these are clinically relevant: A 3% error in dose at a given point is clinically significant; A 3 mm position discrepancy in a high gradient region is likely to be clinically acceptable. However, simple dose-difference metrics fail in high gradient regions while distance-to-agreement (DTA) metrics fail in low gradient regions.

The **gamma index** γ (Low et al., 1998 [242]) elegantly combines both criteria into a single metric.

1.2 Mathematical Definition of the Gamma Index

Definition A.1 (Gamma Index)**.** *For a reference point $\vec{r}_r$ with reference dose $D_r(\vec{r}_r)$, the gamma value is defined as:*

$$\boxed{\gamma(\vec{r}_r) = \min_{\vec{r}_c} \{\Gamma(\vec{r}_r, \vec{r}_c)\}} \tag{1.336}$$

*where the minimization is over all points $\vec{r}_c$ in the calculated distribution, and the **gamma function** Γ is:*

$$\Gamma(\vec{r}_r, \vec{r}_c) = \sqrt{\underbrace{\left(\frac{|\vec{r}_c - \vec{r}_r|}{\Delta d}\right)^2}_{\textit{Spatial term}} + \underbrace{\left(\frac{D_c(\vec{r}_c) - D_r(\vec{r}_r)}{\Delta D}\right)^2}_{\textit{Dose term}}} \tag{1.337}$$

where Δd is the distance-to-agreement (DTA) criterion in mm (typically 3mm), ΔD is the dose difference criterion in percent (typically 3%), $D_r(\vec{r}_r)$ is the reference dose at point $\vec{r}_r$, and $D_c(\vec{r}_c)$ is the calculated dose at point $\vec{r}_c$.

1.3 Geometric Interpretation

The gamma function $\Gamma(\vec{r}_r, \vec{r}_c)$ defines an **ellipsoidal acceptance surface** in the combined dose-space coordinates. Consider a 4D space (x, y, z, D) where positions are measured in units of Δd and dose in units of ΔD:

1.4 Pass/Fail Criterion

A point **passes** the gamma test if:

$$\gamma(\vec{r}_r) \leq 1 \tag{1.338}$$

This means there exists at least one point $\vec{r}_c$ in the calculated distribution that satisfies both criteria simultaneously within the tolerance.

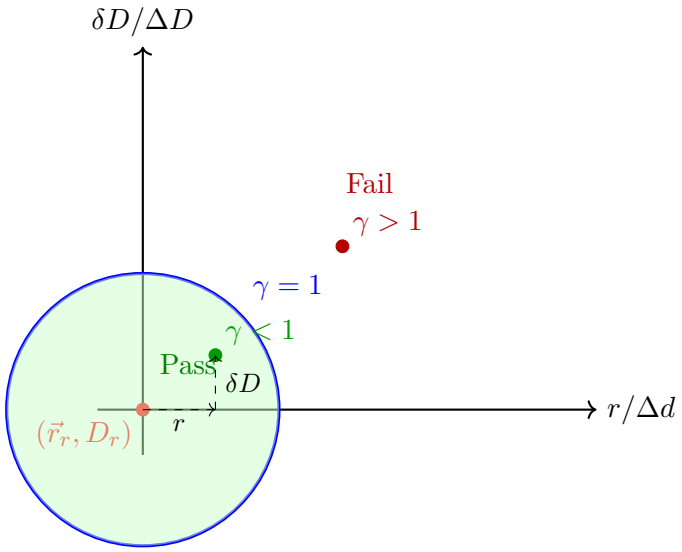


Figure 1.119: Geometric interpretation of gamma index. The unit circle defines the acceptance region ($\gamma \leq 1$). Points inside pass, points outside fail.

1.5 Gamma Pass Rate Calculation

The **gamma pass rate** is the primary clinical metric:

$$\boxed{\text{Pass Rate} = \frac{\text{Number of points with } \gamma \leq 1}{\text{Total number of evaluated points}} \times 100\%} \tag{1.339}$$

Clinical acceptance thresholds vary by application: the standard 3%/3mm criteria require at least 95% pass rate for routine clinical acceptance, while the more stringent 2%/2mm criteria are employed for stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS). Research-level agreement typically employs 1%/1mm criteria.

1.6 Computational Implementation of Gamma Index

1.7 Naive Algorithm (O(N^2) complexity)

The direct computation searches all calculated points for each reference point:

$$\gamma(\vec{r}_r) = \min_{\vec{r}_c \in \mathcal{V}_c} \sqrt{\frac{|\vec{r}_c - \vec{r}_r|^2}{\Delta d^2} + \frac{(D_c(\vec{r}_c) - D_r(\vec{r}_r))^2}{\Delta D^2}} \tag{1.340}$$

where $\mathcal{V}_c$ is the set of all voxels in the calculated distribution.

For typical 3D dose grids (256^3 voxels), this is $\sim 2.8 \times 10^{14}$ operations!

1.8 Search Radius Optimization

Since $\gamma > 1$ indicates failure, we can limit the search radius:

$$|\vec{r}_c - \vec{r}_r| \leq \Delta d \times \gamma_{\max} \tag{1.341}$$

For $\gamma_{\max} = 2$, the search radius is $2\Delta d = 6$mm (for 3mm criterion).

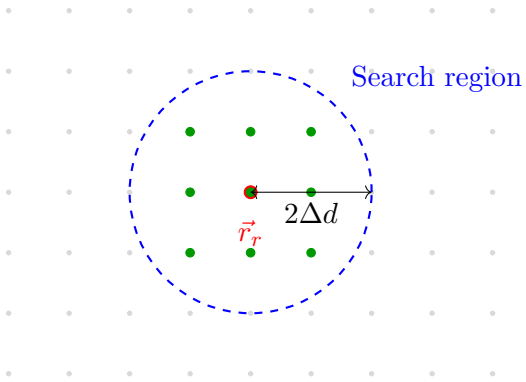


Figure 1.120: Search radius optimization: only points within $2\Delta d$ of the reference point need to be evaluated.

1.9 Fast Gamma Algorithm (Wendling et al.)

The Wendling algorithm [243] uses a more efficient approach that pre-computes dose differences $\delta D(\vec{r}) = D_c(\vec{r}) - D_r(\vec{r})$, creates sorted lists organized by dose difference magnitude, exploits spatial locality to prune the search space, and employs early termination when a passing gamma value ($\gamma \leq 1$) is found. This reduces complexity to approximately $O(N \log N)$.

1.10 Implementation in Our Solver

Our implementation uses a vectorized search approach over a local neighborhood. Algorithm 8 describes the 3D gamma index computation.

Algorithm 8 3D Gamma Index Computation

Require: Reference dose $\mathbf{D}_r \in \mathbb{R}^{N_x \times N_y \times N_z}$, calculated dose $\mathbf{D}_c$, distance criterion Δd [mm], dose criterion ΔD [%], voxel size $(\delta x, \delta y, \delta z)$ [mm]
Ensure: Gamma matrix $\mathbf{\Gamma}$, pass rate P

$D_{\max} \leftarrow \max(\mathbf{D}_r)$ ▷ Maximum reference dose
$\Delta D_{\text{abs}} \leftarrow (\Delta D/100) \cdot D_{\max}$ ▷ Absolute dose tolerance
$R \leftarrow \lceil 2 \cdot \Delta d / \min(\delta x, \delta y, \delta z) \rceil$ ▷ Search radius in voxels
Initialize $\mathbf{\Gamma} \leftarrow \infty$ (same shape as $\mathbf{D}_r$)
for each offset $\vec{o} = (o_x, o_y, o_z)$ with $\|\vec{o}\|_\infty \leq R$ **do**
    $r^2 \leftarrow (o_x \delta x)^2 + (o_y \delta y)^2 + (o_z \delta z)^2$ ▷ Physical distance squared
    $\mathbf{D}_c^{\text{shift}} \leftarrow \text{shift}(\mathbf{D}_c, \vec{o})$ ▷ Shifted calculated dose
    for each voxel (i, j, k) **do**
        $\delta D \leftarrow \mathbf{D}_c^{\text{shift}}(i, j, k) - \mathbf{D}_r(i, j, k)$ ▷ Dose difference
        $\Gamma_{ijk} \leftarrow \sqrt{\dfrac{r^2}{\Delta d^2} + \dfrac{\delta D^2}{\Delta D_{\text{abs}}^2}}$
        $\mathbf{\Gamma}(i, j, k) \leftarrow \min\left(\mathbf{\Gamma}(i, j, k), \Gamma_{ijk}\right)$
    end for
end for
// Compute pass rate above threshold
$\mathcal{M} \leftarrow \{(i, j, k) : \mathbf{D}_r(i, j, k) \geq 0.1 \cdot D_{\max}\}$ ▷ Dose threshold mask
$P \leftarrow 100 \cdot |\{(i, j, k) \in \mathcal{M} : \mathbf{\Gamma}(i, j, k) \leq 1\}| / |\mathcal{M}|$
return $\mathbf{\Gamma}, P$

1.11 Global vs Local Gamma

1.12 Global Gamma (Recommended)

In **global gamma**, the dose criterion ΔD is defined relative to a *global* reference dose (typically maximum dose or prescribed dose):

$$\Delta D_{\text{global}} = \frac{\Delta D_{\%}}{100} \times D_{\max} \tag{1.342}$$

This gives consistent tolerance across the entire dose distribution.

1.13 Local Gamma

In **local gamma**, the dose criterion is relative to the *local* dose:

$$\Delta D_{\text{local}}(\vec{r}_r) = \frac{\Delta D_{\%}}{100} \times D_r(\vec{r}_r) \tag{1.343}$$

This is more stringent in high-dose regions but can be problematic in low-dose regions (division by small numbers).

1.14 Dose Threshold

To avoid noise in low-dose regions, a **dose threshold** is typically applied:

$$\text{Evaluate } \gamma(\vec{r}_r) \text{ only if } D_r(\vec{r}_r) \geq D_{\text{threshold}} \tag{1.344}$$

Common threshold values include 10% of maximum dose as the default, 50% of maximum dose for target-focused analysis, or absolute dose thresholds such as 10 cGy.

B Source Model and Beam Fluence Calculation

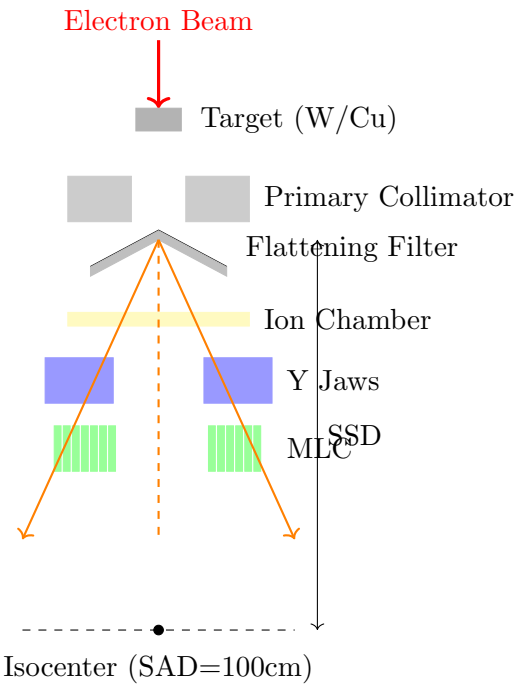


Figure 2.121: Schematic of a medical linac head showing the main components that shape the photon beam.

2.1 Beam Energy Spectrum

The photon beam from a linac has a continuous energy spectrum. For a nominal "6 MV" beam, the spectrum spans from near zero to approximately 6 MeV, with a mean energy around 2 MeV. The spectral parameterizations follow Sheikh-Bagheri and Mohan models.

2.2 Angular Fluence Distribution

The primary fluence exiting the linac head follows:

$$\Psi_0(\vec{r}, \hat{\Omega}) = \frac{\Psi_{\text{SAD}}}{|\vec{r} - \vec{r}_{\text{source}}|^2} \cdot \cos\theta \cdot A(\vec{r}_{\text{MLC}}) \cdot T(\vec{r}_{\text{jaws}}) \tag{2.345}$$

where Ψ_{SAD} is the fluence at source-axis distance (SAD), $\vec{r}_{\text{source}}$ is the virtual source position, θ is the angle from the central axis, $A(\vec{r}_{\text{MLC}})$ is the MLC aperture function (0 or 1), and $T(\vec{r}_{\text{jaws}})$ is the jaw transmission function.

2.3 MLC Aperture Modeling

2.4 Multi-Leaf Collimator Geometry

The MLC consists of opposing leaf banks that can independently move to shape the beam aperture. Each leaf is characterized by its width w (typically 5mm or 10mm at isocenter), leaf tip position x_{tip}, and tongue-and-groove design for interleaf leakage reduction.

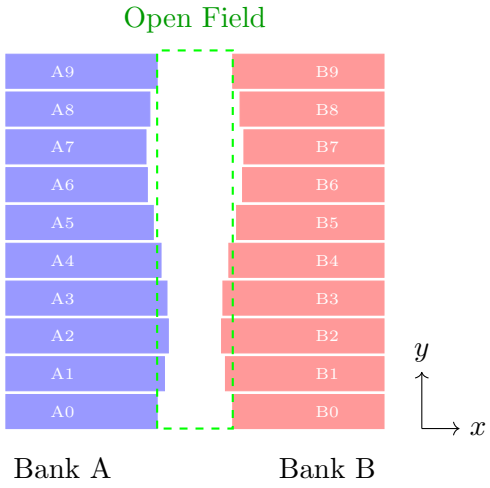


Figure 2.122: MLC leaf bank configuration. Each leaf pair defines an opening in the x-direction. The shape of the field is defined by leaf positions $(x_{A,i}, x_{B,i})$.

2.5 MLC Aperture Function

For DICOM RT Plan data, the aperture function at leaf row i is:

$$A_i(x) = \begin{cases} 1 & \text{if } x_{A,i} < x < x_{B,i} \\ T_{\text{leaf}} & \text{if } x < x_{A,i} \text{ or } x > x_{B,i} \end{cases} \tag{2.346}$$

where $T_{\text{leaf}} \approx 0.01$ is the leaf transmission factor.
The total MLC transmission at position (x, y) is:

$$A_{\text{MLC}}(x, y) = A_{i(y)}(x) \tag{2.347}$$

where $i(y)$ is the leaf index corresponding to y-position.

2.6 VMAT Fluence Integration

2.7 Control Point Interpolation

In VMAT delivery, the gantry rotates continuously while the MLC and dose rate change. The treatment is discretized into **control points**:

$$\Phi_{\text{total}} = \sum_{j=1}^{N_{cp}} \underbrace{w_j}_{\text{MU weight}} \cdot \underbrace{\Phi_j(\vec{r}, \hat{\Omega})}_{\text{Fluence at CP}_j} \tag{2.348}$$

Between control points, the gantry angle follows linear interpolation $\theta_g(t)$, the MLC positions are linearly interpolated $x_{\text{leaf}}(t)$, and the dose rate (MU/deg) is treated as piecewise constant over each interval.

2.8 Cumulative Meterset Weight

The DICOM "Cumulative Meterset Weight" (CMW) gives the integrated dose fraction:

$$\text{CMW}_j = \frac{\sum_{k=1}^{j} \Delta\text{MU}_k}{\text{Total MU}} \tag{2.349}$$

The weight for each control point interval is:

$$w_j = \text{CMW}_{j+1} - \text{CMW}_j \tag{2.350}$$

2.9 Source Building Algorithm

The Source Building Algorithm is a multi-step process which begins with parsing the RTPLAN file to obtain the beam geometry, control point locations, and MLC positions for each control point in the plan. Once these values are obtained, the algorithm calculates the gantry, collimator, and couch angles, constructs the corresponding rotation matrices $\mathbf{R}_g$, $\mathbf{R}_c$, and $\mathbf{R}_t$, and transforms the source location into the patient coordinate system. Following this, ray tracing is performed on each voxel in the CT grid by projecting that voxel onto the BEV, checking whether the voxel intersects an MLC aperture at that projection location and accumulating the total fluence of all rays from that voxel weighted by its cumulative meter set weight. Finally, the angular discretization maps continuous beam directions to S_N quadrature ordinate and the energy spectrum distributes the source among different energy groups.

$$\boxed{Q_{g,m}(\vec{r}) = \sum_{j=1}^{N_{cp}} w_j \cdot A_{\text{MLC}}^{(j)}(\vec{r}) \cdot \frac{\Phi(E_g)}{4\pi} \cdot \frac{\delta(\hat{\Omega} - \hat{\Omega}_j^{\text{beam}})}{|\vec{r} - \vec{r}_{\text{src}}^{(j)}|^2}} \tag{2.351}$$

2.10 Angular Quadrature Assignment

The continuous beam direction $\hat{\Omega}_{\text{beam}}$ must be assigned to discrete S_N ordinates. We use the nearest-neighbor approach:

$$m^* = \arg\min_m |\hat{\Omega}_m - \hat{\Omega}_{\text{beam}}| \tag{2.352}$$

For improved accuracy, the source can be distributed over multiple nearby ordinates using a smoothing kernel:

$$Q_m = \sum_{m'} K(\hat{\Omega}_m, \hat{\Omega}_{m'}) \cdot Q_{m'}^{\text{raw}} \tag{2.353}$$

where K is a Gaussian kernel with width related to the angular quadrature spacing.

2.11 Siddon Ray Tracing for Voxel Coverage

To determine which voxels receive primary fluence, we use **Siddon ray tracing** [53]:

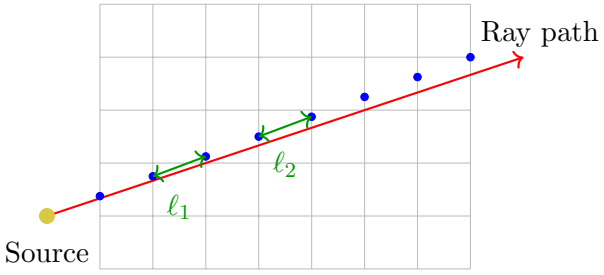


Figure 2.123: Siddon ray tracing algorithm: the path length ℓ_i through each voxel is computed analytically from grid intersection points.

The algorithm computes parametric intersections:

$$t_{x,i} = \frac{x_i - x_{\text{source}}}{\Omega_x} \quad (i = 0, 1, \ldots, N_x) \tag{2.354}$$

$$t_{y,j} = \frac{y_j - y_{\text{source}}}{\Omega_y} \quad (j = 0, 1, \ldots, N_y) \tag{2.355}$$

$$t_{z,k} = \frac{z_k - z_{\text{source}}}{\Omega_z} \quad (k = 0, 1, \ldots, N_z) \tag{2.356}$$

Merging and sorting these t-values gives the voxel boundaries along the ray, and path lengths:

$$\ell_n = (t_{n+1} - t_n) \cdot |\vec{r}_{\text{dest}} - \vec{r}_{\text{source}}| \tag{2.357}$$

C Materials Library Data Structure

3.1 Energy Group Structure

3.1.1 Photon Energy Groups

Photon groups are defined by energy bin edges in MeV. The groups span from the photon cutoff energy (1 keV) to the maximum beam energy (typically 6 MeV for clinical linacs), with logarithmic spacing to adequately sample the rapid variation in cross-sections at lower energies.

Index	$E^{\mathbf{low}}$ **[MeV]**	$E^{\mathbf{high}}$ **[MeV]**
1	0.001	0.01
2	0.01	0.05
⋮	⋮	⋮
G_γ	4.0	6.0

The YAML configuration specifies the photon group boundaries as an array of energy edges in MeV,

Number of photon groups: $G_\gamma = N_{\text{edges}} - 1$

Group-averaged energies are geometric means:

$$E_g = \sqrt{E_g^{\text{low}} \cdot E_g^{\text{high}}} \tag{3.358}$$

3.2 Electron Energy Groups

Electron groups span from the cutoff energy (200 keV) to the maximum beam energy, with logarithmic spacing to capture the rapid variation in stopping power at lower energies.

Index	$E^{\mathbf{low}}$ **[MeV]**	$E^{\mathbf{high}}$ **[MeV]**
1	0.20	0.25
2	0.25	0.32
⋮	⋮	⋮
G_e	4.5	6.0

Number of electron groups: $G_e = N_{\text{edges}} - 1$

3.3 Legendre Expansion Orders

Angular distributions are expanded in Legendre polynomials with truncation orders:

Symbol	**Process**	**Typical Value**
L_γ	Photon-photon scattering (Compton)	4
L_e	Electron-electron scattering (Møller)	8
$L_{\gamma e}$	Photon-electron coupling	1

The scattering cross-section expansion:

$$\sigma_s(\mu) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi} \sigma_{s,\ell} P_\ell(\mu) \tag{3.359}$$

where $\mu = \hat{\Omega} \cdot \hat{\Omega}'$ and $\sigma_{s,\ell}$ are the Legendre moments.

3.4 Material Entry Format

Each material entry in the YAML library contains the fields listed in Table 3.33.

Table 3.33: Material data structure for the cross-section library.

Quantity	**Description**
Material identifier	Unique name (e.g., "SoftTissue-16")
Mass density ρ	Density in g/cm^3
Elemental composition	List of (element, mass fraction) pairs
Photon cross-section data	
$\sigma_t^{(\gamma)}$	Total photon cross-section [G_γ groups]
$\sigma_{s,\ell}^{(\gamma)}$	Legendre scattering moments [$G_\gamma \times G_\gamma \times L_\gamma$]
σ_{Comp}	Compton cross-section [G_γ groups]
σ_{PE}	Photoelectric cross-section [G_γ groups]
σ_{pair}	Pair production cross-section [G_γ groups]
Electron cross-section data	
S_R	Restricted stopping power [G_e groups]
$\sigma_{ED}^{(e)}$	Energy deposition cross-section [G_e groups]
$\sigma_t^{(e)}$	Electron transport cross-section [G_e groups]
$\sigma_{ee,\ell}$	Møller scattering moments [$G_e \times G_e \times L_e$]
$\sigma_{\gamma\to e,\ell}$	Photon-electron coupling [$G_\gamma \times G_e \times L_{\gamma e}$]

References

[1] K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," *Medical Physics*, vol. 35, no. 1, pp. 310–317, 2008.

[2] Y. Li, J. Wang, L. Tan, B. Hui, X. Ma, Y. Yan, C. Xue, X. Shi, E. Drokow, and J. Ren, "Dosimetric comparison between imrt and vmat in irradiation for peripheral and central lung cancer," *Oncology Letters*, Jan. 2018.

[3] K. Bush, I. M. Gagne, S. Zavgorodni, W. Ansbacher, and W. Beckham, "Dosimetric validation of acuros® xb with monte carlo methods for photon dose calculations," *Medical Physics*, vol. 38, p. 2208–2221, Mar. 2011.

[4] O. N. Vassiliev, T. A. Wareing, J. McGhee, G. Failla, M. R. Salehpour, and F. Mourtada, "Validation of a new grid-based boltzmann equation solver for dose calculation in radiotherapy with photon beams," *Physics in Medicine and Biology*, vol. 55, p. 581–598, Jan. 2010.

[5] I. J. Chetty, B. Curran, J. E. Cygler, J. J. DeMarco, G. Ezzell, B. A. Faddegon, I. Kawrakow, P. J. Keall, H. Liu, C. C. Ma, D. W. O. Rogers, J. Seuntjens, D. Sheikh-Bagheri, and J. V. Siebers, "Report of the aapm task group no. 105: Issues associated with clinical implementation of monte carlo-based photon and electron external beam treatment planning," *Medical Physics*, vol. 34, p. 4818–4853, Nov. 2007.

[6] D. W. O. Rogers, "Fifty years of monte carlo simulations for medical physics," *Physics in Medicine and Biology*, vol. 51, p. R287–R301, June 2006.

[7] L. Hoffmann, M. Alber, M. Söhn, and U. V. Elstrøm, "Validation of the acuros ¡scp¿xb¡/scp¿ dose calculation algorithm versus monte carlo for clinical treatment plans," *Medical Physics*, vol. 45, p. 3909–3915, July 2018.

[8] T. Knoos, A. Ahnesjo, P. Nilsson, and L. Weber, "Limitations of a pencil beam approach to photon dose calculations in lung tissue," *Physics in Medicine and Biology*, vol. 40, p. 1411–1420, Sept. 1995.

[9] P. A. Taylor, S. F. Kry, and D. S. Followill, "Pencil beam algorithms are unsuitable for proton dose calculations in lung," *International Journal of Radiation Oncology*Biology*Physics*, vol. 99, p. 750–756, Nov. 2017.

[10] L. Tillikainen, H. Helminen, T. Torsti, S. Siljamäki, J. Alakuijala, J. Pyyry, and W. Ulmer, "A 3d pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media," *Physics in Medicine and Biology*, vol. 53, p. 3821–3839, June 2008.

[11] Y. Tsuruta, M. Nakata, M. Nakamura, Y. Matsuo, K. Higashimura, H. Monzen, T. Mizowaki, and M. Hiraoka, "Dosimetric comparison of acuros xb, aaa, and xvmc in stereotactic body radiotherapy for lung cancer," *Medical Physics*, vol. 41, July 2014.

[12] K. Shao, C. Chai, Y. Yang, W. Chen, and F. Du, "Dosimetric comparison of monte carlo, acuros xb, and anisotropic analytical algorithm for lung cancer plans on halcyon accelerators," *PLOS One*, vol. 20, p. e0334329, Oct. 2025.

[13] F. M. Khan and J. P. Gibbons, *Khan's The Physics of Radiation Therapy.* Lippincott Williams and Wilkins, 5th ed., 2014.

[14] E. B. Podgorsak, *Radiation Oncology Physics: A Handbook for Teachers and Students.* Vienna: International Atomic Energy Agency, 2005.

[15] J. P. Gibbons, J. A. Antolak, D. S. Followill, M. S. Huq, E. E. Klein, K. L. Lam, J. R. Palta, D. M. Roback, M. Reid, and F. M. Khan, "Monitor unit calculations for external photon and electron beams: Report of the aapm therapy physics committee task group no. 71," *Medical Physics*, vol. 41, Feb. 2014.

[16] P. R. Almond, P. J. Biggs, B. M. Coursey, W. F. Hanson, M. S. Huq, R. Nath, and D. W. O. Rogers, "Aapm's tg-51 protocol for clinical reference dosimetry of high-energy photon and electron beams," *Medical Physics*, vol. 26, p. 1847–1870, Sept. 1999.

[17] S. Mahase, D. Julie, K. Salah, P. Christos, and J. Knisely, "A dosimetric evaluation of 6 mv versus 10 mv photon whole-brain radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 105, p. E80–E81, Sept. 2019.

[18] S. Webb, "Intensity-modulated radiation therapy," May 2015.

[19] I. Patel, A. G. Glendinning, and M. C. Kirby, "Dosimetric characteristics of the elekta beam modulator™," *Physics in Medicine and Biology*, vol. 50, p. 5479–5492, Nov. 2005.

[20] C. Hong, S. G. Ju, M. Kim, J. Kim, J. M. Kim, T. Suh, Y. Han, Y. C. Ahn, D. H. Choi, H. Nam, and H. C. Park, "Dosimetric effects of multileaf collimator leaf width on intensity-modulated radiotherapy for head and neck cancer," *Medical Physics*, vol. 41, Jan. 2014.

[21] Z. C. Taşkın, J. C. Smith, H. E. Romeijn, and J. F. Dempsey, "Optimal multileaf collimator leaf sequencing in imrt treatment planning," *Operations Research*, vol. 58, p. 674–690, June 2010.

[22] P. Xia and L. J. Verhey, "Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments," *Medical Physics*, vol. 25, p. 1424–1434, Aug. 1998.

[23] S. V. Spirou and C. S. Chui, "Generation of arbitrary intensity profiles by dynamic jaws or multileaf collimators," *Medical Physics*, vol. 21, p. 1031–1041, July 1994.

[24] D. A. Low, J. M. Moran, J. F. Dempsey, L. Dong, and M. Oldham, "Dosimetry tools and techniques for imrt," *Medical Physics*, vol. 38, p. 1313–1338, Feb. 2011.

[25] N. Linthout, D. Verellen, S. Van Acker, and G. Storme, "A simple theoretical verification of monitor unit calculation for intensity modulated beams using dynamic mini-multileaf collimation," *Radiotherapy and Oncology*, vol. 71, p. 235–241, May 2004.

[26] C. C. Ling, P. Zhang, Y. Archambault, J. Bocanek, G. Tang, and T. LoSasso, "Commissioning and quality assurance of rapidarc radiotherapy delivery system," *International Journal of Radiation Oncology*Biology*Physics*, vol. 72, p. 575–581, Oct. 2008.

[27] A. Mans, M. Wendling, J. Stroom, P. Remeijer, J. Sonke, B. Mijnheer, and M. van Herk, "3d dosimetric verification of vmat by portal dosimetry," *Radiotherapy and Oncology*, vol. 92, p. S25–S26, Aug. 2009.

[28] H. K. Malhotra, S. Raina, J. S. Avadhani, S. deBoer, and M. B. Podgorsak, "Technical and dosimetric considerations in imrt treatment planning for large target volumes," *Journal of Applied Clinical Medical Physics*, vol. 6, p. 77–87, Sept. 2005.

[29] M. Athiyaman, A. Hemalatha, A. Chougule, M. Joan, and H. Kumar, "Estimation of monitor unit through analytical method for dynamic imrt using control points as an effective parameter," *Journal of Radiotherapy in Practice*, vol. 21, p. 343–350, Mar. 2021.

[30] A. Goraj and S. F. de Boer, "Impact of the number of control points has on isodose distributions in a dynamic multileaf collimator intensity-modulated radiation therapy delivery," *Medical Dosimetry*, vol. 37, p. 412–416, Dec. 2012.

[31] S. Webb, "The physical basis of imrt and inverse planning," *The British Journal of Radiology*, vol. 76, p. 678–689, Oct. 2003.

[32] T. Bortfeld, "Imrt: a review and preview," *Physics in Medicine and Biology*, vol. 51, p. R363–R379, June 2006.

[33] Q. Lyu, V. Y. Yu, D. Ruan, R. Neph, D. O'Connor, and K. Sheng, "A novel optimization framework for vmat with dynamic gantry couch rotation," *Physics in Medicine & Biology*, vol. 63, p. 125013, June 2018.

[34] E. Y. Cheung, K. T. Chan, K. T. Cheung, T. S. Cheung, and J. C. Yiu, "Mo11-5 dosimetry comparison between imrt and 3dcrt/ vmat for centrally located non-small cell lung cancer," *Annals of Oncology*, vol. 33, p. S488, July 2022.

[35] P. D. Maguire, G. S. Sibley, S.-M. Zhou, T. A. Jamieson, K. L. Light, P. A. Antoine, J. E. Herndon, M. S. Anscher, and L. B. Marks, "Clinical and dosimetric predictors of radiation-induced esophageal toxicity," *International Journal of Radiation Oncology*Biology*Physics*, vol. 45, p. 97–103, Aug. 1999.

[36] C. X. Yu and G. Tang, "Intensity-modulated arc therapy: principles, technologies and clinical implementation," *Physics in Medicine and Biology*, vol. 56, p. R31–R54, Feb. 2011.

[37] J. L. Bedford and A. P. Warrington, "Commissioning of volumetric modulated arc therapy (vmat)," *International Journal of Radiation Oncology*Biology*Physics*, vol. 73, p. 537–545, Feb. 2009.

[38] M. Kang, J. Dai, M. Huang, Y. Xu, and J. An, "Su-e-t-619: Dosimetric comparison between static imrt and vmat for cervical cancer," *Medical Physics*, vol. 40, p. 348–348, June 2013.

[39] S. S. Vedam, "Management of respiratory motion in radiation oncology," *Medical Physics*, vol. 30, p. 2562–2562, Sept. 2003.

[40] B. F. Hasson, C. Ma, L. Wang, D. E. Wazer, J. E. Reiff, J. E. Reiff, B. J. Fisher, T. W. Speer, J. H. Brashears, J. P. Christodouleas, A. E. Dragun, V. Budach, T. Holmes, F. T. Troicki, J. Poli, F. T. Troicki, J. Poli, L. G. Jensen, L. K. Mell, S. Mose, E. Limbergen, C. B. Saw, N. R. Patel, M. L. Wong, N. J. Yue, J. H. Brashears, C. Heese, R. Rengan, C. R. Thomas, R. Rubin, R. Rubin, J. E. Reiff, J. E. Reiff, N. R. Patel, M. L. Wong, J. A. Chalal, G. G. Grabenbauer, L. G. Jensen, B. S. Rose, A. J. Mundt, N. J. Yue, R. Rubin, T. W. Speer, T. W. Speer, R. Rubin, R. Rubin, R. Rubin, T. W. Speer, D. J. Indelicato, R. H. Sagerman, L. T. Komarnicky-Kocher, C. A. Knowlton, M. K. Mackay, J. A. Chalal, D. Michalski, and M. S. Huq, *AAPM Task Group*, p. 1–1. Springer Berlin Heidelberg, 2013.

[41] W.-Z. Chen, "Impact of dose calculation algorithm on radiation therapy," *World Journal of Radiology*, vol. 6, no. 11, p. 874, 2014.

[42] B. Sarin, B. Bindhu, P. Raghu Kumar, S. Sumeesh, and B. Saju, "Dosimetric accuracy of acuros® xb and aaa algorithms for stereotactic body radiotherapy (sbrt) lung treatments: evaluation with primo monte carlo code," *Journal of Radiotherapy in Practice*, vol. 22, Dec. 2022.

[43] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros¡scp¿cts¡/scp¿: A fast, linear boltzmann transport equation solver for computed tomography scatter – part i: Core algorithms and validation," *Medical Physics*, vol. 45, p. 1899–1913, Apr. 2018.

[44] A. Ahnesjö, M. Saxner, and A. Trepp, "A pencil beam model for photon dose calculation," *Medical Physics*, vol. 19, p. 263–273, Mar. 1992.

[45] W. P. M. Mayles, A. E. Nahum, and J.-C. Rosenwald, *Handbook of Radiotherapy Physics: Theory and Practice, Second Edition*. CRC Press, Dec. 2021.

[46] C. Zhou, N. Bennion, R. Ma, X. Liang, S. Wang, K. Zvolanek, M. Hyun, X. Li, S. Zhou, W. Zhen, C. Lin, A. Wahl, and D. Zheng, "A comprehensive dosimetric study on switching from a type-b to a type-c dose algorithm for modern lung sbrt," *Radiation Oncology*, vol. 12, May 2017.

[47] D. Georg, T. Knöös, and B. McClean, "Current status and future perspective of flattening filter free photon beams," *Medical Physics*, vol. 38, p. 1280–1293, Feb. 2011.

[48] A. Fogliata, G. Nicolini, A. Clivio, E. Vanetti, and L. Cozzi, "Dosimetric evaluation of acuros xb advanced dose calculation algorithm in heterogeneous media," *Radiation Oncology*, vol. 6, July 2011.

[49] N. Papanikolaou, J. J. Battista, A. L. Boyer, C. Kappas, E. Klein, T. R. Mackie, M. Sharpe, and J. Van Dyk, "Tissue inhomogeneity corrections for megavoltage photon beams," Tech. Rep. Report No. 85 (Task Group 65), American Association of Physicists in Medicine, 2004.

[50] A. Tai, Z. Liang, E. Gore, and X. Li, "Dosimetric impact of respiration gating on stereotactic body radiation therapy of centrally located lung tumors," *International Journal of Radiation Oncology*Biology*Physics*, vol. 78, p. S780, Nov. 2010.

[51] T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-mv x rays," *Medical Physics*, vol. 12, p. 188–196, Mar. 1985.

[52] P. Zhang, H. Shu, A. Simon, P. Haigron, and R. De Crevoisier, "A new pencil beam model for photon dose calculations," in *2013 6th International Conference on Biomedical Engineering and Informatics*, p. 428–432, IEEE, Dec. 2013.

[53] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional ct array," *Medical Physics*, vol. 12, p. 252–255, Mar. 1985.

[54] J. R. Palta, T. R. Mackie, and Z. Chen, "Intensity-modulated radiation therapy—the state of the art," *Medical Physics*, vol. 30, p. 3265–3265, Dec. 2003.

[55] R. Mohan, C. Chui, and L. Lidofsky, "Energy and angular distributions of photons from medical linear accelerators," *Medical Physics*, vol. 12, p. 592–597, Sept. 1985.

[56] D. Sheikh-Bagheri and D. W. O. Rogers, "Monte carlo calculation of nine megavoltage photon beam spectra using the beam code," *Medical Physics*, vol. 29, p. 391–402, Feb. 2002.

[57] A. Ahnesjö, M. Saxner, and A. Trepp, "A pencil beam model for photon dose calculation," *Medical Physics*, vol. 19, p. 263–273, Mar. 1992.

[58] A. Ahnesjö, M. Saxner, and A. Trepp, "A pencil beam model for photon dose calculation," *Medical Physics*, vol. 19, p. 263–273, Mar. 1992.

[59] K. Ito, Y. Ishikawa, S. Teramura, N. Kadoya, Y. Katsuta, S. Tanaka, K. Takeda, K. Jingu, and T. Yamada, "Development of a collapsed cone convolution/superposition dose calculation algorithm with a mass density-specific water kernel for magnetic resonance-guided radiotherapy," *Journal of Radiation Research*, vol. 64, p. 496–508, Mar. 2023.

[60] A. Van Esch, L. Tillikainen, J. Pyykkonen, M. Tenhunen, H. Helminen, S. Siljamäki, J. Alakuijala, M. Paiusco, M. Iori, and D. P. Huyskens, "Testing of the analytical anisotropic algorithm for photon dose calculation," *Medical Physics*, vol. 33, p. 4130–4148, Oct. 2006.

[61] A. Fogliata, "Abstract id: 377 the photon dose calculation algorithms," *Physica Medica*, vol. 56, p. 284–285, Dec. 2018.

[62] D. W. O. Rogers, "Fifty years of monte carlo simulations for medical physics," *Physics in Medicine and Biology*, vol. 51, p. R287–R301, June 2006.

[63] J. Sempau, J. M. Fernández-Varea, E. Acosta, and F. Salvat, "Experimental benchmarks of the monte carlo code penelope," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 207, no. 2, pp. 107–123, 2003.

[64] I. Kawrakow, "Accurate condensed history monte carlo simulation of electron transport. i. ¡scp¿egs¡/scp¿nrc, the new ¡scp¿egs4¡/scp¿ version," *Medical Physics*, vol. 27, p. 485–498, Mar. 2000.

[65] J. Unkelbach and D. Papp, "The emergence of nonuniform spatiotemporal fractionation schemes within the standard bed model," *Medical Physics*, vol. 42, p. 2234–2241, May 2015.

[66] S. Hissoiny, B. Ozell, H. Bouchard, and P. Després, "¡scp¿gpumcd¡/scp¿: A new gpu-oriented monte carlo dose calculation platform," *Medical Physics*, vol. 38, p. 754–764, Jan. 2011.

[67] R. Gao, B. Lou, Z. Xu, D. Comaniciu, and A. Kamen, "Flexible-cmgan: Towards precise 3d dose prediction in radiotherapy," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 715–725, IEEE, June 2023.

[68] R. Gao, M. Diallo, H. Liu, A. Magliari, J. Sackett, W. Verbakel, S. Meyers, R. Mcbeth, and M. Zarepisheh, "Automating RT planning at scale: High quality data for AI training," *arXiv preprint arXiv:2501.11803*, 2025. AAPM 2025 Challenge Dataset.

[69] A. Fogliata and L. Cozzi, "Dose calculation algorithm accuracy for small fields in non-homogeneous media: The lung sbrt case," *Physica Medica*, vol. 44, p. 157–162, Dec. 2017.

[70] T. Knöös, E. Wieslander, L. Cozzi, C. Brink, A. Fogliata, D. Albers, H. Nyström, and S. Lassen, "Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations," *Physics in Medicine and Biology*, vol. 51, p. 5785–5807, Oct. 2006.

[71] C. Yan, A. G. Combine, G. Bednarz, R. J. Lalonde, B. Hu, K. Dickens, R. Wynn, D. C. Pavord, and M. Saiful Huq, "Clinical implementation and evaluation of the acuros dose calculation algorithm," *Journal of Applied Clinical Medical Physics*, vol. 18, p. 195–209, Aug. 2017.

[72] L. E. Henke, J. R. Olsen, J. A. Contreras, A. Curcuru, T. A. DeWees, O. L. Green, J. Michalski, S. Mutic, M. C. Roach, J. D. Bradley, P. J. Parikh, R. Kashani, and C. G. Robinson, "Stereotactic mr-guided online adaptive radiation therapy (smart) for ultra-central thorax malignancies: Results of a phase 1 trial," *Advances in Radiation Oncology*, vol. 4, p. 201–209, Jan. 2019.

[73] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, p. 686–707, Feb. 2019.

[74] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, p. 422–440, May 2021.

[75] G. Delaney, S. Jacob, C. Featherstone, and M. Barton, "The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines," *Cancer*, vol. 104, no. 6, pp. 1129–1137, 2005.

[76] A. Ahnesjö, "Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media," *Medical Physics*, vol. 16, p. 577–592, July 1989.

[77] F. Verhaegen and J. Seuntjens, "Monte carlo modelling of external radiotherapy photon beams," *Physics in Medicine and Biology*, vol. 48, p. R107–R164, Oct. 2003.

[78] O. M. Dona Lemus, M. Cao, B. Cai, M. Cummings, and D. Zheng, "Adaptive radiotherapy: Next-generation radiotherapy," *Cancers*, vol. 16, p. 1206, Mar. 2024.

[79] E. Borderias-Villarroel, M. Huet Dastarac, A. M. Barragán-Montero, R. Helander, M. Holmstrom, X. Geets, and E. Sterpin, "Machine learning-based automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness," *Medical Physics*, vol. 50, p. 4480–4490, Apr. 2023.

[80] C. McIntosh, L. Conroy, M. C. Tjong, T. Craig, A. Bayley, C. Catton, M. Gospodarowicz, J. Helou, N. Isfahanian, V. Kong, T. Lam, S. Raman, P. Warde, P. Chung, A. Berlin, and T. G. Purdie, "Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer," *Nature Medicine*, vol. 27, p. 999–1005, June 2021.

[81] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, p. 2318–2331, Oct. 2017.

[82] H. S. Tan, K. Wang, and R. McBeth, "Deep evidential learning for radiotherapy dose prediction," *Computers in Biology and Medicine*, vol. 182, p. 109172, Nov. 2024.

[83] X. Zhang, P. He, Y. Li, X. Liu, Y. Ma, G. Shen, Z. Dai, H. Zhang, W. Chen, and Q. Li, "Dr-only carbon-ion radiotherapy treatment planning via deep learning," *Physica Medica*, vol. 100, p. 120–128, Aug. 2022.

[84] C. McIntosh and T. G. Purdie, "Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning," *Physics in Medicine and Biology*, vol. 62, p. 415–431, Dec. 2016.

[85] A. Babier, B. Zhang, R. Mahmood, K. L. Moore, T. G. Purdie, A. L. McNiven, and T. C. Y. Chan, "Openkbp: The open-access knowledge-based planning grand challenge and dataset," *Medical Physics*, vol. 48, p. 5549–5561, June 2021.

[86] Z. Feng, L. Wen, P. Wang, B. Yan, X. Wu, J. Zhou, and Y. Wang, *DiffDP: Radiotherapy Dose Prediction via a Diffusion Model*, p. 191–201. Springer Nature Switzerland, 2023.

[87] L. T. Hien, P. T. Hieu, and D. N. Toan, "An efficient 3d convolutional neural network for dose prediction in cancer radiotherapy from ct images," *Diagnostics*, vol. 15, p. 177, Jan. 2025.

[88] Z. Sun, Y. Yan, Y. Chen, G. Yao, J. Wang, W. Hu, Z. Lu, and S. Yan, "A physics-informed deep learning model for predicting beam dose distribution of intensity-modulated radiation therapy treatment plans," *Physics and Imaging in Radiation Oncology*, vol. 34, p. 100779, Apr. 2025.

[89] Y. Li, W. Cai, F. Xiao, X. Zhou, J. Cai, L. Zhou, W. Dou, and T. Song, "Simultaneous dose distribution and fluence prediction for nasopharyngeal carcinoma imrt," *Radiation Oncology*, vol. 18, July 2023.

[90] S. Limin, D.-T. Pham, and S. Alam, "Multi-agent deep reinforcement learning for mix-mode runway sequencing," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, p. 586–593, IEEE, Oct. 2022.

[91] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, p. 234–241. Springer International Publishing, 2015.

[92] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770–778, IEEE, June 2016.

[93] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2261–2269, IEEE, July 2017.

[94] S. Liu, J. Zhang, T. Li, H. Yan, and J. Liu, "Technical note: A cascade 3d u-net for dose prediction in radiotherapy," *Medical Physics*, vol. 48, p. 5574–5582, Sept. 2021.

[95] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, p. 1748–1758, IEEE, Jan. 2022.

[96] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[97] R. Labaca-Castro, *Generative Adversarial Nets*, p. 73–76. Springer Fachmedien Wiesbaden, 2023.

[98] V. Kearney, J. W. Chan, T. Wang, A. Perry, M. Descovich, O. Morin, S. S. Yom, and T. D. Solberg, "Dosegan: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation," *Scientific Reports*, vol. 10, July 2020.

[99] A. Babier, R. Mahmood, A. L. McNiven, A. Diamant, and T. C. Chan, "Knowledge-based automated planning with three-dimensional generative adversarial networks," *Medical Physics*, vol. 47, p. 297–306, Nov. 2019.

[100] J. Cao, Z. Wang, H. Guo, H. Cheng, Q. Zhang, and R. Xu, "Spiking denoising diffusion probabilistic models," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, p. 4900–4909, IEEE, Jan. 2024.

[101] S. A. Faroughi, N. M. Pawar, C. Fernandes, M. Raissi, S. Das, N. K. Kalantari, and S. Kourosh Mahjour, "Physics-guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: Fluid and solid mechanics," *Journal of Computing and Information Science in Engineering*, vol. 24, Jan. 2024.

[102] M. Witte and J.-J. Sonke, "A deep learning based dynamic arc radiotherapy photon dose engine trained on monte carlo dose distributions," *Physics and Imaging in Radiation Oncology*, vol. 30, p. 100575, Apr. 2024.

[103] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, pp. 3179–3189, 2018.

[104] G. Jhanwar, N. Dahiya, P. Ghahremani, M. Zarepisheh, and S. Nadeem, "Domain knowledge driven 3d dose prediction using moment-based loss function," *Physics in Medicine & Biology*, vol. 67, p. 185017, Sept. 2022.

[105] M. Cheng, S. Jolly, W. O. Quarshie, N. Kapadia, F. D. Vigneau, and F.-M. S. Kong, "Modern radiation further improves survival in non-small cell lung cancer: An analysis of 288,670 patients," *Journal of Cancer*, vol. 10, no. 1, p. 168–177, 2019.

[106] P. Burkoň, M. Slávik, T. Kazda, P. Pospíšil, T. Procházka, M. Vrzal, and P. Šlampa, "Stereotactic body radiotherapy – current indications," *Klinicka Onkologie*, vol. 32, Feb. 2019.

[107] L. Kepka and J. Socha, "Dose and fractionation schedules in radiotherapy for non-small cell lung cancer," *Translational Lung Cancer Research*, vol. 10, p. 1969–1982, Apr. 2021.

[108] E. Csiki, M. Simon, J. Papp, M. Barabás, J. Mikáczó, K. Gál, D. Sipos, and Á. Kovács, "Stereotactic body radiotherapy in lung cancer: a contemporary review," *Pathology and Oncology Research*, vol. 30, Feb. 2024.

[109] D. Toussie, L. A. Ginocchio, B. T. Cooper, L. Azour, W. H. Moore, G. Villasana-Gomez, and J. P. Ko, "Radiation therapy for lung cancer," *Clinics in Chest Medicine*, vol. 45, p. 339–356, June 2024.

[110] J. Zeng and S. S. Lo, "Sbrt in five fractions," *International Journal of Radiation Oncology*Biology*Physics*, vol. 97, p. 652–653, Mar. 2017.

[111] R. Dagan and R. J. Amdur, "Optimal fractionation for sbrt for spinal metastasis," July 2014.

[112] D. N. Sharma and G. K. Rath, "Brachytherapy for medically inoperable lung cancer," *The Lancet Oncology*, vol. 10, p. 1141–1142, Dec. 2009.

[113] K. D. Westover, J. Seco, J. A. Adams, M. Lanuti, N. C. Choi, M. Engelsman, and H. Willers, "Proton sbrt for medically inoperable stage i nsclc," *Journal of Thoracic Oncology*, vol. 7, p. 1021–1025, June 2012.

[114] O. T. Oladeru, E. J. Vaios, C. Eyler, B. N. Noe, C. Fernandez Del-Castillo, C. D. Weekes, C. Ferrone, K. D. Lillemoe, M. Qadan, A. Parikh, J. W. Clark, D. P. Ryan, J. N. Allen, T. S. Hong, and J. Y. Wo, "Outcomes following liver sbrt for metastatic pancreatic cancer.," *Journal of Clinical Oncology*, vol. 37, p. 418–418, Feb. 2019.

[115] C. M. Montplet, S. A. Chaler, D. J. Bruggeman, M. Bueno, and M. Duch, "Ep-1343 dosimetric validation of acuros xb in water," *Radiotherapy and Oncology*, vol. 103, p. S510, May 2012.

[116] M. T. Htet, T. Sanghangthum, and T. Tawonwong, "Dosimetric evaluation in lung sbrt between aaa and axb algorithms," *Health and Technology*, vol. 15, p. 853–858, Dec. 2024.

[117] P. A. Taylor, S. F. Kry, and D. S. Followill, "Pencil beam algorithms are unsuitable for proton dose calculations in lung," *International Journal of Radiation Oncology*Biology*Physics*, vol. 99, p. 750–756, Nov. 2017.

[118] T. Knoos, A. Ahnesjo, P. Nilsson, and L. Weber, "Limitations of a pencil beam approach to photon dose calculations in lung tissue," *Physics in Medicine and Biology*, vol. 40, p. 1411–1420, Sept. 1995.

[119] S. Senan, N. J. Haasbeek, E. F. Smit, and F. J. Lagerwaard, "Stereotactic radiotherapy for centrally located early-stage lung tumors," *Journal of Clinical Oncology*, vol. 25, p. 464–464, Feb. 2007.

[120] A. Mampuya, M. Nakamura, Y. Hirose, T. Ishigaki, T. Mizowaki, and M. Hiraoka, "Ep-1550: Dosimetric comparison of the two dose reporting modes of acuros xb and aaa for lung sbrt," *Radiotherapy and Oncology*, vol. 119, p. S718, Apr. 2016.

[121] S. E. Braslavsky, A. M. Braun, A. E. Cassano, A. V. Emeline, M. I. Litter, L. Palmisano, V. N. Parmon, and N. Serpone, "Dose (energy or photon dose): From: Glossary of terms used in photocatalysis and radiation catalysis sponsoring bodies: Iupac physical and biophysical chemistry division and iupac organic and biomolecular chemistry division: see more details on p. 1011.," Mar. 2016.

[122] C. M. Montplet, S. A. Chaler, D. J. Bruggeman, M. Bueno, and M. Duch, "Ep-1343 dosimetric validation of acuros xb in water," *Radiotherapy and Oncology*, vol. 103, p. S510, May 2012.

[123] M. v. Zijtveld and J. P. A. Marijnissen, "Analytical anisotropic algorithm (AAA) for GPU-based radiotherapy dose calculation," in *Graphics Processing Unit-Based High Performance Computing in Radiation Therapy*, pp. 243–257, CRC Press, 2015.

[124] Y. Tsuruta, M. Nakata, M. Nakamura, Y. Matsuo, K. Higashimura, H. Monzen, T. Mizowaki, and M. Hiraoka, "Dosimetric comparison of Acuros XB, AAA, and XVMC in stereotactic body radiotherapy for lung cancer," *Medical Physics*, vol. 41, no. 8Part1, p. 081715, 2014.

[125] K. Bush, L. Wang, and E. Mok, "Su-e-t-485: Validation of acuros xb dose calculations in sbrt lung planning with monte carlo methods," *Medical Physics*, vol. 39, p. 3816–3817, June 2012.

[126] K. Shao, C. Chai, Y. Yang, W. Chen, and F. Du, "Dosimetric comparison of Monte Carlo, Acuros XB, and anisotropic analytical algorithm for lung cancer plans on Halcyon accelerators," *PLOS One*, vol. 20, no. 10, p. e0334329, 2025.

[127] W. Luo, F. Yin, C. Ma, J. Fan, and H. Song, "Su-ff-t-172: Dosimetric comparison of mapcheck with monte carlo simulations," *Medical Physics*, vol. 34, p. 2440–2440, June 2007.

[128] A. Allisy, W. A. Jennings, A. M. Kellerer, J. W. Müller, H. H. Rossi, and S. M. Seltzer, "Fundamental quantities and units for ionizing radiation," *Reports of the International Commission on Radiation Units and Measurements*, vol. os-31, p. 1–1, Dec. 1998.

[129] M. D. Leising, G. H. Share, E. L. Chupp, and G. Kanbach, "Gamma-ray limits on na-22 production in novae," *The Astrophysical Journal*, vol. 328, p. 755, May 1988.

[130] J. F. Ward, *Ionizing Radiation Induced DNA Damage: Identities and DNA Repair*, p. 135–138. Springer US, 1986.

[131] S. Brunner, D. Varga, R. Bozó, R. Polanek, T. Tőkés, E. R. Szabó, R. Molnár, N. Gémes, G. J. Szebeni, L. G. Puskás, M. Erdélyi, and K. Hideghéty, "Analysis of ionizing radiation induced dna damage by superresolution dstorm microscopy," *Pathology and Oncology Research*, vol. 27, Nov. 2021.

[132] E. Bruselas *et al.*, "Hypoxia in radiotherapy: From radiobiological mechanisms to therapeutic implications," *Frontiers in Oncology*, vol. 11, p. 607, 2021. Topic on hypoxia and radioresistance.

[133] M. Srivastava and S. C. Raghavan, "Dna double-strand break repair inhibitors as cancer therapeutics," *Chemistry & Biology*, vol. 22, p. 17–29, Jan. 2015.

[134] B. Mahaney, K. Meek, and S. Lees-Miller, "Repair of ionizing radiation-induced dna double-strand breaks by non-homologous end-joining," *Biochemical Journal*, vol. 417, p. 639–650, Jan. 2009.

[135] H. H. Eisenlohr, "Introduction to radiological physics and radiation dosimetry," *International Journal of Radiation Applications and Instrumentation. Part A. Applied Radiation and Isotopes*, vol. 38, p. 163, Jan. 1987.

[136] C. Bardiot, "Mesurer merce cunningham," *Humanités numériques*, Jan. 2020.

[137] E. Fermi, "Quantum theory of radiation," *Reviews of Modern Physics*, vol. 4, p. 87–132, Jan. 1932.

[138] R. E. Huffman, *Photon Cross Sections*, p. 103–120. Elsevier, 1992.

[139] R. Bull, "Stopping powers for electrons and positrons," *International Journal of Radiation Applications and Instrumentation. Part D. Nuclear Tracks and Radiation Measurements*, vol. 11, p. 273, Jan. 1986.

[140] M. Williams, "Computational methods of neutron transport," *Annals of Nuclear Energy*, vol. 12, p. 101, Jan. 1985.

[141] K. A. Gifford, J. L. Horton, T. A. Wareing, G. Failla, and F. Mourtada, "Comparison of a finite-element multigroup discrete-ordinates code with monte carlo for radiotherapy calculations," *Physics in Medicine and Biology*, vol. 51, p. 2253–2265, Apr. 2006.

[142] Varian Medical Systems, *Eclipse Photon and Electron Algorithms Reference Guide*. Varian Medical Systems, version 13.6 ed., 2018. Available: https://jpneylon.github.io/ABR/PDFs/Add_052418/EclipseAlgorithms13.6_RefGuide.pdf.

[143] F. Salvat, J. M. Fernández-Varea, J. Sempau, *et al.*, "PENELOPE-2006: A code system for Monte Carlo simulation of electron and photon transport," in *Workshop proceedings*, vol. 4, p. 7, Citeseer, 2006.

[144] J. H. Hubbell, W. J. Veigele, E. Briggs, R. Brown, D. Cromer, and d. R. Howerton, "Atomic form factors, incoherent scattering functions, and photon scattering cross sections," *Journal of physical and chemical reference data*, vol. 4, no. 3, pp. 471–538, 1975.

[145] D. T. Cromer, "Anomalous dispersion corrections computed from self-consistent field relativistic Dirac–Slater wave functions," *Acta Crystallographica*, vol. 18, no. 1, pp. 17–23, 1965.

[146] O. Iwamoto, N. Iwamoto, S. Kunieda, F. Minato, S. Nakayama, Y. Abe, K. Tsubakihara, S. Okumura, C. Ishizuka, T. Yoshida, *et al.*, "Japanese evaluated nuclear data library version 5: JENDL-5," *journal of nuclear science and technology*, vol. 60, no. 1, pp. 1–60, 2023.

[147] D. E. Cullen, M. Chen, J. Hubbell, S. Perkins, E. Plechaty, J. Rathkopf, and J. Scofield, "Tables and graphs of photon-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated photon data library (EPDL)," tech. rep., Lawrence Livermore National Lab., CA (USA), 1989.

[148] W. Heitler, *The quantum theory of radiation*. Courier Corporation, 1984.

[149] D. Brusa, G. Stutz, J. Riveros, J. Fernández-Varea, and F. Salvat, "Fast sampling algorithm for the simulation of photon Compton scattering," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 379, no. 1, pp. 167–175, 1996.

[150] D. E. Cullen, J. H. Hubbell, and L. Kissel, "EPDL97: the evaluated photo data library97 version," tech. rep., Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 1997.

[151] F. Sauter, "Über den atomaren photoeffekt in der k-schale nach der relativistischen wellenmechanik diracs," *Annalen der Physik*, vol. 403, no. 4, pp. 454–488, 1931.

[152] J. Baró, M. Roteta, J. Fernández-Varea, and F. Salvat, "Analytical cross sections for Monte Carlo simulation of photon transport," *Radiation physics and chemistry*, vol. 44, no. 5, pp. 531–552, 1994.

[153] Y.-S. Tsai, "Pair production and bremsstrahlung of charged leptons," *Reviews of Modern Physics*, vol. 46, no. 4, p. 815, 1974.

[154] H. Davies, H. Bethe, and L. Maximon, "Theory of bremsstrahlung and pair production. II. Integral cross section for pair production," *Physical Review*, vol. 93, no. 4, p. 788, 1954.

[155] I. Kawrakow and D. Rogers, "The EGSnrc code system," *NRC Report PIRS-701, NRC, Ottawa*, 2021.

[156] L. Lorence Jr, J. Morel, and G. Valdez, "Physics guide to CEPXS: a multigroup coupled electron-photon cross-section generating code," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 1989.

[157] J. M. Fernández-Varea, F. Salvat, M. Dingfelder, and D. Liljequist, "A relativistic optical-data model for inelastic scattering of electrons and positrons in condensed matter," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 229, no. 2, pp. 187–218, 2005.

[158] C. Møller, "Zur theorie des durchgangs schneller elektronen durch materie," *Annalen der Physik*, vol. 406, no. 5, pp. 531–585, 1932.

[159] H. Bhabha, "The scattering of positrons by electrons with exchange on Dirac's theory of the positron," *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, vol. 154, no. 881, pp. 195–206, 1936.

[160] S. M. Seltzer, "Cross sections for bremsstrahlung production and electron-impact ionization," in *Monte Carlo transport of electrons and photons*, pp. 81–114, Springer, 1988.

[161] S. Perkins and D. Cullen, "The Livermore electron impact ionization data base," tech. rep., Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 1989.

[162] F. Salvat and J. Fernández-Varea, "Semiempirical cross sections for the simulation of the energy loss of electrons and positrons in matter," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 63, no. 3, pp. 255–269, 1992.

[163] F. Salvat, "Bethe stopping-power formula and its corrections," *Physical Review A*, vol. 106, no. 3, p. 032809, 2022.

[164] F. Salvat and P. Andreo, "SBETHE: Stopping powers of materials for swift charged particles from the corrected Bethe formula," *Computer Physics Communications*, vol. 287, p. 108697, 2023.

[165] F. Rohrlich and B. Carlson, "Positron-electron differences in energy loss and multiple scattering," *Physical review*, vol. 93, no. 1, p. 38, 1954.

[166] U. Fano, "Atomic theory of electromagnetic interactions in dense materials," *Physical Review*, vol. 103, no. 5, p. 1202, 1956.

[167] M. Inokuti and D. Y. Smith, "Fermi density effect on the stopping power of metallic aluminum," *Physical Review B*, vol. 25, no. 1, p. 61, 1982.

[168] R. M. Sternheimer, "The density effect for the ionization loss in various materials," *Physical Review*, vol. 88, no. 4, p. 851, 1952.

[169] T. Lijian, H. Qing, and L. Zhengming, "Analytic fitting to the Mott cross section of electrons," *Radiation Physics and Chemistry*, vol. 45, no. 2, pp. 235–245, 1995.

[170] M. Boschini, C. Consolandi, M. Gervasi, S. Giani, D. Grandi, V. Ivanchenko, P. Nieminem, S. Pensotti, P. Rancoita, and M. Tacconi, "An expression for the Mott cross section of electrons and positrons on nuclei with Z up to 118," *Radiation Physics and Chemistry*, vol. 90, pp. 39–66, 2013.

[171] I. Kawrakow, "Improved modeling of multiple scattering in the voxel Monte Carlo model," *Medical physics*, vol. 24, no. 4, pp. 505–517, 1997.

[172] S. M. Seltzer, "An overview of ETRAN Monte Carlo methods," *Monte Carlo transport of electrons and photons*, pp. 153–181, 1988.

[173] G. Moliere, "Theorie der streuung schneller geladener teilchen i. einzelstreuung am abgeschirmten coulomb-feld," *Zeitschrift für Naturforschung A*, vol. 2, no. 3, pp. 133–145, 1947.

[174] W. Koepf, "Hypergeometric summation," *Vieweg, Braunschweig/Wiesbaden*, vol. 5, no. 6, 1998.

[175] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.

[176] M. Landesman and J. Morel, "Angular Fokker-Planck decomposition and representation techniques," *Nuclear Science and Engineering*, vol. 103, no. 1, pp. 1–11, 1989.

[177] J. Morel, "Fokker-Planck calculations using standard discrete ordinates transport codes," *Nuclear Science and Engineering*, vol. 79, no. 4, pp. 340–356, 1981.

[178] J. Morel, "On the validity of the extended transport cross-section correction for low-energy electron transport," *Nuclear Science and Engineering*, vol. 71, no. 1, pp. 64–71, 1979.

[179] C. R. Drumm, W. C. Fan, L. Lorence, and J. Liscum-Powell, "An analysis of the extended-transport correction with application to electron beam transport," *Nuclear science and engineering*, vol. 155, no. 3, pp. 355–366, 2007.

[180] A. Hébert, *Applied reactor physics*. Presses inter Polytechnique, 2009.

[181] Y. Azmy, E. Sartori, E. W. Larsen, and J. E. Morel, "Advances in discrete-ordinates methodology," *Nuclear computational science: A century in review*, pp. 1–84, 2010.

[182] J. E. Morel, "A hybrid collocation-Galerkin-Sn method for solving the Boltzmann transport equation," *Nuclear Science and Engineering*, vol. 101, no. 1, pp. 72–87, 1989.

[183] S. M. Seltzer and M. J. Berger, "Bremsstrahlung energy spectra from electrons with kinetic energy 1 keV–10 GeV incident on screened nuclei and orbital electrons of neutral atoms with Z= 1–100," *Atomic data and nuclear data tables*, vol. 35, no. 3, pp. 345–418, 1986.

[184] E. Acosta, X. Llovet, and F. Salvat, "Monte Carlo simulation of bremsstrahlung emission by electrons," *Applied Physics Letters*, vol. 80, no. 17, pp. 3228–3230, 2002.

[185] L. Kissel, C. Quarles, and R. Pratt, "Shape functions for atomic-field bremsstrahlung from electrons of kinetic energy 1–500 keV on selected neutral atoms $1 \leq Z \leq 92$," *Atomic data and nuclear data tables*, vol. 28, no. 3, pp. 381–460, 1983.

[186] L. Kim, R. Pratt, S. Seltzer, and M. Berger, "Ratio of positron to electron bremsstrahlung energy loss: an approximate scaling law," *Physical Review A*, vol. 33, no. 5, p. 3002, 1986.

[187] F. Salvat, J. Fernández-Varea, J. Sempau, and X. Llovet, "Monte Carlo simulation of bremsstrahlung emission by electrons," *Radiation Physics and Chemistry*, vol. 75, no. 10, pp. 1201–1219, 2006.

[188] A. Poškus, "Shape functions and singly differential cross sections of bremsstrahlung at electron energies from 10 eV to 3 MeV for Z= 1–100," *Atomic Data and Nuclear Data Tables*, vol. 129, p. 101277, 2019.

[189] W. R. Nelson, H. Hirayama, and D. W. Rogers, "EGS4 code system," tech. rep., Stanford Linear Accelerator Center, Menlo Park, CA (USA), 1985.

[190] F. Salvat and J. M. Fernández-Varea, "Overview of physical interaction models for photon and electron transport used in Monte Carlo codes," *Metrologia*, vol. 46, no. 2, p. S112, 2009.

[191] Collaboration, GEANT and others, "Physics reference manual," tech. rep., CERN, Switzerland, 2016.

[192] A. Naceur, C. Bienvenue, P. Romano, C. Chilian, and J.-F. Carrier, "Extending deterministic transport capabilities for very-high and ultra-high energy electron beams," *Scientific Reports*, vol. 14, no. 1, p. 2796, 2024.

[193] S. Perkins, *Tables and Graphs of Atomic Subshell and Relaxation Data Derived from the LLNL Evaluated Atomic Data Library (EADL), Z = 1-100*. Lawrence Livermore National Laboratory, 1991.

[194] A. Hébert and A. Naceur, "Implementation of the ELECTR module in NJOY," in *EPJ Web of Conferences*, vol. 284, p. 11001, EDP Sciences, 2023.

[195] A. Elbert and A. Laforgia, "An inequality for Legendre polynomials," *Journal of Mathematical Physics*, vol. 35, no. 3, pp. 1348–1360, 1994.

[196] C. Bienvenue and A. Hébert, "High-order diamond differencing schemes for the Boltzmann Fokker–Planck equation in 1D and 2D Cartesian geometries," *Annals of Nuclear Energy*, vol. 171, p. 109032, 2022.

[197] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical methods for physicists: a comprehensive guide*. Academic press, 2011.

[198] F. G. Tricomi, "Sugli zeri dei polinomi sferici ed ultrasferici," *Annali di Matematica Pura ed Applicata*, vol. 31, no. 1, pp. 93–97, 1950.

[199] N. Hale and A. Townsend, "Fast and accurate computation of Gauss–Legendre and Gauss–Jacobi quadrature nodes and weights," *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. A652–A674, 2013.

[200] B. G. Carlson, "A method of characteristics and other improvements in solution methods for the transport equation," *Nuclear science and engineering*, vol. 61, no. 3, pp. 408–425, 1976.

[201] L. J. Lorence, Jr., J. E. Morel, and G. D. Valdez, "Physics guide to CEPXS: A multigroup coupled electron-photon cross-section generating code," Tech. Rep. SAND89-1685, Sandia National Laboratories, 1989.

[202] T. A. Wareing, J. M. McGhee, J. E. Morel, and S. D. Pautz, "Discontinuous finite elementsnmethods on three-dimensional unstructured grids," *Nuclear Science and Engineering*, vol. 138, p. 256–268, July 2001.

[203] L. B. Marks, S. M. Bentzen, J. O. Deasy, F.-M. S. Kong, J. D. Bradley, I. S. Vogelius, I. El Naqa, J. L. Hubbs, J. V. Lebesque, R. D. Timmerman, M. K. Martel, and A. Jackson, "Radiation dose–volume effects in the lung," *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, p. S70–S76, Mar. 2010.

[204] K. Banfill, M. Giuliani, M. Aznar, K. Franks, A. McWilliam, M. Schmitt, F. Sun, M. C. Vozenin, and C. Faivre Finn, "Cardiac toxicity of thoracic radiotherapy: Existing evidence and future directions," *Journal of Thoracic Oncology*, vol. 16, p. 216–227, Feb. 2021.

[205] K. Banfill, A. Abravan, M. van Herk, F. Sun, K. Franks, A. McWilliam, and C. Faivre-Finn, "Heart dose and cardiac comorbidities influence death with a cardiac cause following hypofractionated radiotherapy for lung cancer," *Frontiers in Oncology*, vol. 12, Oct. 2022.

[206] Y. M. Chung, S.-H. Park, W.-B. Tsai, S.-Y. Wang, M.-A. Ikeda, J. S. Berek, D. J. Chen, and M. C.-T. Hu, "Foxo3 signalling links atm to the p53 apoptotic pathway following dna damage," *Nature Communications*, vol. 3, Aug. 2012.

[207] T. K. Pandita, H. B. Lieberman, D.-S. Lim, S. Dhar, W. Zheng, Y. Taya, and M. B. Kastan, "Ionizing radiation activates the atm kinase throughout the cell cycle," *Oncogene*, vol. 19, p. 1386–1391, Mar. 2000.

[208] K. A. Boehme, R. Kulikov, and C. Blattner, "p53 stabilization in response to dna damage requires akt/pkb and dna-pk," *Proceedings of the National Academy of Sciences*, vol. 105, p. 7785–7790, June 2008.

[209] D. Nguyen, X. Jia, D. Sher, M.-H. Lin, Z. Iqbal, H. Liu, and S. Jiang, "3d radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture," *Physics in Medicine & Biology*, vol. 64, no. 6, p. 065020, 2019.

[210] A. F. I. Osman and N. M. Tamam, "Attention-aware 3d u-net convolutional neural network for knowledge-based planning 3d dose distribution prediction of head-and-neck cancer," *Journal of Applied Clinical Medical Physics*, vol. 23, no. 7, 2022.

[211] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," 2018.

[212] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," 2020.

[213] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, IEEE, June 2018.

[214] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," 2020.

[215] O. Pastor-Serrano, P. Dong, C. Huang, L. Xing, and Z. Perkó, "Sub-second photon dose prediction via transformer neural networks," *Medical Physics*, vol. 50, pp. 3159–3171, feb 2023.

[216] S. Yan, A. Maniscalco, B. Wang, H. Zhao, D. Nguyen, S. Jiang, and C. Shen, "Geometry-encoded deep learning (geodl) framework for real-time 3d dose verification for online adaptive radiotherapy," *Machine Learning: Health*, vol. 1, p. 015004, jul 2025.

[217] M. F. Kraus, R. Gao, S. Arberet, F. C. Ghesu, A. Kamen, and D. Comaniciu, "Single shot full plan deep learning dose computation for radiation therapy using spherical harmonics," *Medical Physics*, vol. 53, jan 2026.

[218] X. Hou, W. Cheng, J. Shen, H. Guan, Y. Zhang, L. Bai, S. Wang, and Z. Liu, "A deep learning model to predict dose distributions for breast cancer radiotherapy," *Discover Oncology*, vol. 16, feb 2025.

[219] M. Saito, N. Kadoya, Y. Kimura, H. Nemoto, R. Tozuka, K. Jingu, and H. Onishi, "Evaluation of deep learning based dose prediction in head and neck cancer patients using two different types of input contours," *Journal of Applied Clinical Medical Physics*, vol. 25, sep 2024.

[220] I. H. Creemers, J. M. Kusters, P. G. van Kollenburg, L. C. Bouwmans, D. A. Schinagl, and J. Bussink, "Comparison of dose metrics between automated and manual radiotherapy planning for advanced stage non-small cell lung cancer with volumetric modulated arc therapy," *Physics and Imaging in Radiation Oncology*, vol. 9, pp. 92–96, jan 2019.

[221] G. Della Gala, M. L. P. Dirkx, N. Hoekstra, D. Fransen, N. Lanconelli, M. van de Pol, B. J. M. Heijmen, and S. F. Petit, "Fully automated vmat treatment planning for advanced-stage nsclc patients," *Strahlentherapie und Onkologie*, vol. 193, pp. 402–409, mar 2017.

[222] E. M. Quan, J. Y. Chang, Z. Liao, T. Xia, Z. Yuan, H. Liu, X. Li, C. A. Wages, R. Mohan, and X. Zhang, "Automated volumetric modulated arc therapy treatment planning for stage iii lung cancer: How does it compare with intensity-modulated radio therapy?," *International Journal of Radiation Oncology*Biology*Physics*, vol. 84, pp. e69–e76, sep 2012.

[223] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Springer International Publishing, 2015.

[224] A. Babier, B. Zhang, R. Mahmood, K. L. Moore, T. G. Purdie, A. L. McNiven, and T. C. Y. Chan, "OpenKBP: The open-access knowledge-based planning grand challenge and dataset," *Medical Physics*, vol. 48, pp. 5549–5561, jun 2021.

[225] H. Nemoto, M. Saito, N. Kadoya, T. Komiyama, R. Tozuka, H. Onishi, and K. Jingu, "Evaluation of deep learning-based automated radiotherapy planning for early-stage lung cancer using sbrt-vmat: A comparison with manual planning," *Journal of Applied Clinical Medical Physics*, vol. 26, oct 2025.

[226] N. Wang, J. Fan, Y. Xu, L. Yan, D. Chen, W. Wang, K. Men, J. Dai, and Z. Liu, "Clinical implementation and evaluation of deep learning-assisted automatic radiotherapy treatment planning for lung cancer," *Physica Medica*, vol. 124, p. 104492, aug 2024.

[227] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[228] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo, and Y. Liu, "RoFormer: Enhanced transformer with rotary position embedding," *Neurocomputing*, vol. 568, p. 127063, feb 2024.

[229] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016.

[230] R. Gao, M. Diallo, H. Liu, A. Magliari, J. Sackett, W. Verbakel, S. Meyers, R. Mcbeth, M. Zarepisheh, S. Arberet, M. Kraus, F. C. Ghesu, and A. Kamen, "Automating rt planning at scale: High quality data for ai training," 2025.

[231] K. Maass, A. Aravkin, and M. Kim, "A hyperparameter-tuning approach to automated inverse planning," *Medical Physics*, vol. 49, pp. 3405–3415, mar 2022.

[232] H. Wang, X. Bai, Y. Wang, Y. Lu, and B. Wang, "An integrated solution of deep reinforcement learning for automatic IMRT treatment planning in non-small-cell lung cancer," *Frontiers in Oncology*, vol. 13, p. 1124458, feb 2023.

[233] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, apr 2018.

[234] C. Bodnar, F. Di Giovanni, B. P. Chamberlain, P. Liò, and M. M. Bronstein, "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs," 2022.

[235] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[236] I. Paddick, "A simple scoring ratio to index the conformity of radiosurgical treatment plans: Technical note," *Journal of Neurosurgery*, vol. 93, pp. 219–222, dec 2000.

[237] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics*, vol. 25, pp. 656–661, may 1998.

[238] N. Kadoya, Y. Kimura, R. Tozuka, S. Tanaka, K. Arai, Y. Katsuta, H. Shimizu, Y. Sugai, T. Yamamoto, R. Umezawa, and K. Jingu, "Evaluation of deep learning-based deliverable vmat plan generated by prototype software for automated planning for prostate cancer patients," *Journal of Radiation Research*, vol. 64, pp. 842–849, aug 2023.

[239] W. T. Hrinivich, M. Bhattacharya, L. Mekki, T. McNutt, X. Jia, H. Li, D. Y. Song, and J. Lee, "Clinical VMAT machine parameter optimization for localized prostate cancer using deep reinforcement learning," *Medical Physics*, vol. 51, pp. 3972–3984, apr 2024.

[240] N. Shaffer, A. R. Mudireddy, and J. St-Aubin, "A tandem reinforcement learning framework for localized prostate cancer treatment planning and machine parameter optimization," *Medical Physics*, vol. 53, jan 2026.

[241] S. Arberet, R. Gao, M. Kraus, F. C. Ghesu, W. Verbakel, M. Diallo, A. Magliari, V. Karuppusamy, S. Beriwal, REQUITE Consortium, A. Kamen, and D. Comaniciu, "AI end-to-end radiation treatment planning under one second," 2026.

[242] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics*, vol. 25, p. 656–661, May 1998.

[243] M. Wendling, L. J. Zijp, L. N. McDermott, E. J. Smit, J. Sonke, B. J. Mijnheer, and M. van Herk, "A fast algorithm for gamma evaluation in 3d," *Medical Physics*, vol. 34, p. 1647–1654, Apr. 2007.