A NEW RESIDUAL MONTE CARLO METHOD FOR ESTIMATING ERRORS IN DETERMINISTIC TRANSPORT SOLUTIONS

A Dissertation

by

JAN IZAK CORNELIUS VERMAAK

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jim Morel
Co-Chair of Committee,	Jean Ragusa
Committee Members,	Marvin Adams
	Bani Malek
Head of Department,	Michael Nastasi

May 2022

Major Subject: Nuclear Engineering

ABSTRACT

The residual Monte Carlo (RMC) method is also known in the literature as sequential Monte Carlo and reduced-source Monte Carlo. Given a Monte Carlo method for solving a linear equation and an approximate solution to that system, the residual method enables use of essentially the same Monte Carlo algorithm to directly compute the additive error or "defect" associated with the approximate solution. As the size of the defect decreases relative to the size of the solution, the residual Monte Carlo method becomes increasingly efficient relative to the standard Monte Carlo (SMC) method. Here we present a new RMC algorithm for evaluating the space-angle error in Discrete Ordinates radiation transport solutions, and provide computational examples demonstrating that it can be far more efficient than SMC for this purpose.

DEDICATION

I dedicate this dissertation to my wife Sandra and our two beautiful kids, Zak and Sharne, for whom I would sacrifice anything. I also dedicate this to my father who has been an inspiration to me all my life.

ACKNOWLEDGMENTS

I would firstly like to thank Jim Morel for his patience and care. In a time when my life was very turbulent, sad and scary he was there to provide soothing words and assurances. He took me into a world of computational physics that I have come to love instead of fear. I will never forget how everyone discouraged me from writing a really big code, whilst he simply said “Do it! Why not?". Secondly, I would like to thank Marvin Adams and Jean Ragusa. They pulled me into this environment with plenty of enthusiasm. Their drive and knowledge is always inspiring me to be better, and they are always willing to share their experience.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Jim Morel [advisor], Professor Jean Ragusa [co-advisor], and Professor Marvin Adams of the Department of Nuclear Engineering, as well as Professor Bani Mallick of the Department of Statistics.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

This graduate study was self-financed.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENTS iv

CONTRIBUTORS AND FUNDING SOURCES v

TABLE OF CONTENTS vi

LIST OF FIGURES x

LIST OF TABLES xiv

1. INTRODUCTION 1

    1.1 The Linear Boltzmann Equation (LBE) 1
        1.1.1 The scalar flux 2
        1.1.2 The multigroup approximation 3
        1.1.3 The residual, r, the defect, ψ_D, and the residual transport equation 5
    1.2 Overview of popular methods to solve the Linear Boltzmann transport equation 6
        1.2.1 The Monte Carlo method 7
        1.2.2 The Discrete Ordinates method 8
            1.2.2.1 The Method of Characteristics (MOC) variant 9
            1.2.2.2 The Diffusion Approximation 9
    1.3 The focal point of this research 10
    1.4 Literature review 11
    1.5 Overview of chapters 13

2. INTRODUCTION TO THE MONTE CARLO METHOD 15

    2.1 Source sampling 15
        2.1.1 Sampling a particular source 16
        2.1.2 Sampling the group and/or energy 17
        2.1.3 Sampling the position 19
        2.1.4 Sampling the direction 20
        2.1.5 Conclusion of source sampling 21
    2.2 The stochastic transport process 22
        2.2.1 Step 1: Transporting a particle to a material interaction 22
            2.2.1.1 Step 1a - Distance to interaction within a cell 23

2.2.1.2 Step 1b - Distance to a cell surface (surface interaction) 24
2.2.1.3 Conclusion of step 1 25
2.2.2 Step 2: Randomly sampling a reaction type 25
2.2.3 Step 3: Applying scattering 26
2.3 Contributing to volumetric tallies 29
2.4 Determining the statistical uncertainty of a tally 31
2.5 Modifications to allow for multigroup cross sections 33
2.5.1 The energy change 33
2.5.2 Sampling the scattering cosine 34
2.5.2.1 Rejection sampling of the Legendre expansion 34
2.5.2.2 A generalized Gauss-quadrature from Legendre expansions of σ_s . 36
2.6 Modifications to tallies to provide finite element representations 37
2.7 Modifications to tallies to provide spherical harmonic expansions of the angular flux 40
2.8 The overall process 41
2.9 Batches of particles 42

3. INTRODUCTION TO THE DISCRETE ORDINATES METHOD 43

3.1 Angular discretization 43
3.1.1 Modification of the scattering source 43
3.1.2 Application of an angular quadrature 45
3.2 Spatial discretization 46
3.2.1 A DFEM representation of a continuum-variable 47
3.2.2 The weak form of the discrete ordinates equations 47
3.2.3 A cell-by-cell system to solve 49
3.3 The cell-by-cell system in a global-system perspective 52
3.4 The system solution technique known as a "sweep" 53
3.5 Combining all angles and groups 54
3.6 Operator form of the scattering and external source 56
3.6.1 The discrete-to-moment operator 56
3.6.2 The moment-to-discrete operator 57
3.6.3 The scattering operator 58
3.6.4 The operator-form of the scattering- and external-source 58
3.7 The complete operator-form of the discrete ordinates equations 58
3.8 Algorithms to solve the discretized DO-system 59
3.8.1 Richardson iteration 60
3.8.2 Sweep-based GMRES 60
3.9 Convergence acceleration techniques 62
3.10 The fundamental drawbacks of the discrete ordinates method 63

4. THE RESIDUAL MONTE CARLO SOURCE SAMPLING 67

4.1 Using the absolute value of the residual source as a source PDF 69
4.2 Computing phase-space integrals of the residual source for boundary-faces and cells 69
4.2.1 The angular nature of the approximate solution $\tilde{\psi}$ 70
4.2.2 The finite element nature of the approximate solution $\tilde{\psi}$ 71

4.2.2.1 Continuous representations of $\tilde{\phi}_g$ 71
4.2.2.2 Discontinuous representations of $\tilde{\phi}_g$ 72
4.2.3 Integrals of the absolute value of multivariate functions with Monte Carlo integration 73
4.2.3.1 Analyzing multivariate Monte Carlo integration of the absolute value of the residual 75
4.2.3.2 Batches of particles as a means to ascertain the effect of source uncertainty 76
4.3 Building the residual source CDF 77
4.4 Rejection sampling the residual within a cell or on a cell-face pair 79
4.4.1 Rejection sampling within a cell 80
4.4.2 Rejection sampling on a cell-face pair 81
4.5 Summary of our residual Monte Carlo source sampling method 82
4.5.1 Construction of the nested CDF 83
4.5.2 Sampling the residual source 84

5. SIMULATIONS 86

5.1 Definitions 86
5.1.1 Quantity of Interest (QOI) 86
5.1.2 Simulation load 86
5.1.3 Efficiency $\eta^{Y\to X}$ 87
5.1.4 Source particle Sampling Time ST_X 87
5.1.5 Source particle Transport Time TrT_X 87
5.1.6 Source particle Total simulation Time TT_X 87
5.1.7 Overall effectiveness $E^{Y\to X}$ 87
5.1.8 Key Performance Indicators (KPIs) 88
5.2 Problem 1 - One Dimensional, homogeneous 89
5.3 Problem 2 - One Dimensional, multi-material 96
5.4 Problem 3 - Multidimensional homogeneous pure absorber material with a large distributed source. 99
5.5 Problem 4 - Multidimensional multi-material domain with a duct and a localized source 104
5.6 Problem 5 - A distributed-source variation of problem 4. 110
5.7 Adjoint-based source-biasing for problem 5 119
5.7.1 The adjoint-flux 119
5.7.2 Abstraction of unbiased source sampling 120
5.7.3 Applying the scalar adjoint flux as source biasing on a cell-by-cell basis 122
5.7.4 Results using only the scalar adjoint flux, $\phi^\dagger$, for biasing 123
5.7.5 Angular biasing using the P_1-expansion of $\psi^\dagger$ 127
5.7.6 Results using both the scalar adjoint flux, $\phi^\dagger$, and the angular adjoint flux, $\psi^\dagger$, for biasing 128
5.7.7 Overall insights from adjoint-based biasing 133

6. CONCLUSIONS AND FUTURE WORK 135

6.1 Conclusions 135
6.2 Future work 137
6.2.1 Improvements in source sampling 137
6.2.2 Using mesh-based residual source with parametric surface based geometry and QOIs 137
6.2.3 Different choices for the approximate angular flux other than the scalar flux . 138
6.2.4 The potential of computing energy discretization errors 139

REFERENCES 140

APPENDIX A. RAY TRACING TO CELL SURFACES 145

A.1 Intersection of a line and a plane 145
A.2 Intersection of a line and a strip 147
A.3 Intersection of a plane and a triangle 149

APPENDIX B. BASIS FUNCTION FOR PIECEWISE-LINEAR FEM 150

B.1 2D Polygon cells 150
B.2 3D Polyhedron cells 152

APPENDIX C. AN EXPONENTIAL REPRESENTATION OF A P1 ANGULAR FLUX EXPANSION 156

C.1 An exponential representation 156
C.1.1 Change in reference frame 157
C.1.2 Only working with the shape of the angular flux 157
C.1.3 The corner cases at $r \to 0$ and $r \to 1$ 158
C.1.4 A non-linear system 159
C.1.5 Example distributions 160
C.2 Sampling the exponential distribution 160

LIST OF FIGURES

FIGURE Page

2.1 The source sampling process. 18

2.2 Different input-representations of the energy-dependency of sources, assuming contiguous group sub-domains. [Top] A histogram representation. S'_g is the source strength of group g normalized with $\int_{\mathcal{D}_e} S(E)dE$. [Bottom] An interpolated PDF with points, $S(E_g)$, from the true PDF (the blue curve) supplied at group boundaries. S^*_g is the source strength over $\int_{\mathcal{D}_{e,g}} S(E)dE$ normalized with $\int_{\mathcal{D}_e} S(E)dE$..... 19

2.3 Schematic of the two ways in which a particle track can be traced within a cell, i.e., either from the initial position, $\mathbf{x}_i$, to an interaction, or from $\mathbf{x}_i$ to the cell surface. ... 23

2.4 Schematic of the two rotations applied to $\mathbf{\Omega}_i$ when scattering through a cosine $\cos\theta_L$, applying first the polar rotation about $\mathbf{t}$ to produce $\mathbf{\Omega}_p$, then the random azimuthal rotation of $\mathbf{\Omega}_p$ by an angle φ_L about the original $\mathbf{\Omega}_i$. 28

2.5 Legendre expansion of a test function for different values of L..... 35

2.6 A problematic CDF constructed from a Legendre expansion. 35

2.7 Example application of the generalized Gauss Quadrature..... 37

2.8 Example of a 1D slab simulation with an isotropic cell-source over the first half of the domain. The cell-averaged tallies and DFEM tallies are shown. 39

3.1 Example cell dependency for 2D polygons where cell 1 is dependent on cell 0 at face f..... 51

3.2 Example topological sorting of a 3×3 mesh. 53

3.3 Ray effects for a pure absorber problem with varying levels of angular discretization. [Top left] 4 azimuthal angles. [Top right] 12 azimuthal angles. [Bottom left] 36 azimuthal angles. [Bottom right] 108 azimuthal angles. 64

3.4 Sample DO simulation showing visual ray effects for different amounts of azimuthal angles in the angular quadratures..... 65

4.1 Schematic representation of the processes involved in obtaining a residual source sample..... 83

5.1 One dimensional homogeneous material problem geometry. 89

5.2 Scalar flux solutions and computed defects for problem 1. Here, $\tilde{\phi}$ was taken as zero. 90

5.3 Scalar flux solutions and computed defects for problem 1. The top 2 plots are for $\tilde{\phi}$ being a cell-constant representation, $Q0$, computed from a linear discontinuous DO simulation with an S_2 quadrature. The bottom 2 plots are for $\tilde{\phi}$ taken directly as the linear discontinuous DO solution. 92

5.4 Scalar flux solutions and computed defects for problem 1. Here, $\tilde{\phi}$ is a linear continuous representation, $Q1$, computed from a linear discontinuous DO simulation with an S_2 quadrature. 93

5.5 RMC defect, using a discontinuous $Q0$ representation for $\tilde{\phi}$ [Left], and a discontinuous $Q1$ representation for $\tilde{\phi}$ [Right], with contributions split into positive and negative components. Both solutions were obtained with a DO simulation using an S_2 angular quadrature. 95

5.6 One dimensional test problem setup with multiple-materials. The dimensions are in mean-free-paths. 97

5.7 Scalar flux solutions and computed defects for problem 2. Here, $\tilde{\phi}$ is a linear continuous representation, $Q1$, computed from a linear discontinuous DO simulation with an S_2 quadrature. 98

5.8 Multidimensional pure absorber material test problem with a distributed isotropic source in a large portion of the domain. 100

5.9 DO solution of problem 3 using an S_2 angular quadrature. 101

5.10 Standard Monte Carlo solution of problem 3 using 10×10^6 source particles. 102

5.11 The RMC defect solution, for a DO simulation using an S_2 angular quadrature, of problem 3 using 10×10^6 source particles. 102

5.12 Cell-wise residual source for problem 3. 103

5.13 Multidimensional multi-material test problem with a localized isotropic source. 104

5.14 DO solution of problem 4 using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant. 105

5.15 Standard Monte Carlo solution of problem 4 using 10×10^6 source particles. 106

5.16 The RMC defect solution, for a DO simulation using a GLC angular quadrature (6 polar- and 6 azimuthal-angles per octant), of problem 4 using 10×10^6 source particles. 106

5.17 Cell-wise residual source for problem 4, showing a logarithmic scale [Left] and a linear scale [Right]. 109

5.18 Vector-based visualizations of the P_1 expansion of the residual source moments of problem 4. 110

5.19 Variation of problem 4 where the localized source is now distributed laterally. 112

5.20 DO solution of problem 5 using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant. 113

5.21 Standard Monte Carlo solution of problem 5 using 10×10^6 source particles. 114

5.22 The RMC defect solution, for a DO simulation using a GLC angular quadrature (6 polar- and 6 azimuthal-angles per octant), of problem 5 using 10×10^6 source particles. 115

5.23 Cell-wise residual source for problem 5. 117

5.24 Vector-based visualizations of the P_1 expansion of the residual source moments of problem 5. 118

5.25 The scalar adjoint flux for problem 5, showing a logarithmic scale [Left] and a linear scalr [Right]. The solution is based on a DO simulation using a GLC angular quadrature with 48 azimuthal angles and 6 polar angles. 120

5.26 The product of $\phi^\dagger$ and $\int_V |r|dV$ for problem 5. 125

5.27 Particles tracks for SMC and RMC problem 5 of the first 4000 source particles. The RMC is based on a DO solution using a Gauss-Legendre-Chebyshev angular quadrature with 6 polar angles and 6 azimuthal angles. The red dots indicate source particle creationg. [Top-left] SMC unbiased. [Top-right] RMC unbiased. [Bottom-left] SMC biased with $\phi^\dagger$. [Bottom-right] RMC biased with $\phi^\dagger$. N_{tracks} is the total number of linear tracks drawn, N_{surfev} is the total number of surface-events drawn, N_{scatev} is the total number of scattering events drawn and $N_{deathev}$ is the total number of death events recorded. 126

5.28 The anisotropy index for the P_1 expansion of the adjoint flux for problem 5. 132

5.29 Particles tracks for SMC and RMC problem 5 of the first 4000 source particles. The RMC is based on a DO solution using a Gauss-Legendre-Chebyshev angular quadrature with 6 polar angles and 6 azimuthal angles. The red dots indicate source particle creationg. [Left] SMC biased with $\phi^\dagger$ and $\psi^\dagger$. [Right] RMC biased with $\phi^\dagger$ and $\psi^\dagger$. N_{tracks} is the total number of linear tracks drawn, N_{surfev} is the total number of surface-events drawn, N_{scatev} is the total number of scattering events drawn and $N_{deathev}$ is the total number of death events recorded. 133

A.1 Graphical representation of intersection of a line with a plane. 145

A.2 A schematic of the raytracing required for 2D cells. This involves the intersection of 3D line with a strip. 148

A.3 The formation of a tetrahedron from a triangulation of polyhedron faces. 149

B.1 Piecewise-linear basis function on a 2D polygon. 151

B.2 Piecewise-linear basis functions on a triangle and on a quadrilateral. 151

B.3 Connection of the vertex of interest to the tetrahedrons that comprise the cell. 153

B.4 Influence of different vertices on the shape functions of within a tetrahedral portion of the cell. 155

C.1 Examples of the exponential representation of the angular flux for P_1 approximations with a different r-value. 160

LIST OF TABLES

TABLE Page

5.1 Key Performance Indicators for problem 1 using different approximate solutions, $\tilde{\phi}$, as input to the RMC method. All Monte Carlo simulations here were performed using 10^7 source particles. 93

5.2 Key Performance Indicators for problem 2 using different angular quadratures in the production of $\tilde{\phi}$, as input to the RMC method. All Monte Carlo simulations here were performed using 10^7 source particles. 98

5.3 Key Performance Indicators for problem 3. All Monte Carlo simulations here were performed using 10^7 source particles. 103

5.4 Key Performance Indicators for problem 4. All Monte Carlo simulations here were performed using 10^7 source particles. 107

5.5 Key Performance Indicators for problem 5. All Monte Carlo simulations here were performed using 10^7 source particles. 116

5.6 Key Performance Indicators for problem 5 with scalar adjoint flux biasing for QOI-2. All Monte Carlo simulations here were performed using 10^7 source particles. 125

5.7 Key Performance Indicators for problem 5 with scalar adjoint flux and angular adjoint flux biasing for QOI-2. All Monte Carlo simulations here were performed using 10^7 source particles. The sub-table below the main table is a summary of the downward trend of the uncertainty with application of additional biasing. 131

1. INTRODUCTION

The Residual Monte Carlo (RMC) method enables one to directly compute the error associated with an approximate deterministic solution, measured with respect to an "exact" (zero statistical error) standard Monte Carlo solution. The only alternative to residual Monte Carlo is a calculation of the error by first computing the standard Monte Carlo solution and then subtracting the approximate deterministic solution from it. We later show that as the size of the error in an approximate deterministic solution becomes smaller relative to the size of the solution itself, the RMC method becomes increasingly more efficient than the SMC method. The purpose of our dissertation research is to develop and test an RMC algorithm with several new characteristics and compare its efficiency with the SMC method for a series of test problems.

In this introductory text we define the Linear Boltzmann transport equation as well as several associated aspects relevant to this dissertation, including, but not limited to, the scalar flux. We then discuss popular methods for obtaining approximate solutions to this equation and how the defect of these solutions constitute our focal point. This is then followed by a literature review of work on this topic.

1.1 The Linear Boltzmann Equation (LBE)

The linear Boltzmann transport equation [1], henceforth denoted as LBE, is used as a mathematical model for describing many real-world physical phenomena such as neutral particle transport (e.g., neutrons, photons,) charged-particle transport (e.g., relativistic electrons,) and thermal radiation transport, among others. In continuum-form the steady state linear Boltzmann transport equation

[1] is denoted as

$$\begin{aligned}\boldsymbol{\Omega}\cdot\boldsymbol{\nabla}\psi(\mathbf{x},\boldsymbol{\Omega},E)+\sigma_t(\mathbf{x},E)\psi(\mathbf{x},\boldsymbol{\Omega},E)&\\ =\int_{E'}\int_{4\pi}\sigma_s(\mathbf{x},\boldsymbol{\Omega}'\cdot\boldsymbol{\Omega},E'\rightarrow E)\psi(\mathbf{x},\boldsymbol{\Omega}',E')d\boldsymbol{\Omega}'dE' &\qquad (1.1a)\\ +\ q(\mathbf{x},\boldsymbol{\Omega},E)\qquad \mathbf{x}\in\mathcal{D}_s,E\in\mathcal{D}_e,\boldsymbol{\Omega}\in\mathcal{D}_a,&\end{aligned}$$

$$\boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi(\mathbf{x},\boldsymbol{\Omega},E)=\boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi^{inc}(\mathbf{x},\boldsymbol{\Omega},E),\qquad \mathbf{x}\in\partial\mathcal{D}_s,\boldsymbol{\Omega}\in\left\{\boldsymbol{\Omega}\ :\ \boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}\quad (1.1b)$$

where the angular flux is denoted by $\psi\ (particles/cm^2{-}s{-}steradian)$, the spatial coordinate is denoted by $\mathbf{x}\ (cm)$ in the spatial domain, $\mathcal{D}_s$, the direction of particle flow is denoted by the dimensionless unit-vector, $\boldsymbol{\Omega}$, in the angular domain, $\mathcal{D}_a$ (i.e. on the unit sphere, 4π), and the particle energy is denoted with $E\ (eV)$ in energy domain, $\mathcal{D}_e$. Particles interact with a total macroscopic interaction cross section, $\sigma_t\ (cm^{-1})$, and transfer from $E'{\rightarrow}E$ and from $\boldsymbol{\Omega}'{\rightarrow}\boldsymbol{\Omega}$ via the macroscopic scattering/production cross section, $\sigma_s\ (cm^{-1}{-}steradians)$. Finally an inhomogeneous source $q\ (particles/cm^3{-}s{-}steradian)$ may also be present.

On the boundary, $\partial\mathcal{D}_s$, the condition relevant to this research is that ψ is equal a known incident angular flux, ψ^{inc}, relative to the outward pointing boundary normal, $\mathbf{n}_{\partial\mathcal{D}_s}$.

The independent variables in Eq. (1.1) are: space, $\mathbf{x}=[x,y,z]$, angle, $\boldsymbol{\Omega}=[\sin\theta,\cos\varphi,\cos\theta]$, and energy, E. The direction vector, $\boldsymbol{\Omega}$, is defined in terms of the polar angle, θ, and the azimuthal angle, φ, (both following the convention as in [2]) and therefore constitutes only two variables, resulting in a total of six independent variables. All six these variables are in continuum form.

1.1.1 The scalar flux

Any method that solves either the continuum- or multigroup form of the LBE is limited in the ability to produce the actual continuum-form of the true solution, $\psi(\mathbf{x},\boldsymbol{\Omega},E)$. For instance, at a given geometrical point in space and a specific energy value, the angular distribution of ψ can be very complex and would require a potential prohibitively large amount of data points to resolve

with sufficient fidelity. Additionally, analytical solutions of Eq. (1.1) only exist for comparatively simple problems mainly because the equation presents much complexity and a large amount of independent variables. Consequently, for realistic or practical problems, the continuum solution requires discretization of the independent variables so that an approximate solution can be found.

In the scope of this text we will restrict our focus to problems that seek ψ integrated over the entire angle space, which is often referred to as the *scalar flux*,

$$\phi(\mathbf{x}, E) = \int_{4\pi} \psi(\mathbf{x}, \mathbf{\Omega}, E) d\mathbf{\Omega}. \tag{1.2}$$

The main reason the scalar flux is often the only solution sought is that it can be used to determine a reaction rate across the entire domain via any reaction cross section, σ, i.e.,

$$\begin{aligned} \text{Reaction-rate}(\mathbf{x}, E) &= \sigma(\mathbf{x}, E) \int_{4\pi} \psi(\mathbf{x}, \mathbf{\Omega}, E) d\mathbf{\Omega} \\ &= \sigma(\mathbf{x}, E) \phi(\mathbf{x}, E), \end{aligned} \tag{1.3}$$

which in turn can provide a host of other spatially varying quantities (i.e., fission rate, collision heating rate, material activation, etc.).

The scalar flux can, however, not be obtained for the continuum space- and energy domain, and is commonly obtained as an integral over some spatial- and/or energy sub-domain, or in terms of the finite number of coefficients in a finite element representation, or any combination of the prior two.

1.1.2 The multigroup approximation

The most common approach, as part of the process to enable approximate solutions to be obtained, is to first apply the *multigroup approximation*. The approximation requires the definition of numerous non-overlapping energy sub-domains where each energy sub-domain, $\mathcal{D}_{e,g}$, is called an energy group, normally denoted with group index g, and does not necessarily have to be contiguous. With

these sub-domains defined one can then define a group scalar flux,

$$\phi_g(\mathbf{x}) = \int_{\mathcal{D}_{e,g}} \phi(\mathbf{x}, E) dE \qquad \mathcal{D}_{e,g} \subset \mathcal{D}_e. \tag{1.4}$$

Next we seek to have a group cross section (for any type of reaction), σ_g, such that we preserve the continuous-energy reaction rate, i.e.,

$$\sigma_g(\mathbf{x})\phi_g(\mathbf{x}) = \int_{\mathcal{D}_{e,g}} \sigma(\mathbf{x}, E)\phi(\mathbf{x}, E) dE \tag{1.5}$$

or in more common notation

$$\sigma_g(\mathbf{x}) = \frac{\int_{\mathcal{D}_{e,g}} \sigma(\mathbf{x}, E)\phi(\mathbf{x}, E) dE}{\int_{\mathcal{D}_{e,g}} \phi(\mathbf{x}, E) dE}. \tag{1.6}$$

In this definition $\phi(\mathbf{x}, E)$ acts as a weighting function. The major dilemma, however, is that we are developing multigroup cross sections with the concise mission of determining an approximation of the unknown $\phi(\mathbf{x}, E)$. One solution to this dilemma is to replace the flux with a representative weighting spectrum, $w(E)$, such that

$$\sigma_g(\mathbf{x}) = \frac{\int_{\mathcal{D}_{e,g}} \sigma(\mathbf{x}, E) w(E) dE}{\int_{\mathcal{D}_{e,g}} w(E) dE}. \tag{1.7}$$

where w has the properties:

$$w(E) \geq 0$$

$$\int_{\mathcal{D}_e} w(E) dE = 1.$$

It is important to note that it is vital to choose the weighting function as close to the spectrum[1], that will be observed in ϕ, as possible. This is not always an easy task since it will be position dependent, and a material may experience a variety of flux spectra. Multiple methods exist to obtain representative cross sections or to modify cross sections in special regions, however, that is

[1] We refer to the spectrum here as the character of the particle density per unit energy as a function of energy.

beyond the scope of this text. In this text we will simply state that a multigroup form of the LBE can be derived, using the strategy above, and takes the form

$$\begin{aligned}\boldsymbol{\Omega}\cdot\boldsymbol{\nabla}\psi_g(\mathbf{x},\boldsymbol{\Omega})+\sigma_{tg}(\mathbf{x})\psi_g(\mathbf{x},\boldsymbol{\Omega})&\\ =\sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\boldsymbol{\Omega}'\cdot\boldsymbol{\Omega})\psi_{g'}(\mathbf{x},\boldsymbol{\Omega}')d\boldsymbol{\Omega}'\right]&\\ +\ q_g(\mathbf{x},\boldsymbol{\Omega}),\quad \mathbf{x}\in\mathcal{D}_s, g\in[0,N_G{-}1],\boldsymbol{\Omega}\in\mathcal{D}_a,&\end{aligned}\tag{1.8a}$$

$$\boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi_g(\mathbf{x},\boldsymbol{\Omega})=\boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi_g^{inc}(\mathbf{x},\boldsymbol{\Omega}),\qquad \mathbf{x}\in\partial\mathcal{D}_s,\boldsymbol{\Omega}\in\left\{\boldsymbol{\Omega}\ :\ \boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}\tag{1.8b}$$

where N_G is the number of groups. This equation still maintains the continuum form for the spatial- and angular independent variables but is now discrete in energy.

1.1.3 The residual, r, the defect, ψ_D, and the residual transport equation

The form of Eq. (1.8) is a statement of conservation and thus, given the exact solution, ψ_g, the statement will hold true, and the left-hand side of Eq. (1.8a) will equal the right-hand side exactly. In reality it is impossible to obtain ψ_g in continuum-form, however, several methods exist to obtain an *approximate solution*, $\tilde{\psi}_g$.

With a continuous approximate solution in-hand, the statement of conservation will in general not hold at all points in the phase-space domain[2], an aspect leading us to the definition of the *residual*, r_g, which is simply the imbalance of the right- and left-hand sides of Eq. (1.8) when $\tilde{\psi}_g$ is used instead of ψ_g. It takes the form,

$$\begin{aligned}r_g(\mathbf{x},\boldsymbol{\Omega})=\sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\boldsymbol{\Omega}'\cdot\boldsymbol{\Omega})\tilde{\psi}_{g'}(\mathbf{x},\boldsymbol{\Omega}')d\boldsymbol{\Omega}'\right]\ +\ q_g(\mathbf{x},\boldsymbol{\Omega})&\\ -\boldsymbol{\Omega}\cdot\boldsymbol{\nabla}\tilde{\psi}_g(\mathbf{x},\boldsymbol{\Omega})-\sigma_{tg}(\mathbf{x})\tilde{\psi}_g(\mathbf{x},\boldsymbol{\Omega}),\qquad \mathbf{x}\in\mathcal{D}_s,&\end{aligned}\tag{1.9a}$$

[2]Space and angle

$$r_g(\mathbf{x},\mathbf{\Omega}) = \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\Big(\psi_g^{inc}(\mathbf{x},\mathbf{\Omega}) - \tilde{\psi}_g(\mathbf{x},\mathbf{\Omega})\Big), \qquad \mathbf{x}\in\partial\mathcal{D}_s, \mathbf{\Omega}\in\big\{\mathbf{\Omega}\ :\ \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\big\} \tag{1.9b}$$

This residual can be positive or negative, is in continuum and continuous form, and is a useful quantity on the hand of another important definition namely the group angular *flux-defect*, ψ_{Dg}, defined as

$$\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) = \psi_g(\mathbf{x},\mathbf{\Omega}) - \tilde{\psi}_g(\mathbf{x},\mathbf{\Omega}). \tag{1.10}$$

Using this definition we can substitute $\tilde{\psi}_g = \psi_g - \psi_{Dg}$ into Eq. (1.9) to firstly recover all the terms in Eq. (1.8), all of which subsequently cancel to zero because of conservation, and secondly produces terms that can be arranged into the form of a transport equation as

$$\begin{aligned}\mathbf{\Omega}\cdot\boldsymbol{\nabla}\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) + \sigma_{tg}(\mathbf{x})\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) &\\ = \sum_{g'=0}^{N_G-1}\Big[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\mathbf{\Omega}'\cdot\mathbf{\Omega})\psi_{Dg'}(\mathbf{x},\mathbf{\Omega}')d\mathbf{\Omega}'\Big] &\\ +\ r_g(\mathbf{x},\mathbf{\Omega}), \qquad \mathbf{x}\in\mathcal{D}_s,&\end{aligned} \tag{1.11a}$$

$$\mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) = r_g(\mathbf{x},\mathbf{\Omega}), \qquad \mathbf{x}\in\partial\mathcal{D}_s, \mathbf{\Omega}\in\big\{\mathbf{\Omega}\ :\ \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\big\} \tag{1.11b}$$

This equation is simply another multigroup linear Boltzmann transport equation with the residual, r_g, acting analogously to the inhomogeneous source, q_g, and a boundary incident angular flux, and with ψ_{Dg} as the unknown instead of ψ_g. Since r_g can produce positive or negative virtual particles, the solution, ψ_{Dg}, can also take on positive and negative values. For all discussions past this point this transport equation will be referred to as the *residual transport equation*.

1.2 Overview of popular methods to solve the Linear Boltzmann transport equation

In this section we discuss the two most practical methods of solving the LBE, namely the Monte Carlo method and the Discrete Ordinates method.

1.2.1 The Monte Carlo method

Perhaps the oldest and simplest method for solving Eq. (1.1) is the Monte Carlo method dating back to 1947 [3]. The method is based on the fundamental definition of the interaction cross section, σ, which is conceptually a probabilistic property of a material [1, 2], and can therefore be used to determine the average behavior of transported particles, provided that a sufficient amount, N_p, of source particles are stochastically simulated. Methods based on probabilities, such as the Monte Carlo method, are called *probabilistic methods* and although we will provide succinct detail on the Monte Carlo method in later chapters, we will note here that production codes such as MCNP [4] and GEANT-4 [5] are good examples of how the method can be used in modern applications with exceptional accuracy and fidelity.

The Monte Carlo method has the primary benefit of being able to transport particles in continuum form, i.e., without spatial-, angular- or energy-discretization, and consequently can provide accurate insight into the true solution. The solution itself, however, can only be determined in representable form. The most common representable-form obtained from production Monte Carlo codes are based on integration over the independent variables, i.e., integration over a specific volume, angle-space, and energy-interval. This form is manifested in the concept of a tally [4, 5], which can be modified (as will be shown later) with regards to the spatial variable to provide a discontinuous finite element representation in space and/or harmonic expansions in angle.

Despite the accuracy of the Monte Carlo method its main drawback is the stochastic nature of the solution, which is subject to statistical uncertainty. This uncertainty manifests as an error which can be shown, via the *central limit theorem*, to be a normal distribution around the true value of the solution (at any given point in phase-space). The standard deviation, S, of this error can also be shown to reduce with a factor of $N_p^{-1/2}$. This fact is problematic because therein lies the inherent cost of the Monte Carlo method, i.e., in order to decrease S by a factor of 2 the number of particles simulated needs to increase by a factor of 4, a difficulty that is exasperated by particles that weakly contribute to tallies (i.e. tallies behind shielding materials, tallies of small volume, etc.).

1.2.2 The Discrete Ordinates method

The multigroup form of the LBE, i.e., Eq. (1.8), can also be solved by applying a suitable spatial- and angular-discretization. The Discrete Ordinates [1] method applies an angular quadrature to integrate the angular scattering source-term in the LBE, consequently the name "Discrete Ordinate" stems from the discrete angles used in the quadrature, after which the method requires the solution of the LBE for only this set of discrete angles. Therefore, with N_D number of angles in the quadrature, Eq. (1.8) is split into N_D equations that are coupled via the scattering source. These equations in turn require a suitable spatial discretization to provide a set of equations that forms a linear system to be solved deterministically, resulting in an approximate solution. In general, methods based on a discrete equation form (as opposed to a continuum form) are called *deterministic methods*.

The benefit of having a fully discrete system is that iterative schemes can be devised that leverage several powerful and efficient linear algebra techniques (commonly known as a *sweep*), however, the large dimensionality of this system often requires copious amounts of computer memory. Refinement in the inherent discretizations infringe on these memory constraints and is unfortunately necessary to minimize the truncation error associated with each of the discretizations. The challenging aspect of such refinements is to assess when a particular refinement is sufficient.

The required refinement in spatial discretization is often the easiest to identify since one can assess the effect of refinement in areas where the solution is rapidly varying. Many techniques can be applied to provide concise information on where spatial refinement is required, however, such techniques often assess the effect by using multiple simulations at different levels of refinement. In this aspect we identify the first component of desirable information regarding discretization error, which is to know what the approximate error is of a given spatial discretization without requiring refinement.

Of particular concern regarding the angular discretization used in the Discrete Ordinates method is the presence of "ray effects" [1] which are especially prevalent if the scattering ratio, $c = \sigma_s/\sigma_t$, of the material is low and the source is localized. These ray effects normally manifest as nonphysical oscillatory errors in the scalar flux solution where the magnitude of the oscillations increases

with radial distance from the source and the frequency is dependent on the angular quadrature. The simplest remedy for these angular errors is to increase the order of the angular quadrature (i.e., increasing N_D), however, similar to the spatial discretization, it is often difficult to ascertain what level of angular refinement is actually sufficient, especially considering that it can vary with position and direction.

1.2.2.1 The Method of Characteristics (MOC) variant

A variation of the Discrete Ordinate method can be obtained by using rays at specific spatial intervals, along the ordinate directions of the angular quadrature, to integrate the angular flux within a given cell. These rays are an alternative to the concept of the sweep in the Discrete Ordinates method and is essentially a different way of handling the left-hand side of Eq. (1.8). The equation is solved along a characteristic ray (via a simple transformation of the gradient term), hence the name "Method of Characteristics". The iterative scheme with respect to the scattering source, however, is often the same as the Discrete Ordinates method. Additionally, the method is also troubled with the same ray effects as encountered in the Discrete Ordinates method.

The MOC is a very popular and efficient method for 2D nuclear reactor simulations, however, its adoption in practical 3D problems is scarce, mostly because of the increased ray-tracing cost. The resulting representation of the scalar flux, in terms of the spatial discretization, can be considered to be equivalent to the Discrete Ordinates method.

1.2.2.2 The Diffusion Approximation

The diffusion approximation arises from the assumption that the angular flux is at most linearly dependent on angle. Using the current $\mathbf{J}$, defined as

$$\mathbf{J}(\mathbf{x}, \mathbf{\Omega}, E) = \int_{4\pi} \mathbf{\Omega}\, \psi(\mathbf{x}, \mathbf{\Omega}, E)\, d\mathbf{\Omega}, \tag{1.12}$$

the linear in angle assumption can then be used to derive a relationship between the group current, $\mathbf{J}_g$ and the scalar flux, ϕ_g, which takes the form of Fick's law

$$\mathbf{J}_g(\mathbf{x}, \mathbf{\Omega}) = -D_g(\mathbf{x})\boldsymbol{\nabla}\phi_g(\mathbf{x}) \tag{1.13}$$

where D_g is the group diffusion equation. With this relationship the multigroup diffusion equation can be derived as

$$-\boldsymbol{\nabla}\cdot\Big(D_g(\mathbf{x})\boldsymbol{\nabla}\phi_g(\mathbf{x})\Big) + \sigma_{rg}(\mathbf{x})\phi_g(\mathbf{x}) = \sum_{\substack{g'=0 \\ g'\neq g}}^{N_g-1}\Big[\sigma_{s0,g'\to g}(\mathbf{x})\phi_g(\mathbf{x})\Big] + q_g(\mathbf{x}), \tag{1.14}$$

where σ_{rg} is the group removal cross section and $\sigma_{s0,g'\to g}$ is isotropic scattering cross section from group g' to group g.

The primary advantage of the diffusion approximation is that it eliminates the need for a sweep algorithm because the angular dependence has been removed. Additionally, since the equation is a diffusion-reaction equation the resulting system of equations, after discretization in space, forms a matrix that is symmetric positive definite (SPD). Such systems are very well studied and can be solved numerically with the preconditioned *conjugate gradient* algorithm which is known to be the most efficient algorithm for such systems.

The approximate solution produced by solving the multigroup diffusion equation still requires a quantification of discretization errors just like other deterministic methods. In this case, the spatial discretization error and the error introduced by the linear in angle assumption are the aspects of the approximate solution that needs to be quantified.

1.3 The focal point of this research

Given a Monte Carlo method for solving the *residual transport equation* and an approximate solution to an associated system of equations, produced by methods discussed in section 1.2.2, the residual Monte Carlo (RMC) method enables the use of essentially the same Monte Carlo algo-

rithms for solving the continuum form of the transport equation, i.e., the standard Monte Carlo (SMC) method, with the only difference being that we can now directly compute the additive error or "defect" associated with the approximate solution.

The motivation to do so is that this defect is measured relative to the solution obtained from standard Monte Carlo (SMC) in the limit of an infinite number of Monte Carlo histories (zero statistical error). When the approximate solution is simply zero, the defect is the solution itself, and RMC is identical to SMC. A defect in a quantity of interest (QOI) is obtained from a SMC calculation by first computing the quantity of interest (QOI) and then subtracting the approximate QOI from the SMC QOI. If one is to compute the defect in a QOI with a statistical deviation of p-percent, the defect need only be computed in a RMC calculation with this same percent deviation, but the QOI must be computed in a SMC calculation with a percent deviation equal to p times the defect divided by the QOI. Thus, as the size of the defect relative to the QOI itself approaches zero, the percent deviation required for the SMC calculation approaches zero. This is the basis for expecting that residual Monte Carlo can be signifcantly more efficient than standard Monte Carlo. This property of the RMC method is the focal point of this research.

The major difficulty to overcome is the definition of a continuum residual which is required to ensure that the RMC defect solution is the defect of the approximate discrete solution relative to the continuum solution. This is a particularly hard problem on the hand of Eq. (1.9), which requires the continuum angular flux.

1.4 Literature review

In the previous sections we referred to the flux-error as the flux-defect to avoid confusion with statistical error. For the purpose of being consistent with the cited literature, we will refer to the flux-error as simply the error unless explicitly stating the statistical error.

The topic of RMC, as it relates to this research, can easily be confused with the classical application of RMC to achieve exponential convergence. We want to be clear that the topic of this research is not to achieve exponential convergence but to apply the RMC method to the continuum

transport equation such that we can determine the error between the solution of the discretized transport equation and the solution to the continuum transport equation, thereby quantifying discretization errors.

As background information, we now briefly summarize how RMC is used to obtain exponential convergence. For a fully discrete system, for which one can define a discrete residual, one can perform multiple sequential RMC simulations that successively compute the error in the discrete approximate solution. At each step the error is added to the approximate solution to obtain an updated residual for the next step. Each RMC simulation is tasked with determining the statistical uncertainty of the error as a fixed-fraction of the estimated error, thereby leading to exponential convergence. Since a reduction in the L_2-norm of the error occurs at each step the corresponding norm of the residual source also reduces at each step, hence leading to the RMC method also being called the *Reduced Source Method*[6].

The first reference we found for the concept of residual Monte Carlo on discrete systems date back to that of Halton in 1962 [7] and later in 1994 [8], which termed the process *Sequential Monte Carlo*. The application of this technique to transport-like equations is later seen in work by Evans et al. [9] which operated on the fully discretized thermal radiative diffusion equation.

Application of the method to continuum systems, also achieving exponential convergence, has been demonstrated on simple 1D problems in the 1990s to the early 2000s [6, 10, 11, 12, 13], however, significant difficulties were encountered with multidimensional problems.

The work by Peterson et al. [14] succinctly summarized the difficulties of achieving exponential convergence for multidimensional continuum transport systems. The major difficulty to overcome is that the Monte Carlo method can indeed solve the transport equation in continuum form but it cannot provide a continuum representation of the solution. Hence, the RMC solution of the error has to be projected onto a finite-dimensional solution space which itself incurs error, causing the sequential process to stall in terms of error convergence. The work by Peterson et al.[14] involved only one dimension in space and one dimension in angle which allowed them to define a two dimensional discretized phase-space with independent variables x for space and μ for an-

gle. Then, after detecting a convergence stall they applied mesh refinement on the 2D phase-space based on the size of discontinuities. This h-refinement technique then allowed the exponential convergence to continue down to machine precision, however, such a refinement technique for 2D and 3D spatial dimensions is exceedingly complex.

An additional concern with the RMC source definition in most of the literature (like [14]) is that the residual source definition uses the angular flux along the angular quadrature directions. This is problematic because multidimensional DO problems can easily require the storage of thousands of angles, requiring prohibitive amounts of memory.

We were unable to find references in literature that was purely concerned with finding the error between discrete solutions and continuum solutions for multidimensional multi-material problems. The literature we found on multidimensional problems (i.e., [13]) also did not define a continuum residual which is difficult to do mostly because the angular flux solution is only available as either the values at angular quadrature directions or as harmonic expansion coefficients. The latter creates a disconnect between the angular fluxes used to obtain the coefficients and those that can be represented by the expansion.

We believe we have found a means to define a continuum residual which can be used to determine the error in the scalar flux such that there is no disconnect between the scalar flux of the approximate solution and the error of that scalar flux.

1.5 Overview of chapters

The remaining chapters of this text are as follows. Chapter 2 details the standard Monte Carlo method. We begin this chapter by detailing the process of source sampling including the process required to initialize source sampling, which allows for greater efficiency. The different types of sources we describe are limited to point-, boundary- and cell-sources. We then describe the stochastic transport process where particles are transported between interactions and/or surfaces. Finally, the remainder of the chapter discusses the process of developing solutions from the transport process and other essential utilities.

Since the focal point of this research involves determining the defect of popular deterministic methods we detail the deterministic method used in this research in Chapter 3, which is the Discrete Ordinates method.

We then continue to Chapter 4 which details the processes involved in defining and sampling the residual source. We introduce our fundamental assumption of using only the scalar flux from a DO solution as the approximate solution, which allows us to define a residual that is continuous in angle. We then explore different spatial representations and the relevant modifications to the residual definitions. This is followed by the finer intricacies and initializations required for the source sampling to function.

With our computational methods defined we detail several simulations in Chapter 5. The simulations range from homogeneous 1D problems to multidimensional multi-material problems with geometries that are particularly challenging to any DO method.

We finish with a conclusion in Chapter 6 where we discuss the overall outcomes of the research and identify further work.

2. INTRODUCTION TO THE MONTE CARLO METHOD

We now introduce the Monte Carlo method for particle transport. We start with the sampling of the source which is discussed in detail. This will be followed by the particle transport process, from its birth at a source to its eventual "death" where it is either absorbed or leaves the simulation domain. We will assume (for discussion purposes) that, in the tracking process, that particles can only experience two interactions namely, absorption via an absorption cross section, σ_a, or scatter via a scattering cross section, σ_s. This is followed by a description of how the tracks, created by the particle's transport, are used to contribute to volumetric tallies. Finally, we add additional detail on how some common conditions are fulfilled, i.e., multigroup transport, tally uncertainties, and modification to tallies.

Throughout the discussions that follow we will repeatedly state that a discrete cumulative distribution function (CDF) is sampled. We note here the following meaning. Given discrete CDF, c, with N discrete entries. Each entry is denoted as c_i, with index $i \in \{0, \ldots, N-1\}$, and needs to comply with the requirements $c_i \in [0, 1]$ and $c_{i+1} \geq c_i$ (monotonically increasing). If a pseudo-random number, $\theta_R \in [0, 1)$ is generated then

$$\text{entry } i \text{ is the sampled-entry if } \begin{cases} 1 \leq i \leq N-1 \text{ and } c_{i-1} \leq \theta_R < c_i, \text{ or} \\ i = 0 \text{ and } 0 \leq \theta_R < c_i. \end{cases} \tag{2.1}$$

2.1 Source sampling

The first step of starting the stochastic simulation of particles is the sampling of one or more sources. In this research we will limit the standard Monte Carlo method to using one of three types of sources: i) isotropic point-sources, ii) isotropic boundary-sources, and iii) isotropic cell-sources.

The latter two have a strong connection with the geometry of cells used in the spatial discretization with the cell-source following the convention of being constant within a cell.

The ultimate goal of source sampling is to determine a source particle's position, $\mathbf{x}_i$, direction, $\mathbf{\Omega}_i$, and either its energy, E, if continuous energy cross-sections are used, or its energy-group, g, if multigroup cross sections are used. In other words, we need $(\mathbf{x}_i, \mathbf{\Omega}_i, E)$ or $(\mathbf{x}_i, \mathbf{\Omega}_i, g)$ depending on whether we use continuous- or multigroup cross-sections, respectively.

Figure 2.1 below, graphically represents all the stochastic sampling procedures required for each source type relevant to our standard Monte Carlo method.

2.1.1 Sampling a particular source

A problem can comprise several sources of different type, with each source-type requiring specialized sampling procedures. The sampling process starts with the sampling of discrete source-index CDF, c_s, with a total number of N_s entries (i.e., the number of sources) where each entry, with index $i \in \{0, \ldots, N_s{-}1\}$, is

$$c_{si} = \frac{\sum_{x=0}^{i} Q_x}{Q_{total}}, \tag{2.2}$$

where Q_x is the strength of source x, given by

$$Q_x = \sum_{g=0}^{G-1} Q_{x,g}, \tag{2.3}$$

$$Q_{x,g} = \begin{cases} S_{point,x,g}, & \text{for a point-source,} \\ \int_{S_{\partial\mathcal{D}}} S_{bndry,x,g}\, dS, & \text{for a boundary-source,} \\ \int_{V_{\mathcal{D}}} S_{cell,x,g}\, dV, & \text{for a cell-source,} \end{cases} \tag{2.4}$$

$$Q_{total} = \sum_{x=0}^{N_s-1} Q_x, \tag{2.5}$$

$S_{point,x,g}$, is a point-source strength with units $[s^{-1}]$, $S_{bndry,x,g}$ is a boundary-source strength with units $[cm^{-2}s^{-1}]$, and $S_{cell,x,g}$ is a cell-source strength with units $[cm^{-3}s^{-1}]$, all for a source energy-group g.

With a source index x established, we next sample a group index g, from a discrete source-specific energy-group-index CDF, $c_{sx,e}$, with a total number of G entries where each entry, with index g, is

$$c_{sx,eg} = \frac{\sum_{g'=0}^{g} Q_{x,g'}}{Q_x}. \tag{2.6}$$

The multigroup format of the source-strength, for a specific source ($S_{point,x,g}$, $S_{bndry,x,g}$ or $S_{cell,x,g}$), is a commonly used format for specifying energy dependent sources, but it can also be generalized to sources that emit particles with discrete energies, i.e., sources that mimic radioactive-decay, where the energies are discrete rather than continuous within a group. For this research we will only focus on the multigroup format, and although there is no requirement that the structure of the energy sub-domains (i.e., groups) of a source be the same as those used in any tallies, cross sections, or other sources in the problem, we will use the same multigroup structure throughout this text for improved comprehension of the core ideas.

2.1.2 Sampling the group and/or energy

In production codes (such as MCNP[4]) the group strength can be supplied in two ways: i) a histogram of group-integrated probability densities (top of Figure 2.2), or ii) the group boundary-points on the energy-dependent PDF (bottom of Figure 2.2). If a histogram is supplied, once the relevant histogram bin is sampled (from a CDF built from S'_g-values[1]), the energy is uniformly sampled within the energy-group's sub-domain $\mathcal{D}_{e,g}$ which, for contiguous $\mathcal{D}_{e,g}$ would simply be between the group-boundaries. If the boundary-points are supplied (for contiguous $\mathcal{D}_{e,g}$), i.e., $S(E_g)$ and $S(E_{g-1})$, an interpolated PDF can be created (red-dash curve in Figure 2.2). This

[1] S'_g is the source strength, $S(E)$, integrated over $\mathcal{D}_{e,g}$ and normalized with $\int_{\mathcal{D}_e} S(E)dE$

PDF can then be integrated piecewise, producing S_g^*-values[2], to construct a group CDF which is sampled, to identify the group. Within a group the analytic interpolated-PDF (which is linear) is then integrated to provide an invertable CDF, sampled to provide the energy within the interval.

In this research, we used only the histogram-type of specification.

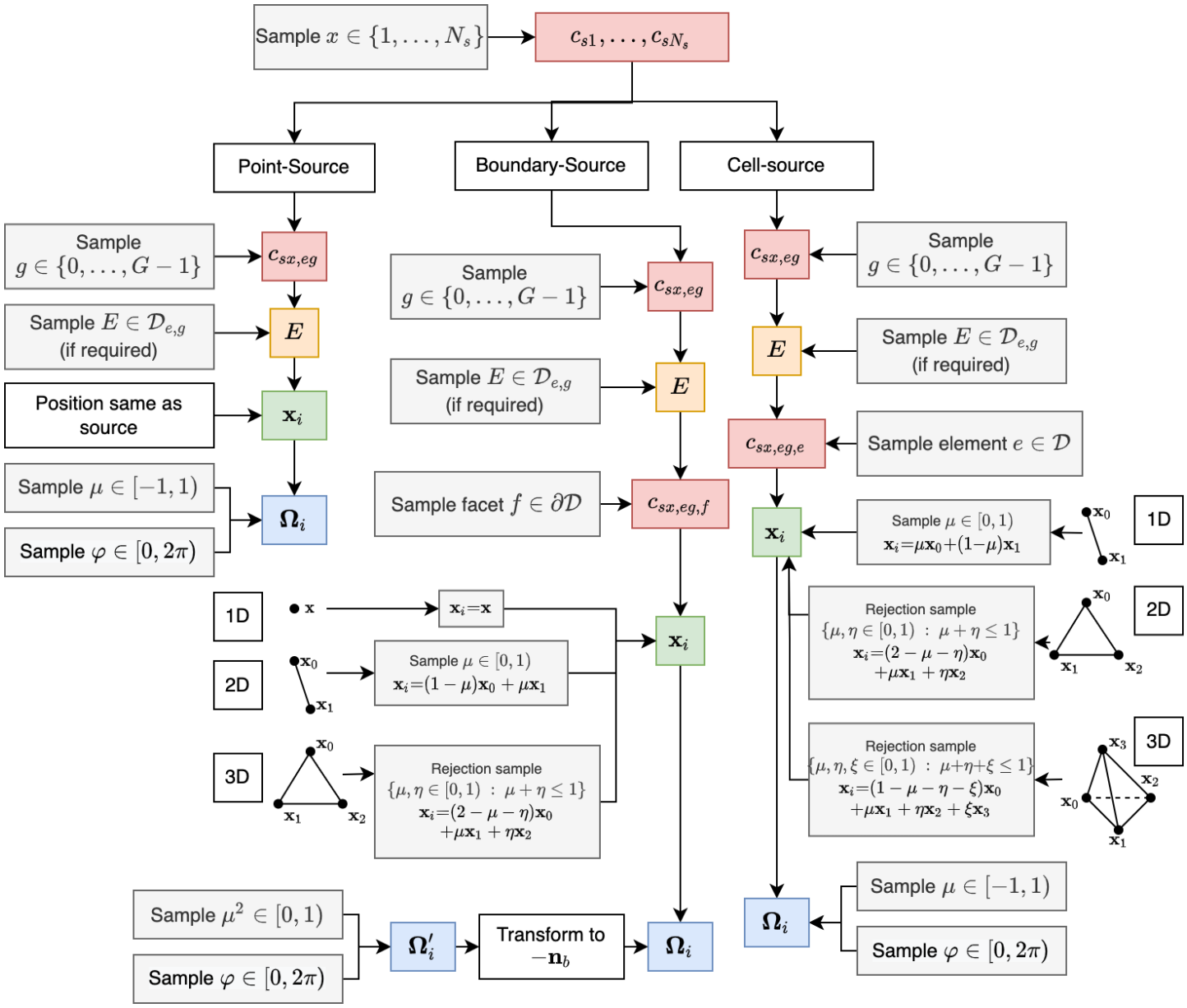

Figure 2.1: The source sampling process.

[2] S_g^* is the interpolated PDF integrated over $\mathcal{D}_e$ and normalized with $\int_{\mathcal{D}_e} S(E)dE$

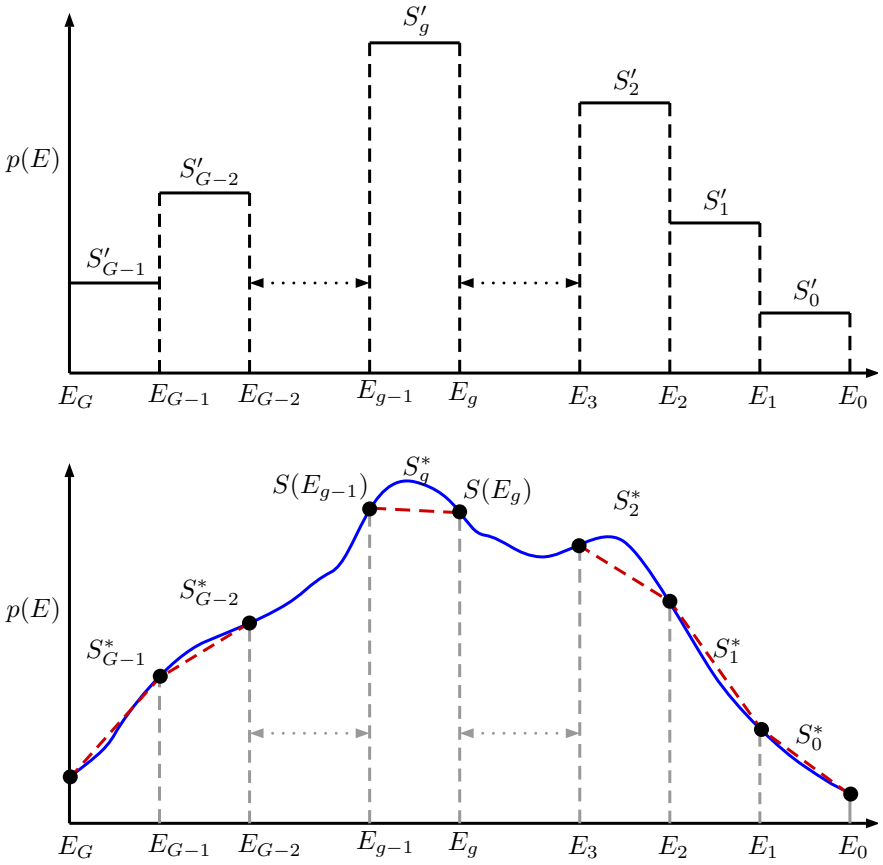

Figure 2.2: Different input-representations of the energy-dependency of sources, assuming contiguous group sub-domains. [Top] A histogram representation. S'_g is the source strength of group g normalized with $\int_{\mathcal{D}_e} S(E)dE$. [Bottom] An interpolated PDF with points, $S(E_g)$, from the true PDF (the blue curve) supplied at group boundaries. S^*_g is the source strength over $\int_{\mathcal{D}_{e,g}} S(E)dE$ normalized with $\int_{\mathcal{D}_e} S(E)dE$.

2.1.3 Sampling the position

For a point-source there is no sampling required since the source-particle's position will be the source location, $\mathbf{x}_s$.

For boundary-sources the process requires the sources' geometry to be decomposed into *elementary facets*. For a linear 2D element the fundamental facet will be a line and for a linear 3D element each face can be decomposed into elementary triangles. The lengths and areas, respectively, of these facets can be used to construct a discrete CDF from which a specific facet can be sampled. Once a facet is identified a position with the facet can be sampled as indicated in

Figure 2.1.

For cell-sources a similar process is followed. The cells are decomposed into elementary sub-elements. For linear 1D elements the sub-element is a line, for linear 2D elements the sub-elements are triangles (formed with the cell-centroid), and for linear 3D elements the sub-elements are tetrahedrons (formed with the face and cell-centroids). The volumes are then used to construct a discrete CDF from which a specific sub-element is sampled followed by the sampling of the position as indicated in Figure 2.1.

2.1.4 Sampling the direction

For this research we only considered isotropic sources.

The isotropic point-source and cell-source share the same direction sampling technique. First the polar cosine $\mu = \cos\theta$ is sampled uniformly in $[-1, 1)$ using a pseudo-random number θ_R as

$$\mu = 2\theta_R - 1. \tag{2.7}$$

This is followed by the use of another pseudo-random number which is used to sample the azimuthal angle, φ, uniformly in $[0, 2\pi)$ as

$$\varphi = 2\pi\theta_R, \tag{2.8}$$

after which the direction vector is assigned as

$$\mathbf{\Omega}_i = \begin{bmatrix} \sqrt{1-\mu^2}\cos\varphi \\ \sqrt{1-\mu^2}\sin\varphi \\ \mu \end{bmatrix}. \tag{2.9}$$

For a boundary-source there are only two differences, the first being the sampling of the polar

cosine which is performed as a cosine-law sampling [18],

$$\mu = \sqrt{\theta_R}, \tag{2.10}$$

and the second difference is the fact that the direction vector needs to be rotated, from a reference coordinate system, to a coordinate system aligned with the negative of the boundary normal, $-\mathbf{n}_b$ at the sampled position (inwardly pointing). In later discussions we will discuss how such rotations are applied.

2.1.5 Conclusion of source sampling

With a source particle's position, direction, and energy- or group sampled, the source particle is subject to transport which follows, however, it is important to note that, at the time of creation, a source particle needs to know in which cell it is located, an aspect that can be identified from the sub-facets or elements created. This information will be tracked as the particles traverses the domain.

It should also be noted that the source sampling procedure requires many pieces of information that needs to be compiled during initialization-time, i.e. the decomposition of cells, the integration of source strengths, and the creation of CDFs, all which will allow efficient repeated sampling of source particles. We will not detail such processes since it can be achieved in multiple ways and distracts from the overall algorithm.

2.2 The stochastic transport process

The transport process can be summarized with the following steps:

Step 1: Transporting a particle to a material interaction. This step needs to determine if a particle will have an interaction within a cell or on its surface.

Step 2: Randomly sampling a reaction type. A sample is made of the CDF constructed from all the different reaction types (i.e., absorption, scattering, etc.). Optionally, this could also be precluded by the sampling of a CDF to identify a component of a material.

Step 3: Applying scattering. If the particle encountered a scattering reaction the final direction and energy needs to be sampled and/or computed.

2.2.1 Step 1: Transporting a particle to a material interaction

Figure 2.3 below will be used as a reference for this discussion. The red track denotes a track where the distance to an interaction is within the cell, while the blue track denotes a track where the distance to an interaction is beyond the distance to the surface of the cell.

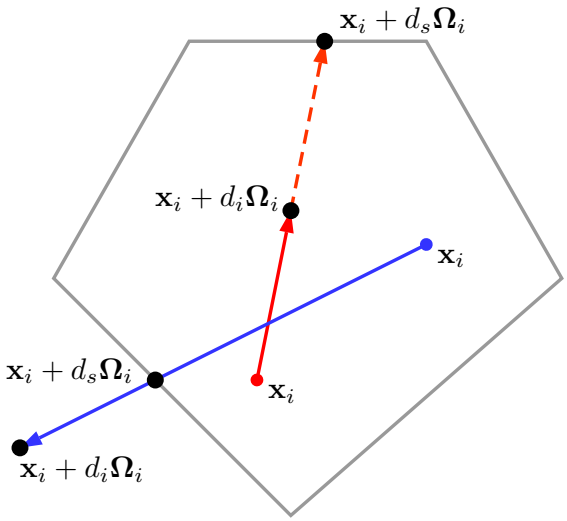

Figure 2.3: Schematic of the two ways in which a particle track can be traced within a cell, i.e., either from the initial position, $\mathbf{x}_i$, to an interaction, or from $\mathbf{x}_i$ to the cell surface.

2.2.1.1 Step 1a - Distance to interaction within a cell

Suppose a particle is created within a cell. Given the particle's initial position, $\mathbf{x}_i$, direction, $\mathbf{\Omega}_i$, and energy, E_i, we first want to sample a random distance to interaction, d_i, along the track $\mathbf{x}_i + s\mathbf{\Omega}_i$, where $\{s \ : \ s \in \mathbb{R}, s \in [0, \infty)\}$ and $d_i \in s$. In order to do this we need to relate a random number θ_R to a distance traveled, d. We start with using the definition of the cross section, for which the uncollided particle density, $n(s)$, along $\mathbf{\Omega}_i$ will vary according to the equation

$$\frac{dn(s)}{ds} = -\sigma_t n(s). \tag{2.11}$$

The solution of this ordinary differential equation is

$$n(s) = n(0)e^{-\sigma_t s}, \tag{2.12}$$

where $n(0)$ is particle density at the start of the track.

The cumulative probability of interacting in a distance d, also known as the interaction cumu-

lative distribution function (CDF), $c_\sigma(d)$, is then

$$c_\sigma(d) = \frac{n(0) - n(d)}{n(0)} = 1 - e^{-\sigma_t d}. \tag{2.13}$$

which is a monotonically increasing function in the space $c_\sigma(d) \in [0, 1)$.

This CDF can be solved using the *inverse distribution method* since the inverse of this CDF is obtainable analytically. Therefore, to sample the distance to an interaction we set $d = d_i$ and replace $c_\sigma(d)$ with a pseudo-random number θ_R, after which we solve for d_i. Consequently,

$$d_i = -\frac{1}{\sigma_t} \ln(1 - \theta_R). \tag{2.14}$$

This equation can therefore be used to randomly sample a distance to interaction in any material as long as it is homogeneous. The obvious question then arises: In a discretized system of cells, what is done to account for the finite size of a cell when the material changes between cells (i.e., an inhomogeneous domain)?

2.2.1.2 Step 1b - Distance to a cell surface (surface interaction)

Conceptually, a spatially discretized domain comprises volumes defined by surfaces. Therefore, in order to determine whether d_i is beyond the current cell's sub-domain, we require the distance to the cell-surface, $d_s \in s$, along the track $\mathbf{x}_i + s\mathbf{\Omega}_i$ which also describes a line in 3D space. Finding d_s is therefore a matter of finding the intersection of a 3D line with a 3D surface regardless of what spatial dimension the simulation is using. To name a few examples, a 1D slab cell would require the intersection of a line with a plane, a 2D linear polygon[3] would require the intersection of a line

[3]In this text a linear polygon is a cell type where each edge comprises a line with only 2 vertices, the edges are orientated counter-clockwise (right-hand-rule), and each edge can be used to form a counter-clockwise orientated triangle with the cell-centroid such that the triangle normal is in the same direction as the right-hand-rule representation of the edges.

with a strip[4], and a 3D linear polyhedron[5] would require the intersection of a line with a triangle.

Fully detailed algorithms to find d_s for multiple cell types, are presented in appendix A.

2.2.1.3 Conclusion of step 1

If $d_i < d_s$ the particle will interact with the material within the cell, alternatively, if $d_i \geq d_s$, the particle will have a surface interaction. When a material interaction occurs, the particle is transported to $\mathbf{x}_f = \mathbf{x}_i + d_i\mathbf{\Omega}_i$ and step 1 is complete. If the particle has a surface interaction it will be transported to the surface, i.e., $\mathbf{x}_f = \mathbf{x}_i + d_s\mathbf{\Omega}_i$, where cell-scope will be transferred to the cell adjacent to the current cell (at the interaction surface) and step 1 will be repeated ($\mathbf{x}_f$ becoming $\mathbf{x}_i$), or, if it interacts with a surface that is on a domain boundary, it will be killed or reflected (the latter only if the boundary is of a reflecting type).

2.2.2 Step 2: Randomly sampling a reaction type

The distance to interaction is sampled in step 1 using the material total interaction cross section, σ_t, which could comprise a number of reactions. The most typical reactions are to be absorbed via an absorption cross section, σ_a, or to scatter via a scattering cross section, σ_s. For simplicity of discussion we will assume that a particle can experience a total number of N_I reactions via cross sections $\sigma_1, \ldots, \sigma_{N_I}$, where

$$\sigma_t = \sum_{i=1}^{N_I} \sigma_i.$$

These discrete cross sections can then be used to construct a discrete CDF, $\mathbf{c}_{xs}$, where each entry is given by

$$c_{xs,0} = 0, \quad c_{xs,j} = \sum_{i=1}^{j} \frac{\sigma_i}{\sigma_t}, \qquad j \in \{0, \ldots, N_I\}.$$

[4]In this text we define a strip using two vertices, $\mathbf{v0}$ and $\mathbf{v1}$, to form a vector $\mathbf{v}$, and a normal vector $\mathbf{n}$. A point, $\mathbf{x}$, is on the strip if $\mathbf{n}\cdot(\mathbf{x} - \mathbf{v0}) = 0$ and $0 \leq (\mathbf{x} - \mathbf{v_0})\cdot\mathbf{v} \leq ||\mathbf{v}||$.

[5]In this text we define a 3D linear polyhedron as a polyhedron where each face is a linear polygon and each triangle in each face can be used to construct a non-inverted tetrahedron (using the cell-centroid as an extra vertex).

This CDF can then be sampled to identify the relevant reaction.

If the sampled reaction is the absorption reaction, σ_a, the particle is killed. Alternatively, the most common interaction is to scatter, however, other reactions such as those creating additional particles can occur. In such cases the additional particles are created (according to the underlying physics of the reaction) and banked for later retrieval. Banked particles are retrieved and transported when the current particle is killed. For the purpose of this text we will only consider the simplistic case where a particle can either be absorbed or scattered.

It should be noted that, in production codes, the material is often comprised of different isotopic elements each with its own set of cross sections. This is often necessary because it is rare to have available nuclear data on a material basis. In this research we manufactured cross sections on a material-basis and therefore did not have to deal with this difficulty, however, this capability merely implies that another CDF be constructed from the individual isotopes' σ_t-values and then subsequently sampled to identify the appropriate cross-section set.

2.2.3 Step 3: Applying scattering

Scattering cross sections are always accompanied by some information regarding the angular distribution. For continuous energy, elastic scattering, cross sections this information is a probability density function (PDF) as a function of the scattering cosine in the center-of-mass (COM) reference frame[6], μ_C, at a specific energy. Since these PDFs are stored at a number of discrete energies, the PDF at the required energy must be obtained using interpolation between the discrete energies (with subsequent normalization). With a PDF established, a suitable CDF[7] is constructed and sampled with a pseudo-random number to obtain μ_C from which we can determine the scattering angle in the COM reference frame, $\theta_C = \cos^{-1}(\mu_C)$.

For elastic scattering reactions of neutrons (as an example) the scattering angle in the laboratory

[6]Ordinarily these PDFs are constant as a function of μ_C for most energies except high energies, where quantum mechanical effects dominate.

[7]Most production codes create a discrete CDF at a resolution equal to or greater than that stored for the PDF, using interpolated values of the PDF.

reference frame, θ_L, is related to the COM reference frame through the use of the atomic mass ratio of the target nucleus mass to the particle mass, A. This relation, adapted slightly from [2], is

$$\theta_L = \tan^{-1}\left(\frac{\sin\theta_C}{\frac{1}{A}+\mu_C}\right), \tag{2.15}$$

and allows us to subsequently compute the true scattering cosine, $\mu_L = \cos\theta_L$, about the current direction of flight, $\mathbf{\Omega}_i$.

We can also, at this point, determine the particles energy after the scattering reaction, E_f, which can be derived from particle kinematics or any relevant physics relations. For simplicity this relation is again adapted from [2] for neutrons, and is given by

$$E_f = \left[\frac{(1+\alpha)+(1-\alpha)\mu_C}{2}\right]E_i, \tag{2.16}$$

where $\alpha = \left(\frac{A-1}{A+1}\right)^2$. In production simulations, where the full compliment of complex interactions can occur the change in energy needs to consider many aspects. For neutrons the change in energy can involve inelastic scattering, thermal free-gas scattering, material motion related up-scattering, etc.

The final requirement to determine the particle's new direction is to establish the azimuthal angle, φ_L, of the scattering reaction relative to $\mathbf{\Omega}_i$. The probability distribution for φ_L is isotropic in $[0, 2\pi]$ for all physically known cases[8] and can therefore be sampled as

$$\varphi_L = 2\pi\theta_R. \tag{2.17}$$

We can now use these angles to rotate $\mathbf{\Omega}_i$ to the scattered direction, $\mathbf{\Omega}_f$, by defining two axes of rotation. The first axis of rotation assists us in emulating the polar angle of the scattering and since the azimuthal rotation is arbitrary we can use any direction for this axis as long as it is perpendicular

[8] If the cosine is known, φ_L will always be relative to the chosen laboratory reference frame, which inherently can be rotated arbitrarily about the current direction of flight without changing the cosine.

to $\mathbf{\Omega}_i$. Using the upward pointing vector, $\hat{k} = [0, 0, 1]$, we can define the polar rotation axis (or tangent vector), $\mathbf{t}$, as

$$\mathbf{t} = \begin{cases} \dfrac{\mathbf{\Omega}_i \times \hat{k}}{||\mathbf{\Omega}_i \times \hat{k}||}, & \mathbf{\Omega}_i \cdot \hat{k} < 1 - \epsilon \\ \hat{i}, & \mathbf{\Omega}_i \cdot \hat{k} > 1 - \epsilon \end{cases}, \tag{2.18}$$

where ϵ is an arbitrary tolerance such that $||\mathbf{\Omega}_i \times \hat{k}|| \neq 0$, and $\hat{i} = [1, 0, 0]$. The second axis of rotation is simply $\mathbf{\Omega}_i$ since, after the polar rotation is applied, the polar-rotated vector, $\mathbf{\Omega}_p$, needs to be rotated azimuthally. This process is depicted in Figure 2.4 below.

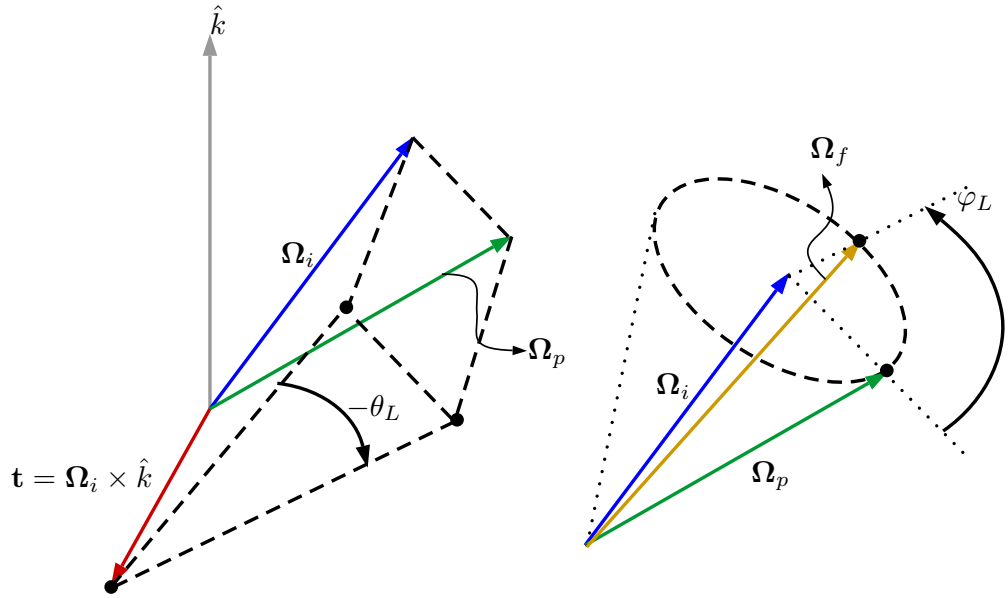

Figure 2.4: Schematic of the two rotations applied to $\mathbf{\Omega}_i$ when scattering through a cosine $\cos\theta_L$, applying first the polar rotation about $\mathbf{t}$ to produce $\mathbf{\Omega}_p$, then the random azimuthal rotation of $\mathbf{\Omega}_p$ by an angle φ_L about the original $\mathbf{\Omega}_i$.

We now seek a means to rotate any vector, $\mathbf{v}$, about an axis defined by a unit normal, $\mathbf{a}$, by a right hand-rule oriented angle, θ_R. From first principles we can derive a formula known as

Rodrigues' formula[15], which is

$$\begin{aligned}\mathbf{v}_{rotated} &= \cos\theta_R\ \mathbf{v} + (\mathbf{a}\cdot\mathbf{v})(1-\cos\theta_R)\ \mathbf{a} + \sin\theta_R\ (\mathbf{a}\times\mathbf{v}) \\ &= \text{RotationAboutAxis}(\mathbf{v},\mathbf{a},\theta_R)\end{aligned} \tag{2.19}$$

where $\mathbf{v}_{rotated}$ is the resulting rotated vector and "RotationAboutAxis" a function name. Now we can apply two rotations, firstly the polar rotation of $\mathbf{\Omega}_i$ about the axis $\mathbf{t}$ by an angle θ_L,

$$\mathbf{\Omega}_p = \text{RotationAboutAxis}(\mathbf{\Omega}_i, \mathbf{t}, -\theta_L), \tag{2.20}$$

and then the azimuthal rotation of $\mathbf{\Omega}_p$ about the axis $\mathbf{\Omega}_i$ by an angle φ_L,

$$\mathbf{\Omega}_f = \text{RotationAboutAxis}(\mathbf{\Omega}_p, \mathbf{\Omega}_i, \varphi_L). \tag{2.21}$$

2.3 Contributing to volumetric tallies

There are numerous ways to estimate quantities in a Monte Carlo simulation. Particles crossing a specific surface can be used to determine the average current or flux on a surface, this process involves so-called *current and surface-crossing estimators*. Collisions can be counted, a process called a *collision estimator*. Finally, the average flux within a volume can be determined using a *track-length estimator*.

For this research we have little need for the current and surface-crossing estimators since we do not require the *current or flux on surfaces*. The *collision estimator* will also not be used because it has an efficiency concern, i.e., the probability of a source particle contributing to the estimation is dependent on σ_t. This leaves just the *track-length estimator* where the scalar flux can be estimated directly. Additionally, the track-length estimator can be modified to incorporate particle weights (i.e., when applying biasing or variance reduction) and to provide quantities such as the spherical harmonic expansion coefficients for the angular flux and integrals weighted with finite element

basis functions.

To comprehend how a track-length estimator is used, we will now discuss how it used to estimated the scalar flux. The process is as follows: As one repeatedly applies the stochastic particle transport process above, the particles trace tracks across the domain each time it is transported from a position $\mathbf{x}_i$ to $\mathbf{x}_f$. At each instance, where this position pair is updated, we can contribute to the volumetric tally, within which the track took place, by computing a track-length, ℓ, as

$$\ell = ||\mathbf{x}_f - \mathbf{x}_i||_2. \tag{2.22}$$

From first principles we can derive that if each track-length, ℓ_i, is contributed to tally T then the average scalar flux in the tally, ϕ_T, can be estimated from

$$\phi_T = \frac{1}{V_T} \int_{V_T} \int_{4\pi} \psi(\mathbf{x}, \mathbf{\Omega}) \, d\mathbf{\Omega} \, dV \approx \frac{1}{N_p V_T} \sum_{i=1}^{N_\ell} \ell_i, \tag{2.23}$$

where N_p is the total number of particles simulated, N_ℓ is the total number of tracks traced within the tally's spatial domain, and V_T is the tally volume. Incorporating energy, requires the average track-length of all particles with energies within the group energy sub-domain $\mathcal{D}_{e,g}$, hence

$$\phi_{T,g} \approx \frac{1}{N_p V_T} \sum_{i=1}^{N_{\ell,g}} \ell_i, \tag{2.24}$$

is the group average flux, where $N_{\ell,g}$ is the total number of tracks traced, by particles of energy $E \in \mathcal{D}_{e,g}$, in the tally. This energy grouping used in tallies is required even if the particles are transported in continuum form (continuum energy, angle and space) since the probability, of a particle having an energy exactly at the energy required, is essentially infinitely small.

The final addition to the notion of a tally is the inclusion of a track-weight (or *particle weight*), w_i, such that

$$\phi_{T,g} \approx \frac{1}{N_p V_T} \sum_{i=1}^{N_\ell} w_i \ell_i. \tag{2.25}$$

Note here that weights can be used for biased sampling, variance reduction and for the computation of several other quantities related to the track-length, and does not necessarily change the value of a tally at the conclusion of a simulation but can greatly influence the statistical uncertainty.

2.4 Determining the statistical uncertainty of a tally

The mean tally value in Eq. (2.24) is subject to statistical uncertainty and, as mentioned in the introduction, the central limit theorem can be used to show that the statistical distribution of the sample mean is a normal distribution about the true mean[1, 18]. For purposes of this discussion we will consider a tally in abstract-form with individual contributions to the tally simply being samples denoted as x_i, the true mean, $\bar{x}$, and a sample mean, $\hat{x}$, being

$$\hat{x} = \frac{1}{N_c} \sum_{i=1}^{N_c} x_i. \tag{2.26}$$

where N_c is the number of contributions or sample size. The unbiased sample variance, S_x^2, is given by

$$S_x^2 = \frac{1}{N_c - 1} \sum_{i=1}^{N_c} (x_i - \hat{x})^2, \tag{2.27}$$

where the $\frac{1}{N_c - 1}$ term is often set to $\frac{1}{N_c}$ when N_c is large enough. This form closely follows the standard definition of the variance, however, it is not entirely practical to use because it requires the storage of all the x_i values. A more practical form can be obtained by multiplying out the square term in Eq. (2.27) and performing some manipulations (assuming a large sample size) to obtain

$$S_x^2 = \frac{1}{N_c} TQ_1 - \frac{1}{N_c^2} (TQ_2)^2, \tag{2.28}$$

where

$$TQ_1 = \sum_{i=1}^{N_c} x_i^2 \quad \text{and} \quad TQ_2 = \sum_{i=1}^{N_c} x_i.$$

This form therefore requires us to make contributions to both the tally quantities, TQ_1 and TQ_2 when a tally score is made (i.e., when a track is traced both $w\ell$ and $(w\ell)^2$ is contributed).

The final step is to relate the variance of the sample mean about the true mean, $S^2_{\hat{x}}$, to the sample variance, S^2_x. Fortunately, this can easily be shown from statistical theory[1, 18] to be

$$S^2_{\hat{x}} = \frac{1}{N_c} S^2_x, \tag{2.29}$$

which provides us with the final equation to provide us with the tally standard deviation,

$$S_{\hat{x}} = \sqrt{\frac{1}{N_c}\left(\frac{1}{N_c} TQ_1 - \frac{1}{N_c^2}(TQ_2)^2\right)}. \tag{2.30}$$

In this context a tally value should always be presented in the form $\hat{x} \pm S_{\hat{x}}$, which translates to our earlier definition of a flux tally as

$$\text{Tally result: } \phi_{T,g} \pm S_{\phi_{T,g}}$$

where

$$S_{\phi_{T,g}} = \frac{1}{N_p V_T}\sqrt{\frac{1}{N_p}\left(\frac{1}{N_p} TQ_{T,g,1} - \frac{1}{N_p^2}(TQ_{T,g,2})^2\right)} \tag{2.31}$$

and

$$TQ_{T,g,1} = \sum_{i=1}^{N_{\ell,g}} (w_i \ell_i)^2 \tag{2.32}$$

$$TQ_{T,g,2} = \sum_{i=1}^{N_{\ell,g}} w_i \ell_i. \tag{2.33}$$

Of special note, in the equations above, is that the accumulations are done over $N_{l,g}$ but the standard deviation is computed using N_p. This is because each source particle j, where $j \in [1, N_p]$, always theoretically contributes to all tallies in the problem but only the subset $i \in [1, N_{l,g}]$ contribute non-zero track-lengths.

2.5 Modifications to allow for multigroup cross sections

Methods that use multigroup cross sections often apply a reduced order model for the angular dependence of the scattering cross section. The most popular model for this is the Legendre expansion of $\sigma_{s,g'\to g}(\mathbf{\Omega}'\to\mathbf{\Omega})$, or $\sigma_{s,g'\to g}(\mu_L)$, at any position in space, where μ_L is the scattering cosine in the laboratory reference frame. The expansion is

$$\sigma_{s,g'\to g}(\mu_L) \approx \sigma^L_{s,g'\to g}(\mu_L) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi}\sigma_{s\ell,g'\to g}P_\ell(\mu_L), \tag{2.34}$$

where L is the expansion order (also referred to as the *scattering-order*), the P_ℓ functions are Legendre polynomials of order ℓ, and $\sigma_{s,g'\to g,\ell}$ are expansion coefficients defined as

$$\sigma_{s\ell,g'\to g} = \int_{-1}^{1} \sigma_{s,g'\to g}(\mu_L)P_\ell(\mu_L)d\mu_L. \tag{2.35}$$

These expansion coefficients are normally known or supplied as input for any method that uses it and greatly reduces the size of the data set required to store the angular distribution information. Additionally, the Legendre polynomials have advantageous properties with regards to spherical harmonic expansions of the angular flux (i.e., orthogonality, special identities, etc.), which are used by nearly all deterministic methods.

2.5.1 The energy change

Unlike the continuous energy scattering process, where the energy change is determined from the sampled cosine, multigroup cross section require us to use the isotropic scattering moments to build a CDF for each group g', denoted as $c_{g'}$, and with g-indexed elements as

$$c_{g'\to g} = \frac{\sum_{j=0}^{g} \sigma_{s,g'\to j,0}}{\sum_{j=0}^{N_G-1} \sigma_{s,g'\to j,0}}. \tag{2.36}$$

Given that a particle will be at energy group g', this CDF is then first sampled with a pseudo-random number to determine the destination energy group g, and with the combination of g' and g we can then construct the Legendre expansion in Eq. 2.34.

2.5.2 Sampling the scattering cosine

There is a problematic aspect to the sampling process related to a Legendre expansion of σ_s, which stems from the fact that the multigroup approximation changes the δ-function nature of continuous energy scattering kinematics into a shape, as a function of μ_L, that normally resembles a combination of a step and a ramp function. This shape is normally sufficiently well approximated by low orders of the Legendre expansion as shown in Figure 2.5, where the isotropic scattering and peak at the correct cosine value are well approximated. The problem with this form, as it relates to the Monte Carlo method, is that we cannot construct a CDF from this information because it results in a CDF as shown in Figure 2.6, which violates the required properties of a CDF (i.e., strictly positive and monotonically increasing).

Two solutions to this problem are generally known[18]. The first is the use of rejection sampling, and the second is the use of discrete cosines from a generalized Gauss-quadrature that preserves the Legendre expansion moments.

2.5.2.1 Rejection sampling of the Legendre expansion

Using rejecting sampling on the Legendre expansion is a conceptually simple process given that we can, at initialization-time, determine the maximum probability density, p_{max}. We can then sample a random cosine, $\mu_r \in [-1, 1]$ after which we sample a random probability density $p_r \in [0, p_{max}]$ and also evaluate the expansion at the sampled cosine, $\sigma^L_{s,g'\to g}(\mu_r)$. The sampled cosine is accepted if $p_r \leq |\sigma^L_{s,g'\to g}(\mu_r)|$. This sampling process has, however, a very low sampling efficiency since the area under the curve can be very small (see Figure 2.5) as compared to the rectangle created with p_{max}.

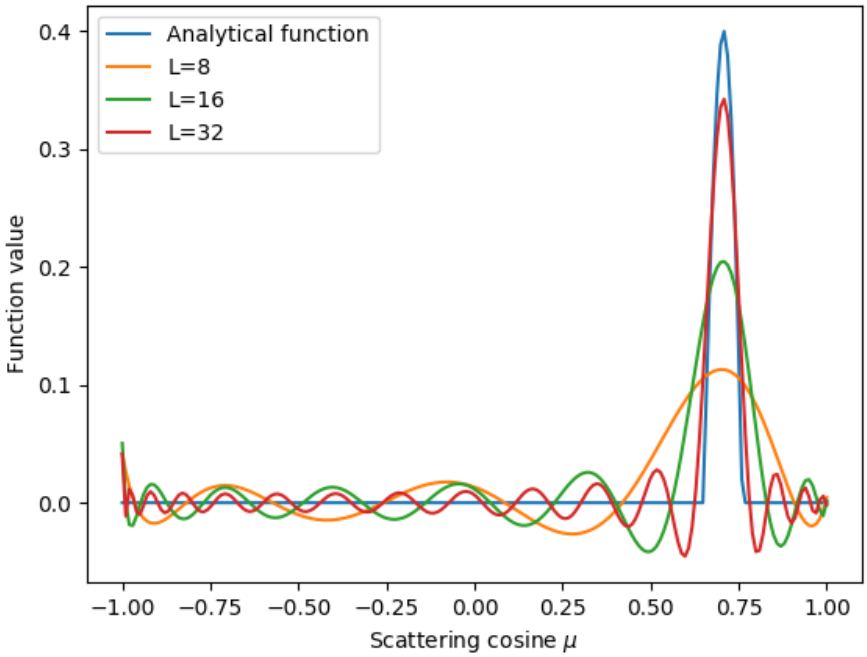

Figure 2.5: Legendre expansion of a test function for different values of L.

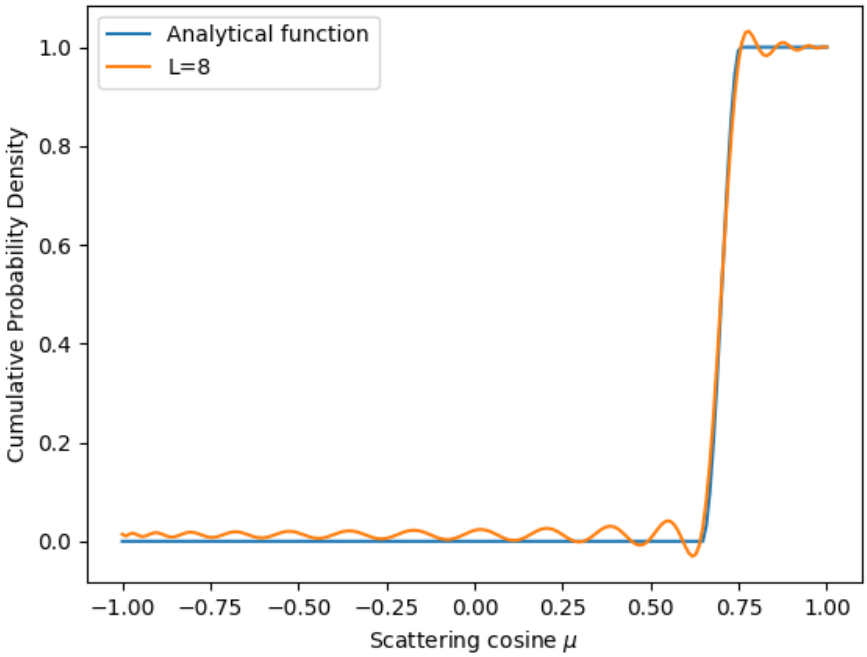

Figure 2.6: A problematic CDF constructed from a Legendre expansion.

Techniques are available to improve rejection sampling efficiency [18] which all involve using a bounding-function, $h(\mu)$, to $\sigma^L_{s,g'\to g}$ such that $h(\mu) \geq \sigma^L_{s,g'\to g}(\mu)$, either over the entire span of $\mu \in [-1, 1]$ or sub-intervals of this span.

Ultimately, rejection sampling is not as efficient as the next method we introduce, therefore we will not detail the intricacies of rejection sampling $\sigma^L_{s,g'\to g}$ further.

2.5.2.2 A generalized Gauss-quadrature from Legendre expansions of σ_s

A very good solution for the CDF-related problem of the Legendre expansions is to replace the expansion with a generalized Gauss-quadrature as was done by Sloan[16]. The quadrature replaces the continuous form of $\sigma^L_{s,g'\to g}$ with a discrete form,

$$\sigma^d_{s,g'\to g}(\mu_L) = \sum_{i=1}^{N_{GQ}} w_i^q \delta(\mu_L - \mu_{s,i}), \tag{2.37}$$

where N_{GQ} is the number of discrete cosines used in the quadrature, w_i^q are the weights (required to be positive), and $\mu_{s,i}$ are the discrete scattering cosines such that $-1 \leq \mu_{s,i} \leq 1$. Additionally the quadrature needs to be moment preserving, i.e.,

$$\sigma_{s,g'\to g,\ell} = \int_{-1}^{1} \sigma_{s,g'\to g}(\mu_L) P_\ell(\mu_L) d\mu_L = \sum_{i=1}^{N_{GQ}} w_i^q P_\ell(\mu_{s,i}). \tag{2.38}$$

The Modified Chebyshev Algorithm (MCA) as depicted by Golub and Fischer[17] provides a succinct means to compute the recurrence coefficients of an orthogonal polynomial (w.r.t. to the weight function $\sigma_{s,g'\to g}$), given the moments $\sigma_{s\ell,g'\to g}$, that is then used to compute the weights, w_i^q, and discrete cosines, $\mu_{s,i}$, in Eq. (2.38).

Since P_ℓ is a polynomial with a maximum order L, the quadrature rule will produce N_{GQ} quadrature points such that $2N_{GQ} - 1 = L$.

Figure 2.7 below shows an example application to an analytical function for $\sigma_{s,g'\to g}$. Since $L = 7$, only 4 quadrature points ($N_{GQ} = 4$) are generated at cosines that will allow the preservation

of the first $L + 1 = 8$ moments.

In order to build a discrete cosine CDF, c_{μ_L}, for the scattering cosine μ_L we simply have to use the cosines and weights in Eq. (2.37).

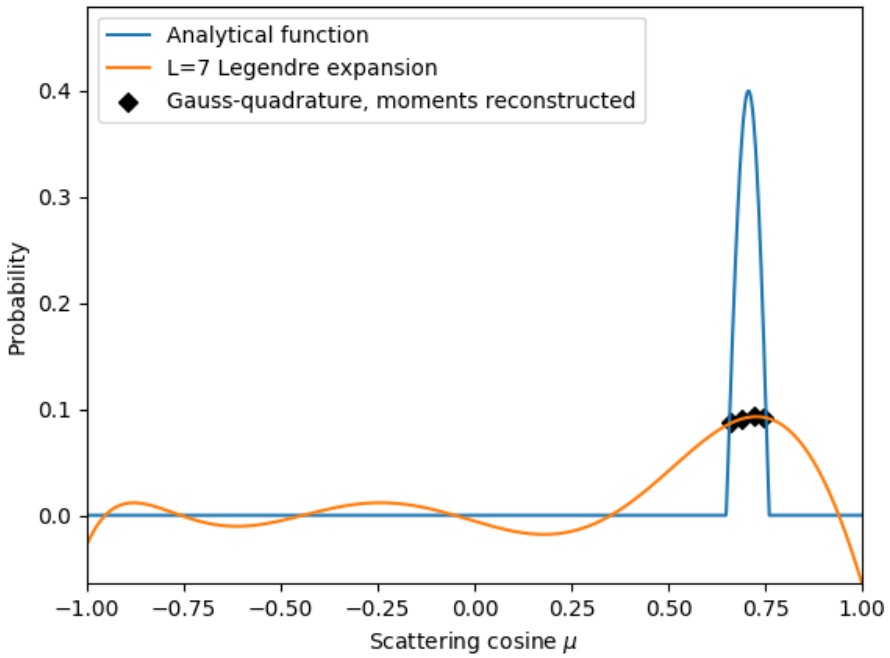

Figure 2.7: Example application of the generalized Gauss Quadrature.

2.6 Modifications to tallies to provide finite element representations

Using the finite element method for spatial discretization allows the scalar flux to be represented as the summation of discrete coefficients, ϕ_j, and associated basis functions, $b_j(\mathbf{x})$, in the form

$$\phi(\mathbf{x}) \approx \phi_h(\mathbf{x}) = \sum_{j=1}^{N_b} \phi_j b_j(\mathbf{x}), \tag{2.39}$$

where ϕ_h is the expanded flux and N_b is the number of basis functions. In order to find the ϕ_j coefficients we define N_b number of equations as volumetric integrals, weighted with basis function

b_i, such that

$$\int_V b_i(\mathbf{x})\phi_h(\mathbf{x})dV = \int_V b_i(\mathbf{x})\phi(\mathbf{x})dV, \tag{2.40}$$

after which we substitute in Eq. (2.39) for ϕ_h to get

$$\sum_{j=1}^{N_b} \phi_j \int_V b_i(\mathbf{x})b_j(\mathbf{x})dV = \int_V b_i(\mathbf{x})\phi(\mathbf{x})dV \tag{2.41}$$

which forms a system in the form $A\boldsymbol{\phi} = \mathbf{b}$ where the ϕ_j coefficients are the unknowns and

$$A_{ij} = \int_V b_i(\mathbf{x})b_j(\mathbf{x})dV, \text{ and} \tag{2.42}$$

$$b_i = \int_V b_i(\mathbf{x})\phi(\mathbf{x})dV. \tag{2.43}$$

where the b_i values are essentially weighted tallies.

This system can be fairly large and sparse if a continuous finite element method (CFEM) is used for discretization but it will be symmetric positive definite (SPD) and can therefore easily be solved with a preconditioned Conjugate Gradient method. When using a discontinuous finite element method (DFEM), however, the basis functions are only restricted to those on a given cell, the amount of which rarely exceeds ~20 (i.e., polyhedrons). Consequently, instead of one large system to solve, we have a small system for each cell that can easily be solved using Gaussian-elimination, although, in the DFEM case the A-matrix is not group or moment dependent and therefore it is more efficient to obtain the small inverse matrix, A^{-1}, via LU-decomposition, and reuse it for different $\mathbf{b}$ vectors on the same cell but different groups and moments.

The conversion of a tally to a DFEM representation requires first that we define tallies that allow the computation of the coefficients of the $\mathbf{b}$-vector. Such tallies are simply

$$\int_V b_i(\mathbf{x})\phi_{T,g}(\mathbf{x})dV \approx \frac{1}{N_p}\sum_{j=1}^{N_{\ell,g}} \tilde{b}_i w_j \ell_j, \tag{2.44}$$

and are largely the same as the volumetric tallies defined in Eq. (2.25), differing only in that there

is no division by the tally volume, and the addition of the average basis-function value, $\tilde{b}_j$.

These tallies can therefore be computed as part of a normal simulation after which, in a post-processing step, these tallies can be transformed to expansion coefficients by solving the individual system on each cell. Figure 2.8 below shows an example of how a finite element representation improves the visual fidelity of a solution.

The DFEM representation unfortunately has a performance-cost since the tally contributions involves the computation of the average basis function value. This computation in-turn also introduces a form of a weight, lower than the weight used to represent the cell-average (i.e., only a fraction of the volume is contributed), and therefore reduces the rate at which the statistical uncertainty of the DFEM tallies reduce (albeit still $1/\sqrt{N_p}$) as compared to the cell-averaged tallies.

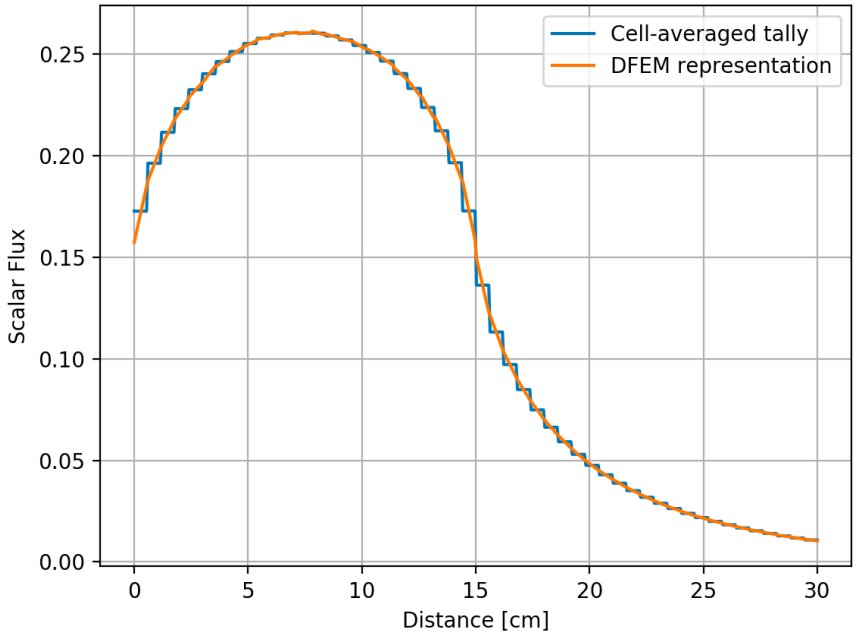

Figure 2.8: Example of a 1D slab simulation with an isotropic cell-source over the first half of the domain. The cell-averaged tallies and DFEM tallies are shown.

2.7 Modifications to tallies to provide spherical harmonic expansions of the angular flux

The angular flux, ψ, at a given energy (or group) and position, can be expanded into spherical harmonic bases as

$$\psi(\mathbf{\Omega}) \approx \psi^L(\mathbf{\Omega}) = \sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi} \sum_{m=-\ell}^{+\ell} \phi_{\ell m} Y_{\ell m}(\mathbf{\Omega}). \tag{2.45}$$

where L is the expansion order, the expansion coefficients, $\phi_{\ell,m}$, are

$$\phi_{\ell m} = \int_{4\pi} \psi(\mathbf{\Omega}) Y_{\ell m}(\mathbf{\Omega}) d\mathbf{\Omega}, \tag{2.46}$$

and the *tesseral spherical harmonics*, $Y_{\ell m}$, are given by

$$Y_{\ell m}(\mathbf{\Omega}) = Y_{\ell m}(\mathbf{\Omega} \mapsto (\theta,\varphi)) = \begin{cases} (-1)^m \sqrt{(2)} \sqrt{\frac{(\ell-|m|)!}{(\ell+|m|)!}} P_\ell^{|m|}(\cos\theta) \sin|m|\varphi & \text{if } m < 0 \\ P_\ell^m(cos\theta) & \text{if } m = 0 \\ (-1)^m \sqrt{(2)} \sqrt{\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) \cos m\varphi & \text{if } m > 0 \end{cases}. \tag{2.47}$$

where $\mathbf{\Omega}$ is mapped to (θ,φ) and the *associated Legendre polynomials*, P_ℓ^m, are defined as a recursion,

$$\begin{aligned} P_0^0 &= 1, \qquad P_1^0 = x, \\ P_\ell^\ell &= -(2\ell-1)\sqrt{1-x^2}\; P_{\ell-1}^{\ell-1}(x) \quad \text{and} \\ (\ell-m)P_\ell^m &= (2\ell-1)x\; P_{\ell-1}^m(x) - (\ell+m-1)P_{\ell-2}^m(x). \end{aligned} \tag{2.48}$$

The expansion coefficients, $\phi_{\ell m}$, can be computed in a flux tally by weighing with the corresponding spherical harmonic. Therefore the tally takes the form

$$\phi_{\ell m,T,g} = \int_{4\pi} \psi_g(\mathbf{x},\mathbf{\Omega}) Y_{\ell m}(\mathbf{\Omega}) d\mathbf{\Omega} \approx \frac{1}{N_p V_T} \sum_{i=1}^{N_{\ell,g}} w_i \ell_i Y_{\ell m}(\mathbf{\Omega}_{\ell,i}) \tag{2.49}$$

where the only new definition required is the direction vector of the i-th track, $\mathbf{\Omega}_{\ell,i}$.

In practice the spherical harmonic weights, which involve the recursive functions above, should never be computed at each tally contribution since this will require multiple unnecessary function

evaluations. Instead, the particle carries its harmonic values in the same way it carries its weight, and subsequently modifies them at each direction change. In this fashion the particle can potentially contribute these harmonic weights to many tallies before it needs to be recomputed.

This modification gains value in the fact that, in many deterministic methods, the angular flux is subject to discretization which can subsequently introduce ray-effects. By supplying the harmonic expansion coefficients, the Monte Carlo method can be used to supply the n-th collided-source by simulating particles only until they have collided n-times, allowing the deterministic method to apply iterative methods to convergence scattering sources.

2.8 The overall process

A summary of the Monte Carlo method is as follows:

1. Source sampling CDFs and quantities are initialized.
2. A loop is started for N_p number of source particles. Within this loop:
   (a) A source particle is sampled as per section 2.1.
   (b) The particle is stochastically transported as per section 2.2 until it is killed (either by being absorbed or by leaving the domain). During its transport journey it contributes to tally quantities TQ_1 and TQ_2 in Eq. (2.31).
   (c) Any banked secondary particles are simulated as per the previous item.
3. Tallies are normalized as per section 2.3 and their uncertainties are computed as per section 2.4.
4. DFEM transforms are applied as per section 2.6.

A good practice for applying the loop in step 2 is to split this loop into multiple batches of particles. After each batch is computed the uncertainty is computed and contributed to a tally fluctuation chart (TFC) which can be used to ascertain the efficiency of a tally.

2.9 Batches of particles

In many cases it is beneficial to subdivide the total simulation, which requires N_p number of source particles, into several smaller simulations each simulating $N_p^{batch,i} : i \in [0, N_{batches}-1]$ number of source particles, such that $N_p = \sum_i N_p^{batch,i}$. These smaller simulations are then referred to as *batches*, and the number of batches is $N_{batches}$.

Simulation-batches offer a multitude of benefits. The most important benefit is that the progression of tally-uncertainties can be tracked in what is known as a Tally Fluctuation Chart (TFC). This involves the application the equations to compute tally estimated standard deviations (the process in section 2.4) at the end of each batch, and storing these values for each batch. At the end of the simulation, after all batches have been simulated, the TFC for each tally comprises the estimated standard deviation for each batch, which can collectively be analyzed to ascertain whether their estimated average is of sufficient accuracy. One indicator, for example, for ascertaining whether N_p is sufficiently large, is to observe the $1/\sqrt{N}$ relationship described in section 2.4. Another important benefit, for parallel simulations, is that batches can be used as a point of synchronization.

Apart from the primary benefits, batched simulations are a means to better understand the performance of a given simulation even if it just provides an opportunity to output information, manipulate data, or to provide intervals at which problem restart-data[9] can be dumped to storage.

Batches require only a single modification to the overall process in section 2.8, being step 2, the loop over N_p. This loop is now split into a nested loop comprising an outer loop over $N_{batches}$, representing each batch, and an inner loop over $N_p^{batch,i}$.

[9]Problem restart-data comprises information about the problem and tallies that will allow the simulation to be stopped and restarted. This mostly involves the storage of the tally accumulators described in section 2.4.

3. INTRODUCTION TO THE DISCRETE ORDINATES METHOD

The Discrete Ordinates (DO) method for solving the multigroup LBE involves the application of an angular discretization as well a spatial discretization.

3.1 Angular discretization

3.1.1 Modification of the scattering source

The scattering source in Eq. 1.8 couples the multigroup LBE, for angle $\mathbf{\Omega}$, to all other angles in the angular phase-space, i.e., angles $\mathbf{\Omega}'$. In continuum form, we cannot derive a set of discrete equations for this coupling and therefore require some form of discretization.

If the angular domain was to be discretized, into N_d amount of discrete angles $\{\mathbf{\Omega}_n \ : \ n \in \{0,\ldots,N_d{-}1\}\}$, then one could theoretically develop scattering cross-sections for the entire set of possible discrete cosines $\{\mu_{s,nm} : \mu_{s,nm} = \mathbf{\Omega}_n \cdot \mathbf{\Omega}_m,\ n,m \in [0,\ldots,N_d{-}1]\}$. Such a scenario however, could require an insurmountable of storage since one would have to store the entire set of discrete angular-fluxes, for each energy group, at each spatial node of the problem. Additionally one would have to store $\sigma_{sg'\to g}$ at all of the discrete cosines, for all groups, and all materials.

To overcome this difficulty, it has long been customary[1] to use a *Legendre expansion* of $\sigma_{sg'\to g}$ as in Eq. 2.34 along with a *spherical harmonic expansion* of ψ_g as in Eq. 2.45. This effectively applies a reduced order model of the scattering source, where the expansions are truncated at a scattering-order, L, to get

$$\begin{aligned}
&\int_{4\pi} \sigma_{sg'\to g}(\mathbf{x},\mathbf{\Omega}'\cdot\mathbf{\Omega})\psi_{g'}(\mathbf{x},\mathbf{\Omega}')d\mathbf{\Omega}' \\
&\approx \int_{4\pi} \sigma^L_{sg'\to g}(\mathbf{x},\mathbf{\Omega}'\cdot\mathbf{\Omega})\psi^L_{g'}(\mathbf{x},\mathbf{\Omega}')d\mathbf{\Omega}' \\
&= \int_{4\pi} \left[\sum_{\ell=0}^{L} \frac{2\ell+1}{4\pi}\sigma_{s\ell,g'\to g}(\mathbf{x})P_\ell(\mathbf{\Omega}'\cdot\mathbf{\Omega})\right]\left[\sum_{\ell'=0}^{L}\frac{2\ell'+1}{4\pi}\sum_{m'=-\ell'}^{\ell'}\phi_{g'\ell'm'}(\mathbf{x})Y_{\ell'm'}(\mathbf{\Omega}')\right]d\mathbf{\Omega}'.
\end{aligned} \tag{3.1}$$

This term can then be simplified substantially using the favorable properties of the combination of the Legendre- and spherical harmonic expansions, the first of which is the *addition theorem of spherical harmonics*[2] which states

$$P_\ell(\mathbf{\Omega}' \cdot \mathbf{\Omega}) = \sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{\Omega}') Y_{\ell m}(\mathbf{\Omega}), \tag{3.2}$$

and the second is the *orthogonality of spherical harmonics* [2], which requires

$$\int_{4\pi} Y_{\ell' m'}(\mathbf{\Omega}') Y_{\ell m}(\mathbf{\Omega}') d\mathbf{\Omega}' = \frac{4\pi}{2\ell' + 1} \delta_{\ell\ell'} \delta_{mm'}, \tag{3.3}$$

where δ is the *Kronecker delta*. If equations (3.2) and (3.3) are substituted into Eq. (3.1) then the scattering source reduces to

$$\int_{4\pi} \sigma_{sg' \to g}(\mathbf{x}, \mathbf{\Omega}' \cdot \mathbf{\Omega}) \psi_{g'}(\mathbf{x}, \mathbf{\Omega}') d\mathbf{\Omega}' \approx \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \left[\frac{2\ell + 1}{4\pi} Y_{\ell,m}(\mathbf{\Omega}) \sigma_{s\ell,g' \to g}(\mathbf{x}) \phi_{g',\ell m}(\mathbf{x}) \right] \tag{3.4}$$

where

$$\phi_{g'\ell m}(\mathbf{x}) = \int_{4\pi} \psi_{g'}(\mathbf{x}, \mathbf{\Omega}) Y_{\ell m}(\mathbf{\Omega}) d\mathbf{\Omega}, \tag{3.5}$$

and is yet to be resolved.

With the scattering moments $\sigma_{s\ell,g' \to g}$ available as input, the only unknowns are the flux moments, $\phi_{g',\ell m}$, which are computed with Eq. (3.5). It is important to note here that the integral over the angular domain is not only important for the computation of the $\phi_{g',\ell m}$ coefficients but also for the computation of the scalar flux in Eq. (1.2). The scalar flux for group g', $\phi_{g'}$ is, however, the same as $\phi_{g'00}$ since $Y_{00}(\mathbf{\Omega})$=1 according to the definition of the spherical harmonics in Eq. (2.47).

We now turn our attention to a method for computing these angular integrals.

3.1.2 Application of an angular quadrature

In order to evaluate the angular integrals in Eq. (3.5), the discrete ordinates method uses an angular quadrature in the form

$$\int_{4\pi} f(\mathbf{\Omega}) d\mathbf{\Omega} \approx \sum_{n=0}^{N_d-1} w_n^Q f(\mathbf{\Omega}_n) \quad n = \{0, \ldots, N_d-1\} \tag{3.6}$$

where f is an arbitrary function of direction, $\mathbf{\Omega}_n$ are the quadrature points on the unit-sphere, N_d is the total number of quadrature points, and w_n^Q are the associated weights. There are several choices for angular quadratures, all with the requirement that it approximates the angular integral as close as possible. Unlike one dimensional polynomial quadratures, angular quadratures for the integrals in Eq. (3.5) are rarely able to provide the exact integral for practical problems mostly because the angular flux, $\psi_{g'}$, is unlikely to be a smooth-function of $\mathbf{\Omega}$.

The most commonly used quadrature in 1D is the Gauss-Legendre quadrature (with an even amount of quadrature points $N_d = 2N$). For 3D problems, quadratures can range from the classical Level Symmetric quadratures [1] ($N_d = N(N+2)$), product quadratures such as the Gauss-Legendre-Chebyshev quadrature, and many others, all with unique properties.

With a quadrature selected, Eq. (3.5) is evaluated, after substituting $f(\mathbf{\Omega}) = \psi_{g'}(\mathbf{x}, \mathbf{\Omega}) Y_{\ell m}(\mathbf{\Omega})$ and $\psi_{ng'} = \psi_{g'}(\mathbf{\Omega}_n)$, as

$$\phi_{g'\ell m}(\mathbf{x}) \approx \sum_{n=0}^{N_d-1} w_n^Q \psi_{g'n} Y_{\ell m}(\mathbf{\Omega}_n) \tag{3.7}$$

where the unknowns are still $\psi_{g'n}$ at the ordinate direction n.

To find these unknowns the discrete ordinates method then requires N_d amount of multigroup equations, which can be defined from Eq. (1.8) by first substituting all $\mathbf{\Omega}$ for $\mathbf{\Omega}_n$ followed by

substituting the scattering source with Eq. (3.4), to get

$$\begin{aligned}\boldsymbol{\Omega}_n \cdot \boldsymbol{\nabla}\psi_{gn}(\mathbf{x}) + \sigma_{tg}(\mathbf{x})\psi_{gn}(\mathbf{x}) = &\sum_{g'=0}^{N_G-1}\sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\left[\frac{2\ell+1}{4\pi}Y_{\ell,m}(\boldsymbol{\Omega})\sigma_{s\ell,g'\to g}(\mathbf{x})\phi_{g'\ell m}(\mathbf{x})\right] \\ &+\ q_g(\mathbf{x},\boldsymbol{\Omega}_n), \\ &\mathbf{x}\in\mathcal{D}_s, g\in\{0,\ldots,N_G{-}1\}, n\in\{0,\ldots,N_d{-}1\}.\end{aligned} \tag{3.8}$$

At this point it will also be convenient to replace the double indexed harmonics, $Y_{\ell m}$, with a single indexed harmonic, Y_h, such that $h \mapsto (\ell, m)$ and $h \in \{0, \ldots, N_h{-}1\}$, with N_h the total number of harmonics. We then also define an equivalent of Eq. (3.5),

$$\phi_{g'h}(\mathbf{x}) \approx \sum_{n=0}^{N_d-1} w_n^Q \psi_{g'n} Y_h(\boldsymbol{\Omega}_n), \tag{3.9}$$

to get a simplified equation

$$\Big(\boldsymbol{\Omega}_n \cdot \boldsymbol{\nabla} + \sigma_{tg}(\mathbf{x})\Big)\psi_{gn}(\mathbf{x}) = \sum_{g'=0}^{N_G-1}\sum_{h=0}^{N_h-1}\Big[w_h Y_h(\boldsymbol{\Omega}_n)\sigma_{sh,g'\to g}(\mathbf{x})\phi_{g'h}(\mathbf{x})\Big] \ +\ q_g(\mathbf{x},\boldsymbol{\Omega}_n), \tag{3.10}$$

where $w_h(h \mapsto (\ell, m)) = \dfrac{2\ell+1}{4\pi}$, and all other quantities with subscript h are the equivalent quantities with subscript (ℓm).

This form of the multigroup LBE will henceforth be referred to as the *discrete ordinates equations*.

3.2 Spatial discretization

There are many possible spatial discretizations that can be applied to Eq. (3.10). For this research we will mainly focus on linear DFEM on 1D cells and the piecewise-linear DFEM [20, 21, 22] as a generic spatial discretization in 2D and 3D, where the definition of the basis functions, $b_i(\mathbf{x})$ (i being the node index), are detailed in appendix B. The DFEM-representation of unknowns utilize $N_{b,c}$ number of basis functions or nodes for each cell c, which we can use to develop a

spatially discretized form of Eq. 3.10.

3.2.1 A DFEM representation of a continuum-variable

A DFEM representation of the continuum-form of ψ_{gn} is performed by expanding these variables into a weighted sum of the DFEM basis functions, which are specific to cell c, i.e.,

$$\Big(\psi_{gn}(\mathbf{x})\Big)_{\mathcal{D}_c} \approx \sum_{j=0}^{N_{b,c}-1} \psi_{gncj} b_j(\mathbf{x}), \text{ and} \tag{3.11}$$

where $\mathbf{x} \in \mathcal{D}_c$, $\mathcal{D}_c$ is the volumetric domain of cell c, and the weights ψ_{gncj} are now the discrete unknowns.

For the remainder of this text, any symbol for group-, direction-, cell-, and node-indices can be used to identify a discrete unknown as long as that index denotes the appropriate category, i.e., i is sometimes also used for the node-index, g' is sometimes used for the group-index, etc.

The order of the indices is also used to denote a sequence of entries, helpful when defining vector- or matrix nesting-structures.

3.2.2 The weak form of the discrete ordinates equations

With a discrete representation of variables defined we now turn our attention to discretizing the discrete ordinates equations. In order to simplify the mathematical formulation we will suppress the group- and direction-indices, g and n respectively.

We will also combine the scattering source and the external source into a single, space-dependent source, q, to get

$$\mathbf{\Omega} \cdot \boldsymbol{\nabla} \psi(\mathbf{x}) + \sigma_t(\mathbf{x}) \psi(\mathbf{x}) = q(\mathbf{x}), \tag{3.12}$$

representing the most basic form for our discussions.

We now seek to develop the Galerkin DFEM weak-form of this equation. We do this by first weighting this equation over a trial space, $\mathcal{D}_c$, with the space and weight defined by a basis function

of the piecewise-linear DFEM representation of cell c, $b_i(\mathbf{x})$ where $i \in \{0, \ldots, N_{b,c}-1\}$, for which we test

$$\int_{\mathcal{D}_c} b_i(\mathbf{x}) \Big(\mathbf{\Omega} \cdot \boldsymbol{\nabla} \psi(\mathbf{x}) + \sigma_t(\mathbf{x}) \psi(\mathbf{x}) \Big) dV = \int_{\mathcal{D}_c} b_i(\mathbf{x}) q(\mathbf{x}) dV. \tag{3.13}$$

Since the DFEM basis functions are defined on cells, the volume integrals in this equation are also the integrals over cells, i.e., in this case, cell c out of a total number of N_{cells} cells. We now take the gradient-term and apply integration by parts as well as a subsequent application of Gauss' divergence theorem, to get

$$\begin{aligned} \int_{\partial \mathcal{D}_c} \mathbf{\Omega} \cdot \mathbf{n} \, b_i(\mathbf{x}) \psi(\mathbf{x}) dA - \int_{\mathcal{D}_c} \mathbf{\Omega} \cdot (\boldsymbol{\nabla} b_i(\mathbf{x})) \psi(\mathbf{x}) dV \\ + \int_{\mathcal{D}_c} b_i(\mathbf{x}) \sigma_t(\mathbf{x}) \psi(\mathbf{x}) dV = \int_{\mathcal{D}_c} b_i(\mathbf{x}) q(\mathbf{x}) dV, \end{aligned} \tag{3.14}$$

where $\mathbf{n}$ is the position dependent surface normal and $\partial \mathcal{D}_c$ is the surface of cell c.

We next apply an upwinding scheme to the surface-integral term, in which we replace ψ with an upwinded $\tilde{\psi}$, such that

$$\tilde{\psi}(\mathbf{x}) = \begin{cases} \psi(\mathbf{x}), & \text{if} \quad \mathbf{\Omega} \cdot \mathbf{n} > 0 \\ \psi_{us}(\mathbf{x}), & \text{if} \quad \mathbf{\Omega} \cdot \mathbf{n} < 0, \end{cases} \tag{3.15}$$

where ψ_{us} is the angular flux, at the same direction, either at the boundary, or in the cell adjacent to cell c at the surface. The DFEM weak-form is finally,

$$\begin{aligned} \int_{\partial \mathcal{D}_c} \mathbf{\Omega} \cdot \mathbf{n} \, b_i(\mathbf{x}) \tilde{\psi}(\mathbf{x}) dA - \int_{\mathcal{D}_c} \mathbf{\Omega} \cdot (\boldsymbol{\nabla} b_i(\mathbf{x})) \psi(\mathbf{x}) dV \\ + \int_{\mathcal{D}_c} b_i(\mathbf{x}) \sigma_t(\mathbf{x}) \psi(\mathbf{x}) dV = \int_{\mathcal{D}_c} b_i(\mathbf{x}) q(\mathbf{x}) dV. \end{aligned} \tag{3.16}$$

3.2.3 A cell-by-cell system to solve

With the group-index, g, and direction-index, n, still suppressed the final step of discretization is now to approximate both ψ and q as expansions of the DFEM basis functions, i.e.,

$$\psi(\mathbf{x}) \approx \sum_{j=0}^{N_{b,c}-1} \psi_j b_j(\mathbf{x}), \text{ and} \tag{3.17}$$

$$q(\mathbf{x}) \approx \sum_{j=0}^{N_{b,c}-1} q_j b_j(\mathbf{x}), \tag{3.18}$$

where ψ_j and q_j are unknown constants. These approximations are then inserted into Eq. (3.16) to obtain, after rearrangement,

$$\begin{aligned}
&\sum_{j=0}^{N_{b,c}-1} \left[\int_{\substack{\partial\mathcal{D}_c \\ \mathbf{\Omega}\cdot\mathbf{n}>0}} \mathbf{\Omega}\cdot\mathbf{n}\, b_i(\mathbf{x}) b_j(\mathbf{x}) dA - \int_{\mathcal{D}_c} \mathbf{\Omega}\cdot(\boldsymbol{\nabla} b_i(\mathbf{x})) b_j(\mathbf{x}) dV + \int_{\mathcal{D}_c} b_i(\mathbf{x})\sigma_t(\mathbf{x}) b_j(\mathbf{x}) dV \right] \psi_j \\
&+ \sum_{j=0}^{N_{b,c}-1} \left[\int_{\substack{\partial\mathcal{D}_c \\ \mathbf{\Omega}\cdot\mathbf{n}<0}} \mathbf{\Omega}\cdot\mathbf{n}\, b_i(\mathbf{x}) \psi_{us,j}\, b_j(\mathbf{x}) dA, \right] = \sum_{j=0}^{N_{b,c}-1} \left[\int_{\mathcal{D}_c} b_i(\mathbf{x}) b_j(\mathbf{x}) dV \right] q_j.
\end{aligned} \tag{3.19}$$

For the piecewise-linear basis functions of interest, and their associated cells, we can segregate the surface of a cell into $N_{f,c}$ number of faces, each with an average normal, $\mathbf{n}_{f,c}$, which allows us to redefine the surface-integrals above into a summation of face-based surface-integrals (with $\partial\mathcal{D}_{c,f}$ being the surface of the face). The face terms are then integrals of the basis-functions, $b_i b_j$, over the face-area, taking the form of a *face mass-matrix* for all values of i and j, and is denoted with $M_{c,f}$. Additionally, σ_t is defined as constant over a cell (by manner of convention in this research), and since the volume-integrals are only over the volume of the cell, the integral term containing σ_t can be written as a product of σ_t and a *cell mass-matrix*, M_c. The set of equations

can then be written, in matrix- and tensor-notation, as

$$\left[\sum_{\substack{f=0\\ \mathbf{\Omega}\cdot\mathbf{n}_{f,c}>0}}^{N_{f,c}-1}\left(\mathbf{\Omega}\cdot\mathbf{n}_{f,c}\right)M_{c,f}-\mathbf{\Omega}\cdot\{G\}_c+\sigma_t M_c\right]\boldsymbol{\psi}_c+\sum_{\substack{f=0\\ \mathbf{\Omega}\cdot\mathbf{n}_{f,c}<0}}^{N_{f,c}-1}\left(\mathbf{\Omega}\cdot\mathbf{n}_{f,c}\right)M_{c,f}\boldsymbol{\psi}_{us,cf}=M_c\mathbf{q}_c \tag{3.20}$$

where the notation $\{\}$ denotes a tensor, and

$$(M_{c,f})_{ij}=\int_{\partial\mathcal{D}_{c,f}}b_i(\mathbf{x})b_j(\mathbf{x})dA, \tag{3.21}$$

$$(M_c)_{ij}=\int_{\mathcal{D}_c}b_i(\mathbf{x})b_j(\mathbf{x})dV. \tag{3.22}$$

Also, $\{G\}_c$ is a rank 3 tensor, with its first two indices denoting

$$(\{G\}_c)_{ij}=\int_{\mathcal{D}_c}(\boldsymbol{\nabla}b_i(\mathbf{x}))b_j(\mathbf{x})dV, \tag{3.23}$$

which we will refer to as the *gradient tensor*. The vectors $\boldsymbol{\psi}_c$, $\boldsymbol{\psi}_{us,cf}$, and $\mathbf{q}_c$ are all vectors of general form $\mathbf{v}$ with entries $(\mathbf{v})_i = v_i$, at the nodes $i \in \{0,\ldots,N_{b,c}-1\}$ of the cell's DFEM-representation. The vector $\boldsymbol{\psi}_{us,cf}$ has scope only to face f and may have many zero entries, for example, Figure 3.1 shows a dependency for cell 1 at face f as $\boldsymbol{\psi}_0 = [\psi_{00},\ldots,\psi_{05}]^T$, $\boldsymbol{\psi}_1 = [\psi_{10},\ldots,\psi_{15}]^T$ and $\boldsymbol{\psi}_{us,1f} = [0,0,0,0,\psi_{02},\psi_{01}]^T$, the latter indicating $\psi_{us,1f4}\mapsto\psi_{02}$ and $\psi_{us,1f5}\mapsto\psi_{01}$.

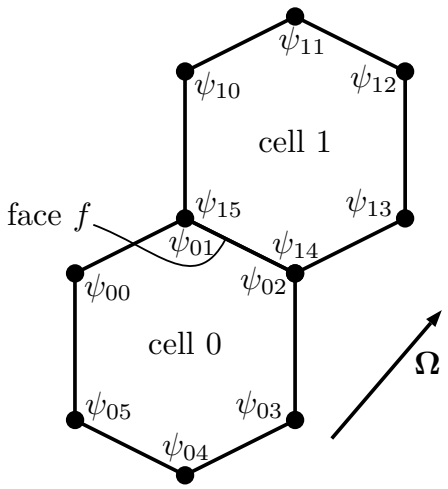

Figure 3.1: Example cell dependency for 2D polygons where cell 1 is dependent on cell 0 at face f.

With all of the entries q_i and $\psi_{us,i}$ known, these equations describe a linear system of equations in the form

$$A_{local,c}\boldsymbol{\psi}_c = \mathbf{b}_{local,c}, \tag{3.24}$$

where

$$A_{local,c} = \left[\sum_{\substack{f=0 \\ \boldsymbol{\Omega}\cdot\mathbf{n}_{f,c}>0}}^{N_{f,c}-1} (\boldsymbol{\Omega}\cdot\mathbf{n}_{f,c})\, M_{c,f} - \boldsymbol{\Omega}\cdot\{G\}_c + \sigma_t M_c \right], \text{ and} \tag{3.25}$$

$$\mathbf{b}_{local,c} = M_c\mathbf{q}_c - \sum_{\substack{f=0 \\ \boldsymbol{\Omega}\cdot\mathbf{n}_{f,c}<0}}^{N_{f,c}-1} L_{local,cf}\, \boldsymbol{\psi}_{us,cf}. \tag{3.26}$$

$$L_{local,cf} = (\boldsymbol{\Omega}\cdot\mathbf{n}_{f,c})\, M_c \tag{3.27}$$

This system is often small and $A_{local,c}$ is dense. The optimal choice of solution is *Gauss-Elimination* (even though it scales as $\mathcal{O}(N_{b,c}^3)$) since it is much more economical than numerical solvers for this system's typical size (rarely $N_{b,c} > 20$).

3.3 The cell-by-cell system in a global-system perspective

The upstream values, $\boldsymbol{\psi}_{us,f}$, for a given cell can potentially always be known if the sequence in which we solve the cell-by-cell systems strictly follows the dependency between cells. If one develops a mapping $c \mapsto k$, mapping cell-index c to k in a manner that allows cell k to be dependent only on cells with indices k', such that $k' < k$, then the cell-sequence $k = \{0, \ldots, N_{cells}-1\}$ represents a *topological sorting* of the cells for a particular direction, $\boldsymbol{\Omega}$.

Under some circumstances, such as certain types of unstructured meshes, cyclic dependencies between cells can exist which does not allow for the creation of a topological sorting. Such cases require specialized iterative methods to enable some resemblance of such a sorting. An example of such a method is detailed in previous work by the author[23], however, for the discussions in this text we will assume that no cyclic dependencies are present.

Along with the cell-index remapping, we can also define the mapping $(i, j) \mapsto (i_G, j_G)$, which maps the cell-local matrix indices (i, j), of matrix A_{local}, to a global system matrix, A_{global}, incorporating all cells, with indices (i_G, j_G). In order to develop this mapping we require an underlying global structure, for which we first define the total number of unknowns for a particular n and g as

$$N_{nodes} = \sum_{c=0}^{N_{cells}-1} N_{b,c}. \tag{3.28}$$

We then construct a vector, $\boldsymbol{\psi}_{gn}$, for all the cells and their associated nodes, at a particular direction n and group g, with definition

$$\boldsymbol{\psi}_{gn} = \{\psi_{gnki} \ : \ \forall k, i \in [0, N_{b,k}-1]\} \tag{3.29}$$

with discrete dimension $(N_{nodes} \times 1)$. With the underlying structure now defined we can create the mapping $i \mapsto i_G$ (which is the same as $j \mapsto j_G$), for each cell, allowing us to write

$$A_{global,gnk}\boldsymbol{\psi}_{gn} + \sum_{\substack{f=0 \\ \boldsymbol{\Omega}\cdot\mathbf{n}_{f,k}<0}}^{N_{f,k}-1} L_{global,gnkf}\ \boldsymbol{\psi}_{gn} = \mathbf{b}_{global,gnk}, \tag{3.30}$$

where $A_{global,gnk}$ and $L_{global,gnkf}$ are sparse matrices with the same square-size as $\boldsymbol{\psi}_{gn}$, i.e., $(N_{nodes} \times N_{nodes})$, and its entries are those in $A_{local,k}$ and $L_{local,kf}$ respectively (at the respective g and n), remapped using $(i,j) \mapsto (i_G, j_G)$. Special care is taken in mapping $L_{local,kf}$ since the cell-node-mapping needs to be unto the cell-node-pair the particular node depends on, which is often not the same cell. The vector $\mathbf{b}_{global,gnk}$ is the same size as $\boldsymbol{\psi}_{gn}$ with entries mapped from $M_{t,k}\mathbf{q}_k$ (again at the respective g and n).

With these definitions in place and $\boldsymbol{\psi}_{gn}$ common between all cells we can define the global system at direction n and group g as

$$L_{gn}\boldsymbol{\psi}_{gn} = \mathbf{b}_{gn} \tag{3.31}$$

where

$$\begin{aligned} L_{gn} &= \sum_{k=0}^{N_{cells}-1} \left[A_{global,gnk} + \sum_{\substack{f=0 \\ \boldsymbol{\Omega}\cdot\mathbf{n}_{f,k}<0}}^{N_{f,k}-1} L_{global,gnkf} \right] \\ \mathbf{b}_{gn} &= \sum_{k=0}^{N_{cells}-1} \mathbf{b}_{global,gnk}. \end{aligned} \tag{3.32}$$

The matrix L_{gn} is a block lower-triangular matrix with the $A_{global,gnk}$-entries on diagonal block k and the $L_{global,gnkf}$ entries all to the left of diagonal block k.

3.4 The system solution technique known as a "sweep"

In order to further comprehend the structure of L_{gn}, consider Figure 3.2 showing the topological sorting of a 3×3 mesh for direction $\boldsymbol{\Omega}$.

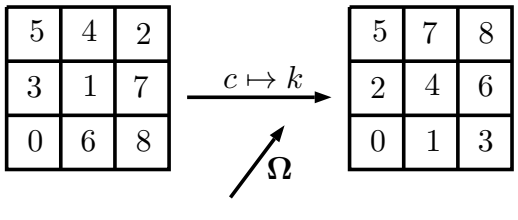

Figure 3.2: Example topological sorting of a 3×3 mesh.

For this discussion only, we will use a shorter notation for $A_{local,k}$ and $L_{local,kf}$ as just A_k and L_{kf}, respectively. The block structure of L_{gn}, using this notation, is then

$$L_{gn} = \begin{bmatrix} A_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ L_{13} & A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ L_{20} & 0 & A_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_{33} & 0 & A_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_{40} & L_{43} & 0 & A_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & L_{50} & 0 & 0 & A_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & L_{60} & L_{63} & 0 & A_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & L_{70} & L_{73} & 0 & A_7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & L_{80} & L_{83} & A_8 \end{bmatrix}$$

which allows one to solve a block-row at a time starting at the top row and sweeping downward. This sweeping action is the origin of the term "sweep" used to describe the action of inverting lower-triangular matrices.

The speed and efficiency of sweeping algorithms in parallel simulations has undergone much progress in the last two decades. The first parallel implementations applied the so-called KBA-partitioning, originating from the method described in [24], where the mesh is partitioned into subsets along a $P_x \times P_y$ overlaid grid, P_i being the number of processors used on the i-th axis. The KBA-method has since been extended to $P_x \times P_y \times P_z$ overlaid grids and was the subject of many theoretical studies [25, 26, 27, 28] and performance analysis. The extended KBA-method has also been demonstrated on semi-structured meshes [28] and later fully unstructured meshes, by cutting the mesh to preserve the orthogonal $P_x \times P_y \times P_z$ sub-domains [23]. Other partitioning schemes such as ParMETIS[29], has also been studied in [23], allowing a more versatile partitioning common to other physics-fields. The latter however, introduces cyclic dependencies in parallel and therefore requires special sweep algorithms.

3.5 Combining all angles and groups

The vectors $\boldsymbol{\psi}_{gn}$ and $\mathbf{b}_{gn}$, and associated matrix L_{gn}-matrix, thus far only encompass a single direction and group. In order to account for all the other directions and groups we abstractly define

a combination of these elements to eventually obtain the operator form of the discrete ordinates method.

We start by grouping together all the directions for group g into vectors $\boldsymbol{\psi}_g$ and $\mathbf{b}_g$ such that

$$\boldsymbol{\psi}_g = \{\psi_{gnki} \ : \ \forall n, k, i \in [0, N_{b,k}-1]\}, \tag{3.33}$$

$$\mathbf{b}_g = \{b_{gnki} \ : \ \forall n, k, i \in [0, N_{b,k}-1]\}, \tag{3.34}$$

which can be regarded as nested vectors, $\boldsymbol{\psi}_g = \{\boldsymbol{\psi}_{gn} \ : \ \forall n\}$ and $\mathbf{b}_g = \{\mathbf{b}_{gn} \ : \ \forall n\}$. We then define the matrix L_g, as a nested matrix containing elements L_{gn} on its diagonal, such that

$$L_g = \text{diag}(\{L_{gn} \ : \ \forall n\}). \tag{3.35}$$

Next we combine all groups together into vectors $\boldsymbol{\psi}$ and $\mathbf{b}$ such that

$$\boldsymbol{\psi} = \{\psi_{gnki} \ : \ \forall g, n, k, i \in [0, N_{b,k}-1]\}, \tag{3.36}$$

$$\mathbf{b} = \{b_{gnki} \ : \ \forall g, n, k, i \in [0, N_{b,k}-1]\}, \tag{3.37}$$

which can again be regarded as nested vectors, $\boldsymbol{\psi} = \{\boldsymbol{\psi}_g \ : \ \forall g\}$ and $\mathbf{b} = \{\mathbf{b}_g \ : \ \forall g\}$. We follow the same principal for a group combined matrix L, as a nested matrix containing elements L_g on its diagonal, such that

$$L = \text{diag}(\{L_g \ : \ \forall g\}). \tag{3.38}$$

The resulting system, is then finally

$$L\boldsymbol{\psi} = \mathbf{b}, \tag{3.39}$$

where L is known as the *transport operator*, the inversion of which requires a series of operations

on the operators contained within L to get

$$\psi = L^{-1}\mathbf{b} \tag{3.40}$$

Since the first two nested levels of L are block diagonal matrices, they are trivially inverted, however, the third nested level contains the discrete entries requiring Gaussian-elimination, on a cell-by-cell system, in a sweeping fashion. For a known $\mathbf{b}$-vector a sweep-operation inverts the transport operator exactly (with no iteration).

3.6 Operator form of the scattering and external source

Starting with the spatial discretization of the discrete ordinates equations we grouped the scattering source and external source into a single source. This is an important step since all iterative algorithms require this form to invert the transport operator L. We now discuss how the vector $\mathbf{b}$ is formed.

3.6.1 The discrete-to-moment operator

From Eq. (3.9) we note that nodal values for ϕ_{gh} can be represented as a quadrature formula based on the nodal values of ψ_{ng}. Therefore, the nodal flux moment at harmonic-index h, for cell k, at the node local-index i, can be computed from

$$\phi_{ghki} = \sum_{n=0}^{N_d-1} w_n^Q Y_h(\mathbf{\Omega}_n)\psi_{gnki}. \tag{3.41}$$

This formulation can be altered into a more succinct form as follows. Suppose we define a flux-moment vector at a given (g,k,i)-set, $\boldsymbol{\phi}_{gki} = \{\phi_{ghki} \ : \ \forall h\}$, and an angular-flux vector at a given (g,k,i)-set, $\boldsymbol{\psi}_{gki} = \{\psi_{gnki} \ : \ \forall n\}$, then $\boldsymbol{\psi}_{gki} \mapsto \boldsymbol{\phi}_{gki}$ as

$$\boldsymbol{\phi}_{gki} = D^{nodal}\boldsymbol{\psi}_{gki}, \tag{3.42}$$

where D^{nodal} is a matrix with dimension $(N_h \times N_d)$ with entries

$$D_{hn}^{nodal} = w_n^Q Y_h(\mathbf{\Omega}_n). \tag{3.43}$$

We now require a means to extend this formalism of D^{nodal} to a global operator D, such that $\boldsymbol{\phi} = D\boldsymbol{\psi}$. In function form we can define

$$D\boldsymbol{\psi} = F_D(\boldsymbol{\psi} \mapsto \boldsymbol{\phi}) \ : \ \phi_{ghki} = \sum_{n=0}^{N_d-1} D_{hn}^{nodal} \psi_{gnki}, \tag{3.44}$$

where F_D is the function-analog for $D\boldsymbol{\psi}$. We now call D the *discrete-to-moment* operator since it maps a set of discrete angular fluxes onto a set of discrete flux moments.

3.6.2 The moment-to-discrete operator

Using Eq. 2.45, when given any flux moment vector, $\boldsymbol{\phi}_{gki}$, for a given group g, cell k and node-index i, we can compute the angular flux vector, $\boldsymbol{\psi}_{gki}$, as

$$\boldsymbol{\psi}_{gki} = M^{nodal} \boldsymbol{\phi}_{gki}, \tag{3.45}$$

where M^{nodal} is a matrix with dimension $(N_d \times N_h)$ with entries

$$M_{nh}^{nodal} = w_h Y_h(\mathbf{\Omega}_n), \tag{3.46}$$

where $w_h(h \mapsto (\ell, m)) = \dfrac{2\ell+1}{4\pi}$. We now also extend this to a global operator M, such that $\boldsymbol{\psi} = M\boldsymbol{\phi}$. In function form we define

$$M\boldsymbol{\phi} = F_M(\boldsymbol{\phi} \mapsto \boldsymbol{\psi}) \ : \ \psi_{gnki} = \sum_{h=0}^{N_h-1} M_{nh}^{nodal} \phi_{ghki}, \tag{3.47}$$

where F_M is the function-analog for $M\boldsymbol{\phi}$. We now call M the *moment-to-discrete* operator since it maps a set of discrete flux-moments to a set of discrete angular fluxes.

3.6.3 The scattering operator

The scattering operator is conceptually formed just from the scattering moments, $\sigma_{sh,g'\to g}$. It maps one vector of flux moments to another. In function form we define

$$S\boldsymbol{\phi} = F_S(\boldsymbol{\phi}' \mapsto \boldsymbol{\phi}) \ : \ \phi'_{ghki} = \sum_{g'=0}^{N_G-1} \sigma_{s,g'\to g,hk}\phi_{g'hki}, \tag{3.48}$$

where F_S is the function-analog for $S\boldsymbol{\phi}$. We now call S the scattering operator.

3.6.4 The operator-form of the scattering- and external-source

At the start of the application of spatial discretization we combined the scattering- and external-source into a single quantity. This quantity is essentially the right-hand side of Eq. 3.10, and if expanded to include all cells, directions and groups can now be written as

$$\mathbf{b} = MSD\boldsymbol{\psi} + \mathbf{q}, \tag{3.49}$$

where $\mathbf{q} = \{q_{gnk} \ : \ \forall g, n, k\}$ and $q_{gnk} = q_g(\mathbf{x}, \boldsymbol{\Omega}_n) \ : \ \mathbf{x} \in \mathcal{D}_k$, but since, according to convention, the external sources are constant within a cell, $q_{gnk} = q_g(\boldsymbol{\Omega}_n)$ for cell k.

3.7 The complete operator-form of the discrete ordinates equations

Combining the operator forms in sections 3.5 and 3.6.4, we get the complete operator-form of the spatially discretized discrete ordinates equations,

$$L\boldsymbol{\psi} = MSD\boldsymbol{\psi} + \mathbf{q}, \tag{3.50}$$

henceforth referred to as the *discretized DO-system*. In this form the system not yet ready to solve for $\boldsymbol{\psi}$ and requires special algorithms.

3.8 Algorithms to solve the discretized DO-system

The inversion of L in Eq. (3.39) can easily be performed if a scattering source is fixed and known since the sweep operation inverts this operator exactly (not requiring iteration). When the scattering source is included we need to solve the system in Eq. (3.50), however, by moving all the terms in Eq. (3.50) containing $\boldsymbol{\psi}$, to the left-hand side such that $(L - MSD)\boldsymbol{\psi} = \mathbf{q}$, we have a system with a matrix $A = L - MSD$ that cannot be inverted directly using sweeping.

With A being a non-symmetric matrix the system can theoretically be constructed and solved using the Krylov sub-space method called GMRES[30], however, there are two major difficulties in doing so. The first is that the system, based on angular fluxes, can be of immense size with a single Krylov-vector containing $(N_G \times N_d \times N_{nodes})$-entries. Therefore, building the entire system, or even a subset of it, could be prohibitively expensive i.t.o. memory. The second difficulty is that the iterative process has a poor convergence rate since some of the eigen-values of A are in the complex-plane. The GMRES method, operating on a fully constructed matrix, has never been applied practically. Instead, a modified matrix-free GMRES method can be used involving sweeping, which we will discuss in sections that follow.

Alternatives to the GMRES method include Algebraic Multi-Grid (AMG) methods, one of which, the Approximate Ideal Restriction (AIR) AMG-method, has been successfully applied in the solution of transport problems[36]. Such methods, however, still face the difficulty of having to build the entire system. Additionally, the performance of AMG methods are still not comparable to methods employing sweeps as stated in the conclusion of [36].

In the sections that follow we will discuss the two practical methods for solving Eq. (3.50) namely Richardson-iteration and sweep-based GMRES.

3.8.1 Richardson iteration

The classical method for solving the system in Eq. (3.50) is known as *source-iteration*[1], or more commonly, *Richardson-iteration*. In this process the scattering source is lagged at iteration (ℓ) such that

$$L\boldsymbol{\psi}^{(\ell+1)} = MSD\boldsymbol{\psi}^{(\ell)} + \mathbf{q}, \qquad \boldsymbol{\psi}^{(0)} = \mathbf{0}. \tag{3.51}$$

At each iteration we essentially have the form of Eq. (3.39) where $\mathbf{b} = MSD\boldsymbol{\psi}^{(\ell)} + \mathbf{q}$. The transport operator is inverted at each iteration $(\ell + 1)$ using sweeping,

$$\boldsymbol{\psi}^{(\ell+1)} = L^{-1}\Big[MSD\boldsymbol{\psi}^{(\ell)} + \mathbf{q}\Big]. \tag{3.52}$$

Additionally, the necessity to store $\boldsymbol{\psi}$ can be negated by noting that $\boldsymbol{\phi}^{(\ell+1)} = D\boldsymbol{\psi}^{(\ell+1)}$ and directly contributing to the flux moments in the form

$$\boldsymbol{\phi}^{(\ell+1)} = D\boldsymbol{\psi}^{(\ell+1)} = DL^{-1}\Big[MS\boldsymbol{\phi}^{(\ell)} + \mathbf{q}\Big]. \tag{3.53}$$

where the right most terms directly build $\boldsymbol{\phi}^{(\ell+1)}$ on a cell-by-cell basis, without necessarily[1] requiring $\boldsymbol{\psi}$ to be stored across all cells, directions and groups.

Richardson iteration is a versatile method that can be implemented with little difficulty but can converge slowly in problems that are scattering dominated (i.e., high σ_s to σ_t ratio).

3.8.2 Sweep-based GMRES

The sweep-based GMRES algorithm has been in practical use for a number of years dating back to 1991[37], after which numerous studies have been published [38, 39]. The method is as follows.

[1]For a given cell, the cell-by-cell system can be constructed for a subset of direction-group pairs spanning a subset of all the directions as well as a subset of all the groups. This is normally done for machine related efficiency (aka caching-effects) and will require the temporary storage of the associated angular fluxes.

By manipulating Eq. (3.50) as follows

$$\begin{aligned} & L\boldsymbol{\psi} - MSD\boldsymbol{\psi} = \mathbf{q} \\ \text{multiply } L^{-1} \rightarrow: \quad & \boldsymbol{\psi} - L^{-1}MSD\boldsymbol{\psi} = L^{-1}\mathbf{q} \\ \text{multiply } D \rightarrow: \quad & D\boldsymbol{\psi} - DL^{-1}MSD\boldsymbol{\psi} = DL^{-1}\mathbf{q} \\ D\boldsymbol{\psi} = \boldsymbol{\phi}: \quad & (I - DL^{-1}MS)\boldsymbol{\phi} = DL^{-1}\mathbf{q}, \end{aligned} \tag{3.54}$$

we get a familiar system-form $A_\phi \boldsymbol{\phi} = \mathbf{b}_\phi$, where $\mathbf{b}_\phi$ is computed with a single sweep, storing the result in a flux-moment form. The application of the GMRES algorithm, in a matrix-free fashion, then requires one to be able to only compute the action of matrix A_ϕ on a Krylov-vector, $\mathbf{v}_K$, as $A_\phi \mathbf{v}_K$. The latter matrix-action is computed at each GMRES-iteration.

The matrix-action can be defined as

$$A_\phi \mathbf{v}_K = \left(F_{Av}(\mathbf{v}_K \mapsto \mathbf{v}'_K) \ : \ \mathbf{v}'_K = \mathbf{v}_K - \mathbf{v}^*_K, \ \mathbf{v}^*_K = DL^{-1}MS\mathbf{v}_K \right) \tag{3.55}$$

where F_{Av} is the function-analog of $A_\phi \mathbf{v}_K$ and $DL^{-1}MS\mathbf{v}_K$ is equivalent to solving the system $L\boldsymbol{\psi}_K = MS\mathbf{v}_K$ where $\left(\boldsymbol{\psi}_K \mapsto \mathbf{v}^*_K \ : \ \mathbf{v}^*_K = D\boldsymbol{\psi}_K \right)$.

This algorithm, like Richardson-iteration, has the benefit of not requiring the storage of full-size angular flux vectors since it operates on flux-moment vectors, $\mathbf{v}_K$. It also allows an effective acceleration of convergence[38], as compared to Richardson-iteration, by using a Krylov-subspace to project a solution that minimizes the residual at every iteration.

The size of the Krylov-subspace can be a topic for concern since computer memory can be exhausted if too many Krylov-vectors are used to form the subspace. An apt solution to this problem is to use the *restarted* GMRES algorithm[30], which restarts the construction of the Krylov-subspace after a specified amount of iterations, bounding the total memory requirement of the algorithm. One needs to exercise caution, however, when specifying the iteration-limit before restarting, since the algorithm can experience slowed convergence if the subspace is too small. From experience gained, executing many transport simulations of scattering-dominated problems,

we have learned that the iteration-limit should not be decreased below the range of 7 to 9. Any limit below this hampers the algorithm's ability to project a substantially improved solution at each iteration, in addition to wasting an iteration every time a restart occurs. The latter occurs because the residual is recomputed[2] at iteration $(\ell + 1)$, if iteration (ℓ) incurred the restart, using the same Krylov-vector as iteration (ℓ), and can comprise a significant fraction of the total iterations[3].

3.9 Convergence acceleration techniques

There are many techniques to accelerate the converge rates of the algorithms discussed in section 3.8 and although it is beyond the scope of this text to explore all these techniques, we will briefly discuss the two techniques applied to the DO code used in this research.

The first portion of acceleration is the within-group scattering acceleration, which can be accelerated with a technique called Diffusion Synthetic Acceleration (DSA) [31, 32]. DSA operates on the incremental change of the scalar flux between iterations of algorithms in section 3.8, and essentially constructs a system based on the diffusion approximation, using the change in the scalar flux as a source. This system is then solved to add another scalar flux increment, forming the new iterative for the transport algorithm. An implementation of DSA, consistent with the spatial discretization used in this work, is called the Modified Interior Penalty (MIP) formulation of DSA [33] and is implemented in the code used for this research.

The second portion of acceleration relates to group-to-group scattering, which can be problematic for problems scattering from lower energy groups to higher energy groups. The technique employed in this research is Two-Grid Acceleration (TGA)[34] and functions similar to within-group DSA with the exception that a single system is solved for all energy groups.

The combination of these techniques is employed as described in [35].

[2]Recomputing the residual requires $A_\phi \mathbf{v}_K$ to be applied and therefore necessitates a sweep.
[3]A single restart iteration can require 20% of the total iteration time if the restart limit is 5

3.10 The fundamental drawbacks of the discrete ordinates method

Besides the spatial discretization, the angular quadrature applied as part of the angular discretization of the multigroup LBE has a direct effect on the accuracy of the solution. This is because the discrete directions used can result in a phenomenon known as *ray effects*. To illustrate this, visualizations of the scalar flux distribution of a sample 2D problem is shown in Figures 3.3 and 3.4 below, for simulations using different amounts of azimuthal angles in a product quadrature.

Figure 3.3 is a pure absorber problem with a localized source and varying amounts of azimuthal angles. The figure shows how ray-effects reduce with an increase in the number of angles.

In Figure 3.4, the dimensionless problem comprises a 5-unit square domain with an incident isotropic boundary source to left of the domain. The interior of the domain is largely a low $\sigma_t = 0.1$ material with no scattering, mimicking a low density material like air. Additionally, three square volumes are within the domain, one volume with a high $\sigma_t = 4$ and a scattering ratio of 0.1 (the dark square to the most left), and two other volumes with $\sigma_t = 1$ and a scattering ratio of 0.1 (the blocks barely visible in the bottom right of Figure 3.4). The mesh used to discretize the domain is a square orthogonal mesh (40×40 mesh cells), with all cells the same size ($\frac{1}{8} \times \frac{1}{8}$).

As can be seen in Figure 3.4, the ray effects are visually prominent when only 16 azimuthal angles are used and becomes indiscernible at 64 azimuthal angles. The visual change between the 32 and 64 azimuthal angle simulations is an indication of the problem area this research attempts to address. The questions here are apparent:

- Visually the ray effects seem to be of no concern at 64 azimuthal angles, however, are they still present and just not visually discernible?
- Is the simulation using 128 azimuthal angles more accurate than the 64 azimuthal angle simulation given that there is no visible difference?

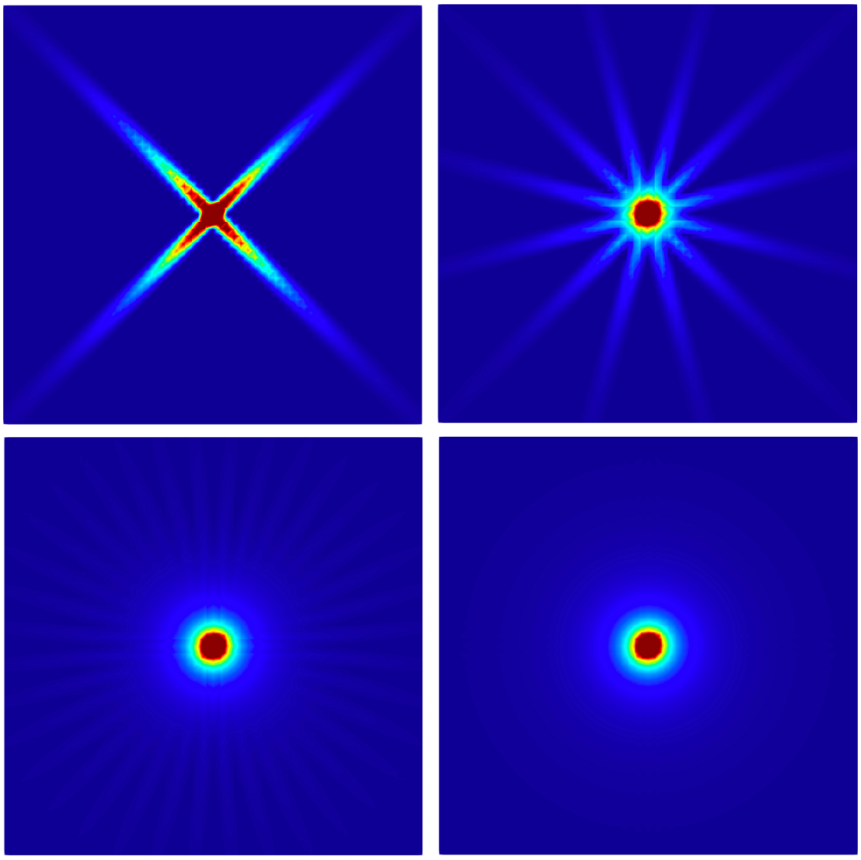

Figure 3.3: Ray effects for a pure absorber problem with varying levels of angular discretization. [Top left] 4 azimuthal angles. [Top right] 12 azimuthal angles. [Bottom left] 36 azimuthal angles. [Bottom right] 108 azimuthal angles.

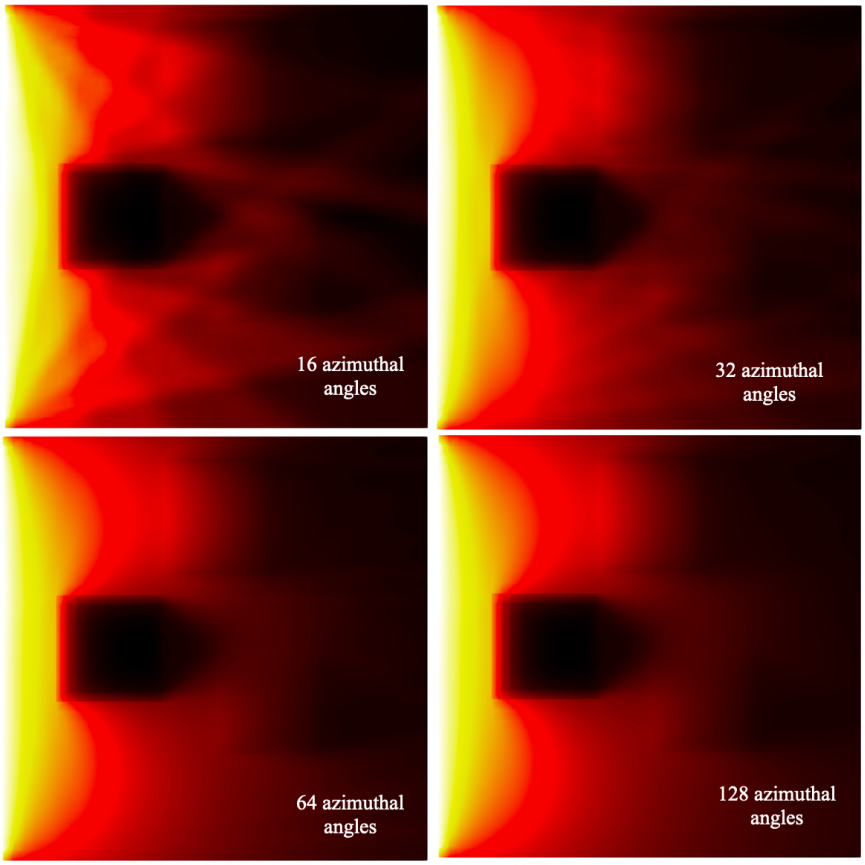

Figure 3.4: Sample DO simulation showing visual ray effects for different amounts of azimuthal angles in the angular quadratures.

The questions posed here are central to our mission. In large simulations, already using thousands of directions, it is often not feasible to consider additional angular refinement (i.e., using even more directions) to judge whether the value of a quantity-of-interest (QOI)[4] has changed. Additionally, given the oscillatory nature of ray effects, the refinement may not capture a change in QOI at all, falsely indicating that the angular refinement is sufficient.

The questions here also extend to the resolution of the spatial discretization. There could've been a QOI within the dark block in Figure 3.4, that is within a flux field that is orders of magnitude smaller than that outside the block. The spatial resolution of the mesh may not have been sufficient to accurately capture the attenuation of the incident flux within the block, resulting in significant

[4]A QOI can be the scalar flux in an area of interest (including a point), a reaction rate in a given volume, or even a visual distribution

error deep within the block. This would be of immense concern if the mesh is already refined to such a level that makes additional refinement prohibitive.

It is therefore clear that we require a means to ascertain the error in the transport solution beyond traditional methods (i.e. successive refinement).

In the introductory chapter we developed the multigroup residual transport equation, Eq. (1.11), which takes the same form as the regular multigroup transport equation in Eq. (1.8) except for the replacement of the external source, q, and boundary incident angular flux, with the multigroup residual, r_g, as well as the replacement of the regular angular flux ψ_g with the angular flux defect ψ_{Dg}.

We recall that any solution to the multigroup LBE, Eq. (1.8), can at-best be an approximate solution, $\tilde{\psi}_g$. To assist with the discussions in this chapter we now repeat the definition of the angular flux defect in Eq. (1.10),

$$\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) = \psi_g(\mathbf{x},\mathbf{\Omega}) - \tilde{\psi}_g(\mathbf{x},\mathbf{\Omega}). \tag{4.1}$$

We also repeat the multigroup residual transport equation,

$$\begin{aligned}\mathbf{\Omega}\cdot\boldsymbol{\nabla}\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) + \sigma_{tg}(\mathbf{x})\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) & \\ = \sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\mathbf{\Omega}'\cdot\mathbf{\Omega})\psi_{Dg'}(\mathbf{x},\mathbf{\Omega}')d\mathbf{\Omega}'\right] & \\ + \; r_g(\mathbf{x},\mathbf{\Omega}), \qquad \mathbf{x}\in\mathcal{D}_s, &\end{aligned} \tag{4.2a}$$

$$\mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\psi_{Dg}(\mathbf{x},\mathbf{\Omega}) = r_g(\mathbf{x},\mathbf{\Omega}), \qquad \mathbf{x}\in\partial\mathcal{D}_s, \mathbf{\Omega}\in\left\{\mathbf{\Omega}\;:\;\mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}, \tag{4.2b}$$

and the definition of the continuum-form of the multigroup residual,

$$\begin{aligned}r_g(\mathbf{x},\mathbf{\Omega}) = \sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\mathbf{\Omega}'\cdot\mathbf{\Omega})\tilde{\psi}_{g'}(\mathbf{x},\mathbf{\Omega}')d\mathbf{\Omega}'\right] \; + \; q_g(\mathbf{x},\mathbf{\Omega}) & \\ - \mathbf{\Omega}\cdot\boldsymbol{\nabla}\tilde{\psi}_g(\mathbf{x},\mathbf{\Omega}) - \sigma_{tg}(\mathbf{x})\tilde{\psi}_g(\mathbf{x},\mathbf{\Omega}), \qquad \mathbf{x}\in\mathcal{D}_s, &\end{aligned} \tag{4.3a}$$

$$r_g(\mathbf{x},\mathbf{\Omega}) = \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\left(\psi_g^{inc}(\mathbf{x},\mathbf{\Omega}) - \tilde{\psi}_g(\mathbf{x},\mathbf{\Omega})\right), \quad \mathbf{x}\in\partial\mathcal{D}_s, \mathbf{\Omega}\in\left\{\mathbf{\Omega}\;:\;\mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}. \tag{4.3b}$$

The transport problem defined in Eq. (4.2) can be solved in continuum-form using the Monte Carlo method. The method detailed in chapter 2 however, only defined source sampling for sources with strictly positive-weighted particles. This is problematic because the residual transport problem now uses r_g which can indeed assume negative-values. Additionally, chapter 2 only defined source-sampling for sources that are constant on a boundary-face or within a cell, both in space and direction, and from Eq. (4.3) we can see that the residual will indeed vary both in space and direction. We therefore require a source-sampling technique that is custom to r_g.

At first one might be tempted to propose rejection-sampling on the entire residual source, however, this will be problematic for several reasons. Firstly, r_g is dependent on 5 independent variables (3 in space, 2 in direction), requiring multivariate rejection sampling in a phase-space domain that could potentially be very large. Secondly, recalling that unmodified rejection sampling requires uniform sampling in the phase-space domain, an elaborate phase-space-dependent bounding-function [1] will have to be constructed to improve the sampling efficiency. This is because r_g itself is dependent on $\tilde{\psi}$ which can vary by orders of magnitude across space and direction.

An alternate option is to construct multiple nested CDFs, similar in nature to the CDFs detailed in the source-sampling section of chapter 2, for boundary- and cell-sources, whereby we sample the relevant source first, then the energy group, followed by the relevant element. Sampling within an element would require a specialized algorithm since this aspect is the fundamental difference (phase-space variation within an element).

Such a nested CDF would comprise firstly a discrete CDF to sample between the boundary residual and the interior-domain residual, followed by a discrete CDF for a given residual-source's energy-group, finally followed by a discrete CDF for the elements constituting the respective residual-source (either boundary-faces or cells).

The fundamental components of these CDFs would be the phase-space integrals for a given group, of the residual PDF over a boundary-face or a cell. The problem of course is that we have

[1] Given a randomly sampled point, x, in the domain of function $f(x)$, a bounding-function, $m(x)$, limits the sampling range of f', which is used to judge if a sample is rejected, i.e., when $f' > f(x)$, the sample is rejected. Instead of sampling f' in the range of $f' \in [0, \max(f(x)))$, the sampling is done in the range $f' \in [0, m(x))$, after which an appropriate correction is made based on $m(x)$.

not yet defined a functional form for either the boundary-face- or cell-PDF. A candidate function, i.e., Eq. (4.3), is not of the proper form for these PDFs since it can be of negative value, and is not normalized.

4.1 Using the absolute value of the residual source as a source PDF

Using the absolute value of Eq. (4.3), $|r_g|$, would allow us to fulfill the strictly-positive requirement of a PDF. Under the perspective that the residual source can be split into a positive- and negative-source, r^+ and r^- respectively, the absolute value here is suitable since the Monte Carlo method is subject to superposition such that

$$\psi_{Dg}(\mathbf{x}, \mathbf{\Omega}) = \psi^+_{Dg}(\mathbf{x}, \mathbf{\Omega}) - \psi^-_{Dg}(\mathbf{x}, \mathbf{\Omega}) \tag{4.4}$$

where a mapping, via the solution of Eq. (4.2) for which $r_g \mapsto \psi_{Dg}$, is made such that

$$\begin{aligned} &r^+_g \mapsto \psi^+_{Dg} : r^+_g \geq 0, \psi^+_{Dg} \geq 0, \text{ and} \\ &|r^-_g| \mapsto \psi^-_{Dg} : r^-_g \leq 0, \psi^-_{Dg} \geq 0, \end{aligned} \tag{4.5}$$

where

$$|r_g| = r^+_g + |r^-_g| \tag{4.6}$$

The strictly positive character of $|r_g|$ still requires normalization for which we need the relevant phase-space integrals over individual boundary-faces and cells.

4.2 Computing phase-space integrals of the residual source for boundary-faces and cells

The relevant phase-space integrals, to normalize the source defined by $|r_g|$, take the form

$$\int_{\mathcal{D}_c} \int_{4\pi} |r_g(\mathbf{x}, \mathbf{\Omega})| \, d\mathbf{\Omega} dV, \text{ and} \tag{4.7a}$$

$$\int_{\partial\mathcal{D}_{c,f}} \int_{4\pi} |r_g(\mathbf{x},\mathbf{\Omega})|\, d\mathbf{\Omega} dA, \tag{4.7b}$$

the latter which is only for a cell-face pair, (c, f), on a boundary (i.e., $\mathbf{x} \in \partial\mathcal{D}_s$).

4.2.1 The angular nature of the approximate solution $\tilde{\psi}$

We now turn our attention to the nature of the available angular information, in order to perform the angular integral portion of Eq. (4.7). At first one might think to use the angular flux as it manifests in the discrete ordinates method, i.e., the discrete angular fluxes at the quadrature directions, $\{\psi_{gn} \,:\, n \in [0, N_G{-}1]\}$, for which we can use the matching quadrature rule to integrate $|r_g|$. The problem with this however, is that these angular fluxes are rarely stored for practical problems, necessitating a different approach.

Another means to obtain angular flux information is to use the flux moments, $\{\phi_{gh} \,:\, h \in [0, N_h{-}1]\}$, used for the DO scattering source and typically stored, whereby an angular flux is reconstructed using the harmonic expansion of the angular flux, i.e., using Eq. (3.45). This unfortunately decouples the approximate solution from the actual angular flux used in the relevant DO simulation because the truncated expansion is a poor representation of the angular flux solved in the cell-by-cell system, Eq. (3.31). This then prevents us from relating ψ_{Dg} to the $\tilde{\psi}$ produced in the DO simulation.

One solution devised in this research is to only use the scalar flux, $\tilde{\phi}_{g0} = \tilde{\phi}_{gh} : h = 0$, produced by the DO simulation. We then assume that $\tilde{\psi}_g(\mathbf{x},\mathbf{\Omega}) = \dfrac{\tilde{\phi}_{g0}(\mathbf{x})}{4\pi}$. The logic, of why we can do this, follows from the fact that the Monte Carlo method cannot produce an exact representation of ψ_{Dg}, only a discretized representation. Conversely, it can produce the exact value of ϕ_{Dg} (in the limit of $N_p{\to}\infty$), defined as

$$\phi_{Dg}(\mathbf{x}) = \int_{4\pi} \psi_{Dg}(\mathbf{x},\mathbf{\Omega}) d\mathbf{\Omega}, \tag{4.8}$$

via track-length estimation. Therefore, ϕ_{Dg} has the same relationship with $\tilde{\phi}_{g0}$, as ψ_{Dg} has with a full fidelity $\tilde{\psi}_g$. We will henceforth refer to the zeroth moment $\phi_{g0} = \phi_{gh} : h = 0$ as just the scalar

flux ϕ_g.

Using only the scalar flux from the DO simulation is the solution we will use for this research, and all further discussions are based on the underlying assumption that $\tilde{\psi}_g(\mathbf{x}, \mathbf{\Omega}) = \dfrac{\tilde{\phi}_g(\mathbf{x})}{4\pi}$. We have left the first approach, i.e., storing the DO simulation's discrete angular fluxes, as further work that needs to be characterized w.r.t. performance and overall memory footprint.

4.2.2 The finite element nature of the approximate solution $\tilde{\psi}$

In the previous section we introduced the assumption that $\tilde{\psi}_g(\mathbf{x}, \mathbf{\Omega}) = \dfrac{\tilde{\phi}_g(\mathbf{x})}{4\pi}$. We will now analogously refer to $\tilde{\psi}_g$ as $\tilde{\phi}_g$ for further discussions.

Depending on whether the finite element representation of $\tilde{\phi}_g$ is based on a continuous- or discontinuous FEM, the form of the residual integrals on specific cells will be different.

4.2.2.1 Continuous representations of $\tilde{\phi}_g$

For continuous representations of $\tilde{\phi}_g$ the solution is expanded into weighted basis functions as

$$\tilde{\phi}_g(\mathbf{x}) = \sum_{i=0}^{N_{b,\mathcal{D}}} \tilde{\phi}_{gi} b_i(\mathbf{x}), \quad \mathbf{x} \in \mathcal{D} \tag{4.9}$$

where the weights, $\tilde{\phi}_{gi}$, are discrete values obtained from solving the discretized DO equations, $N_{b,\mathcal{D}}$ is total number of basis-functions or nodes in the domain, and the basis-functions, $b_i(\mathbf{x})$, are based on the nodes of the problem. For a continuous FEM the nodes are shared between cells, for example, using the piecewise-linear basis functions defined in appendix B, the nodes are on the vertices and the basis functions are non-zero within the cells that share the relevant vertex.

For continuous representations of $\tilde{\phi}_g$, the residual takes the same form as Eq. (4.3) but with the substitution of the angular approximation and the FEM-representations of $\tilde{\psi}_g$, still needed. We

start the substitutions first with the scattering source,

$$\sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\sigma_{sg'\to g}(\mathbf{x},\boldsymbol{\Omega}'\cdot\boldsymbol{\Omega})\tilde{\psi}_{g'}(\mathbf{x},\boldsymbol{\Omega}')d\boldsymbol{\Omega}'\right], \tag{4.10}$$

where we expand the scattering cross-section, $\sigma_{sg'\to g}$, using Eq. 2.34 and substitute the approximation for $\tilde{\psi}_g = \frac{\phi_g}{4\pi}$, the latter which cancels or higher moments of scattering, to get

$$\begin{aligned}&\sum_{g'=0}^{N_G-1}\left[\int_{4\pi}\frac{1}{4\pi}\sigma_{sg'\to g,0}(\mathbf{x})P_0(\mu_s)\frac{\tilde{\phi}_{g'}(\mathbf{x})}{4\pi}d\boldsymbol{\Omega}'\right]\\&=\frac{1}{4\pi}\sum_{g'=0}^{N_G-1}\left[\sigma_{sg'\to g,0}(\mathbf{x})\tilde{\phi}_{g'}(\mathbf{x})\right],\end{aligned} \tag{4.11}$$

noting that $\mu_s = \boldsymbol{\Omega}'\cdot\boldsymbol{\Omega}$ and $P_0(\mu_s) = 1$. We then complete the substitutions in the other terms, for which the residual is then

$$\begin{aligned}r_g(\mathbf{x},\boldsymbol{\Omega}) = &\frac{1}{4\pi}\sum_{g'=0}^{N_G-1}\left[\sigma_{sg'\to g,0}(\mathbf{x})\tilde{\phi}_{g'}(\mathbf{x})\right] \;+\; q_g(\mathbf{x},\boldsymbol{\Omega})\\&-\frac{1}{4\pi}\boldsymbol{\Omega}\cdot\boldsymbol{\nabla}\tilde{\phi}_g(\mathbf{x})-\frac{1}{4\pi}\sigma_{tg}(\mathbf{x})\tilde{\phi}_g(\mathbf{x}),\qquad \mathbf{x}\in\mathcal{D}_s,\end{aligned} \tag{4.12a}$$

$$r_g(\mathbf{x},\boldsymbol{\Omega}) = \boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\left(\psi_g^{inc}(\mathbf{x},\boldsymbol{\Omega})-\frac{1}{4\pi}\tilde{\phi}_g(\mathbf{x})\right),\quad \mathbf{x}\in\partial\mathcal{D}_s,\boldsymbol{\Omega}\in\left\{\boldsymbol{\Omega}\;:\;\boldsymbol{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}. \tag{4.12b}$$

4.2.2.2 Discontinuous representations of $\tilde{\phi}_g$

For discontinuous representations of $\tilde{\phi}_g$ the solution is expanded into weighted basis functions as

$$\tilde{\phi}_g(\mathbf{x}) = \sum_{i=0}^{N_{b,c}}\tilde{\phi}_{gci}b_i(\mathbf{x}),\quad \mathbf{x}\in\mathcal{D}_c \tag{4.13}$$

where the weights, $\tilde{\phi}_{gci}$, are again discrete values obtained from solving the DO equations (this instance being specific to a cell c), $N_{b,c}$ is the number of basis-functions or nodes on cell c, and

the basis-functions, $b_i(\mathbf{x})$ are based on the nodes of cell c. For a discontinuous FEM the nodes are *not* shared between cells, but each cell has its uniques set of nodes, for example, using the piecewise-linear basis functions defined in appendix B, the nodes are on the vertices and the basis functions for cell c are non-zero only within cell c.

Discontinuous representations require a modification to the residual that is related to conservation. Since the solution-representation is discontinuous at the cell boundaries we have to account for the δ-function derivatives now present, the residual therefore is the same as in Eq. (4.12) but with an additional term at all cell boundaries representing an upwinded-scheme,

$$\begin{aligned} r_g(\mathbf{x},\mathbf{\Omega}) = \frac{1}{4\pi}\sum_{g'=0}^{N_G-1}\left[\sigma_{sg'\to g,0}(\mathbf{x})\tilde{\phi}_{g'}(\mathbf{x})\right] \; + \; q_g(\mathbf{x},\mathbf{\Omega}) \\ - \frac{1}{4\pi}\mathbf{\Omega}\cdot\boldsymbol{\nabla}\tilde{\phi}_g(\mathbf{x}) - \frac{1}{4\pi}\sigma_{tg}(\mathbf{x})\tilde{\phi}_g(\mathbf{x}), \qquad \mathbf{x}\in\mathcal{D}_s, \end{aligned} \tag{4.14a}$$

$$r_g(\mathbf{x},\mathbf{\Omega}) = \mathbf{\Omega}\cdot\mathbf{n}_{f,c}\left(\frac{1}{4\pi}\tilde{\phi}_g^N - \frac{1}{4\pi}\tilde{\phi}_g(\mathbf{x})\right), \quad \mathbf{x}\in\partial\mathcal{D}_{c,f}, \mathbf{\Omega}\in\left\{\mathbf{\Omega} \;:\; \mathbf{\Omega}\cdot\mathbf{n}_{f,c}<0\right\}. \tag{4.14b}$$

$$r_g(\mathbf{x},\mathbf{\Omega}) = \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}\left(\psi_g^{inc}(\mathbf{x},\mathbf{\Omega}) - \frac{1}{4\pi}\tilde{\phi}_g(\mathbf{x})\right), \quad \mathbf{x}\in\partial\mathcal{D}_s, \mathbf{\Omega}\in\left\{\mathbf{\Omega} \;:\; \mathbf{\Omega}\cdot\mathbf{n}_{\partial\mathcal{D}_s}<0\right\}. \tag{4.14c}$$

where $\tilde{\phi}_g^N$ is the approximate solution in cell N, which is a neighbor to cell c at face f (of cell c) where $\mathbf{\Omega}\cdot\mathbf{n}_{f,c}<0$, and with $\mathbf{n}_{f,c}$ being the face normal.

Note here that this discontinuous formulation can also account for representations where $\tilde{\phi}$ is constant within a cell, i.e., cell-averaged values or, in other words, where there exist only a single node (node 0) in the cell and an associated unitary basis-function being constant over the entire cell, $b_0(\mathbf{x})=1$.

4.2.3 Integrals of the absolute value of multivariate functions with Monte Carlo integration

The multivariate nature of equations (4.12) and (4.14) embeds a great deal of difficulty in the endeavor of computing integrals of the absolute value of these equations. The integration over angle-space, in particular, makes such integration unwieldy.

A rather simplistic solution to this problem is to use *multivariate Monte Carlo integration*[40]. This method samples the individual domains, $\mathcal{D}_i$, of the elements of a set of independent variables $\{\mathbf{y} \in \mathcal{D}_y : y_i \in \mathcal{D}_i\}$, a total amount of N_y times to obtain the average value of a function $f(\mathbf{y})$ as

$$f_{avg,y} = \frac{1}{N_y} \sum_{j=1}^{N_y} f(\mathbf{y}_j), \tag{4.15}$$

and with the domain volume

$$V_y = \int_{\mathcal{D}_y} d\mathbf{y}, \tag{4.16}$$

the integral of the function can be approximated with

$$I_f = \int_{\mathcal{D}_y} f(\mathbf{y}) d\mathbf{y} \approx V_y f_{avg,y}. \tag{4.17}$$

Following the law of large numbers this approximation becomes exact in the limit of $N_y \to \infty$ and follows an uncertainty profile similar to that of tallies (see chapter 2). Using the central limit theorem it can be shown[40] that the standard deviation, σ_{I_f}, of I_f follows the approximation

$$\sigma_{I_f} \approx V_y \frac{\sigma_{f_{avg}}}{\sqrt{N_y}}, \tag{4.18}$$

where $\sigma_{f_{avg}}$ is the unbiased estimate of the standard deviation of $f_{avg,y}$.

With Monte Carlo integration we can evaluate volume integrals as

$$\int_{4\pi} \int_{\mathcal{D}_c} |r_g(\mathbf{x}, \Omega)| \, dV d\mathbf{\Omega} \approx \frac{4\pi V_c}{N_y} \sum_{j=1}^{N_y} |r_g((\mathbf{x}, \mathbf{\Omega})_j)|, \tag{4.19}$$

and, since the method extends well to area-integrals, we can compute area integrals in the form

$$\int_{\pi} \int_{\partial\mathcal{D}} |r_g(\mathbf{x}, \Omega)| \, dA d\mathbf{\Omega} \approx \frac{\pi A}{N_y} \sum_{j=1}^{N_y} |r_g((\mathbf{x}, \mathbf{\Omega})_j)|. \tag{4.20}$$

Given that these integrals provide approximate values, we require some characterization of the

associated error. Logically, this relates to the concern for how large N_y must be in order for the uncertainty of these integrals to not have a substantial effect on the uncertainty of ϕ_{Dg}.

4.2.3.1 Analyzing multivariate Monte Carlo integration of the absolute value of the residual

One way to analyze the effect of the uncertainty of a given source-group-element set is to evaluate its effect as a perturbation. For example, suppose that the most fundamental discrete-value of a nested residual-source CDF is $\mathbf{R} = \{R_i \ : \ i \in [0, N_R{-}1]\}$, where N_R is the total number of fundamental discrete-values. Also, suppose that these discrete values map to a set of discrete scalar fluxes, $\boldsymbol{\phi}_D$, as $\mathbf{R} \overset{N_p}{\mapsto} \boldsymbol{\phi}_D$ via a Monte Carlo process using N_p number of source particles, where $\boldsymbol{\phi}_D = \{\phi_{Di} \ : \ i \in [0, N_\phi]\}$, and N_ϕ is the total number of angular flux unknowns. In function-form this can be written as

$$\boldsymbol{\phi}_D = F_R(\mathbf{R} \overset{N_p}{\mapsto} \boldsymbol{\phi}_D) \tag{4.21}$$

where F_R is the function-analog to the Monte Carlo process using a residual source. The sensitivity of $\boldsymbol{\phi}_D$ to changes in $\mathbf{R}$ can be determined from the relationship

$$\Delta\boldsymbol{\phi}_D = J(\mathbf{R})\, \Delta\mathbf{R} \tag{4.22}$$

where J is the Jacobian-matrix, evaluated at a given $\mathbf{R}$, with entries

$$J_{ij} = \frac{\partial \phi_{Di}}{\partial R_j}. \tag{4.23}$$

These entries cannot be easily predicted since the influence of a given R_j source-element on a flux variable is governed by complex physical processes, i.e., the particle transport-process which includes the crossing of material interfaces and stochastic material interactions.

An approximation of these entries can be found by using a finite difference approximation such

that

$$\frac{\partial \phi_{Di}}{\partial R_j} \approx \frac{\phi^j_{Di} - \phi_{Di}}{\epsilon_j} \tag{4.24}$$

where $\phi_{Di} \in \boldsymbol{\phi}_{\mathbf{D}}$ and $\phi^j_{Di} \in \boldsymbol{\phi}^j_D = F_R((\mathbf{R}+\epsilon_j \mathbf{e}^{Rj}) \overset{N_R}{\mapsto} \boldsymbol{\phi}^j_D)$. The vector $\boldsymbol{\phi}^j_D$ is the flux-solution given a change ϵ_j of the j-th component of $\mathbf{R}$. The vector $\mathbf{e}^{Rj}$ is the same dimension as $\mathbf{R}$, with all entries zero except for the j-th component, which is 1. Symbolically, $\mathbf{e}^{Rj} = \{e^{Rj}_i \ : \ i \in [0, N_R{-}1], e^{Rj}_i = \delta_{ij}\}$. The coefficients ϵ_j is taken as the predicted standard deviation of the Monte Carlo integration (see Eq. (4.18)), applied to compute R_j.

The development of the Jacobian requires a single primary simulation for the computation of $\boldsymbol{\phi}_D$, followed by N_R simulations to compute the vectors $\{\boldsymbol{\phi}^j_D \ : \ j \in [0, N_R{-}1]\}$. With the Jacobian in-hand we can stochastically sample a number of $\Delta\mathbf{R}$ vectors, each randomly perturbing all elements R_j within an interval about R_j defined by the standard deviation of the integral used to estimate R_j. At each $\Delta\mathbf{R}$ sample, we compute a respective $\Delta\boldsymbol{\phi}_D$ via Eq. 4.22 which allows us to infer the confidence intervals for $\boldsymbol{\phi}_D$, on the hand of the Monte Carlo integration process. We shall denote the integration related confidence interval for value $\phi_{Di} \in \boldsymbol{\phi}_D$ as $\phi_{Di} \pm \sigma^{Di}_{I_f}$. This confidence interval is different from the statistical uncertainty related to the process as shown in Eq. (4.21) which we shall denote as $\phi_{Di} \pm \sigma^{MC}_{Di}$. The integration related uncertainty is then deemed unacceptable if $\sigma^{Di}_{I_f} > \sigma^{MC}_{Di}$.

Clearly, the additional N_R Monte Carlo simulations add considerable cost to the goal of only obtaining $\boldsymbol{\phi}_D$. We therefore never apply this analysis technique in routine simulations.

4.2.3.2 Batches of particles as a means to ascertain the effect of source uncertainty

In section 2.9 we introduced the concept of batches of simulated particles. We also noted that, at the conclusion of each batch, there is an opportunity to manipulate data if needed. Reapplying the Monte Carlo integration of the residual source, at the conclusion of each batch, provides a means to include the integration uncertainty into the uncertainty of the tallies. To comprehend this aspect

consider that the batch value for each tally is given by

$$\phi_{T,b} = \frac{1}{N_{p,b} V_T} \sum_{i=1}^{N_\ell} w_i \ell_i \tag{4.25}$$

where $N_{p,b}$ is the total number of particles simulated at the end of batch b. If we treat $\phi_{T,b}$ as another random variable, sampled $N_{batches}$ times, then the estimation of the variance of $\phi_{T,b}$ is

$$S^2_{\phi_{T,b}} = \frac{1}{N_{batches} - 1} \sum_{b'=1}^{N_{batches}} (\phi_{T,b'} - \hat{\phi}_{T,b})^2 \tag{4.26}$$

where $\hat{\phi}_{T,b}$ is the estimated batch mean for $\phi_{T,b}$ computed as

$$\hat{\phi}_{T,b} = \frac{1}{N_{batches}} \sum_{b'=1}^{N_{batches}} \phi_{T,b'}. \tag{4.27}$$

From the expression for the variance we can then derive a practical form

$$S^2_{\phi_{T,b}} = \frac{N_{batches}}{N_{batches} - 1} \left[\widehat{\phi^2_{T,b}} - \hat{\phi}^2_{T,b} \right], \tag{4.28}$$

where

$$\widehat{\phi^2_{T,b}} = \frac{1}{N_{batches}} \sum_{b'=1}^{N_{batches}} \phi^2_{T,b'}. \tag{4.29}$$

We then use this variance, with sufficient amount of batches, to determine the uncertainty incurred by the Monte Carlo integration, on the batch-average variance, $\frac{1}{N_{batches}} S^2_{\phi_{T,b}}$.

4.3 Building the residual source CDF

We now discuss how a nested CDF is constructed to select a specific cell or cell-face pair to sample, given that the residual source can take on a cell-by-cell nature and a surface nature. We will refer to the residual source based on cells as the *interior residual source*, and the residual source based on cell-surfaces as the *surface residual source*.

We start by computing phase-space integrals of the absolute value of the residual for each cell c, from Eq. (4.19), and each cell-face pair (c, f) (if appropriate), from Eq. (4.20). We do so for each group g. Therefore,

$$R_{gc}^{interior} = \frac{4\pi V_c}{N_y} \sum_{j=1}^{N_y} |r_g((\mathbf{x}, \mathbf{\Omega})_j)| \tag{4.30}$$

is an element of th group-wise interior residual CDF for group g, cell c. The surface elements are

$$R_{gc,f}^{surface} = \frac{\pi A_{c,f}}{N_y} \sum_{j=1}^{N_y} |r_g((\mathbf{x}, \mathbf{\Omega})_j)| \tag{4.31}$$

but are only computed, for a given cell-face pair, if appropriate.

For both the equations above, r_g is defined either by Eq. (4.12) or Eq. (4.14) depending on whether the finite element representation of $\tilde{\phi}$ is respectively continuous or discontinuous. For the remainder of this discussion, we will assume that there exist an $R_{gc,f}^{surface}$-term for each cell-face pair, with only the appropriate pair being non-zero, even though in practice we remove these terms from the CDF to improve performance.

We now build the element-wise interior residual source CDF, $c_g^{interior}$, and element-wise surface residual source CDF, $c_g^{surface}$, as

$$R_{g,total}^{interior} = \sum_{c=0}^{N_{cells}-1} R_{gc}^{interior} \tag{4.32a}$$

$$c_g^{interior} = \left\{ \sum_{c=0}^{c'} \left(\frac{R_{gc}^{interior}}{R_{g,total}^{interior}} \right) \ : \ c' \in [0, N_{cells}-1] \right\}, \tag{4.32b}$$

and

$$R_{g,total}^{surface} = \sum_{c=0}^{N_{cells}-1} \sum_{f=0}^{N_{f,c}-1} R_{gc,f}^{surface} \tag{4.33a}$$

$$c_g^{surface} = \left\{ \sum_{c=0}^{c'} \sum_{f=0}^{f'} \left(\frac{R_{gc,f}^{surface}}{R_{g,total}^{surface}} \right) \ : \ c' \in [0, N_{cells}-1], f' \in [0, N_{f,c}-1] \right\}. \tag{4.33b}$$

Next, we build the group-wise interior residual source CDF, $c^{interior}$, and group-wise surface

residual source CDF, $c^{surface}$, as

$$R_{total}^{interior} = \sum_{g=0}^{N_G-1} R_{g,total}^{interior} \tag{4.34a}$$

$$c^{interior} = \left\{ \sum_{g=0}^{g'} \left(\frac{R_{g,total}^{interior}}{R_{total}^{interior}} \right) \; : \; g' \in [0, N_G-1] \right\}, \tag{4.34b}$$

and

$$R_{total}^{surface} = \sum_{g=0}^{N_G-1} R_{g,total}^{surface} \tag{4.35a}$$

$$c^{surface} = \left\{ \sum_{g=0}^{g'} \left(\frac{R_{g,total}^{surface}}{R_{total}^{surface}} \right) \; : \; g' \in [0, N_G-1] \right\}. \tag{4.35b}$$

Finally we build the top-level CDF, c^R, as

$$c^R = \frac{1}{R_{total}^{interior} + R_{total}^{surface}} \left\{ R_{total}^{interior}, R_{total}^{surface} \right\} \tag{4.36}$$

which only has 2 elements.

The top-level CDF, c^R, allows sampling between the surface sources (cell-face pairs) and cell sources. The group-wise CDFs, $c^{interior}$ and $c^{surface}$, allows the sampling of the relevant group. The element-wise CDFs, $c_g^{interior}$ and $c_g^{surface}$, allows the sampling of a given element in a respective CDF, being either a cell or a cell-face pair. With this nested CDF constructed we can sample an individual cell or cell-face pair, however, we have not yet resolved how to sample within these elements, which is the topic of the next section.

4.4 Rejection sampling the residual within a cell or on a cell-face pair

The complexity of the residual within a cell or a cell-face pair warrants rejection sampling which requires us to sample the range of $|r_g|$ on the relevant element at a randomly sampled pair $(\mathbf{x}, \mathbf{\Omega})$. The absolute value bounds the lower limit of this range to zero, however, we still require the upper limit. Now, in the process of performing the Monte Carlo integrations one can determine and store

an approximation of the maximum value of $|r_g(\mathbf{x}, \mathbf{\Omega})|$, which we will denote as

$$\max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc}^{interior}, \text{ and} \tag{4.37a}$$

$$\max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc,f}^{surface}. \tag{4.37b}$$

Since these are approximate values, they too are sensitive to the number of integration samples, N_y, used during the Monte Carlo integration. This sensitivity is then accounted for in the same fashion as we account for the uncertainty of the Monte Carlo integrations, i.e., by rebuilding the residual source CDF after the execution of each batch, thereby obtaining different maximum values.

With the range of $|r_g|$ known we can apply rejection sampling to a given cell or cell-face pair.

4.4.1 Rejection sampling within a cell

We start the process by sampling a random position, $\mathbf{x}_R$, within a cell. This is done trivially within a 1D element. In 2D and 3D cells the cell is first split into its constituent primitive cells, i.e., triangle for 2D polygons and tetrahedrons for 3D polyhedrons. A primitive constituent cell is then sampled by building a discrete CDF from the constituent volumes/areas. Finally, a position within a primitive constituent cell is sampled as shown in Figure 2.1. With this position established we evaluate the finite element representation of the approximate solution to get, at $\mathbf{x}_R$, its values,

$$\tilde{\phi}_g(\mathbf{x}_R) = \sum_{i=0}^{N_{b,c}-1} \tilde{\phi}_{gci} b_i(\mathbf{x}_R), \qquad \forall g, \tag{4.38}$$

for all groups and then the gradient,

$$\boldsymbol{\nabla}\tilde{\phi}_g(\mathbf{x}_R) = \sum_{i=0}^{N_{b,c}-1} \tilde{\phi}_{gci} \boldsymbol{\nabla} b_i(\mathbf{x}_R), \tag{4.39}$$

for group g only. We then isotropically sample a direction, $\mathbf{\Omega}_R$, by sampling $\mu \in [-1, 1)$ and $\varphi \in [0, 2\pi)$.

We now have all the components required to evaluate Eq. 4.12a or Eq. 4.14a, depending on the FEM used to represent $\tilde{\phi}$, from which we obtain $r_g(\mathbf{x}_R, \mathbf{\Omega}_R)$ and thus $|r_g(\mathbf{x}_R, \mathbf{\Omega}_R)|$. In order to determine if we accept or reject the sampled position and direction we uniformly sample the range of $|r_g|$. Using a pseudo-random number, θ_R, we obtain a range sample, $|r_g|_R^{range}$, as

$$|r_g|_R^{range} = \theta_R \ \max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc}^{interior}. \tag{4.40}$$

A sampled pair $(\mathbf{x}_R, \mathbf{\Omega}_R)$ is accepted if $|r_g|_R^{range} < |r_g(\mathbf{x}_R, \mathbf{\Omega}_R)|$.

If a sample is accepted we now have to account for the actual sign of r_g. One method in which to do this is to assign a marker to the created particles, allowing it to contribute only to tallies associated with positive values of r_g if $r_g(\mathbf{x}_R, \mathbf{\Omega}_R) > 0$, or to tallies associated with the negative values of r_g if $r_g(\mathbf{x}_R, \mathbf{\Omega}_R) < 0$. This method is consistent with the superposition principle discussed in section 4.1.

Another method for accounting for the sign is to assign to the particle a unity-weight with the sign the same as $r_g(\mathbf{x}_R, \mathbf{\Omega}_R)$. In other words the particle weight,

$$w = \text{sgn}(r_g(\mathbf{x}_R, \mathbf{\Omega}_R)), \tag{4.41}$$

is assigned after which the particle can contribute to ordinary flux tallies, i.e., tallies not selective to r^+ or r^-.

One can of course also do both, however, storing three sets of multigroup nodal ϕ_D solutions for routine simulations might require a lot of memory.

4.4.2 Rejection sampling on a cell-face pair

Sampling the position on a cell-face pair is essentially the same as for a cell. We decompose a face into its constituents and sample first which constituent element to use, and then we sample a position within that element.

With a position, $\mathbf{x}_R$, sampled we need to evaluate the approximate solution, $\tilde{\phi}_g$ as we did for a cell, but on a face we require this value only for group g. If the face forms a cell interface, and if the FEM representation of $\tilde{\phi}$ is discontinuous, then we also need to evaluate the approximate solution of the neighboring cell, $\tilde{\phi}_g^N$. Therefore,

$$\tilde{\phi}_g(\mathbf{x}_R) = \sum_{i=0}^{N_{b,c}-1} \tilde{\phi}_{gci} b_i(\mathbf{x}_R), \tag{4.42}$$

and

$$\tilde{\phi}_g^N(\mathbf{x}_R) = \sum_{i=0}^{N_{b,N}-1} \tilde{\phi}_{gNi} b_i(\mathbf{x}_R). \tag{4.43}$$

We then sample the cosine-weighted incident direction, $\mathbf{\Omega}_R$, by sampling, in a reference frame along the negative of the face normal, $\mu^2 \in [0, 1)$ and $\varphi \in [0, 2\pi)$. With this direction established, and if the face is on a boundary, we can obtain the incident angular flux $\psi_g^{inc}(\mathbf{x}_R, \mathbf{\Omega}_R)$.

We now have all the components required to evaluate the surface residuals in Eq. (4.12) or Eq. (4.14), depending on the FEM used to represent $\tilde{\phi}$, from which we obtain $r_g(\mathbf{x}_R, \mathbf{\Omega}_R)$ and thus $|r_g(\mathbf{x}_R, \mathbf{\Omega}_R)|$. We then sample the range of the residual as

$$|r_g|_R^{range} = \theta_R \max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc,f}^{surface}. \tag{4.44}$$

A sampled pair $(\mathbf{x}_R, \mathbf{\Omega}_R)$ is accepted if $|r_g|_R^{range} < |r_g(\mathbf{x}_R, \mathbf{\Omega}_R)|$. We then apply the same sign-accounting methodology used for a cell.

4.5 Summary of our residual Monte Carlo source sampling method

We now summarize the processes involved in the sampling of the residual source. We start with the CDF-construction phase which builds a nested CDF, that can be sampled to identify a cell or cell-face pair, at a specific group g, to be further subjected to sampling. We then summarize the process of actually sampling the source.

In both of these phases we apply a fundamental process which is the sampling of the residual source at a random position, $\mathbf{x}_y$, and direction, $\mathbf{\Omega}_y$, and is described in Figure 4.1. We now refer to this process in the detail of the two phases.

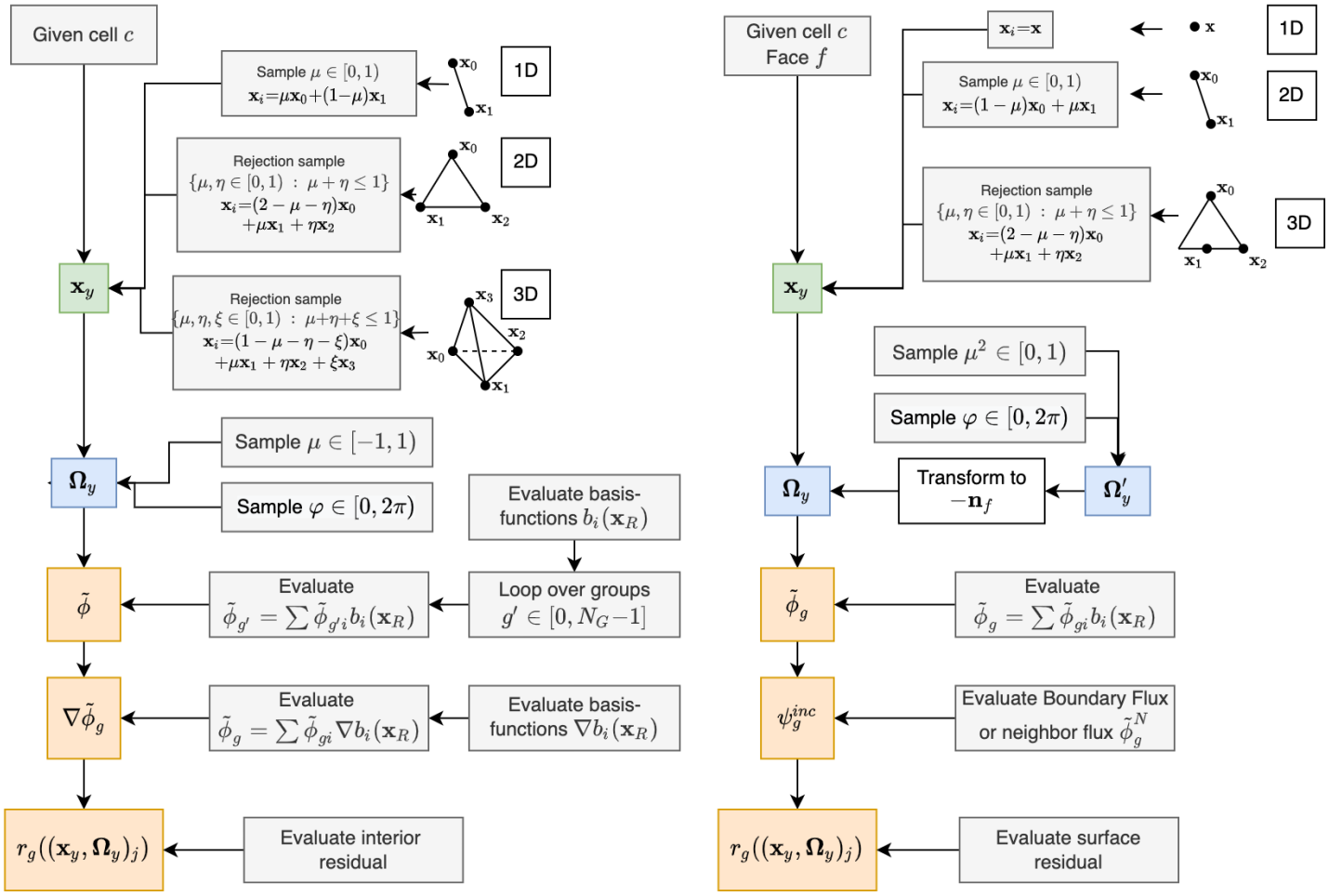

Figure 4.1: Schematic representation of the processes involved in obtaining a residual source sample.

4.5.1 Construction of the nested CDF

The construction process of the nested CDF is done at simulation initialization time as well as at the end of each batch. The process is as follows:

1. For each group, g, in the problem:

   (a) For each cell, c, in the problem:

      i. Sample the cell-interior residual N_y times using either Eq. (4.12a) or Eq. (4.14a), and Figure 4.1. Then compute $R_{gc}^{interior}$ from Eq. (4.30), and determine $\max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc}^{interior}$ from the residual samples.

      ii. For each face, f, of the cell:

         A. If a DFEM representation for $\tilde{\phi}_g$ is not used, and face f is not on a boundary, skip this face.

         B. Sample the cell-face residual N_y times using either Eq. (4.12b) or Eqs. (4.14b) and (4.14c), and Figure 4.1. Then compute $R_{gc,f}^{surface}$ from Eq. (4.31), and determine $\max(|r_g(\mathbf{x}, \mathbf{\Omega})|)_{gc,f}^{surface}$ from the residual samples.

   (b) Compute $R_{g,total}^{interior}$ as the sum of $R_{gc}^{interior} : \forall c$, and $R_{g,total}^{surface}$ as the sum of $R_{gc,f}^{surface} : \forall c, f \in [0, N_{f,c}-1)$.

   (c) Compute the interior element-wise CDF, $c_g^{interior}$, from Eq. (4.32b) and the surface element-wise CDF, $c_g^{surface}$, from Eq. (4.33b).

2. Compute $R_{total}^{interior}$ as the sum of $R_{g,total}^{interior} : \forall g$, and $R_{total}^{surface}$ as the sum of $R_{g,total}^{surface} : \forall g$.

3. Compute the interior group-wise CDF, $c^{interior}$, from Eq. (4.34b) and the surface group-wise CDF, $c^{surface}$, from Eq. (4.35b).

4. Compute the top-level CDF, c^R, from Eq. (4.36).

4.5.2 Sampling the residual source

The sampling phases is as follows:

1. Sample the top-level CDF, c^R, to identify whether a the surface- or interior residual source is to be sampled.

2. If the interior residual source is to be sampled, sample the group-wise CDF, $c_{total}^{interior}$, to identify group g.

   (a) Sample the element-wise CDF, $c_g^{interior}$, to identify a cell c.

   (b) Sample the cell-interior residual, r_g, using either Eq. (4.12a) or Eq. (4.14a), and Figure 4.1.

   (c) Sample the cell-interior residual range, $|r_g|_R^{range}$, from Eq. 4.40.

   (d) Reject the sampled residual if $|r_g|_R^{range} > |r_g|$.

   (e) If the sample is accepted, then flag and/or weight the particle with the sign of r_g and pass the particle to the transport process.

3. If the surface residual source is to be sampled, sample the group-wise CDF, $c_{total}^{surface}$, to identify group g.

   (a) Sample the element-wise CDF, $c_g^{surface}$, to identify a cell-face pair, (c, f).

   (b) Sample the cell-face residual, r_g, using either Eq. (4.12b) or Eqs. (4.14b) and (4.14c), and Figure 4.1.

   (c) Sample the cell-face residual range, $|r_g|_R^{range}$, from Eq. 4.44.

   (d) Reject the sampled residual if $|r_g|_R^{range} > |r_g|$.

   (e) If the sample is accepted, then flag and/or weight the particle with the sign of r_g and pass the particle to the transport process.

We present in this chapter problems and associated simulations that can be used to compare the performance of the Monte Carlo-method using a residual source, which we refer to as the Residual Monte Carlo (RMC) method. The performance of the RMC method will be related to the standard Monte Carlo method (SMC), as described in chapter 2. The SMC method solves the same problem (i.e., geometry, boundary conditions, and sources) as the method that generated the approximate solution $\tilde{\phi}$. In contrast, the RMC method solves a similar problem, with the difference being the source-definition.

5.1 Definitions

5.1.1 Quantity of Interest (QOI)

For a given problem there can be one or more specific quantities constituting the need to simulate the given problem. Such quantities can range from simply obtaining a scalar flux distribution over the entire domain, to the scalar flux in a particular volume or on a surface. Any quantitative or qualitative quantity can therefore be called a *Quantity of Interest* (QOI).

5.1.2 Simulation load

The *simulation load* refers to the required number of source particles, N_p, to be simulated in order to achieve a specific QOI-uncertainty. Since most of the QOIs, in the problems considered in this research, are tally-based, the required uncertainty has a relationship with the number of source particles that follows

$$\sigma^2 = \frac{C}{N_p}, \tag{5.1}$$

where C is some constant, specific to the tally and problem under consideration, σ^2 is the uncertainty of the QOI value, and σ is the standard deviation of the QOI value.

5.1.3 Efficiency $\eta^{Y \to X}$

Given simulation X with N_p^X number of source particles producing a value for a QOI with an uncertainty σ_X^2. The required amount of particles, $N_p^{X'}$, to achieve an uncertainty (for the same QOI) of σ_Y^2 produced by simulation Y, can be determined from the ratio of the respective uncertainties and load factors as

$$\frac{\sigma_X^2}{\sigma_Y^2} = \frac{\frac{C}{N_p^X}}{\frac{C}{N_p^{X'}}} = \frac{N_p^{X'}}{N_p^X}$$

therefore,

$$N_p^{X'} = N_p^X \frac{\sigma_X^2}{\sigma_Y^2}. \tag{5.2}$$

The *efficiency*, $\eta^{Y \to X}$, is then analogous to the ratio of source particles required to achieve this change in uncertainty, and is then

$$\eta^{Y \to X} = \frac{N_p^{X'}}{N_p^X} = \frac{\sigma_X^2}{\sigma_Y^2}. \tag{5.3}$$

5.1.4 Source particle Sampling Time ST_X

For simulation X, the average time required to sample a million source particles, units $[s$ per million$]$.

5.1.5 Source particle Transport Time TrT_X

For simulation X, the average time required to transport a million source particles, units $[s$ per million$]$.

5.1.6 Source particle Total simulation Time TT_X

For simulation X, the average time required to sample and transport a million source particles, units $[s$ per million$]$.

5.1.7 Overall effectiveness $E^{Y \to X}$

If two Monte Carlo methods, X and Y, determine uncertainties for the value of a QOI, σ_X^2 and σ_Y^2, using the same amount of source particles, simulated at rates TT_X and TT_Y, then the *overall*

effectiveness of simulation Y over X is the product of the efficiency, $\eta^{Y \to X}$, and the ratio of TT_X to TT_Y, such that

$$E^{Y \to X} = \eta^{Y \to X} \frac{TT_X}{TT_Y} \tag{5.4}$$

5.1.8 Key Performance Indicators (KPIs)

For all the simulations that will be presented in the sections that follow we will have performance indicators that are common to each scenario. These indicators will be called Key Performance Indicators (KPIs) and are as follows.

1. The standard deviations, σ_{RMC} and σ_{SMC}, of the value of a QOI. This is a reflection of the uncertainty in the value of the QOI.

2. The efficiency of the RMC method compared to the SMC method, $\eta^{RMC \to SMC}$. This is a measure of the "per source particle" efficiency of a given method. It is the factor by which the SMC simulation must adjust N_p to achieve the same uncertainty, σ^2, as that achieved in the RMC simulation.

3. The source particle sampling time, ST_{RMC} and ST_{SMC}. This measures the efficiency of the source sampling technique.

4. The source particle transport time, TrT_{RMC} and TrT_{SMC}. Since both methods use the same transport algorithms this KPI is a measure of the relative distribution of the source (similar sources will have similar transport times).

5. The source particle total simulation time, TT_{RMC} and TT_{SMC}. A measure of the overall speed per source particle. It is the sum of the source sampling time and the transport time.

6. The overall RMC effective compared to the SMC method, $E^{RMC \to SMC}$. A combination of the "per particle" efficiency, and the overall speed gives the overall effectiveness.

5.2 Problem 1 - One Dimensional, homogeneous

Consider the problem as shown in Figure 5.1, which shows a 1D slab with an incident boundary condition on the left boundary. The material is homogeneous with a scattering ratio of 0.5. The QOI for this problem is the right-most scalar flux.

For this simple problem we investigate several aspects of the RMC method. We direct the reader to Figures 5.2, 5.3 and 5.4 which will be in the scope of our discussions in this section. Each figure has two columns or sides. The left column/side displays scalar flux profiles whereas the right column/side displays scalar flux defect profiles. The scalar flux profiles have one profile, determined with SMC, and another computed from a DO simulation (i.e., $\tilde{\phi}$), using varying angular discretizations and a 20 cell linear DFEM spatial discretization. Quantitative data related to these figures is contained in Table 5.1.

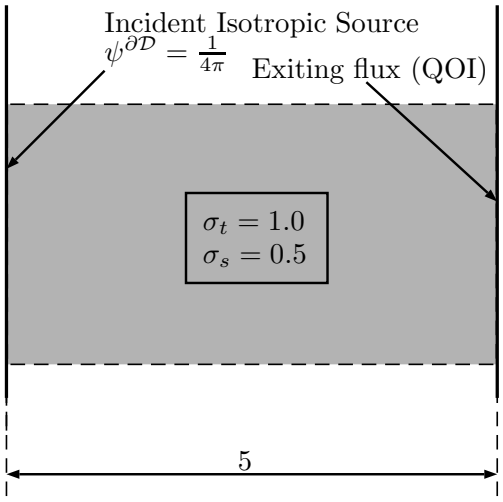

Figure 5.1: One dimensional homogeneous material problem geometry.

The first aspect we want to discuss is the reduction of the RMC method to the SMC method when a zero approximate solution is provided ($\tilde{\phi} = 0$). In this case the RMC's scalar flux defect should exactly recover the SMC solution, both in value and uncertainty, because the source terms are the same (from Eq. (4.3) with $\tilde{\phi}_g = 0$). The left side of Figure 5.2 shows the scalar flux,

computed with SMC, and $\tilde{\phi} = 0$. The right hand side shows the scalar flux defect computed both using the SMC scalar flux solution ($\phi_D = \phi - \tilde{\phi}$) and scalar flux defect using RMC, from which we can see that both methods produce the same scalar flux defect.

The first row of Table 5.1 contains the related KPIs. It firstly shows that the QOI variance is approximately the same for the RMC and SMC methods. This is a good indicator that the methods are solving an identical problem, which it should. Secondly it shows that the source sampling time for SMC is far less than that for RMC, which is not surprising since RMC uses rejection sampling whereas SMC uses the process in Figure 2.1. The transport time for the two methods is approximately the same, indicating again that the source distribution is equivalent. The total source particle simulation time directly reflects the increased sampling time of RMC, indicating that overall the RMC method is slower. Finally, the overall effectiveness, $E < 1$, shows that, because of the increasing sampling cost, RMC is not a good choice when $\tilde{\phi} = 0$.

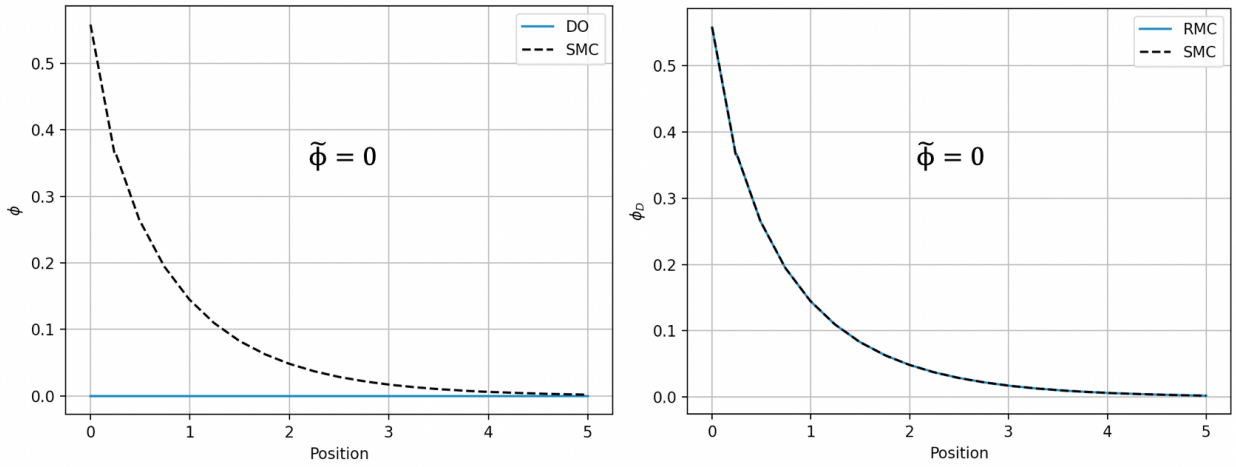

Figure 5.2: Scalar flux solutions and computed defects for problem 1. Here, $\tilde{\phi}$ was taken as zero.

The next aspect under consideration is the effect of different representations of $\tilde{\phi}$. Graphical representations, for $\tilde{\phi}$ obtained from S_2 angular quadrature- DO simulations, is shown in Figure 5.3 and Figure 5.4. Figure 5.3 shows results for a discontinuous cell-constant, $Q0$, representation,

as well as a discontinuous linear, $Q1$, representation, whereas Figure 5.4 shows results only for a continuous $Q1$ representation[1]. The relevant performance data is shown in Table 5.1 where rows 2 to 4 are for the DO simulations using an S_2 quadrature, and rows 5 to 7 are for the DO simulationns using an S_8 quadrature.

The KPIs show that the efficiency, $\eta^{RMC \rightarrow SMC}$, for a given angular quadrature used, is lowest for the discontinuous $Q0$ representations and the highest for the continuous $Q1$ representations. This trend is the same for both the S_2 and the S_8 quadrature. The timings show that the RMC method has considerably higher sampling times as compared to SMC (for all configurations) but with no observable trend when comparing the transport times. The overall effectiveness factor, E, shows that in-general the S_8 angular quadrature performed better than the S_2 angular quadrature. This is not surprising since we hypothesized, in the introduction, that the RMC method should be able to produce a QOI-uncertainty much lower than that of SMC if the true defect is much smaller than the tally value. The latter is indeed the case because the defect in the S_8 solution is much less than that in the S_2 solution.

[1] Obtained from the discontinuous Q1 solution by taking nodal averages

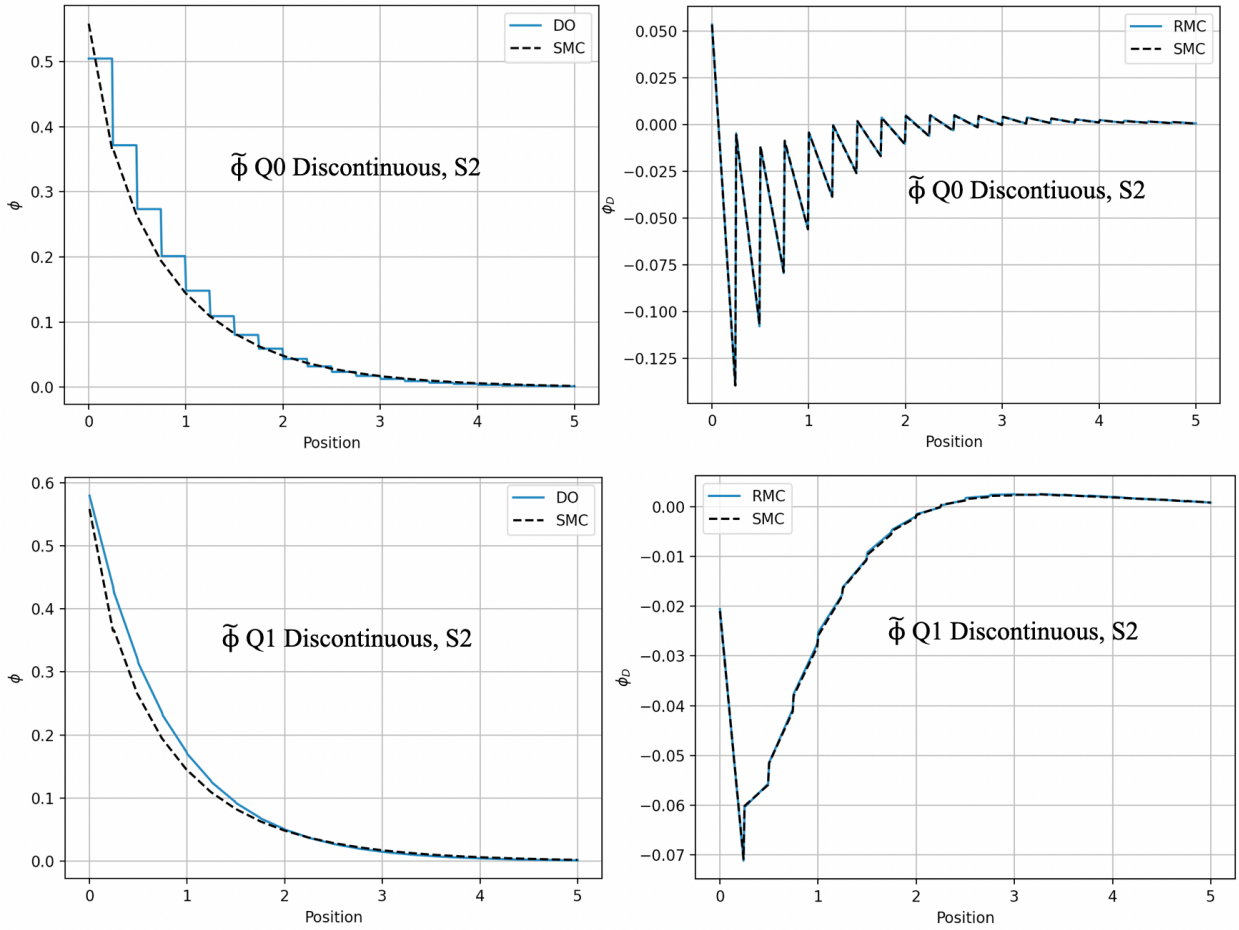

Figure 5.3: Scalar flux solutions and computed defects for problem 1. The top 2 plots are for $\tilde{\phi}$ being a cell-constant representation, $Q0$, computed from a linear discontinuous DO simulation with an S_2 quadrature. The bottom 2 plots are for $\tilde{\phi}$ taken directly as the linear discontinuous DO solution.

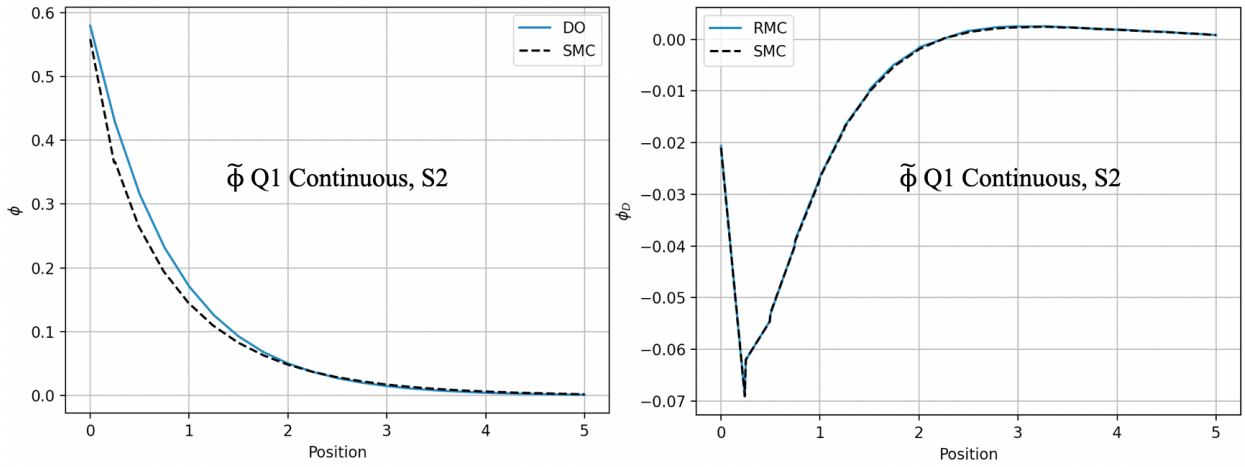

Figure 5.4: Scalar flux solutions and computed defects for problem 1. Here, $\tilde{\phi}$ is a linear continuous representation, $Q1$, computed from a linear discontinuous DO simulation with an S_2 quadrature.

Table 5.1: Key Performance Indicators for problem 1 using different approximate solutions, $\tilde{\phi}$, as input to the RMC method. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	$\boldsymbol{\sigma_{SMC}}$ $\boldsymbol{\sigma_{RMC}}$	$\boldsymbol{\eta^{RMC\to SMC}}$	$\boldsymbol{ST_{SMC}}$ $\boldsymbol{ST_{RMC}}$ **[s per 10^6]**	$\boldsymbol{TrT_{SMC}}$ $\boldsymbol{TrT_{RMC}}$ **[s per 10^6]**	$\boldsymbol{TT_{SMC}}$ $\boldsymbol{TT_{RMC}}$ **[s per 10^6]**	**Rate ratio**	$\boldsymbol{E^{RMC\to SMC}}$
Right-most cell	7.734E-06		0.219	1.649	1.868		
$\widetilde{\Phi} = 0$	7.669E-06	1.017	0.913	1.766	2.679	0.697	0.709
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q0 S2	7.770E-06	0.991	1.181	1.596	2.778	0.673	0.666
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q1 Disc. S2	4.614E-06	2.809	1.370	1.442	2.813	0.664	1.866
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q1 Cont. S2	4.612E-06	2.812	1.382	1.411	2.793	0.669	1.881
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q0 S8	6.787E-06	1.298	1.254	1.522	2.776	0.673	0.874
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q1 Disc. S8	2.926E-06	6.985	1.387	1.390	2.778	0.673	4.698
Right-most cell	-		-	-	-		
$\widetilde{\phi}$ Q1 Cont. S8	2.888E-06	7.170	1.396	1.619	3.015	0.620	4.442

One particular observation to make, is that the discontinuous $Q0$ representation of the approximate solution for the S_2 quadrature has an efficiency less than 1, and is not much better for the S_8 quadrature (not enough to overcome the source sampling cost). This observation is particularly interesting, prompting further investigation. The $Q0$ representation was used before in [41], a conference article, but the authors did not detail the same KPIs as we do here. We argued that the inefficiency could only be explained by the superposition principal of the positive- and negative-valued, r^+ and r^-, respectively. For a more in-depth investigation we then split the respective solutions into positive- and negative-contributions, shown in Figure 5.5. The left side of the figure shows the results for the discontinuous $Q0$ representation and the right side shows the results for the continuous $Q1$ representation.

The source of the inefficiency is clearly seen in this comparison. The magnitude of ϕ_D^+ and ϕ_D^- individually impact their absolute uncertainties (as compared to their relative uncertainties) and since the $Q0$-based RMC solution introduces large positive and negative sources (at the cell interfaces) the associated defect solutions are large, even approaching the magnitude of the solution itself. It is clear that the continuous $Q1$-based RMC solutions have much reduced magnitudes and therefore reduced absolute uncertainty resulting in the better efficiency.

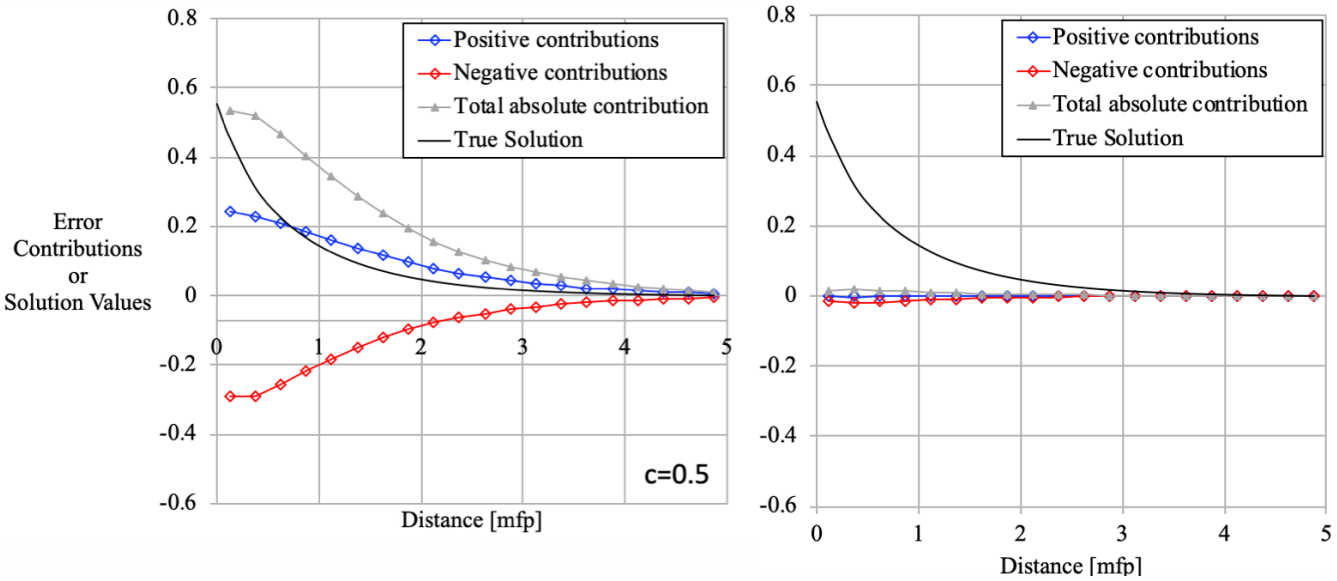

Figure 5.5: RMC defect, using a discontinuous $Q0$ representation for $\tilde{\phi}$ [Left], and a discontinuous $Q1$ representation for $\tilde{\phi}$ [Right], with contributions split into positive and negative components. Both solutions were obtained with a DO simulation using an S_2 angular quadrature.

In this comparison we did not show the discontinuous $Q1$-based RMC solutions since it very closely represented its continuous counterpart. It did however point to the efficiency reducing effects of the cell-surface discontinuities. It would therefore be beneficial if we used a continuous representation for the approximate solution for all our further investigations, with the only surface based source terms involving the boundaries. Fortunately, for any discontinuous approximate solution provided, which we shall denote with $\tilde{\phi}^*$, we can easily compute a continuous representation by taking nodal averages, which will then be used as $\tilde{\phi}$. The RMC obtained defect can then trivially be related to the original approximate solution as

$$\phi_D^* = \tilde{\phi} + \phi_D - \tilde{\phi}^*, \tag{5.5}$$

where ϕ_D^* is the defect of the original discontinuous approximate solution.

The overall insights gained in this section are as follows:

- The RMC method correctly reduces to the SMC method when $\tilde{\phi} = 0$.
- Discontinuities in $\tilde{\phi}$ hinder the efficiency of the RMC method because it increases the magnitude of the respective positive and negative defect solutions, without affecting the net defect solution.
- A continuous $\tilde{\phi}$ should be constructed if the supplied approximate solution is discontinuous in order to minimize discontinuities. The discontinuities at the boundaries are unavoidable.
- The RMC method's effectiveness increases as $\tilde{\phi}$ approaches the true solution but it has a lower bound, dictated by the increased source sampling cost, where SMC will be more effective.

5.3 Problem 2 - One Dimensional, multi-material

Consider the problem as shown in Figure 5.6, which shows a 1D slab comprising three different materials. The material to the left is pure absorbing, $\sigma_t = 0$, but with a homogeneous isotropic source of strength 10. The middle material is a pure scattering material, $\sigma_t = 0.2$. Finally, the material to the right mimics a zero-scattering, low-density material, with $\sigma_t = 0.01$.

For this problem we are interested in both the profile of the scalar flux over the entire domain as well as the exiting flux at the right boundary. Therefore, we have two QOIs, the first being the L_2-norm of the scalar flux defect and the second being the scalar flux at the right most cell. The approximate solution for this problem was generated with a DO simulation using an S_2 and S_8 angular quadrature and a 60 cell linear DFEM spatial discretization.

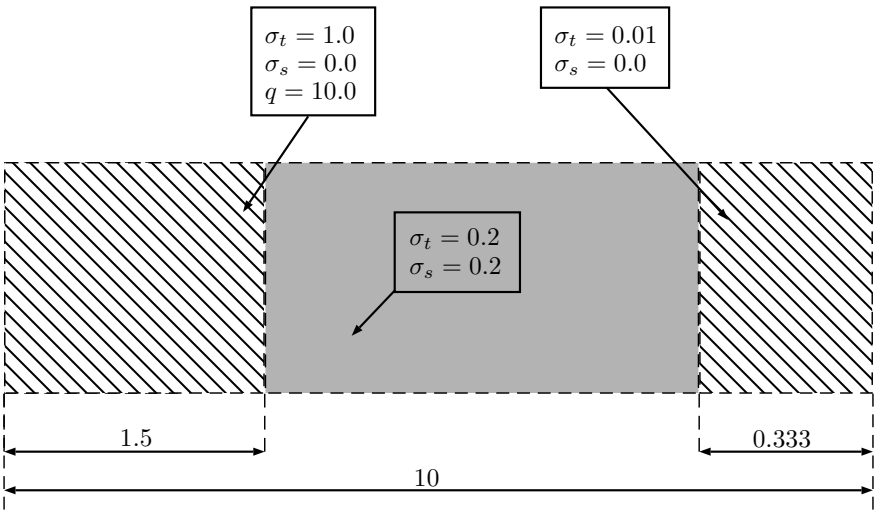

Figure 5.6: One dimensional test problem setup with multiple-materials. The dimensions are in mean-free-paths.

A graphical representation of the solution- and defect-profiles are shown in Figure 5.7. In the figure one can clearly observe the S_2 DO solution being smaller in magnitude in the source region but larger in magnitude outside the source region. The defect solution is well captured as shown on the right of the figure. The relevant performance data is shown in Table 5.2 where the first two rows are relevant to using an S_2 quadrature and the last two rows are relevant to using an S_8 quadrature.

The KPIs for the $L2$-norm QOI show that the efficiency, $\eta^{RMC \to SMC}$, is very large both using an S_2 and S_4 quadrature when generating $\tilde{\phi}$. Consistent with the previous problem, the S_8-based $\tilde{\phi}$ decreases the number of source particles required more than its S_2 counterpart. The same trend is observed for the exiting cell QOI, albeit less of an improvement is observed. This is because the RMC method becomes increasingly efficient as the defect solution becomes smaller.

For all the cases the RMC method again manifests increased source sampling times, ST, but for this problem the transport time, TrT, is also notably increased. This is caused by a redistribution of the source where the RMC source is now effectively concentrated (relatively) further from the boundaries and closer to the middle scattering material where particles survive longer.

The overall effectiveness, E, shows that RMC improves the computation of the defect drasti-

cally when considering the overall scalar flux profile (i.e., the QOI based on the $L2$-norm), and less so for the exiting flux QOI. Regardless, RMC shows a significant improvement.

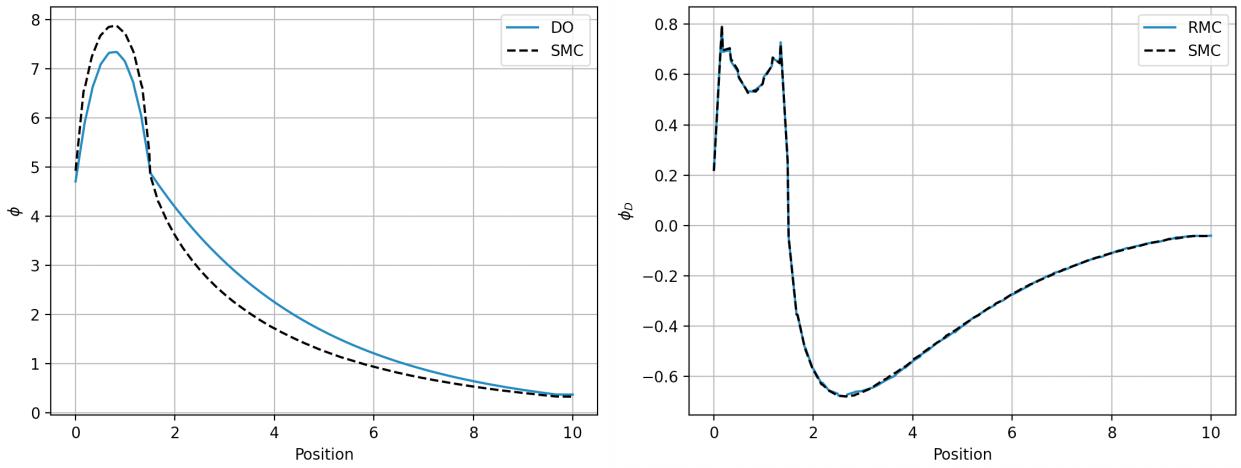

Figure 5.7: Scalar flux solutions and computed defects for problem 2. Here, $\tilde{\phi}$ is a linear continuous representation, $Q1$, computed from a linear discontinuous DO simulation with an S_2 quadrature.

Table 5.2: Key Performance Indicators for problem 2 using different angular quadratures in the production of $\tilde{\phi}$, as input to the RMC method. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	$\boldsymbol{\sigma_{SMC}}$ $\boldsymbol{\sigma_{RMC}}$	$\boldsymbol{\eta^{RMC\rightarrow SMC}}$	$\boldsymbol{ST_{SMC}}$ $\boldsymbol{ST_{RMC}}$ **[s per 10^6]**	$\boldsymbol{TrT_{SMC}}$ $\boldsymbol{TrT_{RMC}}$ **[s per 10^6]**	$\boldsymbol{TT_{SMC}}$ $\boldsymbol{TT_{RMC}}$ **[s per 10^6]**	**Rate ratio**	$\boldsymbol{E^{RMC\rightarrow SMC}}$
L2 norm of ϕ_D	1.260E-03		0.432	4.114	4.545		
$\tilde{\phi}$ Q1 Cont. S2	4.420E-04	8.126	2.100	6.616	8.717	0.521	4.238
Right most cell	4.610E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. S2	2.810E-04	2.691	2.100	6.616	8.717	0.521	1.404
L2 norm of ϕ_D	1.260E-03		-	-	-		
$\tilde{\phi}$ Q1 Cont. S8	2.710E-04	21.617	2.179	6.821	9.000	0.505	10.918
Right most cell	4.610E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. S8	1.750E-04	6.939	2.179	6.821	9.000	0.505	3.505

The overall insights gained in this section are as follows:

- The RMC method drastically improves the computation of a flux-profile defect, with more than an order of magnitude improvement for the S_8 angular quadrature based $\tilde{\phi}$. This indicates the method can be very useful to correct problems with wide spread error/defect.
- RMC on multi-material problems can alter the transport time of source particles significantly, especially since the method delocalizes the source from the original source material.
- RMC is effective for localized QOIs and distributed QOIs, and increases in efficiency with a decrease in the defect (i.e., better approximate solutions).

5.4 Problem 3 - Multidimensional homogeneous pure absorber material with a large distributed source.

Consider the pure absorber problem shown in Figure 5.8. The domain is a 5 unit square with two large regions, one region with a homogeneous isotropic source with a strength of 3 and another with no source. Both regions have $\sigma_t = 1$.

We investigate, with this problem, the ability of the RMC method to correct for angular discretization errors without directly analyzing ray effects. The large distributed source region assists the DO simulation, which used a low order S_2 angular quadrature, in developing a relatively smooth solution in the interior of the domain. This solution is shown in Figure 5.9. The resolution provided by the spatial dicretization, i.e., the number of cells in x and y, was also chosen such that it should not incur spatial discretization error on the same order of magnitude as the angular discretization error (i.e., cell-sizes are $1/10$-th of a mean-free path). The corresponding SMC solution of this problem is shown in Figure 5.10.

The RMC computed defect solution is shown in Figure 5.11. The defect follows a complex but smooth pattern. The S_2 angular discretization is the likely cause of the two ray like structures emanating from the upper corners of the source region, as well as the negative defect area in the

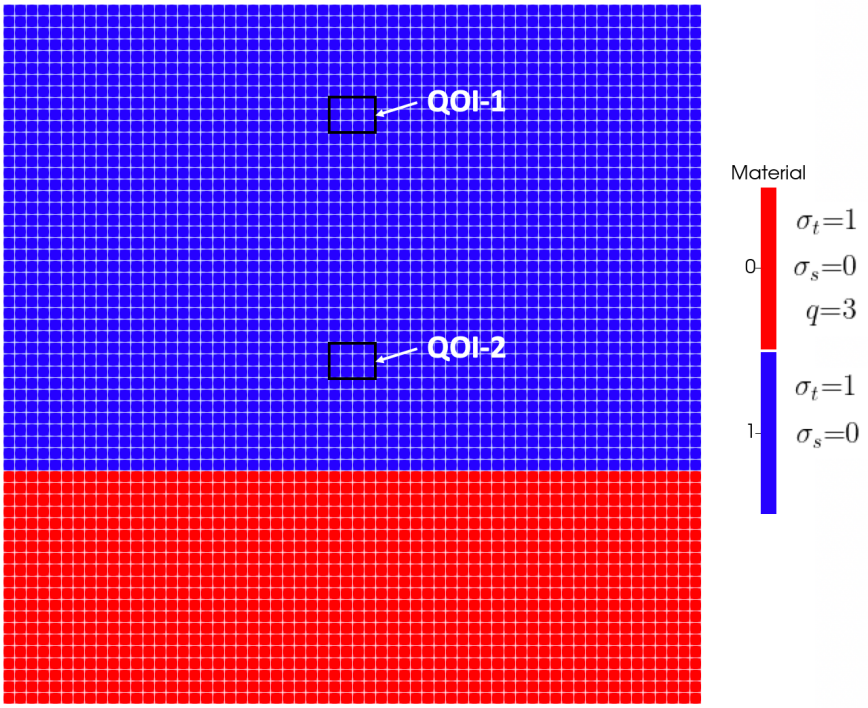

Figure 5.8: Multidimensional pure absorber material test problem with a distributed isotropic source in a large portion of the domain.

center of the domain just above the source region. The relevant performance data is shown in Table 5.3.

First analyzing the case where an S_2 angular quadrature solution is used, the KPIs for all three QOIs show that the RMC method has superior efficiency, however, its efficiency for the $L2$-norm QOI is much less than the localized QOIs, which is contrary to the previous problem. The likely cause for this is that, compared to the localized QOIs, the $L2$-norm QOI includes the uncertainty for areas where the defect is much larger than that of the localized QOIs. Therefore, as the previous problems highlighted, the method has reduced efficiency when the defect solution is large. The efficiency, $\eta^{RMC \rightarrow SMC}$, for QOI-1 is larger than that of QOI-2. This is because QOI-1 resides in an area with a lower magnitude defect and by the same mechanism referenced above, the RMC method exhibits better efficiency.

The KPIs, for the case where an S_{16} angular quadrature solution is used, are similar to that of

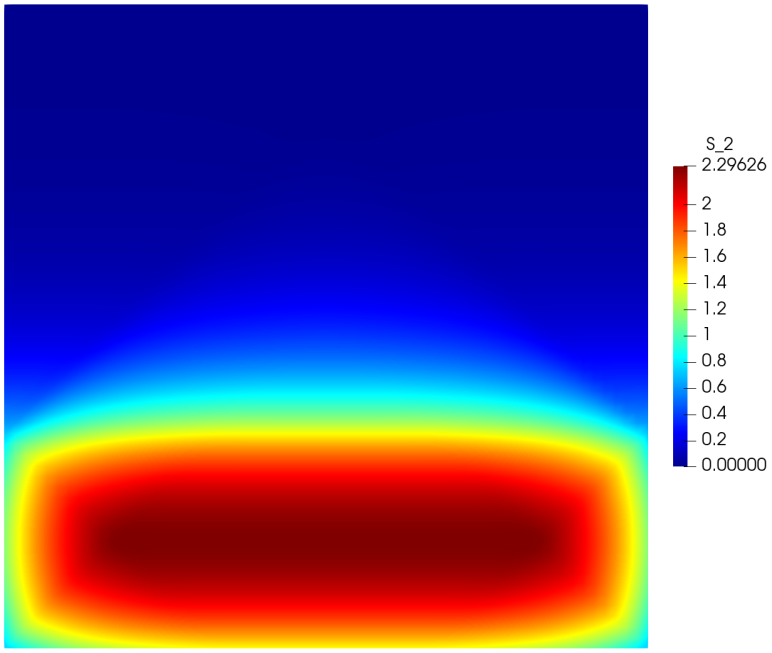

Figure 5.9: DO solution of problem 3 using an S_2 angular quadrature.

the S_2 case except that the uncertainties have further reduced. This is because the overall defect decreased. This effect will reach a limit where the angular discretization will no longer dominate and the spatial discretization will be the dominant form of error.

The increase in the sampling time, ST, for the RMC method is again at the forefront of limiting its effectiveness. The transport time, TrT, is also longer which can be explained by observing the distribution of the cell-wise absolute residual source strength, shown in Figure 5.12. More particles created towards the center of the domain result in the particles having less proximity to the boundaries which increases the likelihood that a particle can be transported across many cells. Even with the considerable source-sampling cost the overall effectiveness, E, for all the QOIs show that the RMC method is more effective than the SMC method.

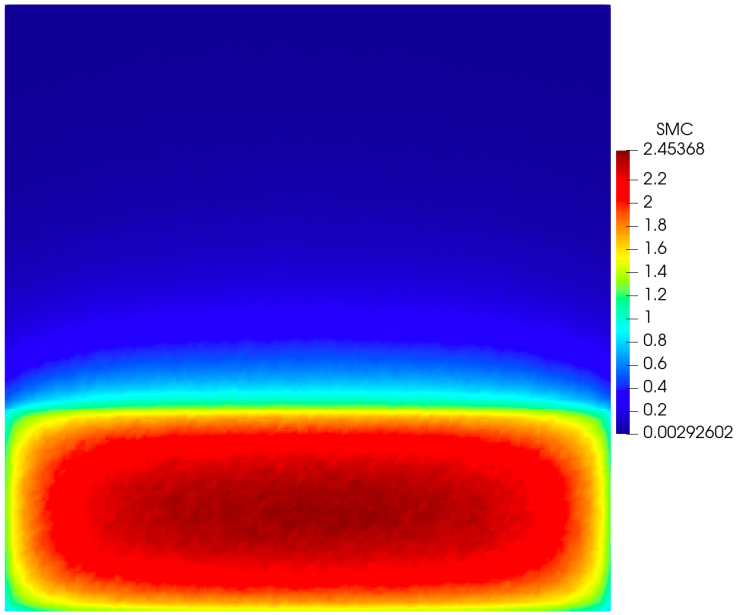

Figure 5.10: Standard Monte Carlo solution of problem 3 using 10×10^6 source particles.

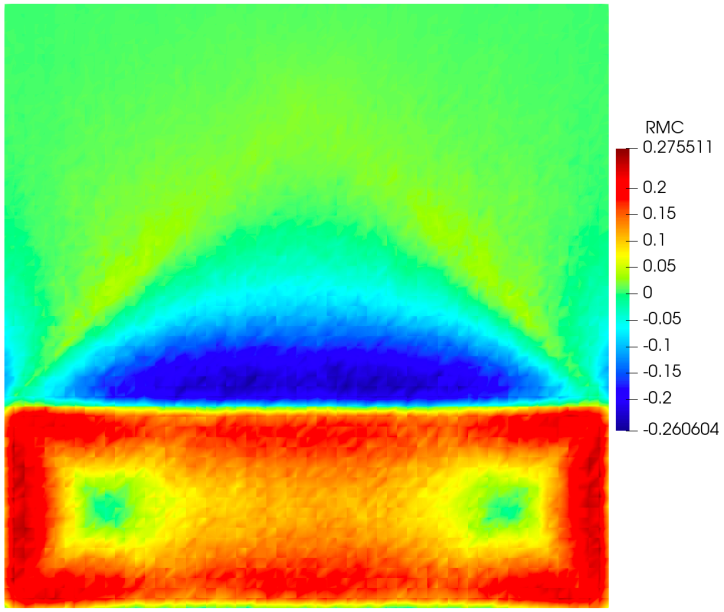

Figure 5.11: The RMC defect solution, for a DO simulation using an S_2 angular quadrature, of problem 3 using 10×10^6 source particles.

Table 5.3: Key Performance Indicators for problem 3. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	σ_{SMC} σ_{RMC}	$\eta^{RMC \to SMC}$	ST_{SMC} ST_{RMC} [s per 10^6]	TrT_{SMC} TrT_{RMC} [s per 10^6]	TT_{SMC} TT_{RMC} [s per 10^6]	**Rate ratio**	$E^{RMC \to SMC}$
L2 norm of ϕ_D	1.390E-02		0.720	19.280	20.000		
$\tilde{\phi}$ Q1 Cont. S2	9.730E-03	2.041	3.782	21.933	25.714	0.778	1.587
QOI-1	8.760E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. S2	2.960E-04	8.758	3.782	21.933	25.714	0.778	6.812
QOI-2	3.410E-03		-	-	-		
$\tilde{\phi}$ Q1 Cont. S2	1.280E-03	7.097	3.782	21.933	25.714	0.778	5.520
L2 norm of ϕ_D	1.390E-02		-	-	-		
$\tilde{\phi}$ Q1 Cont. S16	4.628E-03	9.022	4.124	23.487	27.611	0.724	6.535
QOI-1	8.760E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. S16	2.479E-04	12.488	4.124	23.487	27.611	0.724	9.046
QOI-2	3.410E-03		-	-	-		
$\tilde{\phi}$ Q1 Cont. S16	1.079E-03	9.995	4.124	23.487	27.611	0.724	7.240

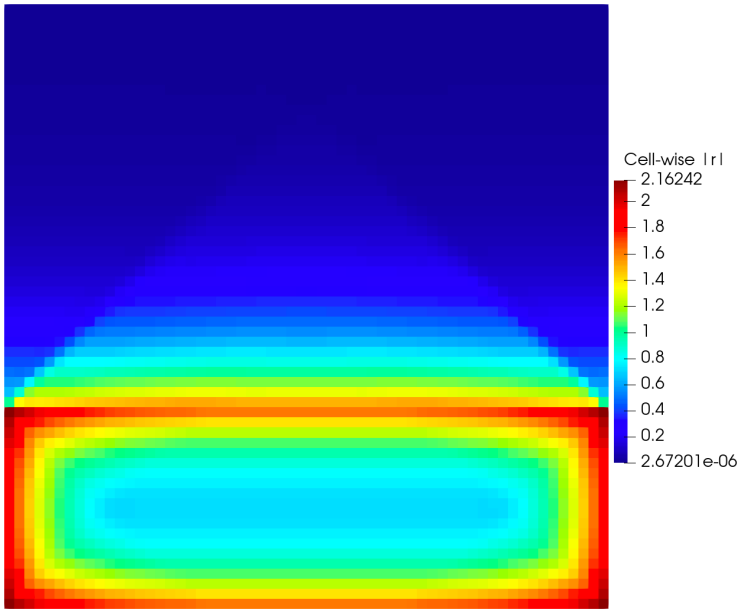

Figure 5.12: Cell-wise residual source for problem 3.

5.5 Problem 4 - Multidimensional multi-material domain with a duct and a localized source

Consider the problem shown in Figure 5.13. The domain has a 5 unit square dimension. At the longitudinal bottom and lateral center is a localized source material containing a distributed isotropic source of strength 3. This material is a high-density pure absorber, $\sigma_t = 6$. On either lateral side of the source are two large blocks of medium-density scattering material, $\sigma_t = 2$ and a scattering ratio of 0.8. The same type of material is present as an "end-piece" at the longitudinal top and lateral center. The remainder of the domain contains a low-density pure scattering material, $\sigma_t = 0.01$, and forms a duct, between the two scattering blocks along the longitudinal direction, and joins the QOI-area at the general longitudinal top area. The first QOI for this problem is the scalar flux in the area just above the termination of the duct and the second QOI is laterally to the side of the first, but outside the direct path of the duct.

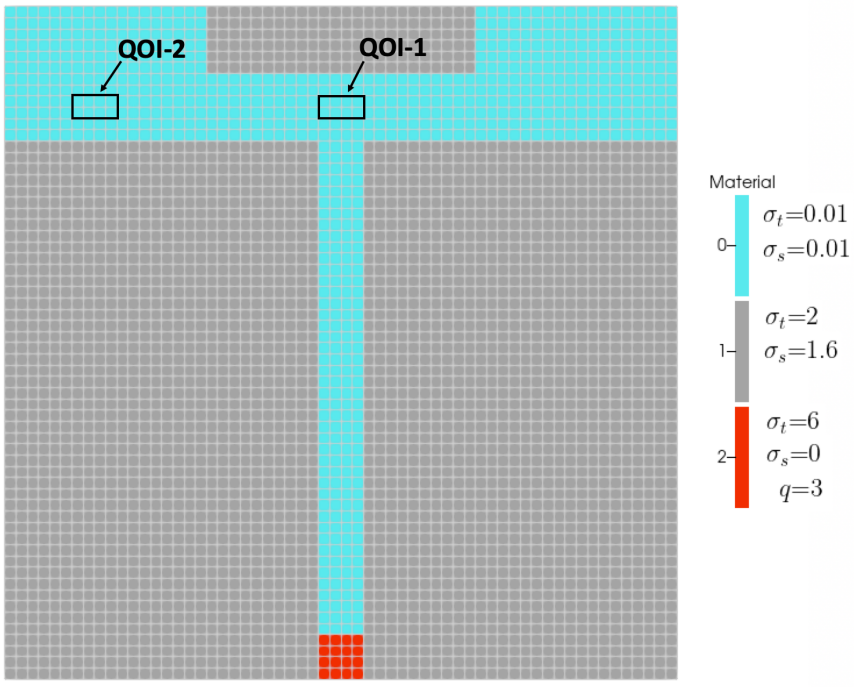

Figure 5.13: Multidimensional multi-material test problem with a localized isotropic source.

The design of this problem is such that we can incur ray effects from a DO simulation. Because the source is localized we can expect rays emanating from its location outward, along the directions of the angular quadrature. The ducted channel also provides a particular challenge since the direct contribution to QOI-1 will be very sensitive to the directions used.

A DO solution, using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant, is shown in Figure 5.14. The corresponding SMC solution is shown in Figure 5.15. The comparison of these two solutions highlight the ray effects that can clearly be seen in Figure 5.14.

The RMC computed defect solution, using the above DO solution, is shown in Figure 5.16. The defect solution surprisingly shows a lot of defect in and around the source material suggesting that spatial discretization error dominates in this area. Away from the source the defect solution follows the ray effects present in the DO solution and within the duct the large discrepancy is highlighted.

The related performance data is shown in Table 5.4. The KPIs for all three QOIs again shows that the RMC method has superior efficiency. There are two particular performance items to discuss; the first is the comparatively high efficiency for QOI-1, and the second is the dramatically increased transport time of the RMC method.

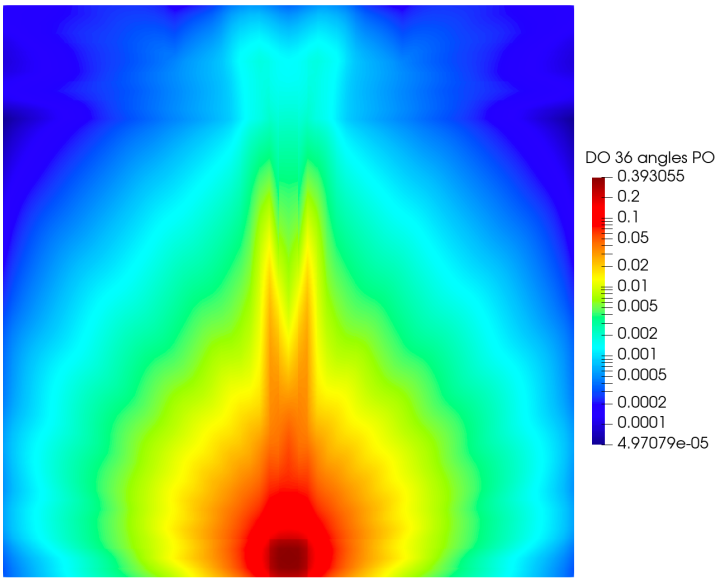

Figure 5.14: DO solution of problem 4 using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant.

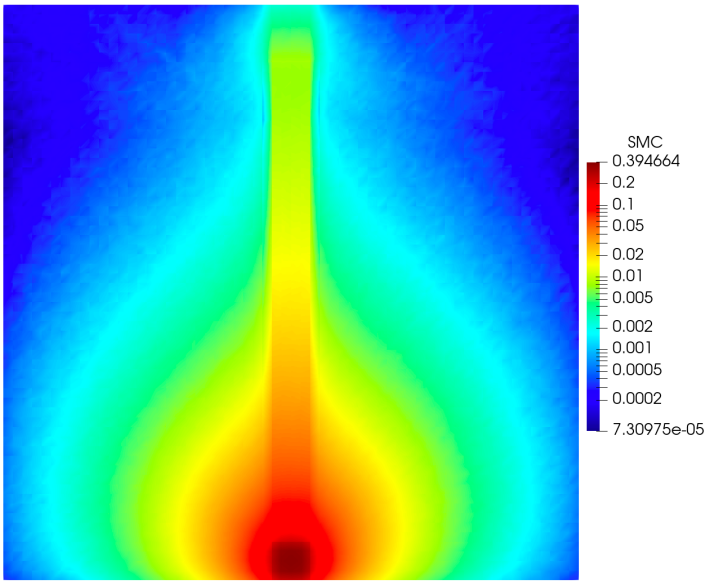

Figure 5.15: Standard Monte Carlo solution of problem 4 using 10×10^6 source particles.

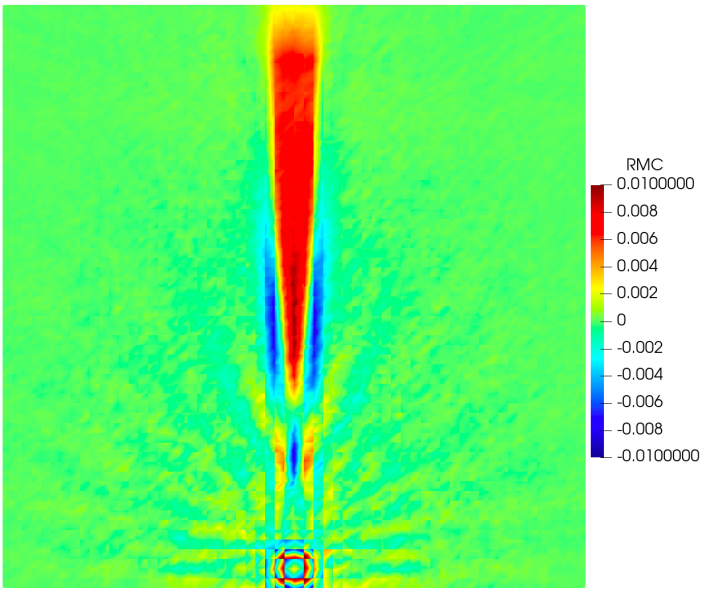

Figure 5.16: The RMC defect solution, for a DO simulation using a GLC angular quadrature (6 polar- and 6 azimuthal-angles per octant), of problem 4 using 10×10^6 source particles.

Table 5.4: Key Performance Indicators for problem 4. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	$\frac{\sigma_{SMC}}{\sigma_{RMC}}$	$\eta^{RMC \to SMC}$	$\frac{ST_{SMC}}{ST_{RMC}}$ **[s per 10^6]**	$\frac{TrT_{SMC}}{TrT_{RMC}}$ **[s per 10^6]**	$\frac{TT_{SMC}}{TT_{RMC}}$ **[s per 10^6]**	**Rate ratio**	$E^{RMC \to SMC}$
L2 norm of ϕ_D	1.420E-04		0.538	14.846	15.385		
$\tilde{\phi}$ Q1 Cont. GLC6,6	5.846E-05	5.900	2.134	23.580	25.714	0.598	3.530
QOI-1	7.590E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC6,6	2.180E-05	12.122	2.134	23.580	25.714	0.598	7.252
QOI-2	1.320E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC6,6	7.379E-06	3.200	2.134	23.580	25.714	0.598	1.915
L2 norm of ϕ_D	1.420E-04		0.538	14.846	15.385		
$\tilde{\phi}$ Q1 Cont. GLC64,6	4.507E-05	9.929	1.997	21.533	23.529	0.654	6.492
QOI-1	7.590E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC64,6	2.122E-05	12.797	1.997	21.533	23.529	0.654	8.367
QOI-2	1.320E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC64,6	6.632E-06	3.961	1.997	21.533	23.529	0.654	2.590

In order to analyze the performance items mentioned above we developed several visualizations. The first is the cell-wise residual source absolute strength, which is the result of applying Eq. 4.30. The visualization of this quantity is shown in Figure 5.17 in both logarithmic- and linear scale, left and right respectively. Now our argument is that the transport time, TrT, can increase in two ways. The first is if particles are created further from the boundary (as we've seen in the previous problems) and the second, which is more likely for this problem, is if proportionately more particles are directly created within highly scattering materials. The linear-scale visualization in Figure 5.17 shows that the residual source is localized around the source material but falls somewhat outside of it. This causes more particles to be born directly into the scattering blocks and results in more transport events that contribute to longer transport times.

We also investigated the angular nature of the residual source in a cell. Again using Monte Carlo integration, we compute the cell-averaged P_1 expansions of the residual source for each cell,

r^{cell}, to obtain the expansion of the residual source in the form

$$r^{cell}(\mathbf{\Omega}) = \frac{1}{4\pi}(r_{00}^{cell} + 3\mathbf{\Omega} \cdot \mathbf{J}_r) \tag{5.6}$$

where r_{00}^{cell} is the zeroth-moment and $\mathbf{J}_r$ is formed from the moments at $\ell = 1$. The anisotropy of r^{cell} can be characterized by computing an *anisotropy index*, I_a, as

$$I_a = \frac{|\mathbf{J}_r|}{r_{00}^{cell}}, \tag{5.7}$$

which varies $I_a \in [0, 1]$ with $I_a = 0$ indicating an isotropic nature and $I_a = 1$ indicating a pure delta-function anisotropic nature.

A visualization of the cell-wise currents $\mathbf{J}_r$ (as arrows), overlaid on a color scale representing the anisotropy index, I_a, is shown in the center of Figure 5.18. On either side of this visualization is a magnification of the source material region, the left is colored by the cell-wise absolute residual source strength and the right is colored by the anisotropy index (same as the center visualization). The left magnification highlights that some source particles are indeed born in the scattering blocks and from both the size of the arrows and the correspondence with the right magnification we can see that they are born nearly isotropically when in the scattering materials. The residual source directly within the source material is more anisotropic. The overall effect here results in more particles proportionally created in the direction of the scattering blocks as opposed to isotropically in the SMC method. These anisotropy effects, in combination with the more distributed nature of the residual source, are the likely cause for the dramatic increase in TrT.

Increasing the resolution of the angular discretization used in the DO simulation, from a Gauss-Legendre-Chebyshev with 6 polar- and azimuthal angles per octant to 64 azimuthal- and 6 polar angles, resulted in a general increase in the per particle efficiency.

We now seek an explanation of why the load factor for QOI-1 is so much larger than that for QOI-2. We continue this investigation in the discussions of the next problem.

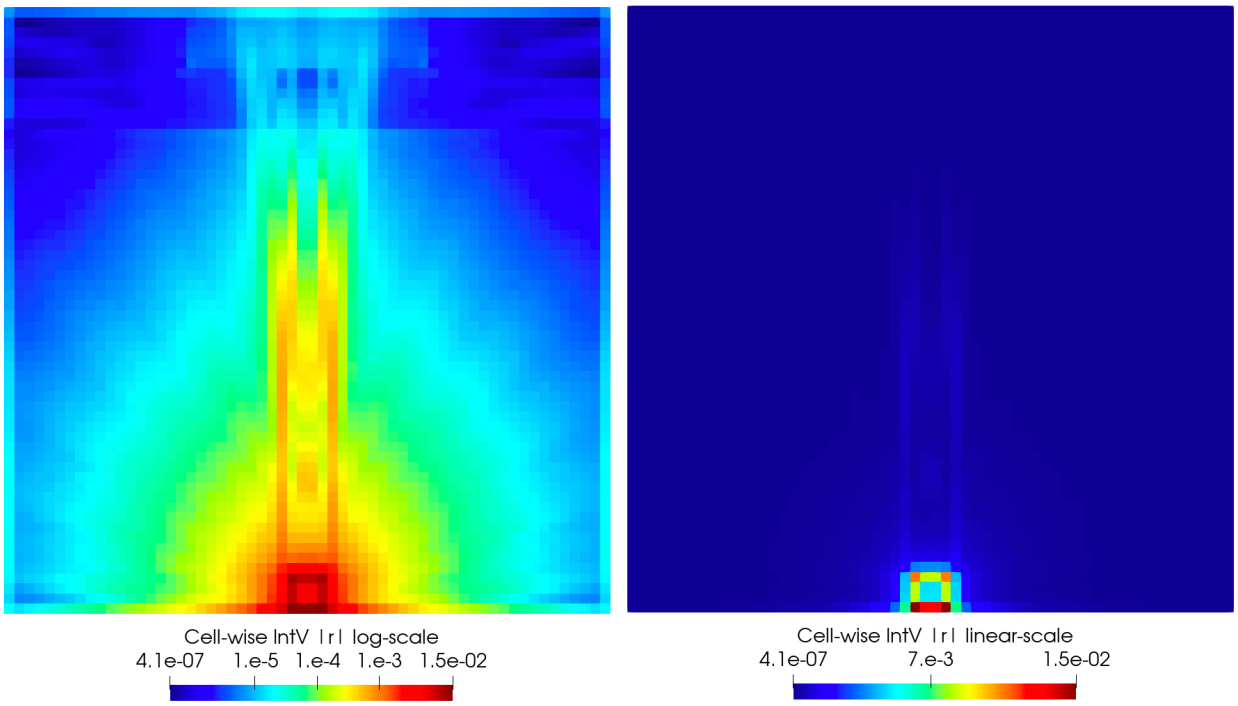

Figure 5.17: Cell-wise residual source for problem 4, showing a logarithmic scale [Left] and a linear scale [Right].

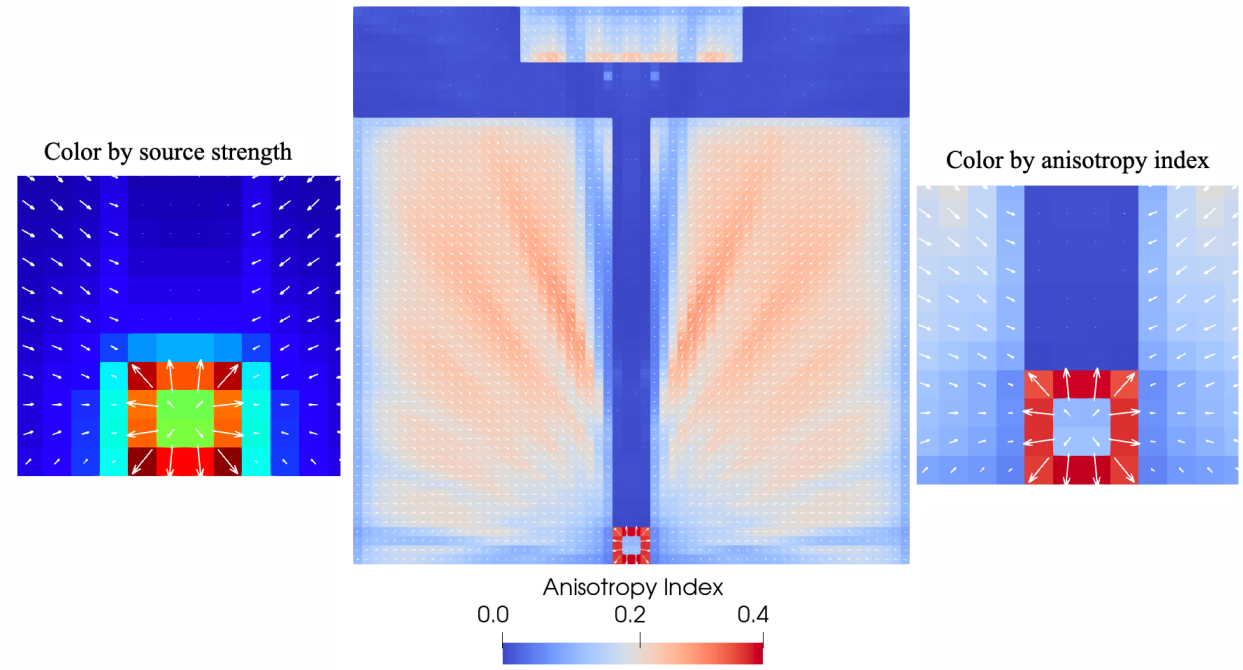

Figure 5.18: Vector-based visualizations of the P_1 expansion of the residual source moments of problem 4.

The overall insights gained in this section are as follows:

- The ducted problem with a source at its end is a challenging problem for DO simulations because of ray effects. The RMC method shows very high effectiveness for QOIs that are sensitive to these ray effects (QOI-1 in this case) with a very high gain factor ($G = 12.122$).
- The residual source can be very anisotropic and in combination with its distribution can alter the transport characteristics quite significantly, resulting in significantly increased transport times, TrT.

5.6 Problem 5 - A distributed-source variation of problem 4.

Consider the problem shown in Figure 5.19. The problem is nearly identical to problem 4 with the exception that the localized source is now distributed laterally.

The problem is designed to test a hypothesis relating to the reduced efficiency of the RMC method for QOI-2 in the previous problem, which has the same geometric position as QOI-2 in this problem. The hypothesis we are testing here is that the reduced effectiveness is a consequence of the similarity in source distribution.

From the SMC perspective of QOI-2, the localized source is behind the scattering blocks and particles likely to reach its volume can just as likely come from multiple scattering tracks within the blocks as it can from the duct. From the RMC perspective of QOI-2, the source is very similarly localized as can be seen in the linear-scale visualization of Figure 5.17. Therefore, the RMC method could be solving a very similar problem as SMC and therefore not exhibit the same drastic improvement as observed for QOI-1.

We verify this argument by using a distributed source that is likely to create an RMC source distribution that will be even more similar to that of the SMC problem (as compared to problem 4). If the source perspective is the likely cause then we should observe the same reduction in efficiency for QOI-2.

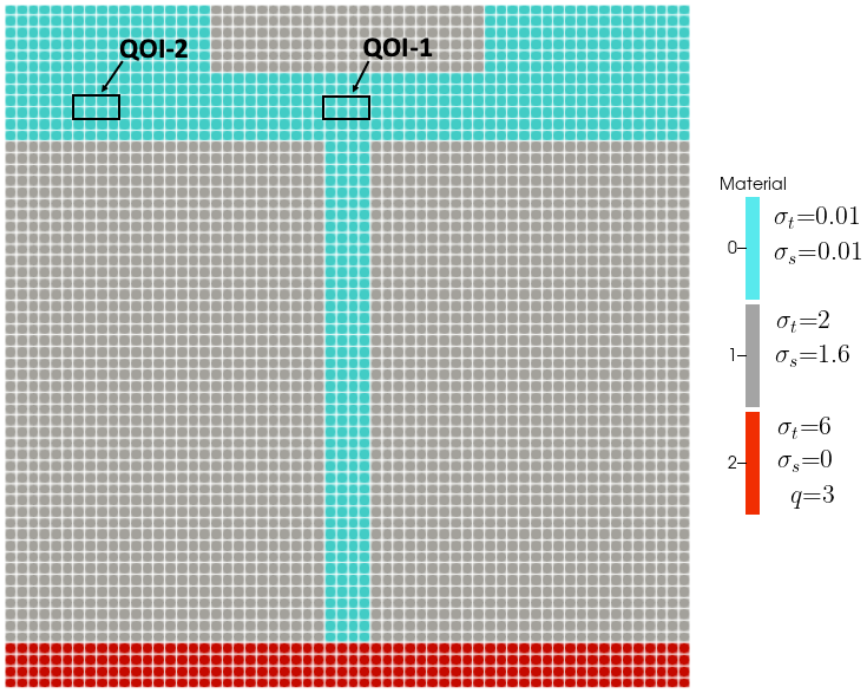

Figure 5.19: Variation of problem 4 where the localized source is now distributed laterally.

A DO solution, again using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant, is shown in Figure 5.20. The corresponding SMC solution is shown in Figure 5.21. The RMC computed defect solution, using the DO solution, is shown in Figure 5.22. As we observed in problem 4, the defect solution again shows significant defect in the vicinity of the source material, hinting at insufficient spatial discretization. The large defect in the duct is similar to that of problem 4.

The related performance data is shown in Table 5.5. The KPIs for all three QOIs again shows enhanced efficiency as compared with SMC, however, for this problem the efficiency, η, for QOI-2 is even lower than that of problem 4, at 1.604. This low efficiency is insufficient to overcome the combined increased source-sampling and transport times, resulting in an overall effectiveness factor, $E < 1$, for QOI-2. The other two QOIs still show a significant gain $E >> 1$. The increased angular resolution of the DO solution shows a further increase in the efficiency with all other KPIs similar.

Apart from the increased source sampling times (which we already know is caused by the rejection sampling), the transport times, TrT, for the RMC method is considerably more than that of the SMC method. This is caused by the residual source being distributed deeper into the scattering blocks. We tested this phenomenon (not shown) by reflecting the problem longitudinally at the bottom boundary, which removes much of the boundary escape for both the RMC and SMC. This then resulted in the transport times being more equal. This phenomenon is an important consideration for the application of the RMC method.

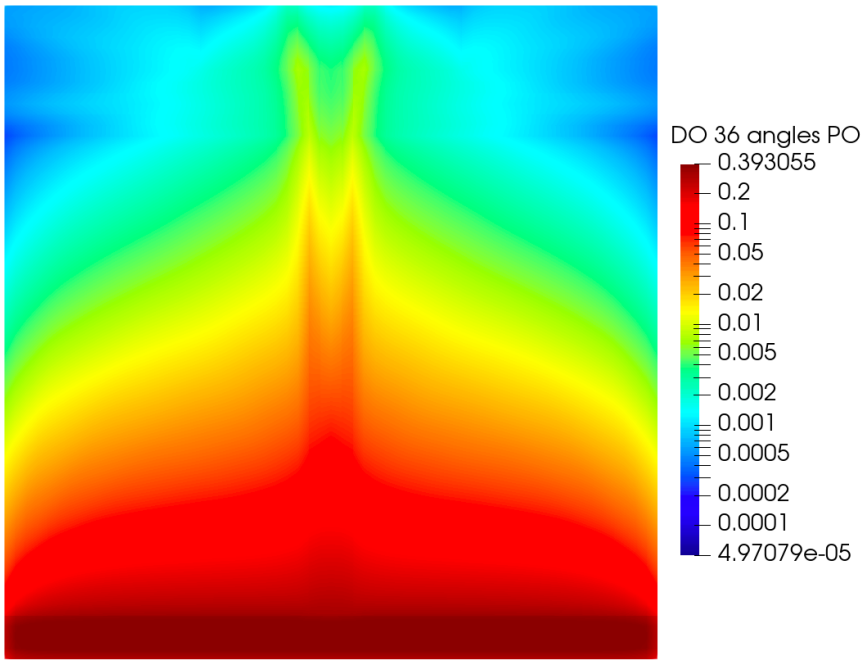

Figure 5.20: DO solution of problem 5 using a GLC angular quadrature with 6 polar- and 6 azimuthal-angles per octant.

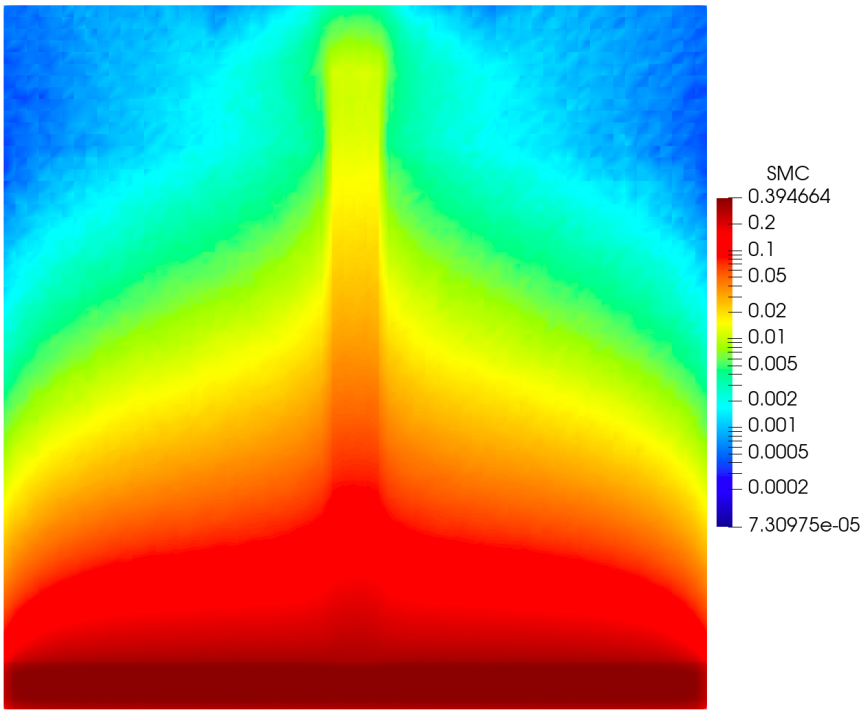

Figure 5.21: Standard Monte Carlo solution of problem 5 using 10×10^6 source particles.

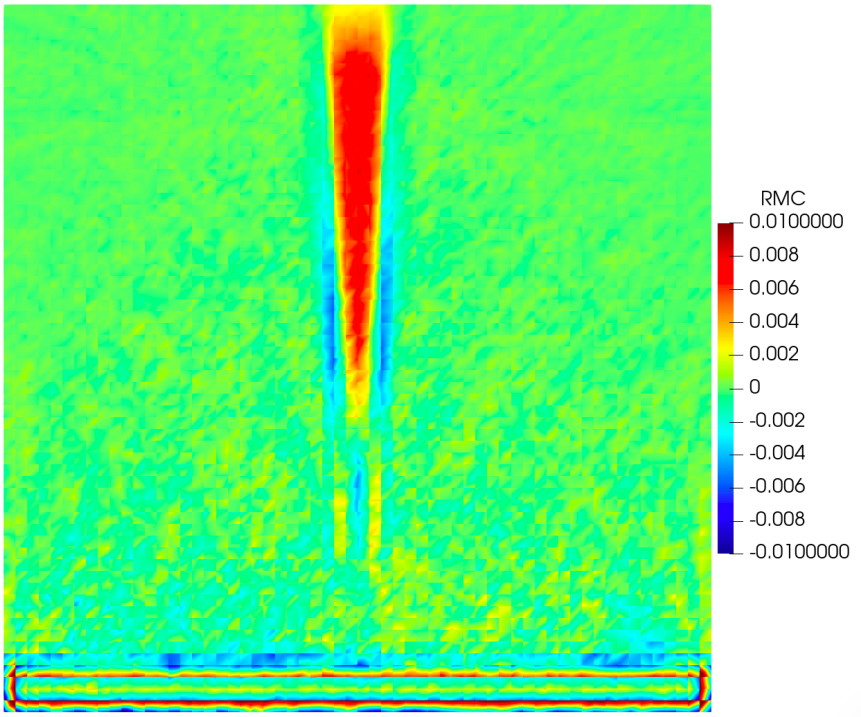

Figure 5.22: The RMC defect solution, for a DO simulation using a GLC angular quadrature (6 polar- and 6 azimuthal-angles per octant), of problem 5 using 10×10^6 source particles.

Table 5.5: Key Performance Indicators for problem 5. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	$\frac{\sigma_{SMC}}{\sigma_{RMC}}$	$\eta^{RMC \to SMC}$	$\frac{ST_{SMC}}{ST_{RMC}}$ **[s per 10^6]**	$\frac{TrT_{SMC}}{TrT_{RMC}}$ **[s per 10^6]**	$\frac{TT_{SMC}}{TT_{RMC}}$ **[s per 10^6]**	**Rate ratio**	$E^{RMC \to SMC}$
L2 norm of ϕ_D	1.560E-03		0.699	9.586	10.286		
$\tilde{\phi}$ Q1 Cont. GLC6,6	5.100E-04	9.356	3.540	25.968	29.508	0.349	3.261
QOI-1	3.540E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC6,6	1.450E-04	5.960	3.540	25.968	29.508	0.349	2.078
QOI-2	8.510E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC6,6	6.720E-05	1.604	3.540	25.968	29.508	0.349	0.559
L2 norm of ϕ_D	1.560E-03		0.699	9.586	10.286		
$\tilde{\phi}$ Q1 Cont. GLC64,6	4.863E-04	10.291	3.766	22.901	26.667	0.386	3.969
QOI-1	3.540E-04		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC64,6	1.343E-04	6.950	3.766	22.901	26.667	0.386	2.681
QOI-2	8.510E-05		-	-	-		
$\tilde{\phi}$ Q1 Cont. GLC64,6	6.639E-05	1.643	3.766	22.901	26.667	0.386	0.634

We now look at the residual source visualization as shown in Figure 5.23. Compared to the source distribution shown in Figure 5.19 we can see that the sources are very similarly distributed. Therefore, as we discussed in the beginning of this section, from the perspective of QOI-2 the source distribution looks the same for the RMC- and SMC method.

Another insight into the increased transport times can be gained by looking at the anisotropy of the residual source which is shown in Figure 5.24, magnified to more closely observe the bottom source region. We can see that the residual source is quite anisotropic in the source material, resulting in a proportionally larger amount of source particles transported in the direction of the scattering media, as compared to the isotropic SMC source. These particle will then experience more transport events and therefore longer average transport time.

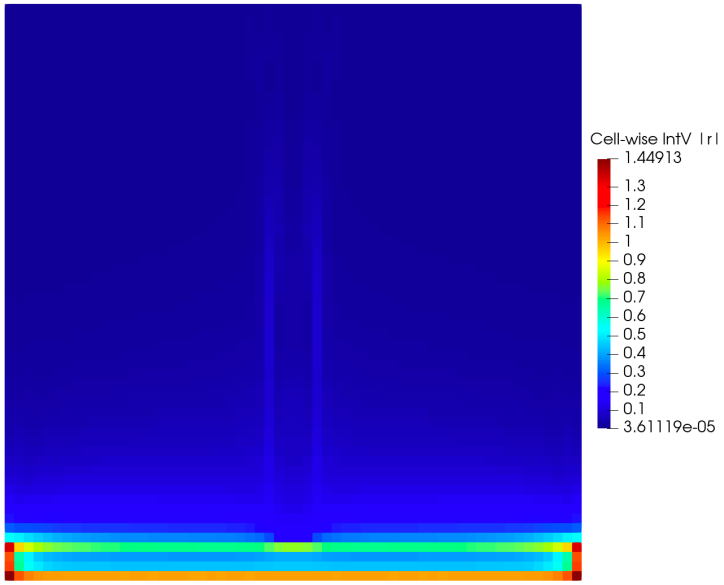

Figure 5.23: Cell-wise residual source for problem 5.

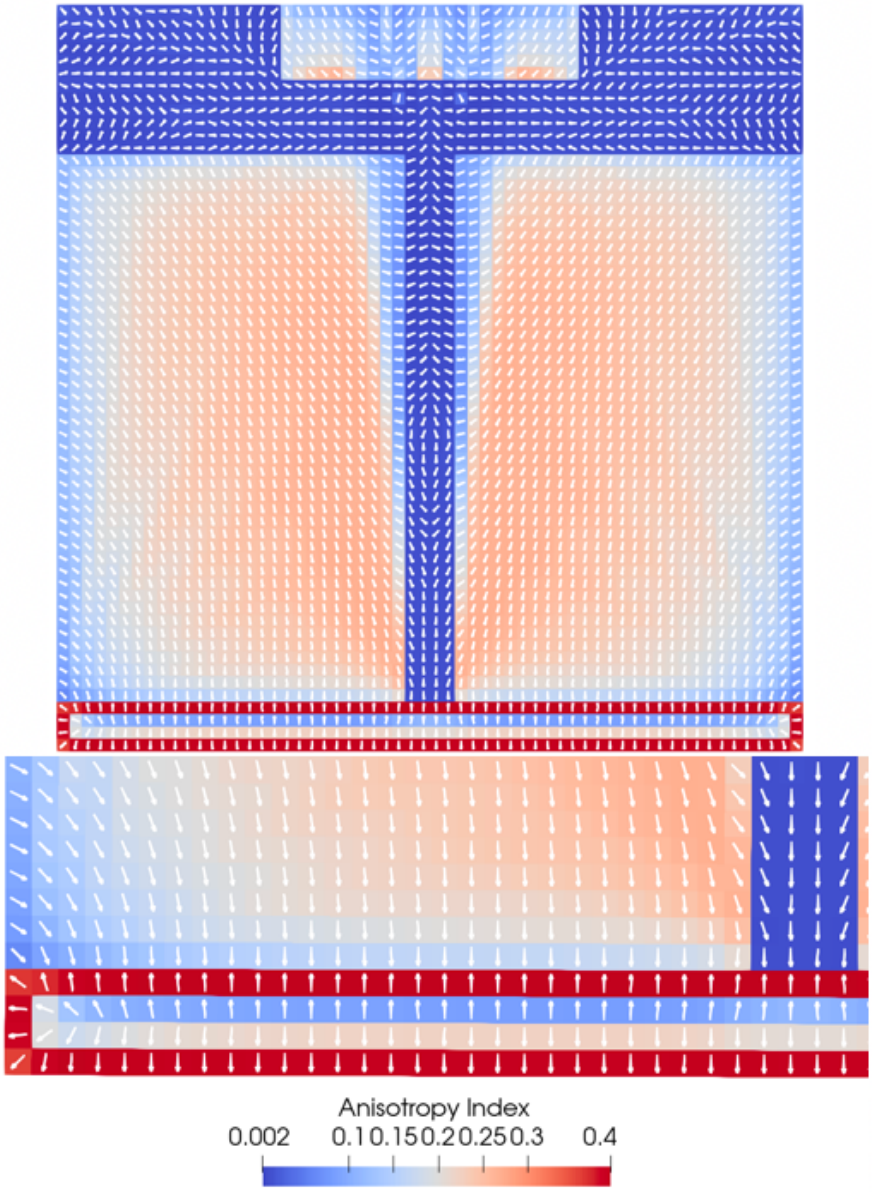

Figure 5.24: Vector-based visualizations of the P_1 expansion of the residual source moments of problem 5.

Insights gained from this section:

- The distribution and anisotropy of the residual can greatly influence the transport times, TrT, and can easily result in reduced overall effectiveness.

- The SMC method can become competitive if the perspective of the QOI is such that the SMC

and RMC problems appear similar. QOIs in areas that are relatively insensitive to spatial- and angular-discretization seem to not benefit much from the RMC method.

5.7 Adjoint-based source-biasing for problem 5

Since the residual source takes on a more distributed nature as compared to the normal problem, we attempted to apply adjoint-based source biasing in order to improve the efficiency of the lowest-efficiency QOI in problem 5. We will not detail how the adjoint problem is formulated which is succinctly detailed in [1]. Instead, we will simply summarize the methods outcome.

It should be noted that certain expectations are inherent to source biasing. Biasing the source with any manner of importance, be it cell-wise and/or angular importance, does only guarantee a higher probability of a source particle scoring within a QOI. It does not guarantee any outcome with regards to changes in the transport-time of the biased source particles, and in some sense one can comprehend that source biasing attempts to preferentially keep particles alive that are more likely to contribute to a QOI. It would therefore not be unexpected to observe an increase in transport time.

5.7.1 The adjoint-flux

The adjoint flux, $\psi^\dagger$, obtained by solving the adjoint problem, is not developed using the propagation of a source in a classical sense but rather uses the QOI as a source-analogue. Interesting properties of the adjoint solution is that the scalar adjoint flux, $\phi^\dagger$, is representative of the overall importance to the QOI and the angular adjoint flux, $\psi^\dagger$, is representative of the angular importance to the QOI. The product of this solution with a source distribution provides a means to bias the source sampling process to proportionally sample more particles from high importance regions.

Figure 5.25 shows the cell-averaged adjoint scalar flux distribution, in both logarithmic- and linear-scale, for problem 5. This adjoint solution was produced by a DO adjoint method using a GLC angular quadrature with 48 azimuthal angles and 6 polar angles. The distribution shows

that the areas closer to the QOI are of greater importance than those further away, which is an intuitive result. The logarithmic distribution also indicates that some importance is given to the duct, however, this is practically invisible in the linear plot.

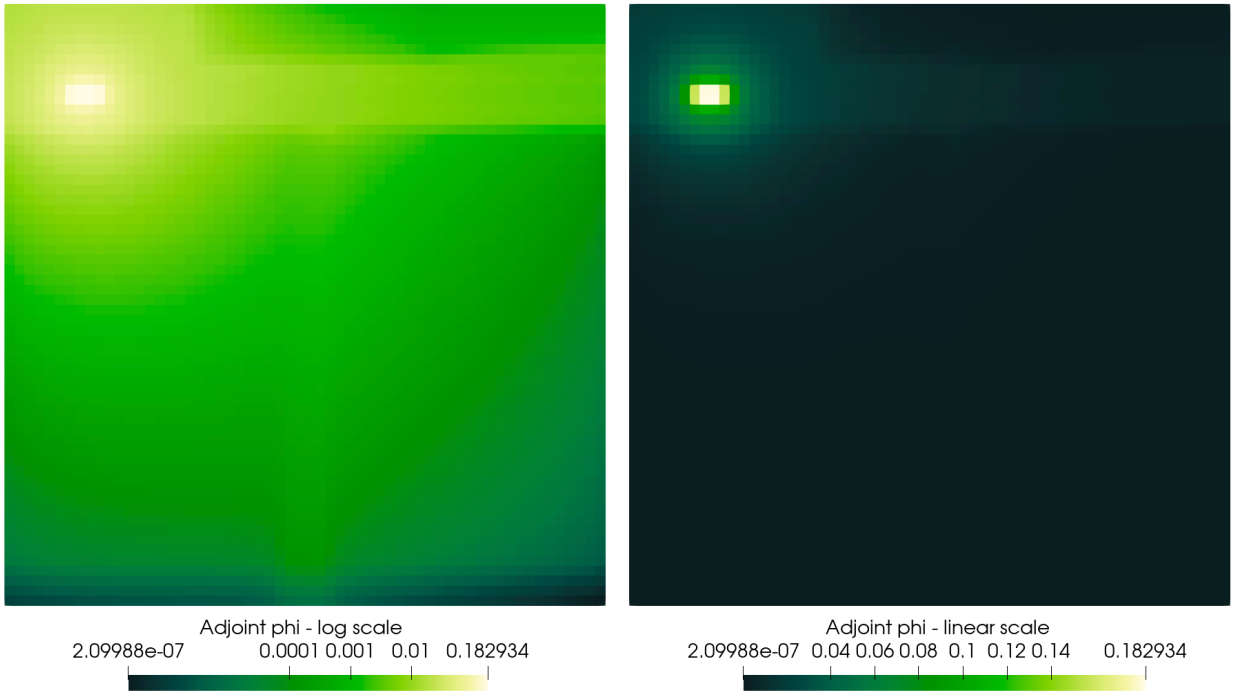

Figure 5.25: The scalar adjoint flux for problem 5, showing a logarithmic scale [Left] and a linear scalr [Right]. The solution is based on a DO simulation using a GLC angular quadrature with 48 azimuthal angles and 6 polar angles.

5.7.2 Abstraction of unbiased source sampling

In order to describe how we apply source biasing, knowing only the importance of a given cell, consider the following abstraction of the discrete PDF used to sample a collection of sources.

Given a source with index x of a total of N_x sources, a group with index g of which there are N_G^x number of groups for source x, and an element with index e of a total of N_E^{xg} elements for source x, group g. An element can be a cell, a cell-face-pair, or a point-source. Therefore

an element index, e, can map to a cell index, c, or a cell-face index-pair, (c, f), or just to 0 for a point-source, the latter which can additionally also map to which cell index c contains the source. Then we define

$$Q_e^{xg} = \begin{cases} S_{xg1} \ : \ e \mapsto 0 \mapsto c & , \text{ if source } x \text{ is a point-source} \\ \displaystyle\int_{\mathcal{D}_{c,f}} S_{xg(c,f)} \, dA \ : \ e \mapsto (c, f) & , \text{ if source } x \text{ is a surface-source} \\ \displaystyle\int_{\mathcal{D}_c} S_{xgc} \, dV \ : \ e \mapsto c & , \text{ if source } x \text{ is a cell-source,} \end{cases} \tag{5.8}$$

as the element-wise source strengths for source x, group g, where $e \in [0, N_E^{xg}-1]$. The group-wise source strengths, Q^x, and the source-wise source strengths, Q, are then

$$Q^x = \left\{ Q_g^x = \sum_{e=0}^{N_E^{x,g}-1} Q_e^{xg} \ : \ g \in [0, N_G^x-1] \right\}, \text{ and} \tag{5.9}$$

$$Q = \left\{ Q_x = \sum_{g=0}^{N_g^x-1} Q_g^x \ : \ x \in [0, N_x-1] \right\}. \tag{5.10}$$

For an unbiased simulation these source strengths can be used to define an unbiased normalized discrete PDF by computing the total source strength

$$Q^{total} = \sum_{x=0}^{N_x-1} \sum_{g=0}^{N_G^x-1} \sum_{e=0}^{N_E^{xg}-1} Q_e^x g \tag{5.11}$$

and then normalizing the source strengths,

$$p_e^{xg} = \frac{Q_e^{xg}}{Q^{total}}, \tag{5.12}$$

$$p^x = \left\{ p_g^x = \sum_{e=0}^{N_E^{x,g}-1} p_e^{xg} \ : \ g \in [0, N_G^x-1] \right\}, \text{ and} \tag{5.13}$$

$$p = \left\{ p_x = \sum_{g=0}^{N_g^x - 1} p_g^x \ : \ x \in [0, N_x - 1] \right\}, \tag{5.14}$$

where p_e^{xg} is the discrete value of the global PDF for source x, group g, element e, p^x is the group-wise discrete PDF, and p is the source-wise discrete PDF.

These PDFs are easily used to compute discrete CDF from which to sample the combination of a source-group-element.

5.7.3 Applying the scalar adjoint flux as source biasing on a cell-by-cell basis

Applying source biasing will give us the biased equivalents, Q^* and p^*, of all unbiased quantities Q and p, as

$$Q^{*x} = \left\{ Q_g^{*x} = \sum_{e=0}^{N_E^{x,g} - 1} \phi^\dagger Q_e^{xg} \ : \ g \in [0, N_G^x - 1] \right\}, \text{ and} \tag{5.15}$$

$$Q^* = \left\{ Q_x^* = \sum_{g=0}^{N_g^x - 1} Q_g^{*x} \ : \ x \in [0, N_x - 1] \right\}. \tag{5.16}$$

For an unbiased simulation these source strengths can be used to define an unbiased normalized discrete PDF by computing the total source strength

$$Q^{*total} = \sum_{x=0}^{N_x - 1} \sum_{g=0}^{N_G^x - 1} \sum_{e=0}^{N_E^{xg} - 1} Q_e^x g \tag{5.17}$$

and then normalizing the source strengths,

$$p_e^{*xg} = \frac{Q_e^{*xg}}{Q^{*total}}, \tag{5.18}$$

$$p^x = \left\{ p_g^{*x} = \sum_{e=0}^{N_E^{x,g} - 1} p_e^{*xg} \ : \ g \in [0, N_G^x - 1] \right\}, \text{ and} \tag{5.19}$$

$$p^* = \left\{ p_x^* = \sum_{g=0}^{N_g^x - 1} p_g^{*x} \ : \ x \in [0, N_x - 1] \right\}, \tag{5.20}$$

where p_e^{*xg} is the discrete value of the global biased PDF for source x, group g, element e, p^{*x} is the group-wise discrete biased PDF, and p^* is the source-wise discrete biased PDF.

Required weight correction:

In order to compensate for the biased sampling, each time a particle's parent element is identified, the weight of that particle needs to be modified to preserve the relative contribution of the particle to any tallies. For instance, if an element is biased such that it is sampled twice as often, then the weight needs to be halved in order to preserve its relative contribution to tallies. For our source biasing scheme this involves the element weight correction, w_{bcorr}^{xge}, as

$$w_{bcorr}^{xge} = \frac{p_e^{xg}}{p_e^{*xg}}, \tag{5.21}$$

which is the ratio of the unbiased discrete PDF evaluation to the biased discrete PDF evaluation for the sampled index.

5.7.4 Results using only the scalar adjoint flux, $\phi^\dagger$, for biasing

Since the adjoint solution has been developed for QOI-2 we only discuss the relevant results for this QOI. In order to make a fair comparison of the effect of the adjoint-based biasing we have to apply the same source biasing to the SMC simulation as we do the RMC simulation.

The relevant performance data is shown in Table 5.6. The first row of the table is a copy of the corresponding data in Table 5.5 since it represents the unbiased simulations. The second row of the table is the scalar adjoint flux biased counterpart of the first row. For both simulations the uncertainties for QOI-2 reduced by approximately 20%. The relative efficiency of the RMC method only increased slightly, from 1.6 to 1.7. The source sampling time for both simulations remained relatively unchanged. This is not surprising since the scalar adjoint flux biasing influences only the selection of the element not the sampling algorithm within the element. This means that the rejection sampling technique within a cell performed the same in the RMC method.

Both simulations showed an increase in transport time, although, the transport time increased more than two-fold for the RMC simulation. The increase in transport time for the SMC simulation is due to the source being sampled further away from the boundaries, however, since the source material is very localized, the effect is not pronounced. The increase in transport time for the RMC simulation is more drastic because the source is more distributed, with large portions in the scattering media. This causes inherently much more particles to be sampled directly within the scattering media. This effect can be visualized by looking at the product of the scalar adjoint flux and the cell-by-cell absolute residual source strength, essentially the product of Figure 5.23 and Figure 5.25, as shown in Figure 5.26. This figure shows that considerably more particles will be created in the scattering medium.

To further analyze the increase in the RMC TrT we show particle tracks, for the first 4000 source particles, in Figure 5.27. The two figures on the left are for SMC and those on the right are for RMC. The top row shows the tracks for the unbiased sources and the bottom row shows the tracks for the source biased with $\phi^{\dagger}$. Source points are denoted with red dots.

As expected, the SMC source distribution is confined to the source material in both biased and unbiased cases. In the biased SMC case we can see that more particles are sampled in the higher y-coordinate but still confined to the source material. This is because the adjoint-importance generally increases as a function of the y-coordinate in this region. The overall effect of biasing for SMC is that it results in more particles entering the scattering medium and therefore we see more particle tracks when SMC is biased, increasing from 19,270 tracks, in the unbiased case, to 30,727 tracks in the biased case. This is the cause of the increase in TrT for SMC, from 9.586 to 11.930.

The unbiased RMC source-points and tracks reflect the residual source shown in Figure 5.23. Since there are significantly more particles born in the scattering medium it is not surprising that the number of tracks is high, 42,445, as compared to the unbiased SMC, which explains why the unbiased RMC TrT is much higher than the unbiased SMC TrT, 25.968 vs 9.586 respectively. The biased RMC source-points and tracks reflect the distribution shown in Figure 5.26 and we observe an even higher amount of source particles created directly in the scattering medium, re-

sulting in a tremendous amount of tracks, 125,355 in total. This explains why the biased RMC TrT increases to 56.822.

Table 5.6: Key Performance Indicators for problem 5 with scalar adjoint flux biasing for QOI-2. All Monte Carlo simulations here were performed using 10^7 source particles.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	σ_{SMC} σ_{RMC}	$\eta^{RMC \to SMC}$	ST_{SMC} ST_{RMC} [s per 10^6]	TrT_{SMC} TrT_{RMC} [s per 10^6]	TT_{SMC} TT_{RMC} [s per 10^6]	**Rate ratio**	$E^{RMC \to SMC}$
QOI-2 unbiased	8.510E-05		0.699	9.586	10.286		
$\tilde{\phi}$ Q1 Cont. GLC6,6	6.720E-05	1.604	3.540	25.968	29.508	0.349	0.559
QOI-2 $\phi^\dagger$ biased	6.993E-05		0.701	11.930	12.632		
$\tilde{\phi}$ Q1 Cont. GLC6,6	5.367E-05	1.698	3.611	56.822	60.433	0.170	0.289

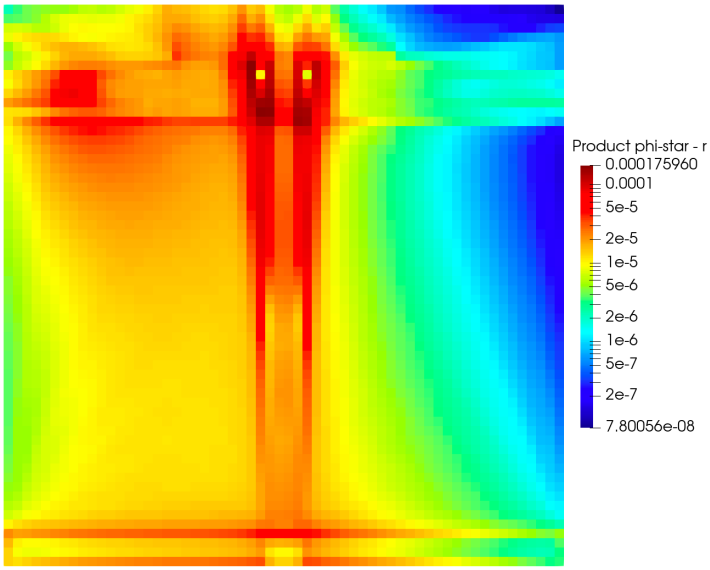

Figure 5.26: The product of $\phi^\dagger$ and $\int_V |r| dV$ for problem 5.

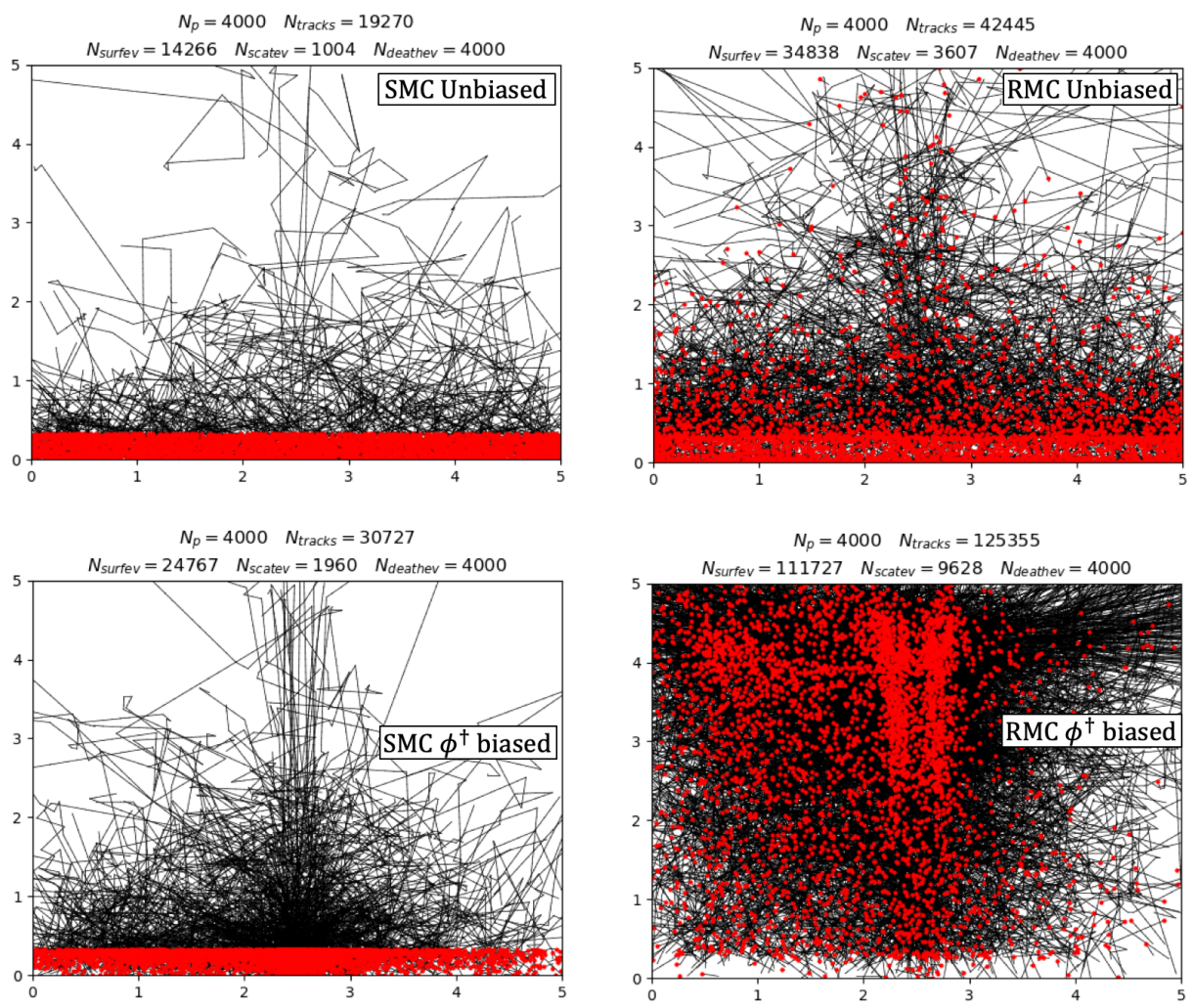

Figure 5.27: Particles tracks for SMC and RMC problem 5 of the first 4000 source particles. The RMC is based on a DO solution using a Gauss-Legendre-Chebyshev angular quadrature with 6 polar angles and 6 azimuthal angles. The red dots indicate source particle creationg. [Top-left] SMC unbiased. [Top-right] RMC unbiased. [Bottom-left] SMC biased with $\phi^\dagger$. [Bottom-right] RMC biased with $\phi^\dagger$. N_{tracks} is the total number of linear tracks drawn, N_{surfev} is the total number of surface-events drawn, N_{scatev} is the total number of scattering events drawn and $N_{deathev}$ is the total number of death events recorded.

5.7.5 Angular biasing using the P_1-expansion of $\psi^\dagger$

The adjoint DO simulation produces the harmonic expansion coefficients of the angular adjoint flux in the same fashion as the non-adjoint simulation. The P_1 order expansions allow us to write

$$\psi^\dagger(\mathbf{\Omega}) \approx \frac{1}{4\pi}(\phi^\dagger + 3\mu|\mathbf{J}^\dagger|) \tag{5.22}$$

where $\phi^\dagger$ and $\mathbf{J}^\dagger$ are formed from the P_1-moments and $\mu = \mathbf{\Omega} \cdot \dfrac{\mathbf{J}^\dagger}{|\mathbf{J}^\dagger|}$. This form of the is not very conducive to angular biasing since it provides a poor approximation of highly forward-peaked angular fluxes. We detail a method, in Appendix C, by which we can use the P_1 moments to construct a moment-preserving exponential representation of the expansion in the form,

$$\psi_e^\dagger(\mathbf{\Omega} \mapsto \mu) \approx \phi^\dagger e^{a+b\mu}, \tag{5.23}$$

where a and b are coefficients, determined as shown in Appendix C. As pointed out earlier, this representation is moment preserving such that

$$2\pi \int_{-1}^{1} \phi^\dagger e^{a+b\mu} d\mu = \phi^\dagger \text{ and}$$
$$2\pi \int_{-1}^{1} \phi^\dagger \mu e^{a+b\mu} d\mu = |\mathbf{J}^\dagger|.$$

We can make this distribution independent of $\phi^\dagger$ by normalizing $\psi_e^\dagger$ with $\phi^\dagger$ to get

$$\psi_{en}^\dagger(\mathbf{\Omega} \mapsto \mu) = e^{a+b\mu}, \tag{5.24}$$

and using $\psi_{en}^\dagger$ as the biasing function.

Applying this function to a biasing scheme requires a modification to the integrals used to determine the source strength within a cell. For SMC the cell-sources are constant and isotropic and since the integral over angle-space of Eq. (5.24) is 1 it does not influence the computation of the source strength. As shown in Appendix C this distribution can be sampled to obtain μ which is

then used in the SMC method to bias source sampling.

For RMC we apply this biasing function to Eq. 4.30 which becomes,

$$R_{gc}^{*interior} = \frac{4\pi V_c}{N_y} \sum_{j=1}^{N_y} \psi_{en}^{\dagger}(\mathbf{\Omega}_j) |r_g((\mathbf{x}, \mathbf{\Omega})_j)|, \tag{5.25}$$

from which we construct the cascade of discrete CDFs as we did before. Note that the biasing function can change the maximum of the integrated function, which we use during rejection sampling, and therefore we need to obtain a new maximum. Rejection sampling within the cell requires us to then evaluate $\psi_{en}^{\dagger}(\mathbf{\Omega})|r_g(\mathbf{x}, \mathbf{\Omega})|$ instead of just $|r_g(\mathbf{x}, \mathbf{\Omega})|$.

Required weight correction:

The required weight correction can be derived as

$$w_{acorr}^{gc} = \frac{R_{gc}^{*interior}}{R_{gc}^{interior} \psi_{en}^{\dagger}}. \tag{5.26}$$

5.7.6 Results using both the scalar adjoint flux, $\phi^{\dagger}$, and the angular adjoint flux, $\psi^{\dagger}$, for biasing

The relevant performance data is shown in Table 5.7. The first row of table is again a copy of the corresponding data in Table 5.5, i.e., the unbiased simulation KPIs. The second row is a copy of the second row in Table 5.6 ($\phi^{\dagger}$ biased simulations) and the third row is the KPIs for scalar- and angular adjoint flux biased simulations. For both the RMC and SMC simulations the uncertainties for QOI-2 reduced, however, the RMC simulation's reduction is nearly negligible. The source sampling times increased for both methods, although for SMC only slightly which can be attributed to minor changes to the localization of the SMC source. The transport times for the SMC simulation is higher with the angular biasing because particles are biased towards the inner scattering medium. The increased transport time for RMC is analyzed in further paragraphs.

SMC exhibits a greater decrease in uncertainty when $\psi^\dagger$ is applied in addition to $\phi^\dagger$, resulting in RMC's relative per-particle efficiency being lower than SMC. We have made many observations to try to understand why using $\psi^\dagger$ is not as effective for the RMC simulation, but we still can make no conclusions. From our observations we determined that an error in the algorithm is unlikely for the following reasons. The adjoint solution used for the biased SMC simulations is the same as that for the biased RMC simulations and the source biasing algorithms are identical. The correct QOI-value is obtained with both biasing techniques which, if the algorithm had an error, would not occur. Finally, simplified problems show that the angular biasing is indeed effective, showing no obvious signs of error. We left the analysis of this aspect as a topic for future work.

The increase in the RMC TrT to 62.557 is a point of interest, recalling that for the $\phi^\dagger$ biased RMC simulation this was 56.822. Intuitively one may think that, since the particles are now directed towards the QOI-region, the transport time would decrease. So why did it increase? To analyze this aspect we again created particle tracks for the first 4,000 source particles, for both SMC and RMC, as shown in Figure 5.29. As can be seen in the figure, the number of SMC- and RMC-tracks increased compared to the tracks when just $\phi^\dagger$ is used. For SMC this is because the source particles are now directed away from the bottom boundary when created and preferentially move in the direction of the scattering medium. For RMC the cause is not so clear.

To ascertain the contribution to TrT of particles created in the top portions of the low-density material (which includes the peak region in Figure 5.26 and everything with $4.0 \leq y \leq 4.5$) we sampled these particles as normal, but killed them directly after creation. This effectively removed the transport time spent on these particles. With this technique we determined that, for RMC, these regions contributed 13.4% of the total transport time with only $\phi^\dagger$ biasing, and 19.2% of the total transport time when biasing with $\phi^\dagger$ and $\psi^\dagger$. Now, when placing a reference point at the lateral central and longitudinally $y = 4.25$ we can observe than only a small percentage of the azimuthal angle space will escape the scattering material. If the angular biasing was pure δ-function in nature then indeed our intuition would be accurate in the fact that we would direct the particles directly at the QOI and see a considerable improvement in the 19.2% portion of the particles created.

However, the angular biasing in the peak regions has an anisotropy-index between 0.4 and 0.6 (see Figure 5.28), which will not direct the particles toward the QOI as well as we expect for an isotropy-index of close to 1.0, especially considering that the residual source is nearly isotropic in that region (see Figure 5.24).

Additionally, a non-intuitive aspect of the angular biasing is the effect, in the scattering medium, at the boundaries. Especially considering that the angular biasing is almost exclusively towards the inner part of the domain (see Figure 5.28). Near the bottom boundary source particles are biased inward quit strongly, with an anisotropy-index of approximately 0.89. Additionally, one sees moderately high anisotropy-indices of approximately 0.6 at the left and right boundary where the residual-source has a comparatively small anisotropy-index < 0.1. The inwardly directed particles have a greater probability of surviving compared to isotropically sampled particles and therefore the contribution to TrT of these inwardly directed particles is likely overthrowing the benefit of the angular biasing at the biased-source peak.

Table 5.7: Key Performance Indicators for problem 5 with scalar adjoint flux and angular adjoint flux biasing for QOI-2. All Monte Carlo simulations here were performed using 10^7 source particles. The sub-table below the main table is a summary of the downward trend of the uncertainty with application of additional biasing.

	QOI std. dev	Efficiency	Sampling Time	Transport Time	Total Time		Overall effectiveness
QOI	$\frac{\sigma_{SMC}}{\sigma_{RMC}}$	$\eta^{RMC \to SMC}$	$\frac{ST_{SMC}}{ST_{RMC}}$ **[s per 10^6]**	$\frac{TrT_{SMC}}{TrT_{RMC}}$ **[s per 10^6]**	$\frac{TT_{SMC}}{TT_{RMC}}$ **[s per 10^6]**	**Rate ratio**	$E^{RMC \to SMC}$
QOI-2 unbiased	8.510E-05		0.699	9.586	10.286		
$\widetilde{\phi}$ Q1 Cont. GLC6,6	6.720E-05	1.604	3.540	25.968	29.508	0.349	0.559
QOI-2 $\phi^\dagger$ bias	6.993E-05		0.701	11.930	12.632		
$\widetilde{\phi}$ Q1 Cont. GLC6,6	5.367E-05	1.698	5.844	54.589	60.433	0.170	0.289
QOI-2 $\phi^\dagger$ $\psi^\dagger$bias	4.897E-05		0.738	10.476	11.215		
$\widetilde{\phi}$ Q1 Cont. GLC6,6	5.350E-05	0.838	4.110	62.557	66.667	0.154	0.129

	σ_{SMC}	σ_{RMC}
No bias	8.510E-05	6.720E-05
$\phi^\dagger$ bias	6.993E-05	5.367E-05
$\phi^\dagger$ $\psi^\dagger$bias	4.897E-05	5.350E-05

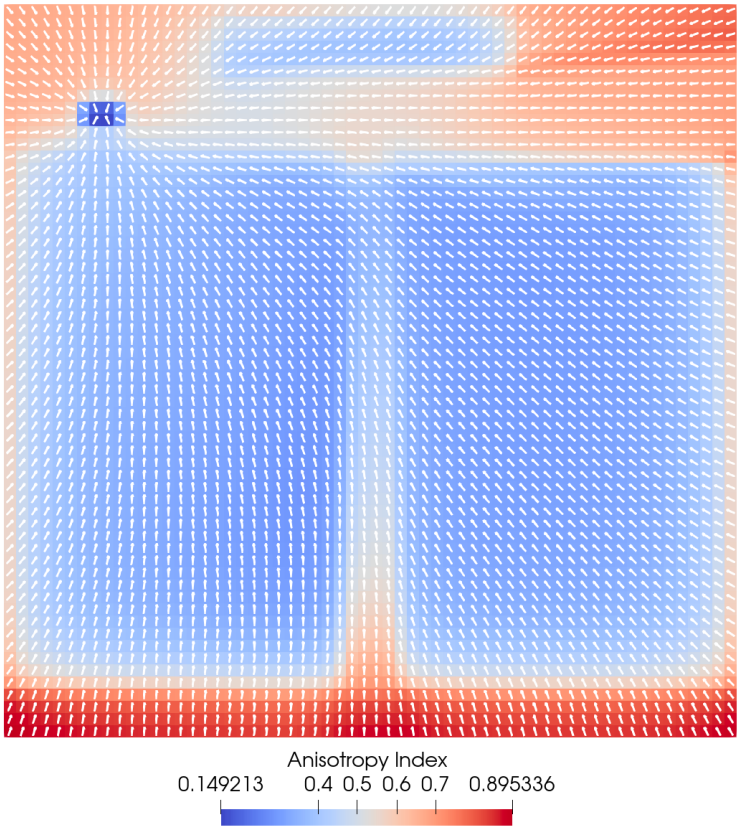

Figure 5.28: The anisotropy index for the P_1 expansion of the adjoint flux for problem 5.

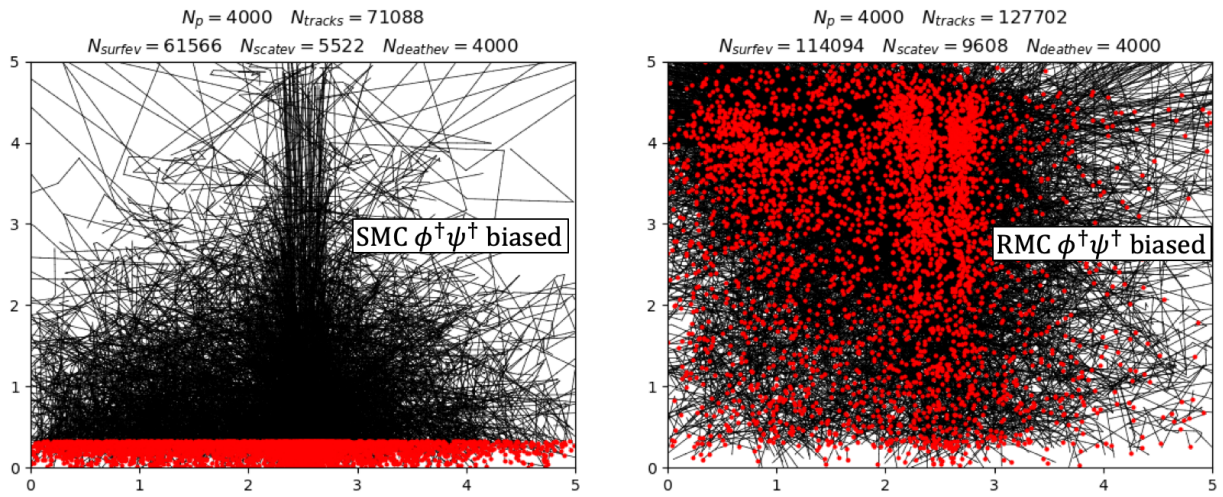

Figure 5.29: Particles tracks for SMC and RMC problem 5 of the first 4000 source particles. The RMC is based on a DO solution using a Gauss-Legendre-Chebyshev angular quadrature with 6 polar angles and 6 azimuthal angles. The red dots indicate source particle creationg. [Left] SMC biased with $\phi^\dagger$ and $\psi^\dagger$. [Right] RMC biased with $\phi^\dagger$ and $\psi^\dagger$. N_{tracks} is the total number of linear tracks drawn, N_{surfev} is the total number of surface-events drawn, N_{scatev} is the total number of scattering events drawn and $N_{deathev}$ is the total number of death events recorded.

5.7.7 Overall insights from adjoint-based biasing

The overall insights gained from adjoint-based source biasing:

- Since the residual-source is domain-wide, biasing the source with $\phi^\dagger$ can potentially create a very different source-distribution compared to the unbiased residual-source. This can greatly impact the total transport time of a given RMC simulation.

- Source biasing, in general, is a variance reduction technique. The response of SMC and RMC to source biasing, just like many variance reduction techniques, could potentially be very different depending on the problem. For this problem SMC responded similar to RMC, when using $\phi^\dagger$ for source biasing, with the goal of reducing QOI uncertainty. Additionally using $\psi^\dagger$ for angular biasing, SMC had an additional reduction in uncertainty much larger than for RMC. Because of the distributed nature of RMC, this problem forced more particles

to be created directly within the scattering medium which resulted in additional transport time per source particle, ultimately rendering RMC less effective than SMC.

- Source biasing makes no guarantees regarding total simulation time. It only increases the probability that a source particle will score a contribution within a QOI. Therefore, source biasing for the purpose of reducing the required run-time needs to be applied judicially.

- The single attempt at variance reduction does not rule out the possibility of using other advantageous variance reduction schemes, however, the exploration of such techniques was left as future work.

6. CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In this dissertation our research involved the following novelties:

1. A continuum residual definition of the multigroup linear Boltzmann equation (with space and angle in continuum form), where we use only the scalar flux from an approximate solution as a continuum definition of the angular flux. This provides a strong connection to the scalar flux component of an approximate solution, produced by a Discrete Ordinates (DO) simulation, and the associated defect in the scalar flux even though we do not use the full fidelity of the angular flux.

2. We created normalized discrete probability distributions of the residual source by using multivariate Monte Carlo integration and related the uncertainty of the Monte Carlo integration to the simulation result by recomputing the source strengths at batched intervals. Therefore the Monte Carlo integration uncertainty reflected in the uncertainty of the simulations. This negated the need for overly biased source particle creation and essentially allowed unity weights for source particles, as opposed to other work in literature which sampled an equal amount of particles in each cell, weighted by their respective residual source strength.

3. We explored different spatial representations of the approximate solution based on the DFEM solution produced by the DO simulation. We first used a cell-constant solution computed as the cell-average of the DFEM approximate solution which created numerous and strong inter-cell discontinuities. We showed that this representation introduces large positive- and negative components of the scalar flux defect even if the cell-averages are exact, resulting in very poor efficiency of the RMC method. We then used the linear-, in 1D, and piecewise-linear (PWL), in 2D and 3D, discontinuous FEM representations of the approximate solution, as produced by a DO simulation, and compared that to the continuous counterparts of the

these representations (determined from nodal averages). This indicated that the continuous FEM representations are the most efficient representations to use. We therefore showed that one does not have to use the exact same solution as produced by a DO simulation, in other words one can make advantageous modifications to the approximate solution, as long as the modifications can be related to the original solution.

4. Using just the continuous FEM representations, we then applied the RMC method to several multidimensional problems. The problems ranged from homogeneous pure absorber problems to multi-material problems with highly scattering media and low optical thickness "ducts" that greatly challenged the DO simulations.

   The L_2-norm of the defect across the entire domain and two localized QOIs was used to judge the efficiency and effectiveness of the RMC method. We defined the RMC efficiency as the ratio of additional particles an SMC method has to simulate to obtain the same uncertainty for a QOI as obtained as the RMC method. This can also be regarded as the "per source particle" efficiency of the RMC method compared to that of the SMC method. Since the RMC method incurred additional source sampling cost and additional average transport times we arrived at the definition of the overall effectiveness of the RMC method as the product of the efficiency and ratio of total simulation time.

5. Across all unbiased problems, the RMC efficiency for all the QOIs indicated that the RMC method has superior efficiency compared to SMC. This was the same for the overall effectiveness across all problems and QOIs except for a single problem where the efficiency gain of a single QOI was offset by significant increases in total simulation time. The latter hinted at the possibility that for some problems not all QOIs will benefit from using the RMC method, unless of course the source sampling time-increase can be improved.

6. The simulations, where the true defect of the DO simulation was comparatively small, exhibited drastic increases in efficiency and were highly effective. This is indicative that the RMC method becomes increasingly effective when the discretization errors are small, making the

method a suitable method for uncertainty quantification.

7. The ducted simulations, where the QOI along the duct is particularly sensitive to angular discretization, showed drastic increases in efficiency and effectiveness. This makes the RMC method a beneficial method to use when angular discretization proves challenging for the DO simulation.

6.2 Future work

6.2.1 Improvements in source sampling

The rejection sampling within a cell, or cell-face pair on a boundary, proves to be comparatively costly. Techniques could be available to improve the rejection efficiency by establishing bounding-functions over the individual domains of the independent variables, i.e., position and angle.

In the current implementation we uniformly sample position and angle simultaneously, then evaluate the residual. Bounding-functions can be used on course intervals of phase-space to limit the size of the individual space- or angular phase-space domain to sample. For example we could break polygonal or polyhedral cells into their constitutive primitive cells, i.e., triangles and tetrahedra respectively, which will establish different and likely less voluminous phase-space domains, per element, to sample.

6.2.2 Using mesh-based residual source with parametric surface based geometry and QOIs

In this research we performed Monte Carlo simulations of a given problem on the same mesh used to perform a DO simulation. This allowed us to project the Monte Carlo solutions onto the same space as the DO solution and therefore enabled comparison of these solutions over the *entire* domain. The unfortunate consequence was that the entire domain was laden with volumetric tallies, which required tally contribution algorithms to be applied every time a particle track occurred within each tally. These tallies were also defined such that we could make a finite element

projection of the domain-wide solution, adding additional layers of processes to the algorithms. The large amount of tallies as well as the finite element aspects of these tallies added a tremendous cost to the transport algorithms. Many practical problems are not concerned with the solution over the entire domain, and especially not a finite element representation of this solution. Instead they require the solution integrated over a localized QOI sub-domain.

The standard mode of operation of production codes like MCNP and GEANT-4 is to represent the geometry with parametric surfaces, logically combined to define cells or regions with homogeneous material composition. The flux solution is then obtained only in a QOI reqion. In this mode, the cells require absolutely no discretization, resulting in a drastic reduction of the number of surfaces in the simulation, as well as tally contribution algorithms that do not have any finite element related processes. These aspects exhibit drastically reduced transport times.

If one is to develop a coarsened version of the DO mesh, maintaining the same material interfaces and overall QOI sub-domains, then one can use this coarsened mesh for the transport phase of Monte Carlo simulations. RMC can then become a very attractive and efficient method for uncertainty quantification of QOIs, because only the residual-source is required to use the refined DO mesh. Additionally, since one is only concerned with volume integrated solutions in QOIs, the finite element processes within tally contribution algorithms can be disabled. In fact, all cell-wise tallies can be disabled, except for those required by the QOIs. The simplified mesh, and simpler tally algorithms, would then result in a drastic increase in the Monte Carlo simulation performance since it reduces the overall transport time.

6.2.3 Different choices for the approximate angular flux other than the scalar flux

Our definition of the continuum residual required the assumption that the angular flux is isotropic. This simplified the residual greatly with respect to the angular dependence and the integral of the scattering source. We have yet to explore the efficiency of the RMC method if we used angular fluxes at the angular quadrature points and the harmonic expansion of the angular flux for the scattering source.

Although the expansion coefficients for harmonic expansions are typically stored for most problems, the angular flux is very memory consuming to store and is less likely to be stored. We have found some work in literature that used these two sets of solutions but they are normally restricted to 1D problems or very simple and small multidimensional problems.

6.2.4 The potential of computing energy discretization errors

In the process of developing discrete CDFs for sampling the residual source we applied cell-wise integrals of the absolute value of the residual source for each energy group. We used multivariate Monte Carlo integration to perform these integrals and therein realized several aspects. The first is that the multigroup structure of the DO solution does not necessarily have to dictate the multigroup structure of the defect solution. The second is that we can use continuous energy cross sections in Monte Carlo integrals, where the energy is uniformly sampled in the DO defined group interval. We can then project a coarse-group DO simulation's approximate solution onto a refined group-structure and determine the defect with respect to the refined group structure. This will allow us to compare reaction rates, in-turn allowing us to determine if any energy discretization errors are significant.

REFERENCES

[1] Lewis E.E, Miller W.F. *Computational Methods of Neutron Transport*, John Wiley & Sons, 1984, ISBN 0-471-09245-2

[2] Duderstadt J.J., Hamilton L.J., *Nuclear Reactor Analysis*, John Wiley & Sons, 1976.

[3] Sood A., *The Monte Carlo Method and MCNP - A Brief Review of Our 40 Year History*, Presented at the International Topical Meeting on Industrial Radiation and Radioisotope Measurement - Aplications Conference, Chicago IL, July 2017. LA-UR-17-26533.

[4] Werner C.J. et al., *MCNP6.2 Release Notes*, Los Alamos National Laboratory, report LA-UR-18-20808 (2018).

[5] Allison J. et al., *Recent Developments in Geant4*, Nuclear Instrumentation Methods, A 835 (2016) 186-225.

[6] Favorite J., Lichtenstein H., *Exponential Monte Carlo convergence of a three-dimensional discrete ordinates solution*, Transactions of the American Nuclear Society, 81, pp. 147-148, 1999.

[7] Halton J., *Sequential Monte Carlo*, Proceedings of Cambridge Philosophical Society, volume 58 pp. 57-78, 1962.

[8] Halton J., *Sequential Monte Carlo techniques for the solution of linear systems*, Journal of Scientific Computations, vol 9 pp. 213-257, 1994.

[9] Evans T.M., Urbatsch T.J., Lichtenstein H., Morel J.E., *A residual Monte Carlo method for discrete thermal radiative diffusion*, Journal of Computational Physics, Volume 189 no. 2 pp539-556, 2003.

[10] Booth T., *Exponential convergence on a continuous Monte Carlo transport problem*, Nuclear Science and Engineering, 127 (3) pp. 338-345, 1997.

[11] Lichtenstein H., *Exponential convergence rates for reduced-source Monte Carlo transport in [x,μ] geometry*, Nuclear Science and Engineering, 133, p. 11, 1999.

[12] Booth T., *Adaptive importance sampling with a rapidly varying importance function*, Nuclear Science and Engineering, 136 (3), pp. 399-408, 2000.

[13] Booth T., *An approximate Monte Carlo adaptive importance sampling method*, Nuclear Science and Engineering, 138 (1), pp. 96-103, 2001.

[14] Peterson J.R., Morel J.E., Ragusa J.C., *Residual Monte Carlo for the one-dimensional particle transport equation*, SIAM Journal of Scientific Computation, vol 38 no. 6, pp. B941-B961, 2016.

[15] Mebius J.E., *Derivation of the Euler-Rodrigues formula for three-dimensional rotations from the general formula for four-dimensional rotations*, arXiv General Mathematics 2007.

[16] Sloan D.P., *A New Multigroup Monte Carlo Scattering Algorithm Suitable for Neutral and Charged-Particle Boltzmann and Fokker-Planck Calculations*, SAND83-7094, PhD Dissertation, May 1983.

[17] Golub G.H., Fisher B., *How to generate unknown orthogonal polynomials out of known orthogonal polynomials*, Numerical Analysis Project manuscript NA-91-06, Computer Science Department - Stanford University, November 1991.

[18] Lux I., Koblinger L., *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press, 1991.

[19] Möller T., Trumbore B., *Fast, Minimum Storage Ray-Triangle Intersection*, Journal of Graphics Tools. 1997.

[20] Bailey T., Adams M.L., *The Piecewise Linear Discontinuous Finite Element Method applied to the RZ and XYZ Transport Equations*, PhD Dissertation, Texas A&M University, May 2008.

[21] Bailey T.S., Chang J.H., Adams M.L., *A Piecewise Linear Discontinous Finite Element spatial discretization of the transport equation in 2D Cylindrical Geometry*, 2009 International Conference on Advances in Mathematics, Computational Methods, and Reactor Physics, 2008.

[22] Bailey T.S., Adams M.L., Yang B., Zika M.R., *A piecewise linear finite element discretization of the diffusion equation for arbitrary polyhedral grids*, Journal of Computational Physics 227 (2008) 3738–3757, 2007

[23] Vermaak J.I.C, Ragusa J.C, Adams M.L., Morel J.E., *Massively parallel transport sweeps on meshes with cyclic dependencies*, Journal of Computational Physics 425(4):109892, January 2021.

[24] Randal S. Baker & Kenneth R. Koch (1998) An Sn Algorithm for the Massively Parallel CM-200 Computer, Nuclear Science and Engineering, 128:3, 312-320, DOI: 10.13182/ NSE98-1

[25] Adams et al., *Provable optimal parallel transport sweeps on regular grids*, International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C). 2013

[26] Adams et al., "Provably Optimal Parallel Transport Sweeps with Non-Contiguous Partitions," in "ANS MC2015-Joint International Conference on Mathematics and Computation (M&C), Supercomputing in Nuclear Applications (SNA) and the Monte Carlo (MC) Method", 2015

[27] T. S. Bailey and R. D. Falgout, "Analysis Of Massively Parallel Discrete-Ordinates Transport Sweep Algorithms With Collisions," Proc. International Conference on Mathematics, Computational Methods & Reactor Physics, Saratoga Springs, May 3-7, CDROM (2009).

[28] Adams et al., *Provable optimal parallel transport sweeps on semi-structured grids*, Journal of Computational Physics, DOI: 10.1016/j.jcp.2020.109234, January 2020.

[29] Karypis G. *ParMETIS - Parallel Graph Partitioning and Fill-reducing Matrix Ordering*, http://glaros.dtc.umn.edu/gkhome/metis/parmetis/overview, date accessed January 30, 2020

[30] Saad Y., Schultz M.H., *GMRES: A Generalized Minimal Residual algorithm for solving non-symmetric linear systems*, SIAM Journal of Scientific and Statistical Computation, volume 7, pages 856-869, 1986.

[31] Edward W. Larsen (1984) Diffusion-synthetic acceleration methods for discrete-ordinates problems, Transport Theory and Statistical Physics, 13:1-2, 107-126, DOI: 10.1080/00411458408211656

[32] Marvin L. Adams & William R. Martin (1992) Diffusion Synthetic Acceleration of Discontinuous Finite Element Transport Iterations, Nuclear Science and Engineering, 111:2, 145-167, DOI: 10.13182/NSE92-A23930

[33] Turcksin, Bruno, and Ragusa, Jean C., "Discontinuous diffusion synthetic acceleration for Ssub n transport on 2D arbitrary polygonal meshes". United States: N. p., 2014. Web. doi:10.1016/J.JCP.2014.05.044.

[34] B. T. Adams & J. E. Morel (1993) A Two-Grid Acceleration Scheme for the Multigroup Sn Equations with Neutron Upscattering, Nuclear Science and Engineering, 115:3, 253-264, DOI: 10.13182/NSE115-253

[35] Hanus M., Ragusa J.C., *Thermal Upscattering Acceleration Schemes for Parallel Transport Sweeps*, Nuclear Science and Engineering, 194:10, 873-893, 2020.

[36] Hanophy J., Southworth B.S., Li R., Manteuffel T., Morel J.E., *Parallel Approximate Ideal Restriction Multigrid for Solving the SN Transport Equations*, Nuclear Science and Engineering, 194:11, 989-1008, 2020.

[37] Oliveira S., Deng Y., *Preconditioned Krylov Subspace Methods for Transport Equations*, Progress in Nuclear Energy 33, 155. 1998

[38] Guthrie B., Holloway J.P., and Patton B.W., *GMRES as a Multi-Step Transport Sweep Accelerator*, Transport Theory Statist. Phys. 28, 83. 1999

[39] Patton B.W., Holloway J.P., *Application of preconditioned GMRES to the numerical solution of the neutron transport equation*, Annals of Nuclear Energy volume 29, issue 2, Pages 109-136, January 2002.

[40] Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., *Numerical Recipes: The Art of Scientific Computing*, Third edition, New York: Cambridge University Press, ISBN 978-0-521-88068-8. 2007.

[41] Yunhaung Zhang and Jim E. Morel, "Transport Error Estimation via Non-Recursive Residual Monte Carlo," PHYSOR 2018: Reactor Physics Paving The Way Towards More Efficient Systems, Cancun, Mexico, April 22-26, 2018, on USB memory stick (2018).

APPENDIX A

RAY TRACING TO CELL SURFACES

A.1 Intersection of a line and a plane

Consider the arrangement depicted in A.1.

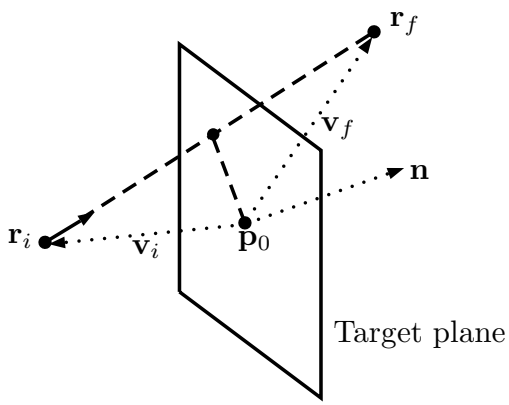

Figure A.1: Graphical representation of intersection of a line with a plane.

3D Line definition 1:

Given a particle's location, $\mathbf{r}_i$, and direction, $\mathbf{\Omega}$, we can create a 3D line of the form

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_i + s\,\mathbf{\Omega} \\ (x, y, z) &= (x_i, y_i, z_i) + s\,[\Omega_x, \Omega_y, \Omega_z] \end{aligned} \tag{A.1}$$

where s is the only unknown parameter required to define $\mathbf{r} = (x, y, z)$.

3D Line definition 2:

If we are supplied with another point on the line, $\mathbf{p}_f$ and then define a weight, $w \in [0, 1]$, which

can define a *line segment*

$$\begin{aligned} \mathbf{r} &= w\mathbf{r}_i + (w-1)\mathbf{r}_f \\ (x,y,z) &= w(x_i,y_i,z_i) + (w-1)(x_f,y_f,z_f) \end{aligned} \tag{A.2}$$

3D Plane definition:

For the equation of a plane we need a refence point, $\mathbf{p}_0$, and a normal, $\mathbf{n}$, after which we can form a vector from point $\mathbf{p}_0$ to $\mathbf{r}_p = (x,y,z)$, an arbitrary point on the plane. The plane is then defined by the relationship

$$\mathbf{n} \cdot (\mathbf{r}_p - \mathbf{p}_0) = 0$$

$$n_x(x - p_{0x}) + n_y(y - p_{0y}) + n_z(z - p_{0z}) = 0. \tag{A.3}$$

Computing the intersection of a line with a plane:

There are several options for computing the intersection. One option is to substitute $\mathbf{r}$ in equation A.1 into $\mathbf{r}_p$ int he equation for the plane (equation A.3) and solve for s. If $s < 0$ then there is no intersection. If $s < ||\mathbf{r}_f - \mathbf{r}_i||_2$ then there is an intersection at point $\mathbf{r}_i + s\ \mathbf{\Omega}$.

A very simple way to simultaneously determine whether a line segment intersects the plane and where the intersection is, is to use the segment definition of the line and compute two vectors; one from $\mathbf{p}_0$ to $\mathbf{r}_i$, and another from $\mathbf{p}_0$ to $\mathbf{r}_f$

$$\mathbf{v}_i = \mathbf{r}_i - \mathbf{p}_0 \quad \text{and} \quad \mathbf{v}_f = \mathbf{r}_f - \mathbf{p}_0$$

after which we compute the projection of these vectors along the normal of the plane by taking the

dot-products

$$D_i = \mathbf{n} \cdot \mathbf{v}_i \quad \text{and} \quad D_f = \mathbf{n} \cdot \mathbf{v}_f.$$

Since we use the same normal for both computations the line is intersecting the plane only if the signs of the dot-products is not equal and w can then be computed as

$$\text{if} \quad \text{sgn}(D_i) \neq \text{sgn}(D_f)$$

$$w = \frac{|D_i|}{|D_i| + |D_f|}$$

A.2 Intersection of a line and a strip

The algorithm to determine the intersection of a line with the face of a polygon, or an intersection with a segment of a polygon, is captured within this utility. A strip is formed by a normal $\mathbf{n}$ and two points $\mathbf{v}_0$ and $\mathbf{v}_1$, and infinite along the direction defined by $(\mathbf{v}_1 - \mathbf{v}_0) \cdot \mathbf{n}$. For further clarification, typical strips are as shown in Figure A.2 below.

Firstly we apply the algorithm in A.1 to determine the intersection point $\mathbf{p}$ with the plane of which the strip is a subset. We then compute 3 vectors for this utility.

$$\mathbf{v}_{01} = \mathbf{v}_1 - \mathbf{v}_0$$

$$\mathbf{v}_{0p} = \mathbf{p} - \mathbf{v}_0$$

$$\mathbf{v}_{1p} = \mathbf{p} - \mathbf{v}_1$$

after which we compute the projection of $\mathbf{v}_{0p}$ and $\mathbf{v}_{1p}$ along the strip's characteristic vector $\mathbf{v}_{01}$ as dot products

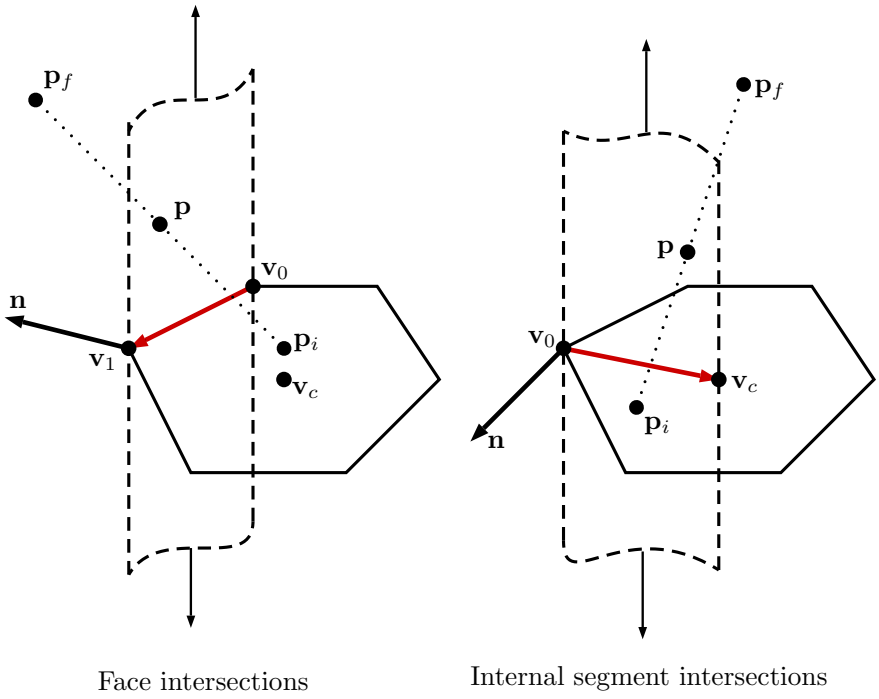

Figure A.2: A schematic of the raytracing required for 2D cells. This involves the intersection of 3D line with a strip.

$$D_0 = \mathbf{v}_{0p} \cdot \mathbf{v}_{01}$$

$$D_1 = \mathbf{v}_{1p} \cdot \mathbf{v}_{01}$$

The intersection point is on the strip if

$$\text{sgn}(D_0) \neq \text{sgn}(D_1)$$

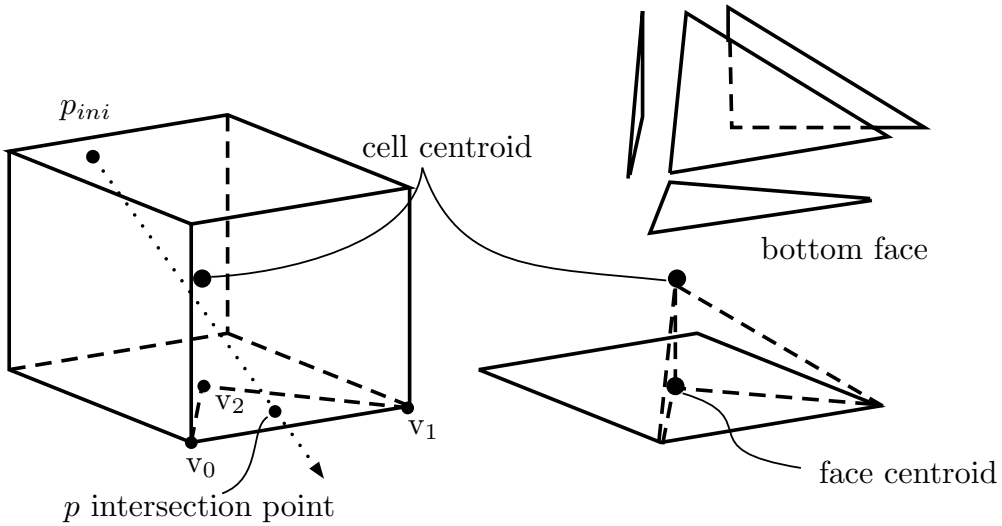

Figure A.3: The formation of a tetrahedron from a triangulation of polyhedron faces.

A.3 Intersection of a plane and a triangle

For this algorithm we use a modified version of the Möller-Trumbore intersection algorithm [19]. We decompose a polyhedron, as shown in Figure A.3 below, and trace from entry to the nearest face-side triangle intersection. The nearest face is considered only if the cell has been marked as having a concavity, otherwise the first triangle intersection is used for calculating the ray length.

APPENDIX B

BASIS FUNCTION FOR PIECEWISE-LINEAR FEM

B.1 2D Polygon cells

The support points for the nodes of a polygon, using piecewise-linear basis functions, are the vertices. The basis functions for each node i are then

$$P_i(x, y) = N_i(x, y) + \beta_i N_c(x, y), \tag{B.1}$$

where the functions N_i and N_c are the standard linear functions defined on triangles. The subscripts i and c refer to the vertices i and center of the polygon, respectively. The β_i value is a weighting constant defined such that

$$\mathbf{x}_c = \sum_{s=0}^{N_s} \beta_s \mathbf{x}_{s,avg}, \tag{B.2}$$

where $\mathbf{x}_c$ is the cell-centroid, $\mathbf{x}_{s,avg}$ is the average coordinate of the two vertices of a side. Naturally it follows that $\beta_s = \dfrac{1}{N_s}$ where N_s is the amount of sides. An example basis function is shown in Figure B.1. It should be noted that a volumetric integral of a single basis function requires integration on each of the sub-triangles of the polygon instead of just a single one. The sub-triangles are formed from each edge of the polygon and the cell-centroid.

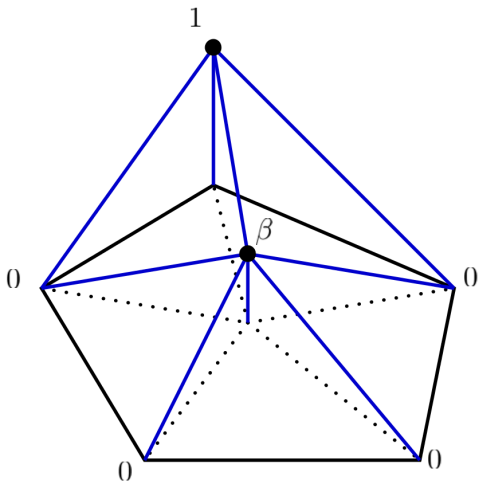

Figure B.1: Piecewise-linear basis function on a 2D polygon.

The method is versatile enough to applied to triangles and quadrilaterals where examples of the shape functions are shown in Figure

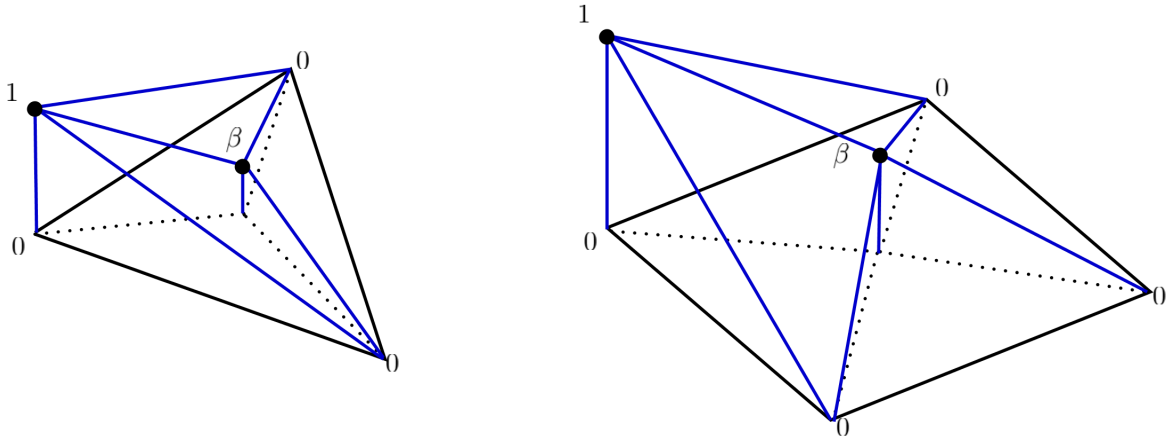

Figure B.2: Piecewise-linear basis functions on a triangle and on a quadrilateral.

B.2 3D Polyhedron cells

The shape function for each vertex/node i of the polyhedron is given by

$$P_i(x,y,z) = N_i(x,y,z) + \sum^{faces\ at\ i} \beta_f N_f(x,y,z) + \alpha_c N_c(x,y,z) \tag{B.3}$$

where the functions $N(x,y,z)$ are the standard linear shape functions defined on a tetrahedron. β_s and α_c is the weight that gives the face centroid, $\mathbf{x}_{fc}$, and cell centroid, $\mathbf{x}_{cc}$, respectively from the sum of the vertices that constitute them. i.e.

$$\mathbf{x}_{fc} = \sum_{v=0}^{N_{vf}} \beta_f \mathbf{x}_v. \tag{B.4}$$

$$\mathbf{x}_{cc} = \sum_{v=0}^{N_{vc}} \alpha_c \mathbf{x}_v. \tag{B.5}$$

where N_{vf} is the number of vertices for the given face and N_{vc} is the number of vertices for the entire cell. Naturally it follows that $\beta_f = \dfrac{1}{N_{vf}}$ and $\alpha_c = \dfrac{1}{N_{vc}}$. The format of equation B.3 is not intuitive at first sight ... it is hard to comprehend the summation over faces “at j", but let us try to clarify this with a diagram (see Figure B.3).

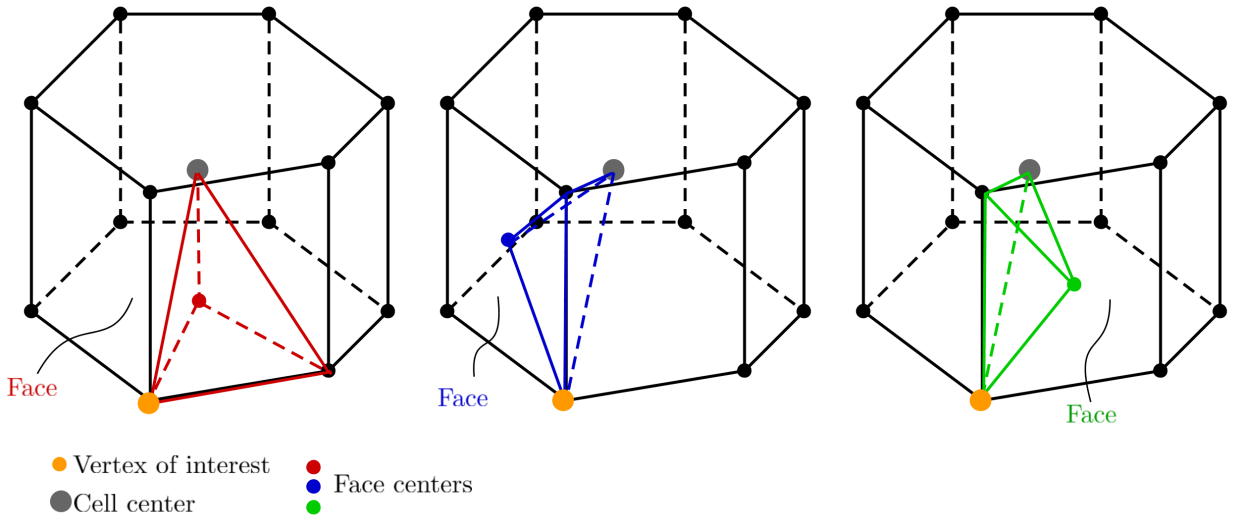

Figure B.3: Connection of the vertex of interest to the tetrahedrons that comprise the cell.

Firstly we split the polyhedron into faces, where each face can be a polygonal face. Each face is then split into a number of sides. A side is a tetrahedron, corresponding to a face, which is formed from each edge of the polygonal face where the vertex collection are the two vertices of the edge, the face center and the cell center. In other words the face center and the two vertices of the edge forms a triangle, and the cell center makes it a tetrahedron.

As was the case with the polygon, all the other vertices j (not i) are connected to the vertex of interest i through the cell center. Again, we don't include the cell center as a point in the simulation so we have to spread its effect through to each of the vertices using the α_c factor. We also have face centered shape functions associated with the division of each polygonal face into sides. On each face, to which vertex i belongs, the shape functions defined on the face centers will protrude into the tetrahedron under consideration (i.e. the tetrahedron associated with vertex i associated with face f, side s). Therefore more clearly we can express the shape functions on a tetrahedron-by-tetrahedron basis

$$P_i^{tet}(x,y,z) = \begin{cases} \alpha_c N_c(x,y,z) & \text{no matter which tetrahedron} \\ +\beta_f N_f(x,y,z) & \text{if vertex } i \text{ is part of the face} \\ +N_i(x,y,z) & \text{if vertex } i \text{ is part of the face-side pair} \end{cases} \tag{B.6}$$

Figure B.4 shows the influence of a shape function (centered on a specific point as denoted by the start of an arrow) from a specific vertex (color). The orange colored vertex's influence is shown on the left most figure where the shape function is then the full equation because all of the conditions are met; i.e. $\alpha_c N_c$ is always present, vertex i is indeed part of the face where this tetrahedron is defined and therefore $\beta_f N_f$ is present, finally it is also part of the side of the tetrahedron and therefore its basic shape function N_i is present.

The middle figure shows the red vertex as a vertex that only has the contribution of $\alpha_c N_c$ and $\beta_f N_f$ because the red vertex is not on the side comprosing the tetrahedron of interest. The right-most figure shows only the contribution of $\alpha_c N_c$ because the blue vertex is not part of any adjacent face.

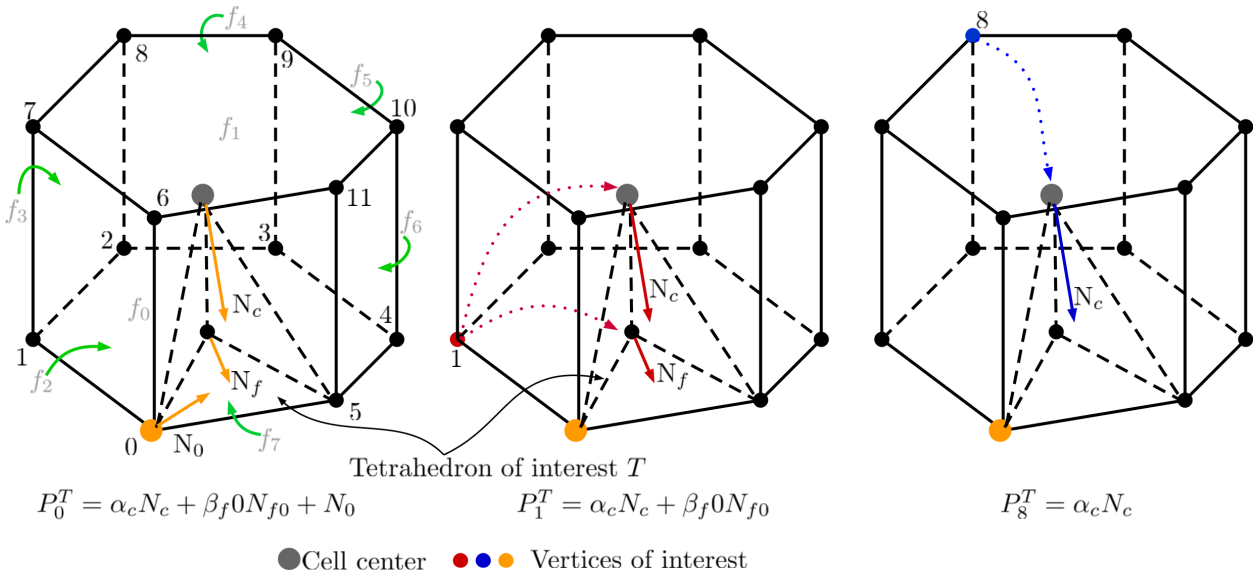

Figure B.4: Influence of different vertices on the shape functions of within a tetrahedral portion of the cell.

APPENDIX C

AN EXPONENTIAL REPRESENTATION OF A P1 ANGULAR FLUX EXPANSION

The P_1 approximation of an anisotropic angular flux is given by

$$\psi(\mathbf{\Omega}) = \frac{1}{4\pi}\phi + \frac{3}{4\pi}\mathbf{\Omega}\cdot\mathbf{J} \tag{C.1}$$

where

$$[\phi_{0,0}, \phi_{1,1}, \phi_{1,-1}, \phi_{1,0}] \mapsto [\phi, J_x, J_y, J_z]$$

$$\mathbf{J} = [J_x, J_y, J_z]$$

which is essentially a linear representation w.r.t. the cosine, μ, of the angle between the desired direction $\mathbf{\Omega}$ and the direction of the current $\mathbf{\Omega}_J = \dfrac{\mathbf{J}}{|\mathbf{J}|}$. For highly anisotropic angular fluxes this P_1 representation is not assured to be positive and hence cannot be used to construct a PDF for random sampling.

C.1 An exponential representation

To circumvent the property of the P_1 approximation possibly being negative, we replace the P_1 approximation by an exponential representation of the form

$$\psi(\mathbf{\Omega}) = e^{a+b\ \mathbf{\Omega}_J\cdot\mathbf{\Omega}} \tag{C.2}$$

where $\mathbf{\Omega}_J$ is the unit direction along $\mathbf{J}$, and a and b are constants such that

$$\int_{4\pi} e^{a+b\ \mathbf{\Omega}_J\cdot\mathbf{\Omega}}\, d\mathbf{\Omega} = \phi \text{ and}$$

$$\int_{4\pi} e^{a+b\ \mathbf{\Omega}_J\cdot\mathbf{\Omega}}\ \mathbf{\Omega}\, d\mathbf{\Omega} = \mathbf{J}.$$

C.1.1 Change in reference frame

For known values of ϕ and $\mathbf{J}$ we need to construct a system of equations to solve for a and b. From the equation above we see that we have more equations than unknowns. One simplification we can apply is to rotate our angular coordinate system to be aligned with $\mathbf{\Omega}_J$ with a rotation matrix R. This rotation is such that

$$\mathbf{\Omega}' = R\,\mathbf{\Omega}, \text{ and}$$
$$\mathbf{J}' = R\,\mathbf{J} = |\mathbf{J}|\mathbf{\Omega}_J,$$

for which we can define μ' as the cosine between $\mathbf{\Omega}'$ and $\mathbf{\Omega}_J$ which reduces the above integrals to one dimensional integrals as

$$2\pi \int_{-1}^{1} e^{a+b\mu'}\, d\mu' = \phi \text{ and}$$
$$2\pi \int_{-1}^{1} e^{a+b\mu'}\, \mu'\, d\mu' = |\mathbf{J}|,$$

which evaluates to

$$\begin{aligned} \frac{4\pi}{b} e^{a}\, \sinh b &= \phi \text{ and} \\ \frac{4\pi}{b^2} e^{a}\, (b\cosh b - \sinh b) &= |\mathbf{J}|. \end{aligned} \tag{C.3}$$

C.1.2 Only working with the shape of the angular flux

Another simplification we can make originates from the observation that the shape of the angular flux (as a function of $\mu' = \mathbf{\Omega}_J \cdot \mathbf{\Omega}'$) is dictated only by the ratio of $|J|$ to ϕ. So if we amend our previous definition of the desired exponential representation (eq. C.2) to rather be

$$\frac{\psi(\mu')}{\phi} = e^{a+b\mu'} \tag{C.4}$$

and redefine the integrals, denoting the properties that the coefficients a and b must support, as

$$\begin{aligned} 2\pi \int_{-1}^{1} e^{a+b\mu'}\, d\mu' &= 1 \text{ and} \\ 2\pi \int_{-1}^{1} e^{a+b\mu'}\, \mu'\, d\mu' &= \frac{|\mathbf{J}|}{\phi}, \end{aligned} \tag{C.5}$$

which evaluates to

$$\begin{aligned} \frac{4\pi}{b} e^{a}\, \sinh b &= 1 \text{ and} \\ \frac{4\pi}{b^2} e^{a}\, (b\cosh b - \sinh b) &= r, \end{aligned} \tag{C.6}$$

where $r = \dfrac{|\mathbf{J}|}{\phi}\ :\ 0 < r < 1$, we now have grounds to construct a system characterized only by constant r and the unknowns a and b.

C.1.3 The corner cases at $r \to 0$ and $r \to 1$

As r tends to zero the angular distribution becomes increasingly isotropic making the value of $b \to 0$, which is problematic in the equations above since b appears as a divisor. To accommodate this case we set a lower bound, r_{low}, for r, below which we explicitly set $b = 0$ and $a = \ln(\frac{1}{4\pi})$.

As r tends to 1 the angular distribution approaches a δ-function, which becomes difficult to solve in its non-linear form. To accommodate this case we set an upper limit, r_{high}, above which we limit r as $r = \min(r, r_{high})$. In this research the exponential distribution is used for angular biasing, allowing us to set r_{high} comparatively low at $r_{high} = 0.9$. This allows the angular biasing to still be effective, on the hand of the P_1 expansion, whilst not resulting in excess weight reductions to account for the biasing.

C.1.4 A non-linear system

The non-linear system, of the form, $F(\mathbf{x}) = 0$, with unknowns $\mathbf{x} = [a, b]$, is developed from Eq. (C.6) and is defined by

$$F(\mathbf{x} \mapsto (a, b)) = \begin{bmatrix} \frac{4\pi}{b} e^a \ \sinh b - 1, \\ \frac{4\pi}{b^2} e^a \ (b \cosh b - \sinh b) - r \end{bmatrix}, \tag{C.7}$$

for which the Jacobian matrix, $F'(\mathbf{x})$, has elements

$$\begin{aligned} F'(\mathbf{x})_{00} &= \frac{\partial F_0}{\partial a} = \frac{4\pi}{b} e^a \ \sinh b, \\ F'(\mathbf{x})_{01} &= \frac{\partial F_0}{\partial b} = \frac{4\pi}{b^2} e^a \ (b \cosh b - \sinh b), \\ F'(\mathbf{x})_{10} &= \frac{\partial F_1}{\partial a} = \frac{4\pi}{b^2} e^a \ (b \cosh b - \sinh b), \\ F'(\mathbf{x})_{11} &= \frac{\partial F_1}{\partial b} = \frac{4\pi}{b^3} e^a \ ((b^2 + 2) \sinh b - 2b \cosh b), \end{aligned}$$

such that

$$F'(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_0}{\partial a} & \frac{\partial F_0}{\partial b} \\ \frac{\partial F_1}{\partial a} & \frac{\partial F_1}{\partial b} \end{bmatrix}.$$

These definitions allow us to formulate a multivariate Newton-method to solve the non-linear system. The solution process is then to first choose a starting value for $\mathbf{x}$ and give it an iteration index $i = 0$, i.e., we choose $\mathbf{x}_0$. Next we evaluate $F(\mathbf{x}_0)$ and $F'(\mathbf{x}_0)$ which we use to compute $\mathbf{x}_1$ from the iterative scheme

$$\mathbf{x}_{i+1} = \mathbf{x}_i - [F'(\mathbf{x}_i)]^{-1} \ F(\mathbf{x}_i). \tag{C.8}$$

We repeat this process until $||F(\mathbf{x}_i)||_2 < \epsilon$.

C.1.5 Example distributions

Figure C.1 below shows the resulting exponential distribution for different values of $r = \frac{|\mathbf{J}|}{\phi}$. Note the δ-function nature at values of r approaching 1 (in Figure C.1 r=0.98).

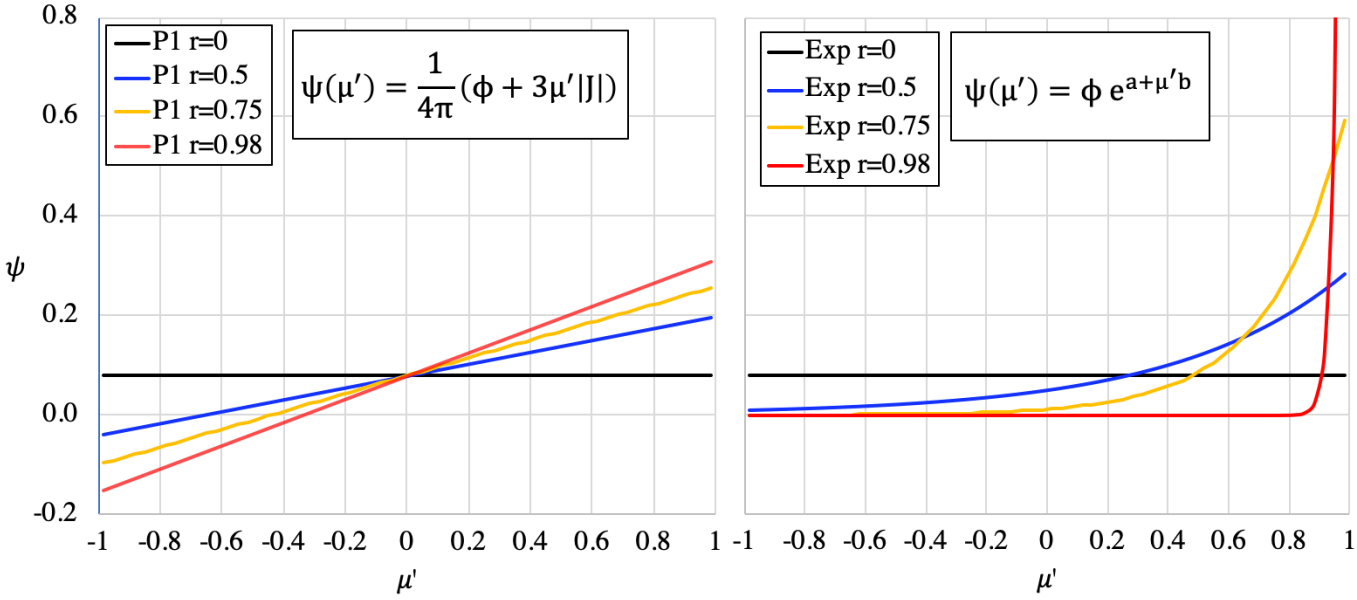

Figure C.1: Examples of the exponential representation of the angular flux for P_1 approximations with a different r-value.

C.2 Sampling the exponential distribution

From our previously defined properties we know that, if the coefficients a and b are known, then

$$2\pi \int_{-1}^{1} e^{a+b\mu'} \, d\mu' = 1,$$

which is equivalent to having a normalized continuous PDF

$$p(\mu') = 2\pi e^{a+b\mu'}. \tag{C.9}$$

The corresponding continuous CDF, $c(\mu')$ is then,

$$\begin{aligned} c(\mu') &= \int_{-1}^{\mu'} p(x)dx \\ &= \frac{2\pi}{b} e^{a-b}(e^{b\mu'+b} - 1) \end{aligned} \tag{C.10}$$

which can be inverted algebraically to get

$$\mu' = \frac{1}{b} \ln\left[e^{-b}\left(\frac{c(\mu)'}{C_0} + 1\right)\right] \tag{C.11}$$

where

$$C_0 = \frac{2\pi}{b} e^{a-b}.$$

This continuous CDF can then be sampled by replacing $c(\mu')$ with a pseudo random number, $\theta_R \in [0, 1)$, and evaluating the above function.