

# Efficient Simulation of Spectral Light Transport in Dense Participating Media and Granular Materials

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik  
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Johannes Meng

aus Mainz

Tag der mündlichen Prüfung: 22. Mai 2017

Erster Gutachter: Prof. Dr.-Ing Carsten Dachsbacher

Zweiter Gutachter: Prof. Dr. Reinhard Klein



Für Cordula und Tobi.



# ACKNOWLEDEMENTS

I wish to thank my advisor, Prof. Dr.-Ing. Carsten Dachsbacher, who has supported and challenged me in innumerable ways and without whom this dissertation would not have been possible.

I would like to express my sincere gratitude to Florian Simon, the best partner in crime I could have ever wished for.

Thank you Prof. Dr. Hartmut Prautzsch, Prof. Dr.-Ing. Boris Neubert, Diana Kheil, Johannes Schudeiske, and all my former colleagues at KIT: Anton Kaplanyan, Christoph Schied, Emanuel Schrade, Eric Heitz, Gábor Liktó, Gregor Mückl, Hauke Rehfeld, Jan Novák, Lorenzo Tessari, Mahmoud Zeidan, Max-Gerd Retzlaff, Marco Ament, Stephan Bergmann, Tamás Szép, Tim Reiner, Thomas Engelhardt, Thorsten Schmidt, Tobias Rapp, and Tobias Zirr.

My time at Disney Research Zürich was as instructive as it was enjoyable, and I am indebted to the co-authors of my paper on granular materials, Marios Papas, Ralf Habel, Steve Marschner, Markus Gross, and Wojciech Jarosz. Thank you for all your support.

I would also like to thank the rendering team at Weta Digital, especially Tomáš Davidovič and Luca Fascione, who are a source of great inspiration to me.

My friends outside the computer graphics community have played a bigger part in the making of this thesis than they may realize. Thank you all, in particular everybody at K12.

Finally, I wish to thank my parents, my brother, my partner, her parents, siblings, and everybody in my extended family, for their love and support.



# LEBENS LAUF

## Beruflicher Werdegang

Seit 2017	Weta Digital Ltd., Wellington, Neuseeland Rendering Research and Development
2012 – 2017	Karlsruher Institut für Technologie Wissenschaftlicher Mitarbeiter
2013	Weta Digital Ltd., Wellington, Neuseeland Praktikum, Rendering Research and Development
2013	Disney Research Zürich, Schweiz Praktikum, Rendering Research and Development
2012	Disney Research Zürich, Schweiz Praktikum, Rendering Research and Development
2007 – 2012	Group on Human Motion Analysis, KIT Studentische Hilfskraft
2004 – 2005	Deutsches Rotes Kreuz Südliche Weinstraße Zivildienst, Rettungssanitäter

## Ausbildung

2006 – 2012	Karlsruher Institut für Technologie Studium der Informatik Vertiefungsrichtungen Robotik und Computergrafik Abschluss: Diplom-Informatiker (1,0 „mit Auszeichnung“)
2010	Hochschule für Gestaltung Karlsruhe Ergänzungsfach Medienkunst
2009	Carnegie Mellon University, Pittsburgh, PA, USA Forschungsaufenthalt
2006	Technische Universität Darmstadt Studium der Mathematik
1995 – 2004	Max-Slevogt-Gymnasium in Landau / Pfalz Abschluss: Abitur (1,5) Slevogt-Medaille für vorbildliches Engagement im künstlerischen Bereich



# ZUSAMMENFASSUNG

Diese Dissertation befasst sich mit der effizienten, wellenlängenabhängigen Lichttransportsimulation für die physikalisch-basierte Bildsynthese mit dichten partizipierenden Medien und granularen Materialien.

Partizipierende Medien sind Stoffe wie Rauch, Feuer, Sand, Haut oder Wachs, in denen Licht gestreut, absorbiert und emittiert werden kann. Granulare Materialien bestehen aus Körnern und können ebenfalls Licht streuen oder absorbieren.

Das Ziel der physikalisch-basierten Bildsynthese ist die Messung der ankommenden Strahldichte  $L$  [ $W/m^2/sr$ ] an einem virtuellen Sensor. Dabei versucht man in der Regel, physikalische Modelle für geometrische Optik, Materialien und Energietransport möglichst akkurat abzubilden. Die Strahldichte an einem beliebigen Punkt im Raum  $\mathbf{x}$  und in Richtung  $\boldsymbol{\omega}$  mit  $\mathbf{x}_s = \mathbf{x} - s\boldsymbol{\omega}$  ist beispielsweise die Lösung der sogenannten Strahlungstransportgleichung (Chandrasekhar, 1960)

$$\begin{aligned} \frac{dL}{ds}(s, \boldsymbol{\omega}_o) = & -\mu_t(s)L(s, \boldsymbol{\omega}_o) \\ & + \mu_e(s)L^e(s, \boldsymbol{\omega}_o) \\ & + \mu_s(s) \int_{4\pi} \phi(\boldsymbol{\omega}_o|\boldsymbol{\omega}_i)L(s, \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i. \end{aligned} \quad (1)$$

Alle Größen hängen von der Wellenlänge des Lichts  $\lambda$  ab. Die Koeffizienten  $\mu$  modellieren Dichte und Größe der Partikel, aus denen das Medium besteht.  $\tau(\mathbf{x}_1, \mathbf{x}_2)$  beschreibt die Auslöschung aufgrund von Ausstreuung und Absorption und  $L^e$  ist die Emission des Materials. Die Phasenfunktion  $\phi(\boldsymbol{\omega}_1|\boldsymbol{\omega}_2)$  ist die Wahrscheinlichkeitsdichte für Streuung von Richtung  $\boldsymbol{\omega}_2$  in Richtung  $\boldsymbol{\omega}_1$ .

Gleichung (1) ist rekursiv formuliert. Eine vollständige Lösung benötigt also die Auswertung eines hochdimensionalen Integrals und ist analytisch nur in sehr einfachen Szenen möglich. Daher nutzen die Methoden, die in dieser Dissertation vorgestellt werden, Monte Carlo-Integration zur näherungsweise Lösung. Dabei wird der Integrand an endlich vielen, zufällig gewählten Stellen ausgewertet. Das üblicherweise eingesetzte *Path Tracing* (Kajiya, 1986) versucht mittels Raytracing, den Weg des Lichts rückwärts von einem virtuellen Sensor bis zu einer Lichtquelle zu rekonstruieren, indem es zufällig Lichtpfade erzeugt. Ein Lichtpfad ist dabei ein  $N$ -Tupel  $X = (x_0, \dots, x_{N-1})$ , wobei jeder Vertex  $x_i$  einen Interaktionspunkt, z.B. an Oberflächen oder bei Streueignissen im Volumen, darstellt.

Der Beitrag eines Pfades zum Ausgabebild wird durch eine Messfunktion  $f(X, \lambda)$  beschrieben, die im Wesentlichen eine rekursiv expandierte Form von Gleichung (1) ist. Man kann nun den Monte Carlo-Schätzer für den Wert eines Pixels schreiben als

$$I(\lambda) \approx \frac{1}{M} \sum_{j=1}^M \frac{f(x_j, \lambda)}{p(x_j|\lambda)}, \quad (2)$$

wobei  $M$  die Stichprobengröße ist und  $p(X|\lambda)$  die Wahrscheinlichkeitsdichte des Pfades  $X$  darstellt. Um diese spektrale Energiedichte in eine Farbe mit je einer roten, grünen und blauen Komponente umzuwandeln, beispielsweise für die Darstellung auf einem Monitor, muss man sie noch mit Gewichtungsfunktionen, den sogenannten *color-matching*-Funktionen, gewichten und integrieren. Dies kann wieder über Monte Carlo-Integration geschehen, wodurch sich insgesamt der Schätzer

$$I = \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \int_{\Lambda} I(\lambda) \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} d\lambda \approx \frac{1}{M} \sum_{j=1}^M \begin{pmatrix} \bar{r}(\lambda_j) \\ \bar{g}(\lambda_j) \\ \bar{b}(\lambda_j) \end{pmatrix} \frac{f(x_j, \lambda_j)}{p(x_j, \lambda_j)} \quad (3)$$

für den Pixelwert  $I$  ergibt. Weil hierbei alle Terme explizit von der Wellenlänge abhängen, spricht man auch von *spektralem Rendering*.

## 1 Beiträge

Diese Dissertation enthält Beiträge aus drei Bereichen.

Farbwerte Texturen werden üblicherweise als RGB-Tripel angegeben. Für spektrales Rendering müssen diese in spektrale Energiedichtefunktionen umgewandelt werden, die den Termen aus Gleichung (1) entsprechen. Kapitel III beinhaltet eine Methode, die diese Konvertierung durchführen kann (Meng, Simon, et al., 2015). Sie erzeugt niederfrequente Spektren, die natürliche Reflektanzen gut nachbilden und ist außerdem im Gegensatz zu vorigen Verfahren wie (Smits, 1999) nicht auf einen Eingabefarbraum festgelegt.

Die weiteren Beiträge dieser Dissertation beschäftigen sich konkreter mit der Darstellung von Materialien, die für den Lichttransport besonders große Herausforderungen darstellen.

Dazu zählen insbesondere dichte partizipierende Medien mit geringer Absorption, in denen oft mit hohem Rechenaufwand lange Lichtpfade erzeugt werden müssen und die daher bei der Monte Carlo-Integration anfällig für starkes Bildrauschen sind. In Kapitel IV wird daher eine Technik für die Varianzreduktion in solchen Medien vorgestellt, dass ein existierendes Verfahren (Křivánek and d'Eon, 2014) erweitert, um es robuster gegenüber starker Hintergrundbeleuchtung zu gestalten (Meng, Hanika, et al., 2016).

Schließlich enthält Kapitel V Forschungsergebnisse zu granularen Materialien wie Sand und Zucker. In einem ersten Teil wird dabei ein Verfahren erläutert, das mit sehr wenig Speicheraufwand Objekte erzeugen kann, die aus Milliarden von Körnern

bestehen. Darauf aufbauend wird außerdem ein Algorithmus für die Simulation des Lichttransports in granularen Materialien beschrieben, der die Laufzeiten der Lichttransportsimulation um mehrere Größenordnungen verringern kann (Meng, Papas, et al., 2015).

## 2 Berechnung spektraler Energiedichtefunktionen

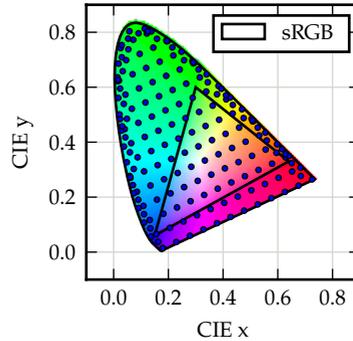
Die Umkehrung von Gleichung (3), also die Berechnung einer spektralen Energiedichte aus einem RGB-Tripel, ist nicht eindeutig: im Allgemeinen existieren unendlich viele Spektren zu einem Farbeindruck.

Verfahren für die Konvertierung eines RGB-Tripels in ein volles Spektrum treffen daher Annahmen, mit denen sie den Lösungsraum einschränken. Eine sinnvolle Annahme ist dabei Nichtnegativität. Man kann die Verfahren außerdem für Reflektanzspektren optimieren, die in der Natur normalerweise glatte, also niederfrequente Funktionen sind (Maloney, 1986).

Mit einer entsprechenden Fehlerfunktion kann man Optimierungsverfahren nutzen, um also eine möglichst glatte, nichtnegative Funktion  $S(\lambda)$  zu finden, die zum gegebenen Eingabefarbwert passt. Dies ist aber rechenintensiv und daher während der eigentlichen Lichtsimulation nicht praktikabel.

Die häufig eingesetzte Technik von Smits (Smits, 1999) nutzt aus diesem Grund sieben vorberechnete Funktionen: Je eine für die rote, grüne und blaue Primärfarbe des Eingabefarbraums sRGB, deren Komplementärfarben Cyan, Magenta und Gelb, sowie den Weißpunkt. Zur Laufzeit werden diese Funktionen lediglich gewichtet addiert. Dieses Verfahren ist sehr schnell, aber unterstützt nur Eingaben aus dem sRGB-Farbraum.

In Kapitel III wird daher eine Methode vorgestellt, die nahezu beliebige Eingabefarben verarbeiten kann und dabei eine schnelle und genaue Interpolation zur Laufzeit erlaubt (Meng, Simon, et al., 2015). Die Methode nutzt wie (Smits, 1999) vorberechnete Basisspektren, die aber in Form eines regulären Gitters über den Raum der sichtbaren Farben verteilt sind. Das Gitter ist in der nebenstehenden Abbildung im direkten Vergleich mit dem Gamut des sRGB-Farbraumes zu sehen. Die Ausgabespektren sind niederfrequent und bilden natürlich vorkommende Reflektanzspektren gut ab. Sie eignen sich daher besonders für die physikalisch-basierte Lichttransportsimulation.



### 3 Varianzreduktion mit Zero-Variance-Methoden

Bei der Lichttransportsimulation mit Monte Carlo-Integration zeigt sich die Varianz des Schätzers als Rauschen im Bild. Es ist theoretisch möglich, das Rauschen des Schätzers komplett zu eliminieren. Algorithmen, die versuchen, dies für partizipierende Medien zu ermöglichen, werden in der *Zero-Variance Random Walk*-Theorie erforscht. In der Praxis ist dieses Ziel aber nicht erreichbar, weil die Methoden bereits das Ergebnis der Integration kennen müssen.

Man kann allerdings das Rauschen durch den Einsatz approximativer Lösungen reduzieren (Dwivedi, 1982a). Dabei werden vereinfachende Annahmen über die Geometrie und die Beschaffenheit des Mediums getroffen, sodass ein analytisches Modell für die Lösung des Lichttransports existiert. Konkret wird die Szene als homogener Halbraum angenähert. Man muss nun noch die Trennebene, die den Halbraum begrenzt, möglichst gut in die Szene einpassen.

Křivánek and d'Eon (2014) orientieren die Ebene an der Oberflächennormale an dem Punkt, an dem der Lichtpfad in das Medium eintritt. Dies funktioniert gut, solange die einfallende Beleuchtung nicht stark variiert. Befindet sich allerdings eine helle Lichtquelle hinter der sichtbaren Geometrie, so erhöht diese Technik die Varianz massiv.

Kapitel IV stellt eine Methode vor, deren Ziel es ist, solche schwierigen Szenen besser handhaben zu können (Meng, Hanika, et al., 2016). Sie ist eine Erweiterung der Methode von Křivánek and d'Eon (2014) und kombiniert verschiedene Möglichkeiten für die Platzierung der Trennebene mit Multiple Importance Sampling (Veach, 1997), um möglichst viele Fälle robust handhaben zu können. Insbesondere richtet sie die Trennebene zu wichtigen Lichtquellen aus, wobei automatisch eine möglichst günstige Platzierung ausgewählt wird.

### 4 Effiziente Modellierung granularer Materialien

Granulare Materialien bestehen aus vielen einzelnen, meist dicht gepackten Körnern. Solche Aggregate können aus Milliarden einzelner Körner bestehen, weshalb die Modellierung nicht rein manuell erfolgen kann. Eine Lösung ist es, auf Physiksimulationen zurückzugreifen, die allerdings äußerst rechen- und speicheraufwendig sind.

In Kapitel V wird daher ein Verfahren vorgestellt, das Körner innerhalb eines vorgegebenen Objektes prozedural instanzieren kann (Meng, Papas, et al., 2015). Es nutzt vorberechnete, kachelbare Kugelpackungen, die effizient auf Strahlschnitte getestet werden können. Während der eigentlichen Instanziierung können Körner zufällig verändert werden, beispielsweise um mehrere Korntypen zu mischen oder ihre Orientierung zu ändern.

Dieses Verfahren ermöglicht Simulationen in fast unbegrenzt großen Räumen voller Körner. Körner und Mischverhältnisse lassen sich verändern, sodass Szenen ohne weitere Vorberechnungen an ein gewünschtes Aussehen angepasst werden können.

Dicht gepackte granulare Materialien stellen außerdem ein äußerst schweres Szenario für Path Tracer dar. Einerseits sind Lichtpfade in solchen Materialien von besonders vielen Streuungen an Grenzflächen und im Inneren der Körner betroffen. Andererseits können übliche Techniken zur Varianzreduktion in granularen Materialien oft nicht erfolgreich eingesetzt werden. Dies führt zu extrem langen Laufzeiten.

Da die tatsächliche Geometrie von granularen Materialien aber nur nahe der Oberfläche eine entscheidende Rolle spielt, stellt Kapitel V ein Verfahren vor, das im Inneren ein approximierendes homogenes partizipierendes Medium annimmt (Meng, Papas, et al., 2015). Die dafür notwendigen Parameter können einerseits aus statistischen Eigenschaften der Kugelpackungen und andererseits aus Vorberechnung mit einzelnen Körnern abgeleitet werden.

Die Approximation durch homogene partizipierende Medien erlaubt weit effizienteres Path Tracing sowie den Einsatz weiterer Methoden für die Varianzreduktion. Insgesamt kann unsere Methode Beschleunigungen um mehrere Größenordnungen erzielen.



# CONTENTS

LEBENS LAUF	VII
ZUSAMMENFASSUNG	IX
1 Beiträge	x
2 Berechnung spektraler Energiedichtefunktionen	xi
3 Varianzreduktion mit Zero-Variance-Methoden	xii
4 Effiziente Modellierung granularer Materialien	xii
I INTRODUCTION	1
1 Contributions	2
1 Efficient and general spectral upsampling	2
2 Biased sampling in participating media	3
3 Efficient rendering of granular materials	3
2 Overview of this dissertation	4
3 List of publications	5
II BACKGROUND	7
1 Probability theory and Monte Carlo simulation	7
1 Measures	7
2 Random experiments, random variables, and probability	8
3 Joint, marginal, and conditional densities	9
4 Expected value	10
5 Bias and consistency	11
6 Variance and covariance	12
7 Chebyshev inequality and the law of large numbers	14
8 Sampling from a probability distribution	14
9 Monte Carlo integration	16
	xv

10 Importance sampling	17
11 Multiple importance sampling	17
2 Radiative transfer	19
1 Radiometric quantities	19
2 Scattering at surfaces	22
3 Participating media	23
4 Emitters	28
5 Irradiance due to emission or scattering	29
6 Sensors	31
7 Light paths and measurement	32
3 Color	34
1 Spectral energy density functions	34
2 Perception of color	35
3 Color spaces	36
4 Monte Carlo-simulation of light transport	38
1 Spectral rendering	38
2 Monte Carlo-estimator for physically-based rendering	39
3 Random walks	39
4 Monte Carlo rendering algorithms	41
<b>III GRID-BASED SPECTRAL UPSAMPLING</b>	<b>43</b>
1 Introduction	43
2 Overview of our upsampling method	45
3 Precomputation	46
1 Optimization procedure	46
2 Grid topology	47
4 Grid-based upsampling	48
5 Results	50
1 Upsampling speed	51
2 Upsampling precision	51
3 Comparison to measured spectra	51
4 Visual evaluation of rendered results	52
6 Conclusion	52

IV	IMPROVED DWIVEDI SAMPLING	55
1	Motivation	55
1	Previous work on sub-surface scattering for skin	56
2	Dwivedi sampling	57
1	Singular eigenfunction expansions	58
2	Approximate importance field	61
3	Finding the diffusion length $\nu_0$	61
4	Biased distributions for Dwivedi sampling	62
5	Slab orientation	64
3	Improving biased sampling	64
1	Closest surface point	64
2	Incident illumination	65
3	Multiple importance sampling	65
4	Selection probability for <i>Classical</i> sampling	66
5	Selection probabilities for biased sampling	67
6	Committing to <i>Incident Illumination</i> sampling	67
4	Results	67
1	Ear scene	68
2	Candle scene	68
3	Buddha scene	68
4	Analysis	72
5	Discussion and Future Work	72
V	A MULTI-SCALE ALGORITHM FOR RENDERING GRANULAR MATERIALS	73
1	Introduction	73
2	Previous work	75
1	Densely packed random media	75
2	Light transport in granular aggregates	76
3	A procedural model for granular materials	77
1	Packed bounding spheres	77
2	Ray intersection	78
4	Teleportation transport	80
1	Free-path distribution	81

2	Teleportation scattering distribution function	82
3	Delta forward scattering	85
4	Grain albedo	85
5	Algorithm	85
5	RTE parameters for granular materials	86
1	Combined free-flight distribution	87
2	Validation	89
6	A multi-scale algorithm for rendering granular materials	92
1	Automatic switching criterion	92
7	Results	94
1	Time to unit variance	94
2	Overview	95
3	Sand Castle scene	96
4	Spices scene	96
5	Snowman scene	97
8	Conclusion and future research	99
1	Diffusion	99
2	Removing EPT	99
3	Packing density	99
4	Heterogeneous mixtures	100
5	Spectral light transport	100
VI	CONCLUSION	101
	GRID BASIS SPECTRA	103
	LIST OF SYMBOLS	115
	BIBLIOGRAPHY	117
	INDEX	127

# I

## INTRODUCTION

This dissertation is concerned with the efficient Monte Carlo simulation of spectral light transport for physically-based rendering of both dense participating media and granular materials.

Physically-based rendering has become an important tool in many different areas of computer graphics: the photorealistic depiction of buildings gives architects a means to communicate precisely with their customers. In advertising, products can be shown even before they are manufactured, or in situations that would be hard to photograph. Industrial designers can make use of light transport simulation to gauge the look and functionality of new creations. Predictive rendering may be used to determine the properties of materials that are hard to obtain or not yet available, and, perhaps most obviously, the gaming and film industries employ a variety of image synthesis methods to create artificial worlds of astounding detail.

The foundations for physically-based light transport simulation, or, more generally, radiative transfer, were laid decades ago in the context of the simulation of radiation in the galaxy and neutron transport literature. Despite this long heritage and the varied applications, the field still has many unsolved problems. Computing systems become more and more powerful, and storage becomes more abundant by the year, but at the same time artists use this as a chance to create more realistic scenes with higher scene complexity. Therefore, efficiency in both storage and processing requirements remains an active topic of research. Additionally, there are particularly hard rendering problems: scenes in which light paths experience specular or highly glossy interactions, dense participating media, scenes with complex visibility, and others. These problems demand novel solutions, and a lot of contemporary research focuses on them.

The general consensus seems to be that simulation methods based on Monte Carlo integration are the most promising, because the measurement at a virtual sensor can be expressed as a high-dimensional integral. Such methods are based on the concept of random walks, which are performed in virtual scenes to find paths of photon populations, which we also call light paths. These are ultimately used to measure radiative energy. In the context of image synthesis, the measurement is the output image.

The main problem of Monte Carlo light transport simulation is image noise, which results from the fact that the energy measured along a path is a random variable. Even worse, some path configurations are extremely hard or, depending on the path sampling algorithm, even impossible to generate. Many of the recent publications in photorealistic image synthesis therefore focus either on noise reduction or on dealing with particularly difficult path configurations. In Chapters IV and V, we will present techniques that can be categorized as noise reduction methods.

The dimensionality of the integral that is solved is determined by the radiative transfer equation. Implementations vary in the number of dimensions they consider explicitly, but some of the most common ones are positions on the sensor and in the scene, position on the lens aperture to obtain depth of field and time to obtain motion blur. If the wavelength of the transported light is considered explicitly, one generally speaks of spectral rendering or spectral light transport. Spectral rendering adds color noise to the image, but also has various advantages. For example, wavelength-dependent phenomena such as dispersion at refractive boundaries can be represented properly, since the path construction can be adapted to the wavelength. However, input color values are generally provided as RGB triplets, which must be converted to a spectral representation that suits the renderer. We will present a technique that solves this underconstrained problem in Chapter III.

Radiative transfer is not only concerned with light transport between surfaces, but models scattering, absorption and even emission of radiative energy from volumes. Examples for such materials include the human skin, wax, various foods, sugar, sand, and many more. This is an especially hard problem to solve: not only can volumetric data sets require enormous amounts of disk space, but rendering them can take prohibitively long. This problem intensifies for dense materials, especially those with a low absorption, because extremely long random walks can have a high contribution to the output image. The methods we show in Chapters IV and V are specifically designed to improve render times in highly scattering materials. In Chapter V, we additionally introduce a procedural technique that can generate almost infinitely large granular assemblies but requires only very little storage space.

## 1 Contributions

We include contributions from three separate but thematically connected publications in this dissertation. All of them were published at peer-reviewed conferences (see Section I.3).

### 1.1 Efficient and general spectral upsampling

In spectral rendering, in contrast to tristimulus rendering with RGB color values, an explicit integration of energy over the spectrum of visible light is performed. To author scenes for spectral rendering, it is often desirable to be able to find emission and reflectance spectra corresponding to tristimulus input data. This problem, which is called spectral upsampling, is heavily underconstrained, and generally requires

expensive optimization procedures. The state of the art in the field of image synthesis is a surprisingly old technique by (Smits, 1999), which performs the optimization in a preprocess to achieve efficient run-time behavior.

Our technique, which was published in (Meng, Simon, et al., 2015), improves on this prior work by allowing for arbitrary input color spaces. Additionally, it achieves round-trip conversions from an RGB input to spectrum and back that are exact up to floating-point accuracy, and still incurs a low run-time cost.

## 1.2 Biased sampling in participating media

Monte Carlo rendering relies on random walks to generate paths that connect sensors to light sources. In the presence of participating media, especially high-albedo, dense, isotropically scattering media, these random walks can produce very long paths not unlike Brownian motion. This can cause excessively long rendering times.

Ideally, one would prefer to generate random walks that aim for bright light sources, so that paths with a high contribution also have a higher chance to be sampled.

One way of generating such paths is Dwivedi sampling (Dwivedi, 1982a,b) which changes the sampling distributions to give preference to certain directions. Dwivedi sampling was introduced to computer graphics by Křivánek and d’Eon (2014), who chose to bias the random walk towards the point of entry of the random walk into the medium.

In strongly backlit scenes, where the light source is opposite the point of entry, this algorithm fails: it draws many samples away from the main light source which results in strong noise.

We improve on this algorithm by introducing biasing towards the closest surface point and towards strong light sources (Meng, Hanika, et al., 2016). This enables our technique to handle a wide variety of previously difficult scenes more robustly.

## 1.3 Efficient rendering of granular materials

Granular materials are aggregates of many individual grains, such as sand, salt or even snow. Such aggregates are extremely difficult to handle even for modern Monte Carlo rendering systems. There are multiple reasons for this: complex visibility increases raytracing cost and prevents efficient techniques such as next event estimation or bidirectional path tracing. Additionally, individual grains may have highly specular surfaces and complex scattering interiors. Finally, modeling and storing materials composed of many millions of grains can be prohibitively expensive.

Previous work in the area of rendering granular materials employs a precomputation to accelerate rendering (Moon and Marschner, 2006), but requires the whole medium to be instantiated for this precomputation. This work also does not address the problem of scene authoring: granular aggregates must be generated by placing grains, and these grains must then also be stored. Large aggregates containing billions of grains are not uncommon, but the storage requirements can render such scenes intractable.

In our work (Meng, Papas, et al., 2015), we propose a procedural modeling technique, which decouples the aggregate composition from the grain geometry and allows for quick iterative design. We also propose a theoretical framework called „teleportation transport“, which models the propagation of light in granular materials as a two-step process: propagation of energy inside grains on one hand and between grains on the other. We then use this framework to derive classical volume parameters from per-grain tabulated scattering functions and an analytical model for chord lengths in sphere packings. Finally, we propose a rendering algorithm that adaptively switches between explicit path tracing of grains and a volumetric representation.

Although our algorithm is approximative, it yields visually accurate results and can be shown to speed up convergence by multiple orders of magnitude.

## 2 Overview of this dissertation

The remainder of this dissertation is structured as follows:

In chapter II, we will review fundamental mathematical and physical concepts which are required for following chapters.

Chapter III is concerned with spectral rendering. We will investigate why spectral rendering is important for photorealistic image synthesis, and how wavelength-dependent functions can be created from RGB triplets. The chapter contains material that was previously published in Computer Graphics Forum (Meng, Simon, et al., 2015).

In Chapter IV, we demonstrate a method based on zero-variance random walks. Our technique can lead to a significant reduction in variance, especially in the presence of strong backlights in scenes with high-albedo, dense participating media. The contents of this chapter were previously published in Computer Graphics Forum (Meng, Hanika, et al., 2016).

Chapter V presents novel techniques for modeling and rendering granular materials, such as sand, snow or sugar. The techniques were previously published in Transactions on Graphics (Meng, Papas, et al., 2015). The paper additionally contains a diffusion approximation method, which was used in the doctoral thesis by Papas (2015), but is not included in this dissertation.

Chapter VI briefly reviews the thesis, and is followed by appendices.

### 3 List of publications

#### Publications fully or partially included in this dissertation

- Meng, J., Hanika, J., and Dachsbacher, C. (2016). "Improving the Dwivedi Sampling Scheme". In: *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 35.4, pp. 37–44.
- Meng, J., Papas, M., Habel, R., Dachsbacher, C., Marschner, S., Gross, M., and Jarosz, W. (2015). "Multi-Scale Modeling and Rendering of Granular Materials". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 34.4, 49:1–49:13.
- Meng, J., Simon, F., Hanika, J., and Dachsbacher, C. (2015). "Physically Meaningful Rendering using Tristimulus Colours". In: *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 34.4, pp. 31–40.

#### Additional publications

- Jarosz, W., Meng, J., Papas, M., Habel, R., Dachsbacher, C., and Marschner, S. (2016). *Bidirectional point distribution functions for rendering granular media*. US Patent 9,472,016.
- Schmidt, T.-W., Novák, J., Meng, J., Kaplanyan, A. S., Reiner, T., Nowrouzezahrai, D., and Dachsbacher, C. (2013). "Path-Space Manipulation of Physically-Based Light Transport". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 32.4, 129:1–129:11.



# II

## BACKGROUND

In this chapter, we will introduce basic concepts that are required to follow our original work in the later chapters. Since modern, physically-based rendering is based on Monte Carlo simulation of light transport, we must first briefly review the main mathematical tool, probability theory, in Section II.1. We will then introduce the physical basis foundation of light transport simulation, radiative transfer and specifically the radiative transfer equation, in Section II.2. The human perception of color is another key concept in our work, and we will outline some basic ideas in Section II.3. Finally, we will discuss how these techniques can then be combined into a method that is able to generate images in Section II.4.

### 1 Probability theory and Monte Carlo simulation

In this section, we will introduce some basic concepts of probability theory. We refer the reader to textbooks such as (Haghighat, 2015; Lehn and Wegman, 2006; Pitman, 2006; Rubinstein and Kroese, 2008; Spanier and Gelbard, 1969) for a more in-depth formal introduction.

#### 1.1 Measures

Measures are of prime importance in probability theory because probability functions are, in fact, measures. Intuitively, a measure is a function that assigns a positive size to a set in a well-behaved fashion. Before we can formally introduce measures, however, we must first briefly look at sets.

Given sets  $M_1, \dots, M_n$  we write their *union* as

$$\bigcup_{i=1}^n M_n = \{x | x \in M_1 \vee \dots, \vee x \in M_n\}. \quad (\text{II.1})$$

The *complement* of a subset  $A \subseteq M$  with respect to  $M$  is defined as

$$A^c = M \setminus A = \{x \in M | x \notin A\}. \quad (\text{II.2})$$

The *power set*  $2^M$  associated with any set  $M$  is the set of all subsets, including the empty set  $\emptyset$ , and  $M$  itself. For example, if  $M = \{a, b, c\}$ , then

$$2^M = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\} \quad (\text{II.3})$$

A subset  $\Sigma_M \subseteq 2^M$  is called a  $\sigma$ -*algebra* (Lehn and Wegman, 2006) if three conditions hold. First,  $M$  itself must be in  $\Sigma_M$ . Second, if a set  $A$  is in  $\Sigma_M$ , then so must be its complement  $A^c$ . Third, if  $A_1, \dots, A_n$  are in  $\Sigma_M$ , then so must be  $\bigcup_{i=1}^n A_i$ .

Note that the first two conditions require that  $\emptyset$  be contained in any  $\sigma$ -algebra. This means that the smallest possible  $\sigma$ -algebra is  $\{\emptyset, M\}$ , while the largest possible  $\sigma$ -algebra is the power set  $2^M$ . A pair  $(M, \Sigma_M)$  is called *measurable space*.

Given a measurable space  $(M, \Sigma_M)$ , a *measure* is a function  $\mu$  that associates numbers in  $\mathbb{R}_0^+ \cup \{\infty\}$  with the elements of a  $\sigma$ -algebra such that for all  $A, A_i \in \Sigma_M$

$$\mu(A) \geq 0 \quad (\text{II.4})$$

$$\mu(\emptyset) = 0 \quad (\text{II.5})$$

$$\sum_{i=1}^n \mu(A_i) = \mu(\bigcup_{i=1}^n A_i). \quad (\text{II.6})$$

This definition is exactly what one would expect from measuring in reality. As an example, assume one wants to measure a stick of length 40cm with a ruler that is only 30cm long. In this case, one could measure the first 30cm, make a mark in the stick, and then shift the ruler to measure the remaining 10cm. The results can simply be added to obtain the full length, which is exactly what the third condition expresses.

## 1.2 Random experiments, random variables, and probability

A *random experiment* is a process that is well-defined and repeatable, but yields unpredictable results (Lehn and Wegman, 2006). An example is throwing a six-sided die. The set of all possible outcomes  $\omega$  is usually written as  $\Omega$ . Note that the use of  $\omega$  as an outcome is restricted to this section. In the rest of the thesis,  $\omega$  will usually denote a direction or a solid angle.

*Random events* are simply elements of some  $\sigma$ -algebra over  $\Omega$ , that is, they are sets of outcomes. An event  $\Lambda \subseteq \Omega$  occurs if the associated random experiment is performed and the outcome  $\omega$  is in  $\Lambda$ . The empty set is also called the *impossible event* because it can never occur, and the full set  $\Omega$  is called the *certain event* (Spanier and Gelbard, 1969). Since random events are sets of outcomes, they can be composed using set operations.

The central concept in probability theory is that of probability itself, and we are now equipped to define it properly as a *probability measure*. A probability measure  $P$  is a measure as defined in Sec. II.1.1 that associates random events with numbers in  $[0, 1]$  such that the measure of the certain event  $P\{\Omega\} = 1$ . Note the braces in this notation, which indicate that probability measures operate on sets rather than

individual outcomes.

It is often convenient to associate the outcomes of a random experiment with numbers. A function  $X : \Omega \mapsto \mathbb{R}$  that does this is called a *random variable*. Random experiments tend to give rise to random variables quite naturally. As an example, the numbers on the sides of six-sided die can be seen as a random variable.

Random variables can be used to construct sets for which we can then determine a probability measure. The usual notation for this is

$$F(x) = P\{X \leq x\} = P\{\omega \in \Omega | X(\omega) \leq x\}. \quad (\text{II.7})$$

The function  $F(x)$  is also called the *cumulative distribution function*, or CDF, and it fully describes the probability distribution of the associated random variable. This notation can be extended to include the probability of events where the random variable assumes values in some set  $M \subseteq \mathbb{R}$  as

$$P\{X \in M\}. \quad (\text{II.8})$$

There are *discrete* and *continuous* random variables. Generally, discrete random variables have a CDF that is a step function, while continuous random variables have a differentiable CDF with derivative  $p(x)$ :

$$F(x) = \int_{-\infty}^x p(x) dx. \quad (\text{II.9})$$

The function  $p(x)$  is then called the *probability density function*, or PDF, of the random variable. If we are interested again in the probability of events where the random variable assumes values in some set, we can now write

$$P\{X \in M\} = \int_M p(x) dx. \quad (\text{II.10})$$

In discrete random variables, this is replaced by a summation. In fact, this will be true for most of the following concepts (Lehn and Wegman, 2006; Spanier and Gelbard, 1969), and so we shall restrict this discussion to the continuous case.

### 1.3 Joint, marginal, and conditional densities

The result of an experiment that yields multiple random variables  $(X_1, \dots, X_N)$  can be described using a *joint probability measure*

$$P\{X_1 \leq t_1, \dots, X_N \leq t_N\} = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_N} p(x_1, \dots, x_N) dx_N \dots dx_1, \quad (\text{II.11})$$

where  $p(x_1, \dots, x_N)$  is called the *joint probability density function* of the random variables.

An example is an experiment where the velocity of particles is measured as a

three-dimensional vector. Each component of the vector can be interpreted as a one-dimensional random variable.

Individual random variables can be removed from the joint PDF by integration over all possible values:

$$p(x_2, \dots, x_N) = \int_{-\infty}^{\infty} p(x_1, \dots, x_N) dx_1. \quad (\text{II.12})$$

This is called the *marginal PDF* (Spanier and Gelbard, 1969) of  $X_2, \dots, X_N$  and it describes the distribution of probability mass if nothing is known about  $X_1$ .

If the joint PDF can be written as a product of the marginal PDFs of individual random variables,

$$p(x_1, \dots, x_N) = p(x_1) \cdots p(x_N), \quad (\text{II.13})$$

then the random variables are said to be *independent* (Pitman, 2006). Otherwise, the dependence is described with the the function

$$p(x_2, \dots, x_N | x_1) = \frac{p(x_1, \dots, x_N)}{p(x_1)}, \quad (\text{II.14})$$

which is called the *conditional PDF* of  $X_2, \dots, X_N$  given  $x_1$ . From Eqn. (II.14), it follows that if the random variables  $X_1, X_2$  are independent, then  $p(x_2 | x_1) = p(x_2)$  and  $p(x_1 | x_2) = p(x_1)$ . The conditional CDF of a random variable  $X_1$  given  $(X_2, \dots, X_N)$  is

$$F(t_1 | x_2, \dots, x_N) = \int_{-\infty}^{t_1} p(x_1 | x_2, \dots, x_N) dx_1. \quad (\text{II.15})$$

## 1.4 Expected value

The *expected value* of a random variable  $X$  is the value that one would expect on average if the experiment is repeated many times. It is defined as

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx \quad (\text{II.16})$$

if  $X$  is continuous with PDF  $p(x)$  (Lehn and Wegman, 2006). Since the integral is a linear operation, the following rule holds for constants  $a, b \in \mathbb{R}$ :

$$E[aX + b] = aE[X] + b \quad (\text{II.17})$$

If  $h(X_1, \dots, X_N)$  is a function of  $N$  random variables  $X_i$  assumes values in  $M_i$ , then  $h(X_1, \dots, X_N)$  is itself a random variable, and its expected value is

$$E[h(X_1, \dots, X_N)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \dots, x_N) p(x_1, \dots, x_N) dx_N \cdots dx_1 \quad (\text{II.18})$$

with the joint PDF  $p(x_1, \dots, x_N)$  (Lehn and Wegman, 2006).

As a special case, the random variable  $S_N = \sum_{i=1}^N X_i$  has the expected value

$$\begin{aligned}
 E[S_N] &= \int_{-\infty}^{\infty} \left( \sum_{i=1}^N x_i \right) p(x_1, \dots, x_N) dx_N \cdots dx_1 \\
 &= \sum_{i=1}^N \int_{-\infty}^{\infty} x_i p(x_1, \dots, x_N) dx_N \cdots dx_1 \\
 &= \sum_{i=1}^N \int_{-\infty}^{\infty} x_i \left[ \int_{-\infty}^{\infty} p(x_1, \dots, x_N) \prod_{j \neq i} dx_j \right] dx_i \quad (\text{II.19}) \\
 &= \sum_{i=1}^N \int_{-\infty}^{\infty} x_i p(x_i) dx_i \\
 &= \sum_{i=1}^N E[X_i]
 \end{aligned}$$

Note that we only used the definition of a marginal PDF as given in Eqn. (II.12).

## 1.5 Bias and consistency

Given an estimator  $\widehat{F}$  for the quantity  $q$ , the *bias*  $\theta$  of the estimator with respect to  $q$  is defined as the error

$$\theta(\widehat{F}, q) = E[\widehat{F}] - q. \quad (\text{II.20})$$

An estimator is called *unbiased* if the bias is zero (Rubinstein and Kroese, 2008).

As an example consider an experiment that is repeated  $N$  times. The result are values  $(X_1, \dots, X_N)$ , which can also be interpreted as  $N$  identically distributed random variables (Spanier and Gelbard, 1969). We will write the common expected value as  $E[X_i] = E[X]$  for all  $i \in \{1, N\}$ . The *sample mean* of the experiment can be written with the random variable  $S_N$  from Eqn. (II.19) as  $S_N/N$ . Its expected value is

$$E[S_N/N] = \frac{1}{N} \sum_{i=1}^N E[X] = E[X]. \quad (\text{II.21})$$

With respect to the expected value  $E[X]$ , the bias of this estimator is

$$\theta(S_N/N, E[X]) = E[S_N/N] - E[X] = 0. \quad (\text{II.22})$$

The sample mean is therefore an unbiased estimator for the expected value of  $X$ .

If an estimator is biased, then it will over- or underestimate the value  $q$  systematically. An estimator  $\widehat{F}_N$  is called *consistent* if the error can be made arbitrarily small

by increasing the sample size  $N$  (Spanier and Gelbard, 1969):

$$\widehat{F}_N \text{ consistent} \iff \lim_{N \rightarrow \infty} P\{|\widehat{F}_N - q| \geq \epsilon\} = 0. \quad (\text{II.23})$$

## 1.6 Variance and covariance

The *variance* of a probability distribution is the expected squared deviation from the expected value (Lehn and Wegman, 2006):

$$\begin{aligned} \text{Var}[X] &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 p(x) dx \\ &= \int_{-\infty}^{\infty} x^2 p(x) dx - 2E[X] \int_{-\infty}^{\infty} xp(x) dx + E[X]^2 \int_{-\infty}^{\infty} p(x) dx \quad (\text{II.24}) \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2. \end{aligned}$$

Variance is a measure of how well the expected value represents the underlying probability distribution. Since variance is a square function, it is sometimes more convenient to use the *standard deviation*  $\sigma$  instead:

$$\sigma[X] = \sqrt{\text{Var}[X]}. \quad (\text{II.25})$$

Scaling a random variable by  $a \in \mathbb{R}$  and shifting it by  $b \in \mathbb{R}$  will scale the variance with  $a^2$ :

$$\begin{aligned} \text{Var}[aX + b] &= \int_{-\infty}^{\infty} (ax + b - E[ax + b])^2 p(x) dx \\ &= \int_{-\infty}^{\infty} (ax + b - aE[X] - b)^2 p(x) dx \quad (\text{II.26}) \\ &= \int_{-\infty}^{\infty} a^2 \cdot (x - E[X])^2 p(x) dx \\ &= a^2 \text{Var}[X]. \end{aligned}$$

The shift  $b$  does appear in the result because variance is measured relative to the expected value, which is shifted by an equal amount.

Given two random variables  $X$  and  $Y$ , the *covariance* of the two is defined in a similar fashion:

$$\text{Cov}[X, Y] = E[(X - E[X]) \cdot (Y - E[Y])] \quad (\text{II.27})$$

The covariance measures correlation between the two random variables, or in other words, how strongly they exhibit a linear dependency (Rubinstein and Kroese, 2008). If  $\text{Cov}[X, Y] = 0$ , then there is no linear dependency, and the random variables are said to be uncorrelated (Lehn and Wegman, 2006).

The variance of the sum of two random variables  $X$  and  $Y$  is (Lehn and Wegman, 2006)

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]. \quad (\text{II.28})$$

This intuitively makes sense considering that the variance is a quadratic function, and this formula indeed resembles the binomial theorem  $(a + b)^2 = a^2 + b^2 + 2ab$ . Also note that if  $X$  and  $Y$  are uncorrelated, then  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

If  $X$  and  $Y$  are independent, then they are also uncorrelated:

$$\begin{aligned} \text{Cov}[X, Y] &= \iint_{-\infty}^{\infty} (x - E[X]) \cdot (y - E[Y]) p(x, y) dy dx \\ &= \iint_{-\infty}^{\infty} (x - E[X]) p(x) \cdot (y - E[Y]) p(y) dy dx \\ &= \int_{-\infty}^{\infty} (x - E[X]) p(x) dx \cdot \int_{-\infty}^{\infty} (y - E[Y]) p(y) dy \\ &= \left( \int_{-\infty}^{\infty} x p(x) dx - E[X] \cdot \int_{-\infty}^{\infty} p(x) dx \right) \\ &\quad \cdot \left( \int_{-\infty}^{\infty} y p(y) dy - E[Y] \cdot \int_{-\infty}^{\infty} p(y) dy \right) \\ &= \left( \int_{-\infty}^{\infty} x p(x) dx - E[X] \right) \cdot \left( \int_{-\infty}^{\infty} y p(y) dy - E[Y] \right) \\ &= 0 \end{aligned} \quad (\text{II.29})$$

It follows that the variance of the sample mean  $S_N/N$  is

$$\text{Var}[S_N/N] = N \cdot \frac{\text{Var}[X]}{N^2} = \frac{\text{Var}[X]}{N}. \quad (\text{II.30})$$

The *mean squared error* of  $N$  samples is an unbiased estimator for  $\text{Var}[X]$ :

$$\begin{aligned} E \left[ \frac{1}{N} \sum_{i=1}^N N(X_i - E[X])^2 \right] &= \frac{1}{N} \sum_{i=1}^N E [(X_i - E[X])^2] \\ &= \frac{1}{N} \sum_{i=1}^N E [X_i^2 - 2X_i E[X] + E^2[X]] \\ &= \frac{1}{N} \sum_{i=1}^N E[X_i^2] - E[X]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \text{Var}[X_i] + E[X]^2 - E[X]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \text{Var}[X] = \text{Var}[X]. \end{aligned} \quad (\text{II.31})$$

Replacing the expected value  $E[X]$  with the sample mean  $S_N/N$  in the formula above yields the *sample variance*. This is, however, a biased estimator for  $\text{Var}[X]$ , and the *unbiased sample variance* should be used, instead (Spanier and Gelbard, 1969):

$$\frac{1}{N+1} \sum_{i=1}^N (X_i - S_N/N)^2. \quad (\text{II.32})$$

## 1.7 Chebyshev inequality and the law of large numbers

The meaning of the variance as a measure of deviation from the expected value is formalized in the *Chebyshev inequality* (Lehn and Wegman, 2006). It states that if the variance  $\text{Var}[X]$  exists for a random variable  $X$ , then for every  $c > 0$ ,

$$P\{|X - E[X]| \geq c\} \leq \frac{\text{Var}[X]}{c^2}. \quad (\text{II.33})$$

Substituting the sample mean for  $X$  in the Chebyshev inequality, and using an arbitrarily small value  $\epsilon$  for the bound  $c$ , we can apply Eqns. (II.26) and (II.21) to obtain

$$P\{|S_N/N - E[X]| \geq \epsilon\} \leq \frac{\text{Var}[X]}{N \cdot \epsilon^2}. \quad (\text{II.34})$$

In the limit, the right-hand side must vanish, and so we can write

$$\lim_{N \rightarrow \infty} P\{|S_N/N - E[X]| \geq \epsilon\} = 0. \quad (\text{II.35})$$

This is called the weak *law of large numbers* (Lehn and Wegman, 2006). It states that if the sample is large enough, then it is almost certain that the sample mean will be close to the expected value.

## 1.8 Sampling from a probability distribution

A very common problem is generating random numbers that are distributed according to some probability density function  $p(x)$  with CDF  $F(x)$ . We will describe two methods here that use as their input uniform random variables  $Y$  with

$$P\{Y \leq y\} = y \quad \forall y \in [0, 1]. \quad (\text{II.36})$$

In computer graphics, suitable numbers are usually generated using pseudo-random number generators such as the Mersenne Twister (Matsumoto and Nishimura, 1998), or low discrepancy sequences like the Halton sequence (Halton, 1964).

The first sampling method is called *rejection sampling* (Spanier and Gelbard, 1969). It requires the PDF to vanish outside a known interval  $[a, b]$ , and must have a global upper bound

$$x_{\max} \geq p(x) \quad \forall x. \quad (\text{II.37})$$

The first step in this algorithm is to draw uniform random numbers  $Y_1$  and  $Y_2$ . The sample  $Z = a + Y_1(b - a)$  is accepted if

$$p(Z) \geq Y_2 x_{\max}. \quad (\text{II.38})$$

Otherwise, the sample is rejected and the process must be repeated.

To see that this indeed generates samples with the proper distributeion, denote with  $P_i\{Z \leq t\}$  the probability that  $X \leq t$  is accepted after  $i$  rejections. Then,

$$P_0\{Z \leq t\} = \int_a^t \frac{1}{b-a} \frac{p(x)}{x_{\max}} dx = F(x) \frac{1}{(b-a)x_{\max}} = F(x) \frac{1}{\Lambda}, \quad (\text{II.39})$$

where  $1/(b-a)$  is the probability density of  $a + Y_1(b-a)$  and

$$\int_0^{p(x)} \frac{1}{x_{\max}} dy = \frac{p(x)}{x_{\max}} \quad (\text{II.40})$$

is the acceptance probability. Similarly,

$$\begin{aligned} P_{i+1}\{Z \leq t\} &= P_i\{Z \leq t\} \cdot \int_a^b \frac{1}{b-a} \left(1 - \frac{p(x)}{x_{\max}}\right) dx \\ &= P_i\{Z \leq t\} \cdot \left(1 - \frac{1}{\Lambda}\right), \end{aligned} \quad (\text{II.41})$$

which models rejection in the first step. More generally, for all  $i \geq 0$ ,

$$P_i\{Z \leq t\} = F(x) \frac{1}{\Lambda} \left(1 - \frac{1}{\Lambda}\right)^i. \quad (\text{II.42})$$

Thus, the CDF of the samples generated via rejection is

$$P\{Z \leq t\} = F(x) \frac{1}{\Lambda} \sum_{i=0}^{\infty} \left(1 - \frac{1}{\Lambda}\right)^i = F(x), \quad (\text{II.43})$$

and rejection sampling produces samples that are distributed with  $F(x)$ . The method can be employed in two or more dimensions (Spanier and Gelbard, 1969). However, if  $x_{\max}$  is a loose bound, then the acceptance probability can become arbitrarily small. This may lead to long and unpredictable run times.

It is therefore desirable to generate samples with constant effort. This is possible if the CDF is invertible in the sense that a function  $F^{-1}$  exists such that  $F^{-1}(F(x)) = x$  and  $F(F^{-1}(Y)) = Y$ , since the random variable  $F^{-1}(Y)$  is distributed with CDF  $F(x)$  (Spanier and Gelbard, 1969):

$$P\{F^{-1}(Y) \leq x\} = P\{Y \leq F(x)\} = F(x). \quad (\text{II.44})$$

We can perform these operations because the CDF is monotonically increasing.

Note that the second step is simply the definition of a uniform probability distribution.

This method is called the *inversion method*. It can also be used to generate multi-dimensional samples  $(x, y)$  distributed according to some joint probability density function  $p(x, y)$  using the fact that  $p(x, y) = p(x)p(y|x) = p(y)p(x|y)$  (Spanier and Gelbard, 1969).

## 1.9 Monte Carlo integration

Monte Carlo integration (Metropolis and Ulam, 1949) combines two insights. First, the expected value of a random variable can be used to compute the value of an integral:

$$\begin{aligned} E[f(X)] &= \int_{-\infty}^{\infty} f(x) p(x) dx \\ \iff E[f(X)/p(X)] &= \int_{-\infty}^{\infty} f(x) dx. \end{aligned} \tag{II.45}$$

Second, and this is justified by the law of large numbers, we can use the mean of a large enough sample to robustly estimate  $E[X]$ . This results in the *Monte Carlo* estimator

$$\hat{F}_N = \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)}. \tag{II.46}$$

This estimator has the distinct advantage that it only needs to evaluate  $f(x)$  at individual points, and the integral need not be known analytically. This is especially useful for evaluating the high-dimensional integrals of light transport theory. We will cover this in more detail later in this chapter.

As we have seen in Eqn. (II.30), the variance of the Monte Carlo estimator is

$$\text{Var}[\hat{F}_N] = \frac{\text{Var}[f(X)/p(X)]}{N}. \tag{II.47}$$

This means that the standard deviation is inversely proportional to the the square root of  $N$ :

$$\sigma[\hat{F}_N] = \frac{\sigma[f(X)/p(X)]}{\sqrt{N}}. \tag{II.48}$$

In other words, to reduce the error in the estimator by a factor of two, one needs to quadruple the sample size. This is a major limitation, because it means that it becomes more and more impractical to reduce the error in a Monte Carlo estimator.

If the variance, or a good estimator of variance, of a Monte Carlo estimator is plotted over the number of samples  $N$  with logarithmic axes, then the graph should show a

line with slope  $-1$ :

$$\begin{aligned} \log \frac{\text{Var}[f(X)/p(X)]}{\exp N} &= \log \text{Var}[f(X)/p(X)] - \log \exp N \\ &= -N + \log \text{Var}[f(X)/p(X)]. \end{aligned} \quad (\text{II.49})$$

Conversely, if the plots instead show curved lines approaching some value, then the estimator must be biased. This can be a useful tool for finding errors.

## 1.10 Importance sampling

The probability density  $p(X)$  can be chosen freely, as long as it is nonzero for all possible values of  $X$ . In particular, given the probability density

$$p^*(x) = \frac{f(x)}{\int_{-\infty}^{\infty} f(y) dy}, \quad (\text{II.50})$$

the estimator always yields the perfect result:

$$\frac{1}{N} \sum_{i=1}^N \frac{f(x_i) \int_{-\infty}^{\infty} f(y) dy}{f(x_i)} dx = \int_{-\infty}^{\infty} f(y) dy. \quad (\text{II.51})$$

This optimal probability density is not used in practice, since it requires knowledge of the solution. However, it is often possible to find probability densities that are proportional to at least a part of the integrand, which in turn can lead to a decrease in variance. The deliberate choice of a probability density is called *importance sampling* (M. H. Kalos and Whitlock, 2008).

## 1.11 Multiple importance sampling

The combination of multiple sampling techniques can be expressed in the multiple importance sampling framework (MIS) introduced by Veach and L. J. Guibas (1995).

Veach defined the *one sample model* as the estimator

$$\frac{w(x, t) \cdot f(x)}{p(t) \cdot p(x|t)}, \quad (\text{II.52})$$

where one of many random techniques  $t \in \mathcal{T}$  is chosen at random to generate a single sample. If the weights are normalized so that

$$\int_{\mathcal{T}} w(x, t) = 1 \quad \forall x, \quad (\text{II.53})$$

then the expected value of this estimator over all techniques and values of  $x$  is

$$\begin{aligned}
 & \int_{\mathcal{T}} p(t) \int_{-\infty}^{\infty} p(x|t) \frac{w(x,t) \cdot f(x)}{p(t) \cdot p(x|t)} dx dt \\
 &= \int_{\mathcal{T}} \int_{-\infty}^{\infty} w(x,t) \cdot f(x) dx dt \\
 &= \int_{-\infty}^{\infty} f(x) \int_{\mathcal{T}} w(x,t) dt dx \\
 &= \int_{-\infty}^{\infty} f(x) dx,
 \end{aligned} \tag{II.54}$$

and so the one sample model is an unbiased estimator for the integral over  $f(x)$ . The same concept applies for a finite number of techniques, where the normalization constraint Eqn. II.53 is replaced by a sum.

One set of weights that fulfills the normalization is

$$w(x,t) = p(t), \tag{II.55}$$

in which case the estimator looks like the normal Monte Carlo estimator where the fact that multiple techniques are used is completely ignored:

$$\frac{f(x)}{p(x|t)}. \tag{II.56}$$

This combination with *implicit weights* is useful if there are infinitely many techniques, since only one PDF  $p(x,t)$  has to be evaluated. The provably optimal way (Veach and L. J. Guibas, 1995) of combining  $M$  techniques using the one sample model, however, is the *balance heuristic*

$$w(x,t) = \frac{p(x|t)}{\sum_{m=1}^M p(x|t_m)}. \tag{II.57}$$

Note that for this heuristic, the PDF for all techniques must be evaluated in order to compute the weights.

In contrast to the one sample model, one may sometimes want to draw samples from multiple techniques at once. The *multi sample model* facilitates this (Veach, 1997):

$$\sum_i \frac{1}{M_i} \sum_{j=1}^{M_i} \frac{w(x_{i,j}, t_i) f(x_{i,j})}{p(x_{i,j}|t_i)}. \tag{II.58}$$

If the weights are again normalized so that

$$\sum_{i=1}^M w(x, t_i) = 1 \quad \forall x, \tag{II.59}$$

then the estimator is also unbiased:

$$\begin{aligned}
 & \sum_{i=1}^M \frac{1}{M_i} \sum_{j=1}^{M_i} \int_{-\infty}^{\infty} p(x|t_i) \frac{w(x, t_i) f(x)}{p(x|t_i)} dx \\
 &= \int_{-\infty}^{\infty} f(x) \sum_{i=1}^M \frac{1}{M_i} \sum_{j=1}^{M_i} w(x, t_i) dx \\
 &= \int_{-\infty}^{\infty} f(x) \sum_{i=1}^M w(x, t_i) dx \\
 &= \int_{-\infty}^{\infty} f(x) dx.
 \end{aligned} \tag{II.60}$$

If one of the  $M$  techniques is particularly good at generating certain samples (Veach, 1997), then the *power heuristic*

$$w(x, t) = \frac{p^\alpha(x|t)}{\sum_{m=1}^M p^\alpha(x|t_m)} \tag{II.61}$$

yields lower variance with the multi sample model than even the balance heuristic. In practice,  $\alpha = 2$  is generally used.

## 2 Radiative transfer

### 2.1 Radiometric quantities

Radiometric quantities can be used to describe how strong radiation is at a given position, and in which direction it is flowing. They are basic physical concepts, and we will need some of them to build physically based rendering algorithms later. This section contains a brief introduction that will suffice for the understanding of this thesis. However, more comprehensive descriptions can be found in textbooks on optics and radiative transfer (Chandrasekhar, 1960; Ishimaru, 1978; Salah and Teich, 2007).

#### 2.1.1 Solid angle

An angle is defined as circular arc length  $l$  relative to the circle radius,  $r$ :

$$\theta = \frac{l}{r}. \tag{II.62}$$

This can be understood as describing a set of directions in two dimensions. Although angles are measured in units of *radian* [rad], they are really dimensionless, since they are the result of dividing two lengths both measured in m.

In three dimensions, the corresponding quantity is called *solid angle*, which is the spherical surface  $A$  relative to the square sphere radius:

$$\omega = \frac{A}{r^2}. \quad (\text{II.63})$$

Solid angles are measured in units of *steradian* [sr], but again they are really a dimensionless quantity.

### 2.1.2 Speed of light

The speed of light in vacuum  $c_0$  is a fundamental constant in nature. Its value is known with certainty, because the length of a meter is defined to be the distance travelled by light in vacuum during a  $1/299792458$ th of a second (BIPM, 1983). Conversely,

$$c_0 = 299792458 \text{ m s}^{-1}. \quad (\text{II.64})$$

In a dielectric medium, light generally slows down by a factor  $n$ , which is called the material's *index of refraction* (Salah and Teich, 2007):

$$c(n) = \frac{c_0}{n}, \quad (\text{II.65})$$

The absorption in semiconductors or conductors can be expressed by using a complex index of refraction  $n = \eta + i\kappa$ , where the real part models refraction and the complex part models absorption.

### 2.1.3 Photon energy

Photon energy in a material depends on the speed of light  $c$  and the photon wavelength  $\lambda$ , or equivalently its frequency  $\nu = c/\lambda$ :

$$Q(\lambda) = \frac{hc}{\lambda} \quad (\text{II.66})$$

$$Q(\nu) = h\nu, \quad (\text{II.67})$$

where  $h = 6.62607004 \cdot 10^{-34}$  Js is Planck's constant.

In the following, we will omit the explicit dependence on the wavelength in equations.

### 2.1.4 Radiant flux

The radiant energy in a differential time interval  $[t, t + dt]$  can be expressed using the *radiant flux*  $\Phi$  :

$$dQ = \Phi(t) dt. \quad (\text{II.68})$$

The unit of radiant flux is Watt,  $W = \text{J/s}$ . In computer graphics, the explicit time dependency is usually omitted, since one is mainly interested in the energy at equi-

librium state. We will do the same in the following, and use radiant flux as the basic quantity. Radiant flux is also called *radiant power*.

### 2.1.5 Radiant intensity

A sensor that captures only photons in a differential pencil of solid angle  $d\omega$  around a given direction  $\omega$  will sense a flux of

$$d\Phi = I(\omega) d\omega. \quad (\text{II.69})$$

Here,  $I(\omega)$  is called *radiant intensity*, and it is measured in  $[W/sr]$ .

The same concept applies to emitters that have no surface area, such as point light sources, or infinitely distant environment light sources.

### 2.1.6 Irradiance and radiosity

A similar quantity arises if a sensor captures photons that are incident on a differential surface or volume element  $dx$ , but disregards direction:

$$d\Phi = E(x) dx. \quad (\text{II.70})$$

The quantity  $E(x)$  is called *irradiance*. It is measured in  $[W/m^2]$  on surfaces and  $[W/m^3]$  in volumes. If exitant photons are measured rather than incident photons, then the formula can be written in terms of *radiosity* or *radiant exitance*  $B(x)$ :

$$d\Phi = B(x) dx. \quad (\text{II.71})$$

### 2.1.7 Radiance

Measuring flux in a subset of positions and directions leads to the flux

$$d\Phi = L(x, \omega) \cos \theta dx d\omega \quad (\text{II.72})$$

if  $x$  is on a surface, and

$$d\Phi = L(x, \omega) dx d\omega \quad (\text{II.73})$$

if it is in a volume. The cosine term is measured towards the normal  $\mathbf{n}(x)$  of the surface element and accounts for the perceived decrease in area at grazing angles:

$$\cos \theta = |\langle \omega, \mathbf{n}(x) \rangle|. \quad (\text{II.74})$$

Note that we have chosen to write this cosine as an absolute value. This is because for the foreshortening term, it does not matter whether the surface element is hit by the incident ray from the front or the back.

It can be inconvenient to explicitly differentiate between the surface and the volume case all the time. This is one of the reasons the so-called *projected solid angle* measure

is handy. It is defined as

$$d\omega^\perp = \begin{cases} \cos \theta d\omega, & \text{on surfaces} \\ d\omega, & \text{in volumes} \end{cases}. \quad (\text{II.75})$$

The name stems from the fact that the cosine term can be interpreted as projecting the solid angle onto the surface plane (Cohen and Wallace, 1993).

$L(\mathbf{x}, \boldsymbol{\omega})$  is called *radiance* or *specific intensity*, and is measured in  $[W/sr/m^2]$  for surfaces and  $[W/sr/m^3]$  for volumes. Radiance is the quantity that is usually transported in image synthesis systems.

As a consequence of Equations (II.70) and (II.72) the irradiance incident on a differential surface or volume element  $d\mathbf{x}$  due to a beam with radiance  $L(\mathbf{x}, \boldsymbol{\omega})$  and solid angle  $d\omega$  is

$$dE(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\omega}) d\omega^\perp. \quad (\text{II.76})$$

## 2.2 Scattering at surfaces

Recall from Equation (II.72) that a differential beam of solid angle  $d\omega_i$  incident on a differential surface element  $\mathbf{x}$  with area  $d\mathbf{x}$  carries the incident flux

$$L(\mathbf{x}, \boldsymbol{\omega}_i) \cos \theta d\mathbf{x} d\omega_i. \quad (\text{II.77})$$

The reflected radiance of the surface element in direction  $\boldsymbol{\omega}_o$  due to this incident beam can differ by a constant factor, the *bidirectional reflectance-distribution function*  $f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$  (Nicodemus et al., 1977, BRDF), which accounts for directional scattering and absorption:

$$dL(\mathbf{x}, \boldsymbol{\omega}_o) = f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L(\mathbf{x}, \boldsymbol{\omega}_i) \cos \theta d\mathbf{x} d\omega_i. \quad (\text{II.78})$$

As shown in Equation (II.76), this can be written as

$$dL(\mathbf{x}, \boldsymbol{\omega}_o) = f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) dE(\mathbf{x}), \quad (\text{II.79})$$

where the differential irradiance is only due to the incident beam.

Writing Equation (II.79) in terms of physical units, and isolating the unit of the BRDF, we find that it is

$$\frac{W}{sr m^2} \frac{m^2}{W} = \frac{1}{sr}. \quad (\text{II.80})$$

For non-flourescent materials, the BRDF must be normalized over all exitant directions to guarantee energy conservation:

$$\int_{4\pi} f_r(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \cos \theta_o d\boldsymbol{\omega}_o \leq 1. \quad (\text{II.81})$$

It is important to note that if the wavelength is considered explicitly, this constraint

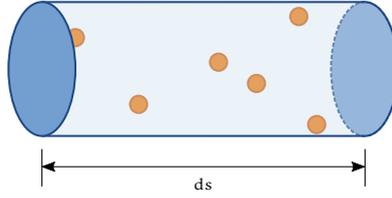


Figure II.1.: A differential volume element with cross-section  $d\sigma$  and length  $ds$ . The volume element contains  $N$  particles, which may scatter or absorb radiant energy.

must hold for each individual wavelength.

Additionally, one generally requires reciprocity (Nicodemus, 1965) so that scattering is symmetric:

$$f_r(\omega_i, \omega_o) \stackrel{!}{=} f_r(\omega_o, \omega_i). \quad (\text{II.82})$$

The BRDF can be textured across a surface, in which case it depends on the position as well as the incident and exitant directions:  $f_r(\mathbf{x}, \omega_i, \omega_o)$ .

Additionally, the concept can be used for transmissive surfaces, in which case the function is called *bidirectional transmittance-distribution function* (BTDF). To model materials which exhibit scattering below the surface, such as skin, one can use a *bidirectional scattering-surface reflectance distribution function* (Nicodemus et al., 1977, BSSRDF).

Sometimes, the acronym BxDF is used as an umbrella term for all such distribution functions. We will write the BxDF as  $f(\omega_i, \omega_o)$ .

## 2.3 Participating media

In radiative transfer, participating media are materials that interact with beams of light by scattering, absorbing or emitting energy.

These concepts are usually introduced by examining a differential, cylindrical volume element of cross-section  $d\sigma$  and length  $ds$  (Chandrasekhar, 1960). Such a volume element is shown in Figure II.1.

### 2.3.1 Outscattering and absorption

Without loss of generality, we will assume that the volume element contains  $N$  discrete particles. Each of the particles may either scatter or absorb photons. The visible cross section of all scattering particles is  $N\sigma_s$ , and the visible cross section of all absorbing particles is  $N\sigma_a$ . All particles are far smaller than the volume element itself:  $N\sigma_s \ll d\sigma$  and  $N\sigma_a \ll d\sigma$ .

In this situation, consider a beam of photons with radiance  $L$  incident on the volume element such that its direction is aligned with the cylinder.

The probability of a photon in the beam hitting a scattering particle is  $N \sigma_s / d\sigma$ . Thus, the change in radiance in the beam due to outscattering is

$$dL = -\frac{N \sigma_s}{d\sigma}. \quad (\text{II.83})$$

Using the *density*  $\rho = N / (d\sigma ds)$ , this can be written as

$$dL = -\rho \sigma_s L ds \iff \frac{dL}{ds} = -\rho \sigma_s L. \quad (\text{II.84})$$

The quantity  $\rho \sigma_s = \mu_s$  is called the *scattering coefficient* and has units of  $1/m$ . Similarly, the change in radiance due to absorption is

$$\frac{dL}{ds} = -\rho \sigma_a L = -\mu_a L. \quad (\text{II.85})$$

with the *absorption coefficient*  $\mu_a$ .

Generally, all of the cross-sections and coefficients depend on the position  $\mathbf{x}$  and the wavelength  $\lambda$ , but we will omit this explicit dependency where it is not required.

### 2.3.2 Emission

Particles in a participating medium such as fire may emit energy. There are multiple ways to model such materials, but we will model them as heated black bodies, similar to soot in reality. As a consequence, they not only emit radiance, but absorb all incident energy, as well. Thus, with the *emission cross-section*  $\sigma_e$  and the *emission coefficient*  $\mu_e$ , the change in radiance due to emissive particles with radiance  $L^e$  is

$$\frac{dL}{ds} = \rho \sigma_e L^e - \rho \sigma_e L = \mu_e L^e - \mu_e L. \quad (\text{II.86})$$

### 2.3.3 Extinction and transmittance

Since absorption, outscattering and emission cause the same term  $-\mu L$ , the respective coefficients are often combined into the *extinction cross-section*  $\sigma_t = \sigma_s + \sigma_a + \sigma_e$  and the *extinction coefficient*  $\mu_t = \mu_s + \mu_a + \mu_e$ . The extinction cross-section is also called *total cross-section* in literature (Chandrasekhar, 1960).

The extinction due to absorptive, scattering and emissive particles along a ray with parameter  $s$  is

$$\frac{dL}{ds}(s) = -\mu_s(s)L(s) - \mu_a(s)L(s) - \mu_e(s)L(s) = -\mu_t(s)L(s). \quad (\text{II.87})$$

The solution of this differential equation is

$$L(s) = \exp\left(-\int_0^s \mu_t(t) dt\right) L(0). \quad (\text{II.88})$$

This can be verified using the fundamental theorem of calculus, which states that

$$\frac{d}{ds} \int_0^s \mu_t(t) dt = \mu_t(s). \quad (\text{II.89})$$

Thus, using the chain rule,

$$\begin{aligned} \frac{dL}{ds}(s) &= \frac{d}{ds} \left[ \exp\left(-\int_0^s \mu_t(t) dt\right) L(0) \right] \\ &= \frac{d}{ds} \left[ -\int_0^s \mu_t(t) dt \right] \exp\left(-\int_0^s \mu_t(t) dt\right) L(0) \\ &= -\mu_t(s) L(s). \end{aligned} \quad (\text{II.90})$$

The function

$$\tau(s) = \exp\left(-\int_0^s \mu_t(t) dt\right) \quad (\text{II.91})$$

is called *transmittance*, because it measures the amount of energy that is transmitted through a slab of material without being scattered or absorbed. Transmittance is a dimensionless quantity.

Transmittance is always less than one. This is easy to see once we realize that physical cross-sections and densities must be positive, and so the integral in the exponent must be positive, as well:

$$\int_0^s \rho(t) \sigma_s(t) dt \geq 0. \quad (\text{II.92})$$

It follows that

$$\begin{aligned} -\int_0^s \rho(t) \sigma_s(t) dt &\leq 0 \\ \iff \exp\left(-\int_0^s \rho(t) \sigma_s(t) dt\right) &\leq 1. \end{aligned} \quad (\text{II.93})$$

This property makes sense if we examine Equation (II.88) again: transmittance greater than one would increase the radiance in the slab of material, but the equation is concerned with extinction effects only. In a homogeneous medium, where  $\mu_t$  is constant, the transmittance reduces to the simple exponential

$$\tau(s) = \exp(-\mu_t s). \quad (\text{II.94})$$

### 2.3.4 Free-flight distribution

In an infinite medium, we may assume that a photon will interact with some scattering or absorbing particle at some distance  $s$ . By design, the transmittance function is the probability

$$\tau(s_0) = P[s > s_0] \quad (\text{II.95})$$

for some distance  $s_0$ . Conversely, the probability for an interaction in  $[0, s_0]$  is

$$1 - \tau(s_0) = P[s \leq s_0]. \quad (\text{II.96})$$

This means that  $1 - \tau(s_0)$  is really a cumulative distribution function, for which we should be able to find a probability density function using differentiation:

$$p(s) = \frac{d}{ds} [1 - \tau](s). \quad (\text{II.97})$$

This is also known as the *free-flight distribution*, since it is the PDF for the distance to the next scattering or absorbing event. This distribution function describes the geometric constellation in a participating medium in the sense that it gives stochastic guarantees about distances between particles.

The differentiation can be carried out using the same tools that we used in Equation (II.90):

$$\frac{d}{ds} [1 - \tau](s) = -\frac{d}{ds} \tau(s) = \mu_t(s) \exp\left(-\int_0^s \mu_t(t) dt\right). \quad (\text{II.98})$$

In the homogeneous case, where  $\mu_t$  is a constant, this simplifies to an exponential probability density function (Lehn and Wegman, 2006):

$$p(s) = \mu_t \exp(-\mu_t \cdot s). \quad (\text{II.99})$$

### 2.3.5 Mean free path

The expected value of the free-flight distribution is

$$\begin{aligned} E[p(s)] &= \int_0^\infty s \cdot \left(-\frac{d}{ds} \tau(s)\right) ds \\ &\stackrel{\text{(i.b.p.)}}{=} -[s \cdot \tau(s)]_0^\infty + \int_0^\infty \tau(s) ds \\ &= -\lim_{s \rightarrow \infty} [s \cdot \tau(s)] + \int_0^\infty \tau(s) ds \\ &\stackrel{\text{(de L'Hôpital)}}{=} \int_0^\infty \tau(s) ds. \end{aligned} \quad (\text{II.100})$$

In a homogeneous medium, this integral can be evaluated analytically:

$$E [p(s)] = \int_0^{\infty} \exp(-\mu_t s) ds = \frac{1}{\mu_t}. \quad (\text{II.101})$$

This means that the reciprocal of the extinction coefficient is the distance that a photon will travel on average before interacting with a scattering or absorbing particle. For this reason, it is also called the *mean free path*, and we will also write it as  $\bar{l}$ . In general, like the volume coefficients, the mean free path may depend on the wavelength.

### 2.3.6 Inscattering

Photons that are scattered into the beam will cause an increase in radiance. In radiative transfer, one generally assumes a stochastic model where a probability distribution  $\phi(\omega_o | \omega_i)$  determines the fraction of photons that are scattered from direction  $\omega_i$  to  $\omega_o$ , if they hit a scattering particle.

Combined with the scattering coefficient that determines if a scattering event takes place in the first place, we obtain the increase in radiance due to inscattering

$$\frac{d}{ds} L(s, \omega_o) = \int_{4\pi} \mu_s(s) \phi(\omega_i, \omega_o) L(s, \omega_i) d\omega_i. \quad (\text{II.102})$$

### 2.3.7 Phase function

The PDF  $\phi(\omega_o | \omega_i)$  is called *phase function*, and it is a probability density function with respect to the outgoing direction  $\omega_o$ :

$$\int_{4\pi} \phi(\omega_o | \omega_i) d\omega_o = 1 \quad \forall \omega_i. \quad (\text{II.103})$$

By this definition, the phase function is only used to describe scattering, but does not include absorption. It is important to note that  $\omega_i$  is the direction of the photon before scattering. The phase function is measured in units of  $1/sr$ .

### 2.3.8 Radiative transfer equation

The *radiative transfer equation* (Chandrasekhar, 1960, RTE) is a combination of the above equations. It describes local changes in radiance due to emission, inscattering, outscattering and absorption. It reads

$$\begin{aligned} \frac{dL}{ds}(s, \omega_o) = & -\mu_t(s)L(s, \omega_o) \\ & + \mu_e(s)L^e(s, \omega_o) \\ & + \mu_s(s) \int_{4\pi} \phi(\omega_o | \omega_i) L(s, \omega_i) d\omega_i. \end{aligned} \quad (\text{II.104})$$

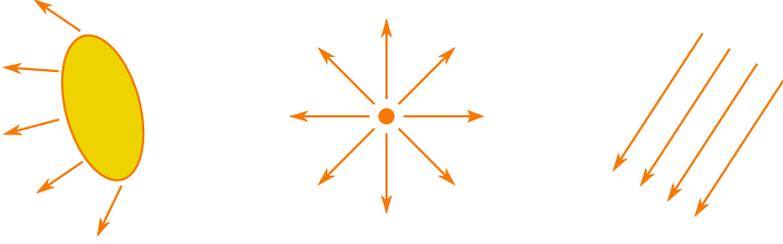


Figure II.2.: Illustration of area, point and distant light sources.

The part of the above equation that is responsible for an increase in radiance is sometimes called *source term*, and it is written as

$$Q(s, \omega_o) = \mu_e(s)L^e(s, \omega_o) + \mu_s(s) \int_{4\pi} \Phi(\omega_o | \omega_i)L(s, \omega_i) d\omega_i. \quad (\text{II.105})$$

With this short-hand, the RTE can be written as

$$\frac{dL}{ds}(s, \omega_o) = -\mu_t(s)L(s, \omega_o) + Q(s, \omega_o). \quad (\text{II.106})$$

## 2.4 Emitters

An emitter is a source of photons. In general, it is defined by its emitted radiance  $L^e(\mathbf{x}, \omega)$ . We will describe area light sources, point lights, distant lights, and environment lights. Emissive volumes were already detailed in Section II.2.3.2. The light sources introduced here are illustrated in Figure II.2.

### 2.4.1 Area lights

Area light sources are the most common type of light source. Such light sources have a surface area, and in general the emitted radiance may be a function of position and exitant direction:  $L^e(\mathbf{x}, \omega)$ . We will assume that area light sources only emit into the positive hemisphere. Also, while it is possible to texture emission across a surface, and to define directional emission profiles, area light sources are often modeled as perfect lambertian emitters where  $L^e(\mathbf{x}, \omega) = L^e$  is constant.

### 2.4.2 Point lights

It is sometimes convenient to approximate small lights as point light sources. Such light sources have no area, and so their emission is given as radiant intensity  $I^e(\omega)$ . The intensity can be constant, or it can vary with direction. A common model is called

*spot light*, where intensity is a cosine cone:

$$I^e(\boldsymbol{\omega}) \propto \langle \boldsymbol{\omega}, \boldsymbol{\omega}_0 \rangle^n. \quad (\text{II.107})$$

### 2.4.3 Distant lights

Distant lights are at infinite distance, but they model light sources that are far away from the scene. Such emitters are parallel, which means that all emitted photons have the same direction. Distant lights have no associated position, but illuminate all points in the scene from a given direction  $\boldsymbol{\omega}_0$ . Their emission is given as intensity, which for standard distant lights is a delta distribution:

$$I^e(\boldsymbol{\omega}) \propto \delta(\boldsymbol{\omega} - \boldsymbol{\omega}_0). \quad (\text{II.108})$$

A special case of distant are *environment lights*, which can be seen as a collection of infinitely many distant light sources, one for each possible direction. Their emission is given in intensity, and it generally is stored in textures that are called *environment maps*. The process of lighting scenes with environment maps is also called Image-based Lighting (Fernando, 2004).

Image-based lighting can be used to illuminate virtual scenes with photographed real environments, which helps recreate realistic lighting situations.

## 2.5 Irradiance due to emission or scattering

We will now construct basic building blocks that will allow us to build full light paths later on. Both the BxDF and the phase function convert irradiance to exitant radiance, and so we are interested in finding the irradiance at differential surface and volume elements at some point  $\mathbf{x}_0$  due to emission or reflection at another point  $\mathbf{x}_1$ . In the following, we will use the distance  $r = \|\mathbf{x}_0 - \mathbf{x}_1\|$  and the light beam direction

$$\boldsymbol{\omega} = \frac{\mathbf{x}_0 - \mathbf{x}_1}{r}. \quad (\text{II.109})$$

The basic quantity we can work with is the differential radiant flux in the incident beam of light,  $d\Phi$ . We can determine it from exitant radiance or radiant intensity at  $\mathbf{x}_1$  depending on whether the source element is a surface element, volume element, point light, or distant light:

$$d\Phi = \begin{cases} L(\mathbf{x}_1, \boldsymbol{\omega}) \cos \theta_1 d\mathbf{x}_1 d\boldsymbol{\omega}, & \text{Surface source} \\ L(\mathbf{x}_1, \boldsymbol{\omega}) \cos \theta_1 d\mathbf{x}_1 d\boldsymbol{\omega}, & \text{Volume source} \\ I(\boldsymbol{\omega}) d\boldsymbol{\omega}, & \text{Point / distant source.} \end{cases} \quad (\text{II.110})$$

Here,  $\cos \theta_1$  is the cosine of the angle between the surface normal at  $\mathbf{x}_1$  and  $\boldsymbol{\omega}$ .

All of the above equations contain the differential solid angle  $d\boldsymbol{\omega}$ . It depends on

the type of the element at  $\mathbf{x}_0$ , and on the distance to the light source:

$$d\omega = \begin{cases} \frac{\cos \theta_0}{r^2} d\mathbf{x}_0, & r \text{ finite} \rightarrow \text{surface} \\ \frac{\mu_s(\mathbf{x}_0)}{r^2} d\mathbf{x}_0, & r \text{ finite} \rightarrow \text{volume} \\ \frac{\cos \theta_0}{1\text{m}^2} d\mathbf{x}_0, & r \text{ infinite} \rightarrow \text{surface.} \end{cases} \quad (\text{II.111})$$

The term  $\cos \theta_0$  is the cosine of the angle between the surface normal at  $\mathbf{x}_0$  and  $-\boldsymbol{\omega}$ , and it projects  $d\mathbf{x}_0$  to a plane perpendicular to  $\boldsymbol{\omega}$ . The scattering coefficient performs a very similar function for target volume elements.

Two things are important with respect to distant lights. First, distant lights are parallel emitters, and so no division by  $r$  is required to determine the solid angle. However, the units still require an explicit division by  $1\text{m}^2$ . Second, we assume that all participating media have finite boundaries, and so we need not include the case of distant lights illuminating volume elements directly in our discussion.

Finally, if the beam passes through a participating medium, the flux is attenuated by transmittance  $\tau(\mathbf{x}_0, \mathbf{x}_1)$ .

The differential irradiance at  $\mathbf{x}_0$  is related to the flux as

$$d\Phi = dE(\mathbf{x}_0) d\mathbf{x}_0, \quad (\text{II.112})$$

We may now substitute the solid angle terms (Equation II.111) into Equation (II.110) to obtain the flux in the differential tube from  $\mathbf{x}_0$  to  $\mathbf{x}_1$ . We find that all resulting terms contain the factor  $d\mathbf{x}_0$ , and so we may equate with Equation (II.112) to obtain general rules for the differential irradiance at  $\mathbf{x}_0$  due to incident illumination from direction  $-\boldsymbol{\omega}$ . The rules all have the form

$$dE(\mathbf{x}_0) = s(\mathbf{x}_0) \cdot \tau(\mathbf{x}_0, \mathbf{x}_1) \cdot \frac{\mathbf{g}(\mathbf{x}_0) \cdot \mathbf{g}(\mathbf{x}_1)}{d(\mathbf{x}_0, \mathbf{x}_1)} \cdot \mathcal{E}(\mathbf{x}_1, \boldsymbol{\omega}). \quad (\text{II.113})$$

In this formula,

- $s(\mathbf{x}_0) = \mu_s(\mathbf{x}_0)$  if  $d\mathbf{x}_0$  is a volume element and 1 otherwise,
- $\mathbf{g}(\mathbf{x}_0) = \cos \theta_0$  if  $d\mathbf{x}_0$  is a surface element and 1 otherwise,
- $\mathbf{g}(\mathbf{x}_1) = \cos \theta_1$  if  $d\mathbf{x}_1$  is a surface element and 1 otherwise,
- $d(\mathbf{x}_0, \mathbf{x}_1) = r^2$  if  $r$  is finite and  $1\text{m}^2$  otherwise,
- $\mathcal{E}(\mathbf{x}_1, \boldsymbol{\omega}) = L(\mathbf{x}_1, \boldsymbol{\omega})$  if  $d\mathbf{x}_1$  is a surface element or volume element, and  $\mathcal{E}(\mathbf{x}_1, \boldsymbol{\omega}) = I(\boldsymbol{\omega})$  if the energy originates from a point light or distant light.

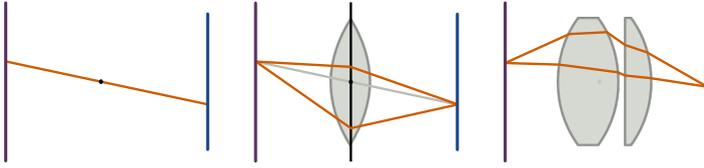


Figure II.3.: Various lens models in use in photorealistic image synthesis: pinhole camera (left), the thin lens approximation (center), and a fully raytraced lens system (right).

### 2.5.1 Geometric term

The factor

$$G(\mathbf{x}_0, \mathbf{x}_1) = \frac{g(\mathbf{x}_0) \cdot g(\mathbf{x}_1)}{d(\mathbf{x}_0, \mathbf{x}_1)} \quad (\text{II.114})$$

in Equation (II.113) is also called *geometric term* because it describes the geometric configuration of the segment  $\mathbf{x}_0\text{--}\mathbf{x}_1$ . It has units of  $[\text{s r} / \text{m}^2]$

The geometric term has a singularity at  $r = 0$ , where it diverges to infinity. Even for very small  $r$ , the geometric term can reach excessively large values. This can cause numerical problems in a rendering program, and so some rendering algorithms try to avoid having to explicitly consider geometric terms by importance sampling them (see Section II.1.10).

## 2.6 Sensors

Sensors detect photons in the scene. They are a concept that is dual to emitters, which emit photons.

Consequently, one can define a corresponding sensor type for each of the emitter types introduced earlier. However, we will restrict sensors to small, planar surface patches, as they can be found in cameras.

A sensor has an associated sensitivity function  $W^e(\mathbf{x}, \boldsymbol{\omega})$ , which determines how sensitive the sensor is to photons incident at  $\mathbf{x}$  with direction  $-\boldsymbol{\omega}$ . Note that we define  $\boldsymbol{\omega}$  to be the direction the photon originates from.

Since our sensors are basically surfaces, the rules for irradiance on surface elements above apply to them, as well. Specifically, the cosine foreshortening term must be considered. Some rendering systems, knowingly or by accident, include this cosine term in the sensitivity function, but we will instead keep it with the geometric term.

### 2.6.1 Optical systems

A simple camera model, which is used frequently, is the pinhole camera model (Barsky et al., 2003). This model corresponds to the *camera obscura*, in which no lens and a very small aperture are used.

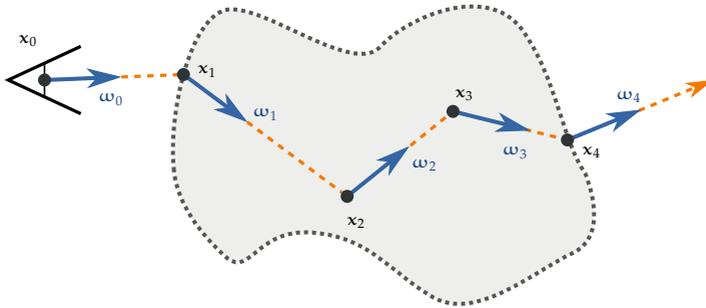


Figure II.4: A light path with six vertices connecting a sensor with an environment light source, so that the last vertex  $x_5$  can be interpreted as being at infinity.

Due to their small aperture, real-world pinhole cameras require bright lighting conditions or long exposures to collect sufficient light. They also render the whole scene in focus, although small apertures lead to diffraction effects that can blur the image (Potmesil and Chakravarty, 1981).

Despite these shortcomings, pinhole cameras are often used in rendering systems, because they are simple to implement and require little processing power.

To model depth of field, or to decrease exposure times, one may employ larger apertures. Those generally require lenses to focus. A simple lens model that is used in computer graphics is the *thin lens approximation* (Barsky et al., 2003). In this approximation, the lens is assumed to be infinitely thin.

To match the look of real cameras, including effects such as chromatic aberrations and distortion, it is possible to model lens systems as a series of finite-size refractive elements with apertures (Hanika and Dachsbacher, 2014; Schrade et al., 2016). Such models yield very realistic results and can be solved efficiently using precomputed polynomials, but they reduce the optical efficiency of the camera system.

## 2.7 Light paths and measurement

A light path is an  $(N + 1)$ -tuple of vertices  $(x_0, \dots, x_N)$ . We will demonstrate some concepts on a concrete light path, which we show in Figure II.4: It starts at a sensor, undergoes a surface interaction, is scattered in a volume twice, again hits a surface, and finally exits towards the environment.

Using the segment irradiances described above, we can now compute the contri-

bution of this path to the spectral sensor measurement:

$$\begin{aligned}
 f_c(\mathbf{x}_0, \dots, \mathbf{x}_5) = & W^e(\mathbf{x}_0, \boldsymbol{\omega}_0) \\
 & \cdot G_{0,1} \cdot f(-\boldsymbol{\omega}_0, \boldsymbol{\omega}_1) \\
 & \cdot G_{1,2} \cdot \tau(\mathbf{x}_1, \mathbf{x}_2) \cdot \Phi(-\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \\
 & \cdot G_{2,3} \cdot \mu_s(\mathbf{x}_2) \cdot \tau(\mathbf{x}_2, \mathbf{x}_3) \cdot \Phi(-\boldsymbol{\omega}_2, \boldsymbol{\omega}_3) \\
 & \cdot G_{3,4} \cdot \mu_s(\mathbf{x}_3) \cdot \tau(\mathbf{x}_3, \mathbf{x}_4) \cdot f(-\boldsymbol{\omega}_3, \boldsymbol{\omega}_4) \\
 & \cdot G_{4,5} \cdot I^e(-\boldsymbol{\omega}_5) \\
 & d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{x}_4.
 \end{aligned} \tag{II.115}$$

For different path configurations, similar terms arise from the segment irradiances.

The function  $f_c(\mathbf{X})$  is also called the *measurement contribution function* for the path  $\mathbf{X}$ . Given the set  $\mathcal{X}(\mathbf{x}_0)$  of all possible light paths that end in  $\mathbf{x}_0$ , we can now write the sensor measurement at  $\mathbf{x}_0$  as

$$\mathcal{J}(\mathbf{x}_0, \lambda) = \int_{\mathcal{X}(\mathbf{x}_0)} f_c(\mathbf{X}, \lambda) d\mathbf{X}. \tag{II.116}$$

Here, we have included the wavelength  $\lambda$  as an explicit parameter to express that the measurement is a spectral energy density.

The above integral is also called the *path integral* in literature, and the set  $\mathcal{X}(\mathbf{x}_0)$  is also referred to as *path space*. While these concepts are usually associated with (Veach, 1997) in computer graphics, they were used much earlier in the context of neutron transport theory (Spanier and Gelbard, 1969, Section 3.3).

### 2.7.1 Adjoint quantities

The radiometric quantities describe the transport of radiative energy, but often is convenient to analyze the adjoint problem, where sensors are considered emitters that radiate *importance*  $W$  into the scene. The emission of importance is governed by the sensitivity function  $W^e(\mathbf{x}, \boldsymbol{\omega})$  (see Section II.2.6), and the emission of light sources corresponds to their sensitivity for importance. The importance field  $W(\mathbf{x}, \boldsymbol{\omega})$  can be understood as a measure of how much of the radiative energy leaving a position  $\mathbf{x}$  in direction  $\boldsymbol{\omega}$  will at some point be detected by a sensor (Peter and Pietrek, 1998).

Since importance is the adjoint quantity to radiance, the equations governing the transport of radiant energy can be applied directly to the adjoint problem (Spanier and Gelbard, 1969), although special care needs to be taken in the evaluation of scattering at dielectric boundaries (Veach, 1997).

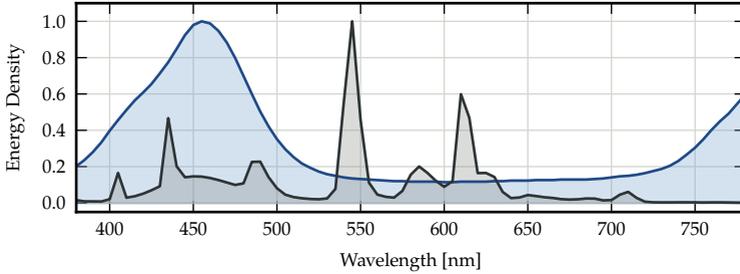


Figure II.5.: The attenuation spectrum of a blue diffuse surface (McCamy et al., 1976; Ohta and Robertson, 2006, Patch 13 Blue) and the emission spectrum of a fluorescent light source (CIE, 2004, Illuminant FL10).

The adjoint radiative transfer equation can be written as

$$\begin{aligned}
 \frac{dW}{ds}(s, \boldsymbol{\omega}_o) &= -\mu_t(s)W(s, \boldsymbol{\omega}_o) \\
 &+ \mu_e(s)W^e(s, \boldsymbol{\omega}_o) \\
 &+ \mu_s(s) \int_{4\pi} \phi(\boldsymbol{\omega}_o | \boldsymbol{\omega}_i) W(s, \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i.
 \end{aligned}
 \tag{II.117}$$

It is interesting to note that the measurement contribution function contains both the sensitivity  $W^e$  and the emitted radiance  $L^e$ . In fact, given the path  $(\mathbf{x}_0, \dots, \mathbf{x}_k)$  and omitting explicit notation of the directional variable, the measurement can be expressed as

$$f_c(\mathbf{x}_0, \dots, \mathbf{x}_k) = W^e(\mathbf{x}_0)L(\mathbf{x}_0) = W(\mathbf{x}_k)L^e(\mathbf{x}_k).
 \tag{II.118}$$

## 3 Color

### 3.1 Spectral energy density functions

A function  $S(\lambda)$  that associates an energy density with individual wavelengths is called a *spectral energy density*, or simply *spectrum*.

Generally, spectral energy densities describe the distribution of energy over multiple wavelengths, but there are two important special cases. *Emission spectra* model the distribution of photons emitted by a light source. They can, in principle, carry arbitrary amounts of energy, and may also exhibit strong peaks.

*Attenuation spectra* are used to model absorption, for example on surfaces in the scene. Ignoring fluorescence, they must be less than 1 for all wavelengths to guar-

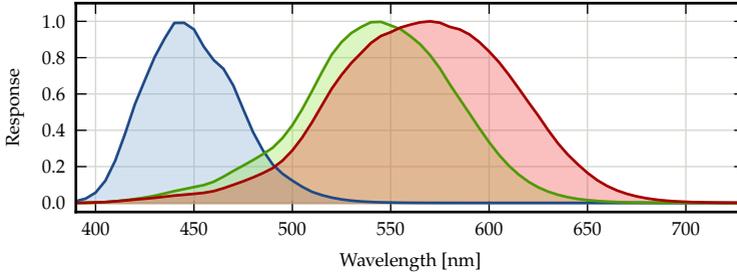


Figure II.6.: The response functions for human cone cells  $s(\lambda)$  (blue),  $m(\lambda)$  (green) and  $l(\lambda)$  (red) cones, as given by (Stockman and Sharpe, 2000).

antee energy conservation. Additionally, analysis of real materials has shown that attenuation spectra tend to be smooth functions, and suprisingly, a band limit can be given (Maloney, 1986). The process of attenuation is mathematically expressed by multiplication of a spectrum with an attenuation spectrum.

Figure II.5 shows the spiky emission spectrum of a standardized fluorescent light source, as well as the much smoother attenuation spectrum of a diffuse, blue surface.

*Color saturation* is a way of describing the width of a spectrum. Unsaturated, gray colors have spectra that are close to constant over the range of visible wavelengths, while very saturated spectra are composed of narrow peaks. This means that a saturated attenuation spectrum can never be very bright, because it must conserve energy (Schrödinger, 1920).

### 3.2 Perception of color

In the human eye, there are usually three distinct cell types that are used for color perception. They are called *cone cells*, and there usually exists one cone cell type for short, medium, and long wavelengths. As a consequence, color perception in humans is generally three-dimensional, although evidence exists that some individuals experience color in four dimensions (Jordan et al., 2010).

A wavelength-dependent response function can be associated with each cone cell type. These functions are usually written as  $\bar{s}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{l}(\lambda)$ , and can be used to determine the response of a cone cell to an incident light spectrum  $S(\lambda)$  as

$$s = \int_{\lambda} S(\lambda) \cdot \bar{s}(\lambda) d\lambda \quad (\text{II.119})$$

$$m = \int_{\lambda} S(\lambda) \cdot \bar{m}(\lambda) d\lambda \quad (\text{II.120})$$

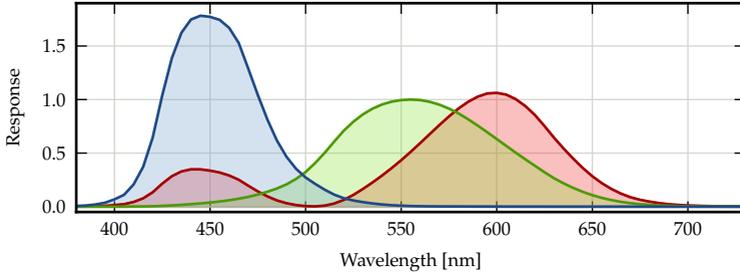


Figure II.7.: The color matching functions  $\bar{z}(\lambda)$  (blue),  $\bar{y}(\lambda)$  (green) and  $\bar{x}(\lambda)$  (red), which define the *CIE XYZ* color space (CIE, 2004).

$$l = \int_{\Lambda} S(\lambda) \cdot \bar{l}(\lambda) d\lambda. \quad (\text{II.121})$$

In Figure II.6, we show the response functions as specified by (Stockman and Sharpe, 2000).

### 3.3 Color spaces

A *color space* is a way of organizing colors, which is described mathematically by an associated *color model*.

Justified by the three-dimensional nature of human color perception, many color spaces are based on three *primary colors* that can roughly be categorized as Red, Green, and Blue. They use linear combinations of these primaries to obtain individual color values. The coefficients of the linear combination are then called *tristimulus values*.

The set of all colors that is representable in a given color space is called the *gamut*, and it depends on the primary colors. An important color space that contains all visible colors is *CIE XYZ* (CIE, 2004; Fairman et al., 1998). The primary colors of *XYZ* are imaginary and cannot be produced by physical lamps (Wyszecki and Stiles, 1982). However, *XYZ* can be used to give tristimulus values for the primary colors of other color spaces, and so provides a way of comparing and converting between them. Light spectra can be converted to *XYZ* using the response functions  $\bar{x}$ ,  $\bar{y}$ , and  $\bar{z}$  (CIE, 2004):

$$X = \int_{\Lambda} S(\lambda) \cdot \bar{x}(\lambda) d\lambda \quad (\text{II.122})$$

$$Y = \int_{\Lambda} S(\lambda) \cdot \bar{y}(\lambda) d\lambda \quad (\text{II.123})$$

$$Z = \int_{\Lambda} S(\lambda) \cdot \bar{z}(\lambda) d\lambda. \quad (\text{II.124})$$

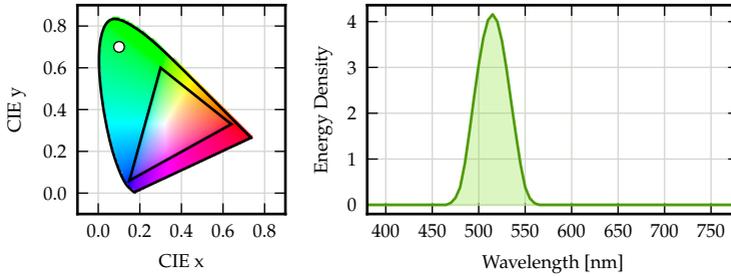


Figure II.8.: A chromaticity diagram showing the triangular gamut of the *sRGB* color space, and the location of the spectrum shown on the right. The horseshoe-shaped part of the chromaticity diagram is the set of all colors visible to the human eye.

These response functions, which we reproduce in Figure II.7, are also called *color matching functions*. The reverse process, determining a spectrum from tristimulus values, is called *spectral upsampling*. We will cover this problem in more detail in Chap. III.

Important additional color spaces include *sRGB* and *Rec.709* (ITU, 2002), which are often used for display on computer screens and similar devices. *Adobe RGB* (Adobe, 2005) has a slightly larger gamut and is useful for applications in photography. *CIE RGB* is a color space with monochromatic primaries (CIE, 2004), which results in an extremely wide gamut. The *Academy Color Encoding Specification (ACES)*, like *XYZ*, includes all visible colors, and its primaries are imaginary. *ACES* is primarily used in the movie industry.

Related to *XYZ*, the color space *xyY* is a way of decoupling color information from brightness, and it is defined as

$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}. \quad (\text{II.125})$$

Here, the color  $xy$  is called the *chromaticity*, and  $Y$  is the *luminance* of a color. The chromaticity determines the saturation and hue of a color. A chromaticity diagram, which plots only  $xy$ , is often used to visualize color spaces. A color space with three primary colors has a triangular gamut in  $xy$ . This can be seen in Figure II.8 for the gamut of *sRGB*.

The outer boundary of the horseshoe, or *spectral locus*, contains all monochromatic colors. Their spectra are mathematically delta distributions, and carry energy only on one wavelength. The line at the bottom contains mixtures between blue and red; it is called the *purple line*. The *white point* is in the center of the diagram, but its location differs between color spaces. For RGB color spaces, it is the location

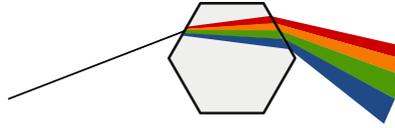


Figure II.9.: Dispersion at the refractive boundaries of an object with wavelength-dependent index of refraction.

associated with  $\text{RGB} = (1, 1, 1)$ . An important white point is *equal-energy white*, which is defined by a constant spectrum  $S(\lambda) = 1$ . The XYZ color space is defined such that  $(x, y) = (1/3, 1/3)$  corresponds to equal-energy white. The standardized emitter *Illuminant E* (CIE, 2004) has an equal-energy spectrum. Other important white points are derived from their color temperatures. For example, the white point *D65* corresponds to the emission spectrum of a black body heated to 6500K (Demtröder, 2015).

## 4 Monte Carlo-simulation of light transport

Photorealistic image synthesis can be understood the process of solving the path integral (II.116) for all points on the sensor. One way to achieve this is using Monte Carlo integration. In this section, we will review the idea of using random walks to generate samples in path space, which can then be used to evaluate the measurement contribution and finally averaged in a Monte Carlo estimator.

### 4.1 Spectral rendering

The output of a photorealistic rendering algorithm is usually a raster image comprised of pixels that carry tristimulus color values. To obtain these tristimulus values, there are two basic approaches.

One very common technique, which is also called *tristimulus rendering*, uses tristimulus color values for all wavelength-dependent quantities throughout the rendering process. This has the advantage that input values such as surface colors need not be given as spectral density functions. However, we have already seen that the simulation of light transport requires nonlinear operations, such as exponentiation for absorption in participating media, and so the output of a tristimulus renderer cannot produce exact results in general.

In contrast to tristimulus rendering, a rendering algorithm may represent wavelength-dependent quantities as spectral energy densities. This approach is called *spectral rendering*.

Spectral rendering generally produces more accurate colors than tristimulus rendering, and enables proper handling of effects such as dispersion at refractive boundaries (see Figure II.9). However, it requires integrating the measurement  $J(\mathbf{x}, \lambda)$  with appropriate response functions before tristimulus values can be accumulated into the image buffer (see Section II.3):

$$J(\mathbf{x}) = \begin{pmatrix} R(\mathbf{x}) \\ G(\mathbf{x}) \\ B(\mathbf{x}) \end{pmatrix} = \int_{\Lambda} J(\mathbf{x}, \lambda) \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} d\lambda. \quad (\text{II.126})$$

## 4.2 Monte Carlo-estimator for physically-based rendering

In the case of spectral rendering, one may write the Monte Carlo estimator as

$$f_c(\mathbf{x}_0) \approx \frac{1}{N} \sum_{i=1}^N \frac{f_c(\mathbf{X}_i, \lambda_i)}{p(\mathbf{X}_i|\lambda_i) \cdot p(\lambda_i)} \begin{pmatrix} \bar{r}(\lambda_i) \\ \bar{g}(\lambda_i) \\ \bar{b}(\lambda_i) \end{pmatrix}, \quad (\text{II.127})$$

where  $p(\mathbf{X}_i|\lambda_i) \cdot p(\lambda_i) = p(\mathbf{X}_i, \lambda_i)$  is the sampling probability density of the pair  $(\mathbf{X}_i, \lambda_i)$ . The basic spectral Monte Carlo rendering algorithm works as follows:

```

1:  $J \leftarrow (0, 0, 0)$ 
2: for  $i \leftarrow 1 \dots N$  do
3:    $\lambda_i \leftarrow \text{SAMPLEWAVELENGTH}()$ 
4:    $\mathbf{X}_i \leftarrow \text{SAMPLEPATH}(\lambda_i)$ 
5:    $\mathbf{c} \leftarrow f_c(\mathbf{X}_i, \lambda_i) / p(\mathbf{X}_i|\lambda_i)$ 
6:    $J \leftarrow J + (\bar{r}(\lambda_i), \bar{g}(\lambda_i), \bar{b}(\lambda_i)) \cdot \mathbf{c} / p(\lambda_i)$ 
7: end for
8:  $J \leftarrow J / N$ 

```

It is important to note here that this being a Monte Carlo estimator, the measurement contribution function is only evaluated for one wavelength. This is the reason why this type of spectral rendering is also called *single-wavelength transport*. It has the advantage that the sampling probability density for the path construction can depend explicitly on the wavelength, which may be used to reduce variance in the estimator.

## 4.3 Random walks

The function `SAMPLEPATH()` in the pseudocode above must generate a path history  $(\mathbf{x}_0, \dots, \mathbf{x}_n)$  for which the measurement contribution can be evaluated. This is usually done using *random walks* (Spanier and Gelbard, 1969), where the history is simply the trajectory of a particle that traverses the scene. Techniques based on

Markov-Chain Monte Carlo (Kelemen et al., 2002; Veach and L. J. Guibas, 1997) that instead mutate existing paths are not within the scope of this thesis.

The probability density functions sampled to obtain a random walk, and their exact physical meaning, depend on the specific algorithm. Particularly, it is of great importance whether the random walk starts at a sensor or a light source. However, random walks in photorealistic image synthesis tend to follow these steps:

- A trajectory is created by spawning a particle. For our purposes, the particle can be understood as either an abstract *photon* carrying radiant energy or an *importon* carrying importance (see Section II.2.7.1). We will write the emission probability density function as  $p_e(\mathbf{x}, \boldsymbol{\omega})$ . In general, this and the following PDFs involved in the random walk may depend on the wavelength  $\lambda$ .
- The particle then changes position according to  $p_t(\mathbf{x}'|\mathbf{x})$ . Usually, raytracing operations are used in photorealistic image synthesis to resolve visibility in this step.
- The particle may then be absorbed through surface or volume interaction, in which case the trajectory is terminated with zero contribution. The absorption probability density is  $p_a(\mathbf{x}, \boldsymbol{\omega})$ .
- A „photon“ particle may encounter a sensor, and an „importon“ particle may encounter a light source. In that case, the random walk is complete, and a measurement may be added to the frame buffer according to Equation (II.127).
- Finally, the particle may undergo scattering and obtain a new direction. This is governed by the PDF  $p_s(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\omega}')$ .

In an *analog random walk*, the sampling probability density functions above closely mirror the underlying physical laws. For example, in a participating medium, the PDF  $p_t(\mathbf{x}, \mathbf{x}') = \mu_t(\mathbf{x}')\tau(\mathbf{x}, \mathbf{x}')$  is proportional to the transmittance. Similarly, the absorption PDF in participating media is  $p_a(\mathbf{x}, \boldsymbol{\omega}) = 1 - \alpha(\mathbf{x})$  in analog random walks, where  $\alpha$  is the single-scattering albedo.

In *non-analog random walks*, the same need not be true. Absorption in particular is often modeled not as termination of a random walk, but by down-weighting the path's contribution accordingly (Coveyou et al., 1967; Spanier and Gelbard, 1969). Conceptually, the random walk is not performed by single particles, but by whole populations of photons or importons.

As an example why non-analog random walks are important, consider a situation as shown in Figure II.10, where photons carrying a radiant power of 1W experience exactly one surface interaction. Photons may be absorbed with PDF  $p_a(\mathbf{x}) = 1 - \alpha$ . Otherwise, the sensor registers the photon. The analog random walk yields an estimator with the expected value

$$(1 - \alpha) \cdot 0 + \alpha \cdot 1 = \alpha, \quad (\text{II.128})$$

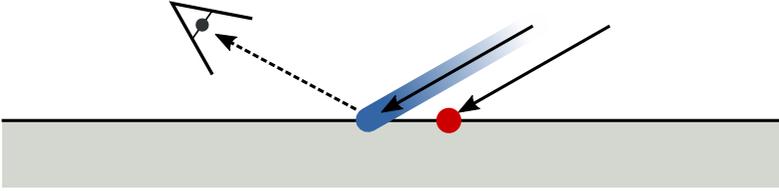


Figure II.10.: In analog random walks (red), path histories are terminated upon absorption. Non-analog random walks (blue) describe populations of particles and can be continued with a lower weight, which may lead to lower variance in a Monte Carlo estimator.

and has variance

$$(1 - \alpha)\alpha^2 + \alpha(1 - \alpha)^2 = (1 - \alpha)\alpha. \quad (\text{II.129})$$

In contrast, the non-analog random walk always counts the same result,  $\alpha$ , which is also the expected value of the associated estimator. Here, the variance is 0. The non-analog random walk is far more efficient because it does not lose any samples to absorption

#### 4.4 Monte Carlo rendering algorithms

A very commonly used technique is *path tracing* (Kajiya, 1986), which constructs paths from the sensor and tries to find connections to light sources. Path tracing places many samples near the sensor, where they contribute. However, path tracers can have severe problems finding small or distant light sources. Ideal point lights cannot be handled by path tracers without extensions.

*Light tracing* (Dutré et al., 1993) constructs paths from light sources, and tries to connect to sensors. Light tracers trivially handle point light sources, but can have trouble finding paths that actually contribute to the output image. In fact, many paths in light tracers tend to be constructed in vain, which leads to wasted sample effort.

The combination of path tracing and light tracing is called *bidirectional path tracing* (Lafortune and Willems, 1993, 1996; Veach and L. Guibas, 1995). Bidirectional path tracing constructs sub-paths from both the sensor and the light sources in the scene, and tries to connect the two. While bidirectional path tracing remedies many of the shortcomings of path tracing and light tracing, it can still lead to a significant waste of samples due to complex visibility in the scene. Additionally, in participating media, connections across long distances often yield very low contribution paths due to the exponential transmittance term.

*Photon mapping* (Jensen, 1996) is a two-pass algorithm which first traces „photons“ from the light sources, storing them in a buffer called photon map. The second pass then builds paths from the sensor and performs a density estimation in the photon map, merging the contribution from multiple light paths. Photon mapping is

particularly useful for rendering refracted patterns of light, or caustics, but requires memory for storing photons. Additionally, photon mapping is a consistent, but biased technique. As such, it introduces systematic error in the image.

*Vertex connection and merging* (Georgiev et al., 2012) and *unified path sampling* (Hachisuka et al., 2012) combine bidirectional path tracing and photon mapping in one consistent theory. *Unified points, beams and paths* (Křivánek, Georgiev, et al., 2014) extends this to volumetric effects.

The main focus of this thesis are participating media and granular materials, which present particularly difficult cases for bidirectional methods. This is why we choose to use path tracing as the main path sampling technique.

We augment this with *next event estimation* (Coveyou et al., 1967), which attempts to connect to known light sources at every path vertex. This can be interpreted as a version of bidirectional path tracing where light paths have only one vertex. Next event estimation has the potential for a significant reduction in variance, particularly for rough surfaces and thin, non-emissive participating media, but care needs to be taken for very short distances: the  $1/r^2$  singularity in the geometric term (see Section II.2.5.1) is known to cause infinite variance (M. Kalos, 1963).

# III

## GRID-BASED SPECTRAL UPSAMPLING

### 1 Introduction

Spectral upsampling, as was mentioned in Section II.3.3, is the problem of determining a spectral energy density so that the response of a sensor matches a given color. Unfortunately, this problem is ill-defined, because there are in general an infinite amount of light spectra matching any given color. This phenomenon, which is called *metamerism*, is rooted in the fact that the integral transformation used to determine the sensor response (see Equation (II.121)) is a projection from an infinite-dimensional to a three-dimensional space.

Though spectral upsampling is not a unique operation, several methods exist that choose one of the many solutions. Generally, if an attenuation spectrum is the target, then it must conserve energy and be smooth (see Section II.3.1).

MacAdam (1935a,b) implicitly defines an XYZ to spectrum conversion by using box spectra. MacAdam's spectra consist of a single box function wrapping around the interval so that they only contain one rising and one falling edge. The method is not limited to a specific input color space, and it produces spectra containing the highest amount of energy possible for any given input color saturation. However, the resulting box spectra are only suitable for rendering to a limited extent, because they may cause unintuitive color shifts in the indirect illumination in fluorescent illumination. This effect can be seen in Figure III.1

Glassner (1995, p.104) gives a family of conversions which take a set of three basis functions as input. This could be the monitor response curves or the first three Fourier bases. These methods can produce negative spectral power, which is a severe problem in physically based light transport simulation, where negative power must be clamped, resulting in uncontrolled color shift.

The upsampling method introduced by Smits (1999) is practical and widely adopted. It is based on a small set of ten precomputed box bases, one for each of the colors Red, Green, Blue, Cyan, Yellow, Magenta, and White. The red, green, and blue spectra are chosen to match the *Rec.709* primaries, while the blue, cyan, and yellow spectra

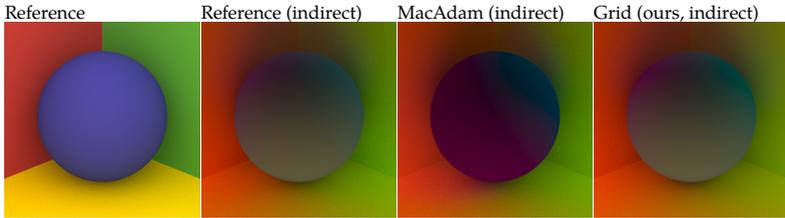


Figure III.1.: MacAdam box spectra may cause strong color shifts in indirect illumination, if the light source is fluorescent. Our smooth spectra reproduce colors much more faithfully. Images showing indirect illumination were scaled  $\times 4$  for display (Meng, Simon, et al., 2015).

are their complements. The white spectrum is computed for the equal-energy white point  $\mathbf{x}_Y = (1/3, 1/3)$ .

The precomputation employs a nonlinear optimization procedure, which attempts to find a spectrum that matches the input color value while minimizing

$$\epsilon = 100 \cdot (\max\{S_i\} - 1) + \sqrt{\sum_{i=1}^9 (S_{i+1} - S_i)^2}, \quad (\text{III.1})$$

where  $S_i$  is the  $i$ -th bin of the resulting spectrum. The first part of this error term favors energy-conserving spectra, while the second part ensures smoothness in the sense that neighboring bins with large differences in value incur a penalty. The smoothness criterion is used because the method is meant to produce spectra that mimic the appearance of naturally occurring reflectance spectra.

It is important to note that the basis spectra found by Smits are not fully energy conserving, since they exceed 1 for some wavelengths. Smits solves this problem by clamping the spectra (Smits, 1999).

The interpolation scheme devised by Smits works as follows. Assume the target color is *Rec.709*  $(R, G, B) = (0.2, 0.3, 0.4)$ . All components are at least 0.2 in magnitude. The algorithm first tries to mix in as much as possible of the precomputed white spectrum by weighting it with a factor of 0.2. The remaining color is  $(0, 0.1, 0.2)$ , which contains only cyan components. The algorithm therefore adds to the output the cyan spectrum with a weight of 0.1. This leaves the value  $(0, 0, 0.1)$ , and so the blue spectrum is also added with a weight of 0.1. In this manner, the algorithm always adds the widest possible spectrum first, proceeding to narrower spectra until the output color is matched.

Because the precomputation uses *Rec.709* primaries, this method cannot properly convert colors that are outside the *Rec.709* gamut. One can employ the same methodology for other color spaces, but as the primaries get more and more saturated, the basis spectra become less and less smooth. The stronger basis peaks are visible in

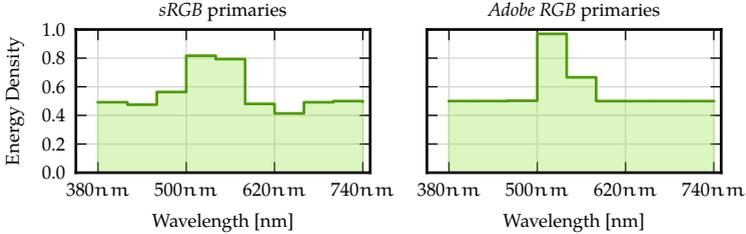


Figure III.2.: Smits spectra for *Adobe RGB* (0.5, 0.8, 0.5) interpolated using *sRGB* (left) and *Adobe RGB* (right) primaries. The more saturated *Adobe RGB* primaries have narrower spectra, which results in a peak in the output (Meng, Simon, et al., 2015).

upsampled spectra, as well. This can be seen in Figure III.2.

In addition to the problems with wide-gamut color spaces, Smits' method does not support round trips: converting a color to spectrum and then back will result in a small, but noticeable color cast. For instance the input  $XYZ = (1, 1, 1)$  results in a maximum spectrum value of 1.00012, and output  $XYZ = (0.99945, 0.99942, 0.99982)$ . This is not the worst case, and the difference seems small. However, with multiple interreflections and in the presence of bright light sources, color shifts can be perceived. We analyze the perceptual error inside and outside this gamut in Figure III.6.

Despite these shortcomings, the novel method we describe in this chapter can be seen an extension to the method of Smits. However, it allows for almost arbitrary input colors and supports perfect round trips.

Wang et. al. (Wang et al., 2004) measure a set of reflectance spectra, and extend this set using Beer's Law. For upsampling, their technique linearly interpolates the basis spectra in three-dimensional  $XYZ$  space. The values for the measured spectra are unfortunately not provided in (Wang et al., 2004). Since the measurements require special hardware, reproducing their results is a challenge. Our method is conceptually similar, but instead of a three-dimensional search, we employ a very simple and efficient two-dimensional scheme.

## 2 Overview of our upsampling method

We are interested in upsampling arbitrary  $XYZ$  input to a spectral energy density function, such that the result is *physically meaningful*. We define a spectrum to be physically meaningful if it resembles natural spectra as much as possible. In particular, the theoretical limit of MacAdam's box spectra are too sharp for our purposes, since spectra derived this way produce unnaturally saturated indirect light with unexpected colour shifts (see Figure III.1). Also, our method must support saturated

input colors, and should not be limited to the *Rec.709* gamut. Finally, our method should be efficient and suitable for evaluation at run-time.

To achieve this, similar to (Smits, 1999), we interpolate precomputed basis spectra to obtain spectra for input values at run time. However, instead of using just seven spectra, we use many more, and distribute them over the space of visible chromaticities in the form of a roughly regular grid. This grid structure allows for efficient bilinear interpolation at run-time in most cases, with a fall-back to barycentric interpolation for highly saturated colors. It also covers almost all colors visible to the human eye, independent of the input color space. The main limiting factor is the fact that we use a finite resolution discretization to represent spectra.

Using basis spectra close to the spectral locus enables us to support very saturated colors, which require spectra with narrow peaks. Nevertheless, our method is able to produce smooth spectra for input color near the white point.

### 3 Precomputation

The goal of the precomputation is to determine basis spectra for all predetermined chromaticities in our grid. Since the grid topology does not make any assumptions on luminance, we impose unit brightness in the sense that  $X + Y + Z = 1$ . This means that interpolated spectra will also have unit brightness. To match the specified output brightness, we can simply scale the output spectrum, since the integral operator converting the spectrum to  $XYZ$  is linear.

#### 3.1 Optimization procedure

We search for smooth spectra that match the input  $(x, y)$  chromaticity. Our spectra are discretized in  $5n_m$  steps, in contrast to the much coarser discretization in (Smits, 1999).

Our error term is similar to the one used by Smits (see Equation III.1). However, we found better convergence using the square norm

$$\epsilon = \sum_{i=1}^{N-1} (S_{i+1} - S_i)^2. \quad (\text{III.2})$$

The optimization is subject to

$$S_i \in [0, 1000] \quad \forall i \quad (\text{III.3})$$

and

$$\sum_i 5n_m \cdot S_i \cdot (\bar{x}_i, \bar{y}_i, \bar{z}_i) = (x, y, 1 - x - y). \quad (\text{III.4})$$

Again,  $S_i$  is the  $i$ -th bin of the output spectrum, and  $(\bar{x}_i, \bar{y}_i, \bar{z}_i)$  are the tabulated color matching functions (see Section II.3.3). Condition (III.4) states that the spectrum must match the input chromaticity. Note that we do not explicitly enforce energy

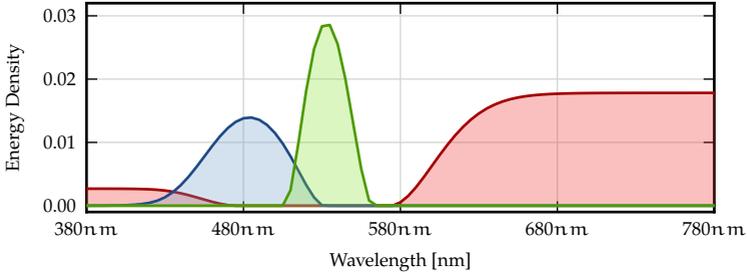


Figure III.3: Three of the smooth basis spectra used in our grid-based upsampling scheme. The full set is reproduced on pages 103ff (Meng, Simon, et al., 2015).

conservation, but merely require the spectra to be bounded. We will detail techniques for ensuring energy conservation later.

While Smits used Matlab (Smits, 1999) for the precomputation, we use Python’s `scipy.optimize` package to perform the optimization. Specifically, we employ the `minimize` solver in SLSQP mode, which is based on (Kraft, 1988). The starting point for our optimization is always an equal energy white spectrum.

The resulting basis spectra are smooth and mostly uni-modal. Two modes can be observed for chromaticities near the purple line. There is some slight noise in the output due to the numerical minimization procedure, and this noise also depends on the starting point. However, we found the process to be overall stable. We reproduce all precomputed spectra on pages 103ff.

In contrast to MacAdam’s box spectra with maximal energy content (MacAdam, 1935b), our spectra are far smoother, and resemble natural reflectance spectra more closely.

### 3.2 Grid topology

To place the grid samples, we first rotate the  $(x, y)$  coordinate system into a new space  $x^*y^*$  where the purple line is horizontal, and the origin corresponds to the equal-energy white point  $(1/3, 1/3)$ .

The location of the grid points is then determined by regular sampling of  $4 \times 6$  cells inside the spectral locus, where we ensure that the most saturated red (lower right corner) lies exactly on a grid point.

This set is extended by a few points on the spectral line which form triangle fans together with the inner grid points. To smooth out interpolation artifacts, we also insert one more point at the center of gravity of all points generated in each boundary cell (see Figure III.4). In total, this algorithm results in 186 grid points, and consequently,

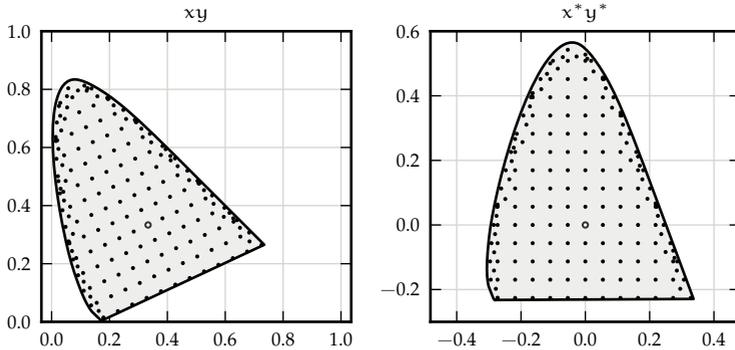


Figure III.4.: The 186 chromaticity sample points used in our method  $xy$  space (left) and  $x^*y^*$  space (right). In  $x^*y^*$ , the equal-energy white point is the origin, and the grid is axis-aligned. This enables us to quickly find the four neighboring basis spectra at run time (Meng, Simon, et al., 2015).

186 basis spectra to be precomputed.

Finally, all grid locations are scaled towards the white point by a small adjustment factor  $1 - \epsilon$ . This step means that some highly saturated to monochromatic colors cannot be converted to spectra with our method. The reason for this scaling is our spectrum discretization. Discretized spectra cannot properly represent peaks narrower than the bin size, which occur close to the spectral locus. This makes the optimization step unstable near the spectral locus. Moving grid points slightly toward the white point avoids these instabilities, while retaining the regular grid structure required for fast interpolation.

The locations of our basis spectrum samples are shown in Figure III.4.

## 4 Grid-based upsampling

The precomputation, though computationally expensive, only needs to be done once. The resulting  $M$  basis spectra  $\{B_1, \dots, B_M\}$  are interpolated linearly to obtain spectra  $S(x, y, \lambda)$  for a given input chromaticity  $(x, y)$ :

$$S(x, y, \lambda) = \sum_{i=1}^M w_i(x, y) B_i(\lambda), \quad (\text{III.5})$$

where the weights  $w_i$  are subject to

$$\sum_{i=1}^M w_i(x, y) = 1 \quad (\text{III.6})$$

$$\sum_{i=1}^M w_i(x, y) x_i = x \quad (\text{III.7})$$

$$\sum_{i=1}^M w_i(x, y) y_i = y. \quad (\text{III.8})$$

Using the fact that the precomputed spectra are normalized so that  $X_i + Y_i + Z_i = 1$ , we can now calculate the chromaticity  $(x', y')$  of the interpolated spectrum:

$$\begin{aligned} x' &= \int_{\Lambda} S(x, y, \lambda) \bar{x}(\lambda) d\lambda = \sum_{i=1}^M w_i(x, y) \int_{\Lambda} \bar{x}(\lambda) B_i(\lambda) d\lambda \\ &= \sum_{i=1}^M w_i(x, y) x_i \\ &= x \end{aligned} \quad (\text{III.9})$$

and

$$\begin{aligned} y' &= \int_{\Lambda} S(x, y, \lambda) \bar{y}(\lambda) d\lambda = \sum_{i=1}^M w_i(x, y) \int_{\Lambda} \bar{y}(\lambda) B_i(\lambda) d\lambda \\ &= \sum_{i=1}^M w_i(x, y) y_i \\ &= y. \end{aligned} \quad (\text{III.10})$$

Therefore, the interpolation yields spectra with the correct chromaticity. The output brightness is

$$\begin{aligned} X' + Y' + Z' &= \sum_{i=1}^M w_i(x, y) X_i + \sum_{i=1}^M w_i(x, y) Y_i + \sum_{i=1}^M w_i(x, y) Z_i \\ &= \sum_{i=1}^M w_i(x, y) \underbrace{(X_i + Y_i + Z_i)}_{=1} = \sum_{i=1}^M w_i(x, y) = 1, \end{aligned} \quad (\text{III.11})$$

but we can obtain the desired target brightness by simple scaling.

To determine the interpolation weights  $w_i$ , we need to distinguish three cases. If the target chromaticity is in one of the inner cells of the regular grid, then we find the four nearest basis spectra using simple division operations and interpolate bilinearly.

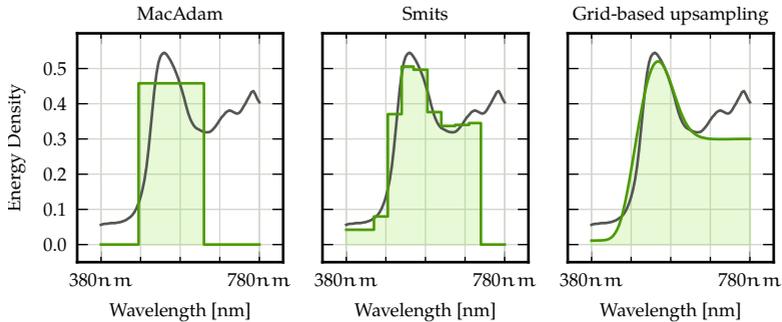


Figure III.5.: Metameric spectra upsampled with three different techniques, overlaid on a measured spectrum for the same color (Meng, Simon, et al., 2015).

Otherwise, we test the border triangles, and, if one of them overlaps  $(x, y)$ , employ barycentric interpolation. Note that condition (III.8) is fulfilled by both bilinear interpolation and barycentric interpolation weights.

Finally, if none of the precomputed cells contains the target chromaticity, we cannot find a matching output spectrum. This only happens for extremely saturated colors that are close to monochromatic, and are unusual for practical scenes. We find this an acceptable shortcoming, since monochromatic spectra cannot be represented in our binned representation in any case.

## 5 Results

In Figure III.5, we show a measured spectrum (McCamy et al., 1976; Ohta and Robertson, 2006, Patch 11 Yellow Green) and metameric spectra obtained for each of the methods we compare.

The MacAdam box spectrum (left) approximates the ground truth, but large wavelength intervals, it carries no energy at all. As we showed in Figure III.1, this can lead to color shifts particularly in the indirect illumination component.

The spectrum created with Smits' method (center) is a far better approximation. However, the method is limited in its ability to produce smooth output by the relatively coarse 10-bin discretization.

Our grid-based upsampling algorithm produces a spectrum with a similar overall shape, since we use a similar optimization criterion for our basis spectra. However, our 80-bin discretization allows for a far smoother result.

It is interesting to note that neither of the techniques is a very good match at the two ends of the spectrum. The reason for this that humans are relatively insensitive to visible light with extremely short or long wavelengths. This is reflected in the color matching functions, particularly in  $\bar{y}(\lambda)$  (see Section II.3.3), which are used during the optimization to guarantee a metameric spectrum.

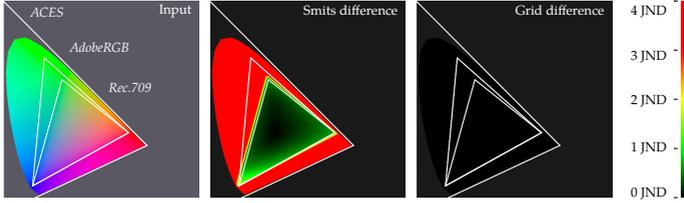


Figure III.6.: Round-trip error for spectra obtained with Smits’ method and our novel grid-based upsampling technique. The difference is visualized as false color, indicating perceptual error as multiples of *just noticeable differences* (Meng, Simon, et al., 2015).

## 5.1 Upsampling speed

We implemented upsampling through box spectra, Smits, and our grid-based method in C. Each method performed 10000 runs, and we measured CPU time on an Intel Core i7-2630QM CPU at 2.00 GHz.

Unsurprisingly, box spectra were fastest, taking 1700m.s to complete the run. Smits’ method required 3342m.s. Similarly, our technique took 3314m.s.

This leads us to conclude that despite the increased quality, there is no noticeable speed impact for using our technique.

## 5.2 Upsampling precision

We compare round-trip error for all supported chromaticity values in Figure III.6 as multiples of *just noticeable differences* (JND). One just noticeable difference is 2.3 CIE 1976  $\Delta E$ , which can be measured as the Cartesian distance in  $L^*a^*b^*$  space (CIE, 2008). The gamuts of *ACES*, *AdobeRGB* and *Rec.709* are shown for better orientation.

Smits’ method is not designed to work outside the *Rec.709* gamut, and trades round-trip precision for validity of reflectances even within this gamut. This is because the algorithm combines two competing goals, optimization to match color and gamut mapping, in only one step.

In contrast, as shown in Equations (III.9) and (III.10), our technique matches target chromaticities perfectly.

## 5.3 Comparison to measured spectra

For the renders in Figure III.7, we used spectra computed with each of the MacAdam, Smits, and grid upsampling methods. The target chromaticities match those of spectra measured for use in color checkers (Ohta and Robertson, 2006). The bottom row shows differences to a spectral render using the original measured spectra.

Our novel technique produces the renders with the least difference, indicating that they match measured reflectance spectra well.

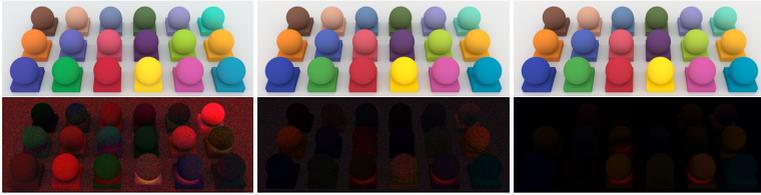


Figure III.7.: Top row: Reflectances in this scene were upsampled using the techniques by MacAdam (left), Smits (center), and our novel grid-based technique (right). Bottom row: Difference images, where the original measured spectra were used for the reference (Meng, Simon, et al., 2015).

## 5.4 Visual evaluation of rendered results

In Figure III.8, we compare the upsampling techniques by MacAdam, Smits, and our new grid upsampling technique in a scene with three Buddha statues, each containing a dense, participating medium. The single-scattering albedo values of the statues are *CIE RGB*  $(1, 0.1, 0.1)$  (red),  $(0, 1, 0.05)$  (green), and  $(0, 0.03, 1)$  (blue), which for each of the upsampling techniques result in spectra that are not energy-conserving. To avoid divergence, we therefore fix the spectra by scaling to maximum 1. The energy-conserving box spectra (left) cause a distinct lack in detail, and the statues look over-saturated. The statues rendered with Smits' spectra (center), are still extremely saturated and have a look that is reminiscent of fluorescent materials. The smooth spectra computed with our grid upsampling method produce a much more natural look (right).

## 6 Conclusion

We have introduced a novel spectral upsampling algorithm that supports all but the most saturated chromaticities. Due to the grid structure of our precomputed base spectra, the method is extremely fast. It allows single-wavelength upsampling, which makes it well suited for spectral Monte Carlo rendering algorithms. The spectra obtained through our method are smooth and present a convincing means for rendering with naturally looking reflectances. Finally, our technique is consistent in the sense that it supports round-trips.

In (Meng, Simon, et al., 2015), we also introduce the concept of the *solid of natural reflectances*, which is the set of colors for which an energy conserving, smooth spectrum exists, and methods for mapping arbitrary input colors into this set. This set is defined using the smoothness criterion we also employ for the grid upsampling technique presented in this chapter. However, since that part of our work is mainly due to the joint first author, F. Simon, it is not included in this thesis.



Figure III.8.: Three Buddha statues containing saturated homogeneous participating media. Upsampling of the single-scattering albedo was performed with the techniques of MacAdam (left), Smits (center), and our grid-based technique (right) (Meng, Simon, et al., 2015).



# IV

## IMPROVED DWIVEDI SAMPLING

### 1 Motivation

Many fast approximations to rendering sub-surface scattering are known, but they tend to show artifacts near concavities and thin features in the geometry. The reason for this is that most analytic methods build on the assumption of simplified geometry, such as semi-infinite slabs. Results will be sub-optimal, and sometimes visually displeasing, whenever this assumption is violated grossly.

Classical Monte Carlo simulation does not suffer from this problem, but tends to be slow in participating media: uninformed random walks may scatter many times before exiting the medium. Yet, one cannot simply terminate random walks just because they are long, because they may contribute significantly to the image. This is especially true in high-albedo participating media.

To solve this problem, Křivánek and d'Eon (Křivánek and d'Eon, 2014) applied the zero-variance random walk theory by Dwivedi (Dwivedi, 1982a,b) to rendering. In this method, geometry is also approximated by semi-infinite slabs to compute an analytic approximation of the light field. However, this light field is not visualized directly. Instead, it is merely used to bias the sampling probability distributions employed for the Monte Carlo random walk. In essence, Dwivedi sampling will prefer directions towards the slab surface, and will move in larger steps if the random walk progresses in the direction of the slab normal. Note that the resulting estimator is unbiased despite the double use of the word bias in literature.

Křivánek and d'Eon (Křivánek and d'Eon, 2014) place the semi-infinite slab so that its normal is aligned with the surface normal at the *Point of Entry*, which is the point where the current random walk first entered the medium (see Figure IV.1 left). This generally leads to much shorter random walks, and thus improves convergence considerably.

However, this strategy is not optimal in certain scenarios. The *Point of Entry* really is an approximation to the closest surface point, but this approximation may be misleading. Additionally, in scenes featuring a strong back light, going back towards the *Point of Entry* can lead to most random walks being drawn away from the main light source. These issues are illustrated in Figure IV.1 (center and right). We observe that

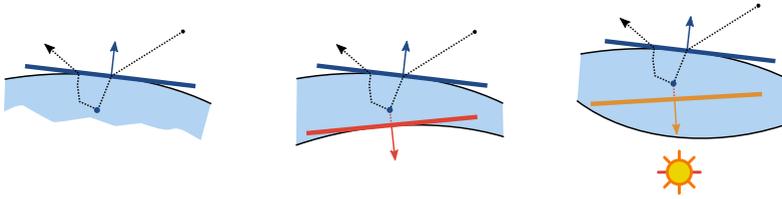


Figure IV.1.: Slab orientation using the normal at the *Point of Entry* (PoE) for different input geometries. The slab (shown in blue) is a good approximation in large geometries with low curvature (left). It tries to approximate the closest surface point, which can fail in the presence of thin geometric features (center) or high surface curvature. Additionally, if there is a strong back light, PoE sampling will drag the random walk away from the light source (right). The slabs used by our proposed techniques are shown in red: *Closest Point* (center) and *Incident Illumination* (right) (Meng, Hanika, et al., 2016).

in such cases other means of biasing the random walk can perform much better (see results in Section IV.4).

As our main contributions,

- we propose replacing the *Point of Entry* sampling strategy with *Closest Point* sampling,
- we propose using *Incident Illumination* sampling for scenarios with a strong backlight, and
- we propose a heuristic approach for combining *Classical*, *Closest Point*, and *Incident Illumination* sampling using multiple importance sampling (Veach and L. J. Guibas, 1995).

## 1.1 Previous work on sub-surface scattering for skin

Rendering skin efficiently and accurately has received a fair amount of attention in literature. The dipole method (Jensen, Marschner, et al., 2001) in combination with irradiance caches (Jensen and Buhler, 2002) was very popular for a long time, and scattering parameters have been derived from measurements (Donner and Jensen, 2005, 2006). Fast approximations in the form of sums of Gaussians have been researched (d'Eon, Luebke, et al., 2007). Diffusion profiles on the surface have been refined from the theoretical view (d'Eon and Irving, 2011) as well as from the practical side (Christensen and Burley, 2015). Those have been refined by using directional dipole models (Frisvad et al., 2014). Production work flows have been simplified considerably by removing the irradiance caches from the pipeline and by moving to importance sampling of diffusion profiles in the ray tracing context (King et al.,

2013). Such a ray tracing scheme, however, does not simulate the full transport inside the tissue but only samples a new exit point on the surface, under the assumption of locally flat geometry, leading to problems with high curvature and near concavities. For example, light bleeding may occur or diffusion profiles may be distorted. To avoid such problems, it is possible to use these closed form solutions to guide Monte Carlo sampling, which we will detail in the next section.

## 2 Dwivedi sampling

In this section, we review the main concepts of Dwivedi sampling (Dwivedi, 1982a,b) and its foundation, zero variance random walk theory.

The concept of zero variance random walks has been explored by various authors, e.g. (T. E. Booth, 1987; Dwivedi, 1982a,b), but a good overview is given in a work by Hoogenboom (Hoogenboom, 2008). The idea is appealing: compute a solution to the RTE (II.104) with Monte Carlo random walks, but in such a way that the estimator has no variance.

This can be done by modifying, or biasing, the sampling distributions for position and direction in the random walk. Note that the word *bias* is overloaded in this context: even though sampling distributions are biased, the resulting estimator is still unbiased in the sense that it has the correct expected value. This is similar to importance sampling the free path for the time dimension in the context of transient rendering (Jarabo et al., 2014).

Dwivedi (1982b) showed that one can construct zero variance random walks by sampling a new position  $\mathbf{x}'$  and direction  $\boldsymbol{\omega}'$  proportionally to the product of phase function  $\phi$ , transmittance  $\tau$ , and importance  $W$ :

$$p(\mathbf{x}', \boldsymbol{\omega}' | \mathbf{x}, \boldsymbol{\omega}) \propto \phi(\boldsymbol{\omega}' | \boldsymbol{\omega}) \tau(\mathbf{x}, \mathbf{x}') W(\mathbf{x}', \boldsymbol{\omega}'), \quad (\text{IV.1})$$

where

$$\tau(\mathbf{x}, \mathbf{x}') = \exp\left(-\int_{\mathbf{x}}^{\mathbf{x}'} \mu_t(\mathbf{x}^*) d\mathbf{x}^*\right). \quad (\text{IV.2})$$

Unfortunately, this requires knowledge of the importance function for every position and direction. This is a far more difficult problem than determining the sensor response for a given sensor, and so true zero variance random walks are mainly of theoretical interest.

However, as Dwivedi points out (Dwivedi, 1982b), they can still be used to reduce variance if an approximate importance function is known. Such an importance function can be computed analytically for slab geometry using singular eigenfunction expansions.

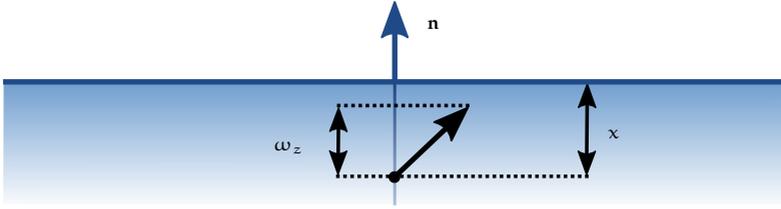


Figure IV.2.: An illustration of slab geometry. The slab normal is  $\mathbf{n} = (0, 0, 1)^\top$ . A ray can be parametrized by its depth  $x$  and the cosine of the angle between its direction and the slab normal,  $\omega_z$  (Meng, Hanika, et al., 2016).

## 2.1 Singular eigenfunction expansions

In this section, we will derive *singular eigenfunction expansions* (SEF) in enough detail to understand the rest of the chapter. The full derivation, which we follow very closely and which includes a proof of full-range completeness, can be found in (McCormick and Kuscer, 1973).

We will derive SEF for the adjoint radiative transfer equation, which describes how importance  $W$  is propagated from a sensor through the scene. The same derivation applies to the standard RTE that is used to describe the propagation of radiance.

SEF are a way of finding a solution to the RTE, and this is used in Dwivedi sampling to change the sampling distributions.

Ignoring time and wavelength dependencies, the full adjoint RTE is

$$\left[ \frac{d}{ds} + \mu_t(\mathbf{x}) \right] W(\mathbf{x}, \boldsymbol{\omega}) = W^e(\mathbf{x}, \boldsymbol{\omega}) + \mu_s(\mathbf{x}) \int_{4\pi} \phi(\boldsymbol{\omega}|\boldsymbol{\omega}') W(\mathbf{x}, \boldsymbol{\omega}') d\boldsymbol{\omega}'. \quad (\text{IV.3})$$

Here,  $W(\mathbf{x}, \boldsymbol{\omega})$  is the importance at position  $\mathbf{x}$  in direction  $\boldsymbol{\omega}$ , and  $W^e(\mathbf{x}, \boldsymbol{\omega})$  describes sensors in terms of emitted importance. Similar to the standard RTE,  $\mu_t(\mathbf{x})$  and  $\mu_s(\mathbf{x})$  are the extinction and scattering coefficients, respectively, while  $\phi(\boldsymbol{\omega}|\boldsymbol{\omega}')$  is the phase function.

To be able to derive an analytic solution for this equation, one introduces a few simplifications.

1. The medium is assumed to be homogeneous, so  $\mu_t$  and  $\mu_s$  are constants.
2. Similar to diffusion methods, the geometric setup is assumed to be a one-dimensional slab with surface at depth  $x = 0$ . Directions in slab geometry can be represented by just the cosine of the angle to the surface normal  $\langle \boldsymbol{\omega}, \mathbf{n} \rangle = \omega_z$ . This setup is shown in Figure IV.2.
3. There is no emission inside the volume. The slab surface can then be seen as a sensor, or importance emitter. This means that  $W^e(x, \omega_z) = 0$  and the RTE is a homogeneous differential equation.

4. The phase function is assumed to be constant, so that there is no preferred direction of scattering.

In total, the adjoint RTE in a homogeneous slab with isotropic scattering can be written as (Chandrasekhar, 1960)

$$\left[ \omega_z \frac{\partial}{\partial x} + \mu_t \right] W(x, \omega_z) = \frac{\mu_s}{2} \int_{-1}^1 W(x, \omega'_z) d\omega'_z. \quad (\text{IV.4})$$

We now separate the spatial and directional variables using the ansatz

$$W(x, \omega_z) = \varphi(\nu, \omega_z) \cdot e^{-\mu_t x/\nu}. \quad (\text{IV.5})$$

The directional part of this solution  $\varphi(\nu, \omega_z)$  is called an eigenfunction. It is interesting to note that the spatial part is always non-negative, and so the directional part must be non-negative, as well, or importance would become negative. Additionally, if importance is non-zero for at least some direction  $\omega_z$ , then  $\varphi(\nu, \omega_z)$  must be greater than zero, which tells us that

$$\int_{-1}^1 \varphi(\nu, \omega'_z) d\omega'_z > 0. \quad (\text{IV.6})$$

Substituting Equation (IV.5) into Equation (IV.4), we obtain

$$\begin{aligned} \left[ \omega_z \frac{\partial}{\partial x} + \mu_t \right] \varphi(\nu, \omega_z) e^{-\mu_t x/\nu} &= \frac{\mu_s}{2} e^{-\mu_t x/\nu} \int_{-1}^1 \varphi(\nu, \omega'_z) d\omega'_z \\ \iff \left[ -\omega_z \frac{\mu_t}{\nu} + \mu_t \right] \varphi(\nu, \omega_z) e^{-\mu_t x/\nu} &= \frac{\mu_s}{2} e^{-\mu_t x/\nu} \int_{-1}^1 \varphi(\nu, \omega'_z) d\omega'_z \\ \iff [\nu - \omega_z] \varphi(\nu, \omega_z) &= \nu \frac{\alpha}{2} \int_{-1}^1 \varphi(\nu, \omega'_z) d\omega'_z, \end{aligned} \quad (\text{IV.7})$$

where  $\alpha = \mu_s/\mu_t$  is the single-scattering albedo. Using Equation (IV.6), we can impose the normalization

$$\int_{-1}^1 \varphi(\nu, \omega'_z) d\omega'_z = 1, \quad (\text{IV.8})$$

and, since  $\varphi$  appears on both sides of the equation, Equation (IV.7) simplifies to

$$[\nu - \omega_z] \varphi(\nu, \omega_z) = \frac{\alpha}{2} \nu \iff \varphi(\nu, \omega_z) = \frac{\alpha}{2} \frac{\nu}{\nu - \omega_z}. \quad (\text{IV.9})$$

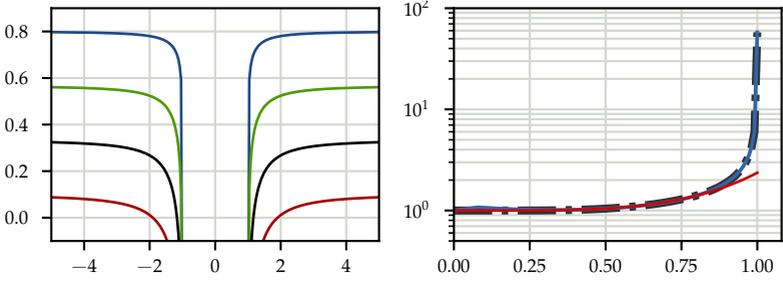


Figure IV.3.: Left: A plot of the left side of Equation (IV.10) for  $\alpha = 0.2$  (blue),  $\alpha = 0.43$  (green),  $\alpha = 0.67$  (black), and  $\alpha = 0.9$  (red). The roots of this function are  $\pm \nu_0$ . Right: the two approximations to  $\nu_0$  for low (red) and high (blue) albedo as given in (McCormick and Kuscer, 1973). They are overlaid on the ground truth (black, dashed) (Meng, Hanika, et al., 2016).

All  $\nu$  that induce valid solutions are called eigenvalues, and must satisfy the normalization constraint (IV.8). Substituting into the normalization constraint, we get

$$\begin{aligned}
 \frac{\alpha}{2} \nu \int_{-1}^1 \frac{1}{\nu - \omega_z} d\omega_z &= 1 \\
 \iff 1 - \frac{\alpha}{2} \nu \log \left( \frac{\nu + 1}{\nu - 1} \right) &= 0 \\
 \iff 1 - \frac{\alpha}{2} \nu \log \left( \frac{1 + 1/\nu}{1 - 1/\nu} \right) &= 0 \\
 \iff 1 - \alpha \nu \tanh^{-1} \left( \frac{1}{\nu} \right) &= 0.
 \end{aligned} \tag{IV.10}$$

Note that these manipulations are not possible if  $\nu \in [-1, 1]$ . As shown in Figure IV.3 (left), the left side of Equation (IV.10) is an even function and it has exactly two roots  $\pm \nu_0$ . These are called discrete eigenvalues.  $\nu_0$  is also called the *diffusion length* in literature (Case et al., 1953). We will show in Section IV.2.3 how to find the diffusion length in practice.

McCormick and Kuscer (McCormick and Kuscer, 1973) show that there are also eigenvalues  $\nu \in [-1, 1]$ . Those are called singular eigenvalues because  $\varphi(\nu, \omega_z)$  is infinite for  $\nu = \omega_z$ . Full-range completeness can also be shown (McCormick and Kuscer, 1973), and so one can write the set of all possible solutions to the RTE as the

linear combination

$$\begin{aligned}
 W(x, \omega_z) &= A(\nu_0) \varphi(\nu_0, \omega_z) e^{-\mu_t x / \nu_0} \\
 &+ \int_{-1}^1 A(\nu) \varphi(\nu, \omega_z) e^{-\mu_t x / \nu} d\nu \\
 &+ A(-\nu_0) \varphi(-\nu_0, \omega_z) e^{-\mu_t x / -\nu_0}.
 \end{aligned} \tag{IV.11}$$

Here,  $A(\nu)$  are coefficients. Equation (IV.11) is the singular eigenfunction expansion of  $W(x, \omega_z)$ .

## 2.2 Approximate importance field

Previous work (Dwivedi, 1982a; Křivánek and d'Eon, 2014) now prunes the expansion aggressively. First, the importance in a slab without detectors must decrease as the distance to the surface goes to infinity. Consequently, coefficients of non-decreasing modes in Equation (IV.11) are set to 0.

Second, the singular part of the expansion decreases faster than the discrete part, since  $\nu_0 > 1$  and the integration domain ends at 1. Thus, assuming  $x$  is large, the solution will be dominated by the discrete part.

This leads to the final approximation

$$W(x, \omega_z) \propto \frac{1}{\nu_0 - \omega_z} e^{-\mu_t x / \nu_0}. \tag{IV.12}$$

Note that we are only interested in proportionality here because we are looking for a probability density function. This also means that the coefficient  $A(\nu_0)$  need never actually be evaluated.

## 2.3 Finding the diffusion length $\nu_0$

As shown in Section IV.2.1, the diffusion length  $\nu_0$  is the solution to Equation (IV.10). It exists for  $\alpha \in [0, 1)$  and can be found numerically using root finding algorithms. Unfortunately, these are so expensive that they cannot be used during rendering. However, Case et al. (1953, p.55) give two series expansions for  $\kappa_0 = 1/\nu_0$ . The first expansion is

$$\begin{aligned}
 \kappa_{0,\text{low}} \approx & 1 - 2e^{-2/\alpha} \left( 1 + \frac{4-\alpha}{\alpha} e^{-2/\alpha} + \frac{24-12\alpha+\alpha^2}{\alpha^2} e^{-4/\alpha} \right. \\
 & \left. + \frac{512-384\alpha+72\alpha^2-3\alpha^3}{\alpha^3} e^{-6/\alpha} \right),
 \end{aligned} \tag{IV.13}$$

and is designed for low albedo materials. The second expansion is geared towards high albedo materials and reads

$$\begin{aligned} \kappa_{0,\text{high}} \approx & \sqrt{3(1-\alpha)} \left( 1 - \frac{2}{5}(1-\alpha) - \frac{12}{175}(1-\alpha)^2 \right. \\ & \left. - \frac{2}{125}(1-\alpha)^3 - \frac{166}{67375}(1-\alpha)^4 \right). \end{aligned} \quad (\text{IV.14})$$

Both approximations are overlaid on the exact solution in Figure IV.3 (right). We switch between the approximations at  $\alpha = 0.56$ . For  $\alpha = 1$ , the equation cannot be satisfied, and so we cannot find  $\nu_0$ . Instead, we use  $\nu_0$  computed for  $\alpha = 0.9999$  whenever  $\alpha \geq 0.9999$ .

## 2.4 Biased distributions for Dwivedi sampling

Combining Sections IV.2 and IV.2.1, and assuming a constant phase function, a reduction in variance can be expected by sampling directions and positions from the distributions

$$p(\omega_z) \propto \frac{1}{\nu_0 - \omega_z} \quad (\text{IV.15})$$

$$p(x'|x) \propto \tau(x, x') e^{-\mu_t x' / \nu_0} \quad (\text{IV.16})$$

The antiderivative of the directional distribution function (IV.15) with respect to  $\omega_z$  is

$$-\log(\nu_0 - \omega_z), \quad (\text{IV.17})$$

and so the normalization must be one over

$$\begin{aligned} \int_{-1}^1 \frac{1}{\nu_0 - \omega_z} d\omega_z &= -\log(\nu_0 - \omega_z) \Big|_{-1}^1 \\ &= \log(\nu_0 + 1) - \log(\nu_0 - 1) \\ &= \log\left(\frac{\nu_0 + 1}{\nu_0 - 1}\right). \end{aligned} \quad (\text{IV.18})$$

We can now write the cumulative distribution function

$$\begin{aligned} F(\omega_z) &= \frac{1}{\log\left(\frac{\nu_0+1}{\nu_0-1}\right)} \cdot -\log(\nu_0 - \omega_z) \Big|_{-1}^{\omega_z} \\ &= \frac{1}{\log\left(\frac{\nu_0+1}{\nu_0-1}\right)} \cdot \log\left(\frac{\nu_0+1}{\nu_0 - \omega_z}\right). \end{aligned} \quad (\text{IV.19})$$

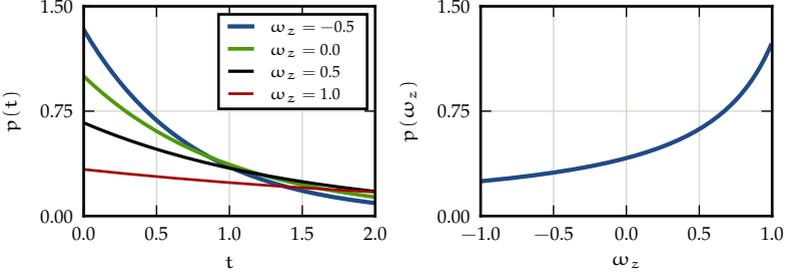


Figure IV.4.: The biased distance sampling (left) and directional sampling (right) distributions used in Dwivedi sampling, both given for  $\nu_0 = 1.5$  (Meng, Hanika, et al., 2016).

Using the inversion method, we obtain the sample warping

$$\begin{aligned}
 F(\omega_z) &\stackrel{!}{=} \xi, \\
 \Leftrightarrow \log\left(\frac{\nu_0 + 1}{\nu_0 - \omega_z}\right) &= \log\left(\frac{\nu_0 + 1}{\nu_0 - 1}\right) \cdot \xi, \\
 \Leftrightarrow \frac{\nu_0 + 1}{\nu_0 - \omega_z} &= \left(\frac{\nu_0 + 1}{\nu_0 - 1}\right)^\xi \\
 \Leftrightarrow (\nu_0 + 1) \left(\frac{\nu_0 - 1}{\nu_0 + 1}\right)^\xi &= \nu_0 - \omega_z \\
 \Leftrightarrow \omega_z &= \nu_0 - (\nu_0 + 1) \left(\frac{\nu_0 - 1}{\nu_0 + 1}\right)^\xi.
 \end{aligned} \tag{IV.20}$$

The free path sampling PDF (IV.16) can be simplified using the assumption that the medium is homogeneous, which means that the transmittance term is an exponential in  $t = \|\mathbf{x}' - \mathbf{x}\|$ . Using  $\mathbf{x}' = \mathbf{x} - \omega_z \mathbf{t}$ , we get

$$p(\mathbf{t}|\mathbf{x}, \omega_z) \propto e^{-\mu_t t} e^{-\mu_t (\mathbf{x} - \omega_z \mathbf{t})/\nu_0} = e^{-(1 - \omega_z/\nu_0) \mu_t t}. \tag{IV.21}$$

The normalized distribution is a simple exponential distribution:

$$p(t|\mathbf{x}, \omega_z) = (1 - \omega_z/\nu_0) \mu_t e^{-(1 - \omega_z/\nu_0) \mu_t t}. \tag{IV.22}$$

This distribution can be sampled as usual. It effectively has a modified extinction coefficient,

$$\mu'_t(\omega_z) = (1 - \omega_z/\nu_0) \mu_t. \tag{IV.23}$$

In Figure IV.4, we illustrate that the directional distribution prefers directions towards the slab normal ( $\omega_z = 1$ ), while the free path distribution has a longer tail (and thus

prefers longer paths) if the direction is oriented along the slab normal.

## 2.5 Slab orientation

One of the assumptions used in Dwivedi sampling is that the geometry can be approximated by a semi-infinite slab. In realistic scenes, this assumption is usually violated in general, but it may hold locally. It is thus important to consider how exactly a slab can be placed into a scene for good performance. Since the positional variable is not present in the final sampling distributions Equations (IV.15) and (IV.22), this amounts to choosing how the slab is oriented, or what its normal is.

Křivánek and d’Eon (2014) use the surface normal at the *Point of Entry* (PoE) as the slab normal, i.e. the point where the random walk first enters the medium (see Figure IV.1, left). The underlying assumption is that the normal at the PoE will usually point away from the medium. If a random walk is biased into this direction, it will probably exit the medium quickly. Křivánek and d’Eon show that this technique can effectively reduce variance in the resulting image. They employ multiple importance sampling to combine it with *Classical* (C) sampling, where the unbiased phase function and free path distribution are used for sampling.

PoE sampling can be problematic in scenes where the medium surface is not approximated well by a slab (see Figure IV.1). If the surface is not really planar, the PoE may not be a good estimate of the closest surface point, and the normal at the PoE may not be a good estimate of the overall surface orientation. To alleviate this problem, we propose biasing the random walk towards the location of the true closest surface point.

Also, the *Point of Entry* technique has difficulties in scenarios where a light source opposite the PoE is strong enough to have a non-negligible contribution to the image. The biasing draws most samples away from the light source in this case, which leads to an increase in variance. We propose to use an additional biasing technique that shifts the random walk towards a point placed on a light source.

## 3 Improving biased sampling

In this section, we detail our proposed improvements to the Dwivedi sampling technique as introduced by Křivánek and d’Eon (2014). We call their technique C+PoE, or *Classical + Point of Entry* sampling.

### 3.1 Closest surface point

Instead of using the normal at the point of entry as the slab normal, we propose using the direction towards the *Closest Surface Point* (CP)  $\mathbf{x}_{CP}$ .

We employ a bounding volume hierarchy to efficiently compute the closest point

transform for the current vertex position  $\mathbf{x}$ . We then use

$$\mathbf{n}_{\text{CP}} = \frac{\mathbf{x}_{\text{CP}} - \mathbf{x}}{\|\mathbf{x}_{\text{CP}} - \mathbf{x}\|} \quad (\text{IV.24})$$

as the slab normal. Note that in contrast to PoE sampling, we do not use the surface normal at the CP directly, because that can be problematic in scenes with highly displaced surface geometry.

We search the CP only once per random walk, specifically at the first vertex inside the medium, and always bias towards the same surface point. However, the direction  $\mathbf{n}_{\text{CP}}$  is updated at every vertex. It is possible to also recompute a new  $\mathbf{x}_{\text{CP}}$  at every vertex, but our tests did not show a benefit that would justify the increase in cost.

### 3.2 Incident illumination

To handle non-uniform illumination better, one could use next event estimation from the volume to connect to light sources. However, this may not be practical due to refractive surfaces such as rough dielectrics (Walter et al., 2007) found on skin (Hanika, Droske, et al., 2015; Holzschuch, 2015), or because the medium is so dense that transmittance to the surface is practically zero.

Instead, we propose sampling one light vertex per random walk, and biasing towards it at each step. Using importance sampling, we can choose light vertices proportional to their emission. We ignore visibility in this step. In scenes with strong backlights, this technique biases most paths towards the brightest light sources in the scene, leading to much better convergence in translucent regions.

The slab direction can be computed in a fashion similar to Equation (IV.24) for area lights or point lights:

$$\mathbf{n}_{\text{DL}} = \frac{\mathbf{x}_{\text{DL}} - \mathbf{x}}{\|\mathbf{x}_{\text{DL}} - \mathbf{x}\|}. \quad (\text{IV.25})$$

For directional light sources such as environment maps, we can use the light direction as the slab normal.

### 3.3 Multiple importance sampling

Similar to Křivánek and d’Eon (2014), we combine the *Classical*, *Closest Point* and *Incident Illumination* sampling techniques using multiple importance sampling with the one sample model and the balance heuristic (see Section II.1.11).

This is imperative with biased sampling techniques: since the transmittance and phase functions do not cancel out of the contribution function with non-perfect importance sampling, low-probability random walks with a high contribution lead to excessively bright pixels, or fireflies. Multiple importance sampling can detect such paths, and reduce their weight automatically.

A new sampling technique is selected randomly at every vertex of the random walk, but biasing techniques have zero probability before the first volume vertex. This means that the first volume vertex is always sampled using *Classical* sampling.

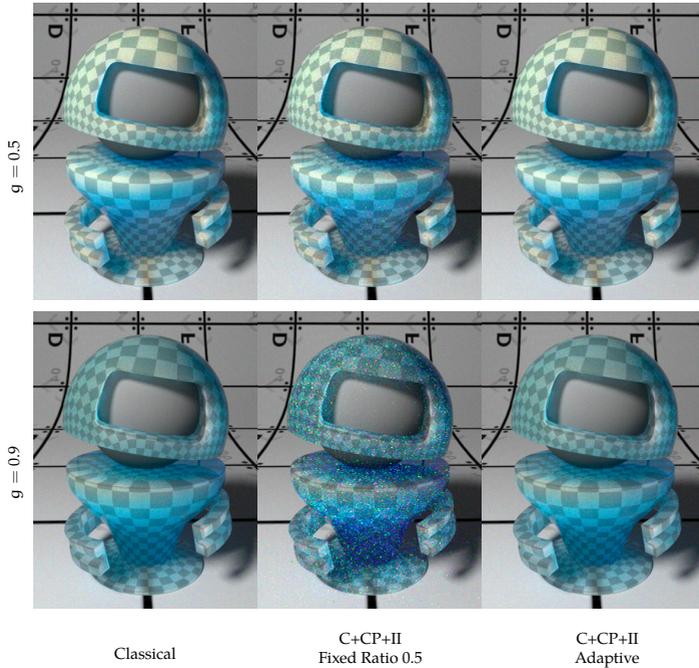


Figure IV.5.: Anisotropic phase functions are problematic with biased sampling, which may cause strong variance (center). Our adaptive heuristic (see Equation IV.26) decreases the probability of selecting a biasing technique as the anisotropy increases, reducing variance (right) (Meng, Hanika, et al., 2016).

### 3.4 Selection probability for *Classical* sampling

In (Křivánek and d’Eon, 2014), a fixed percentage of samples is allocated to *Classical* sampling. We follow this approach for constant phase functions, where we use a probability of 0.1 for selecting *Classical* sampling.

In media with anisotropic phase functions, such as the Henyey-Greenstein phase function (Henyey and Greenstein, 1941) with mean cosine  $g \neq 0$ , directional biasing creates random walks with a low probability density if the slab normal and the ray direction are not aligned. This is demonstrated in Figure IV.5. We use an adaptive criterion that reduces the probability for selecting a biased sampling technique for anisotropic scattering. It basically lets the sampling degrade gracefully to pure *Classical* sampling. We found that

$$p^c = \max \left\{ 0.1, |g|^{1/3} \right\} \quad (\text{IV.26})$$

yields results with low noise over the full range of possible mean cosines. Note that this may differ if other phase functions are used, especially complex, highly anisotropic phase functions. However, it should be noted that biased sampling may generally not be the best choice for media with such phase functions.

### 3.5 Selection probabilities for biased sampling

In a medium with extinction coefficient  $\mu_t$ , the expected number of volume events on distance  $t$  is  $\mu_t \cdot t$ . If a direct random walk of length  $t$  is created using perfect *Classical* sampling, the contribution of that random walk will be weighted by a factor of

$$\alpha^{\mu_t \cdot t}. \quad (IV.27)$$

We use this value to define the probability for selecting one of the three biasing techniques:

$$p_i^b = (1 - w^c) \frac{\alpha^{N_i}}{\sum_{j=0}^3 \alpha^{N_j}}, \quad (IV.28)$$

where  $N_i = \mu_t \cdot t_i$  is the expected number of events to the slab surface of technique  $i$ . In essence, if the medium has a high single-scattering albedo  $\alpha$ , travelling a large distance and across multiple scattering events may be worthwhile.

When *Incident Illumination* biasing is used, random walks tend to get rather long, since they may traverse the whole medium before exiting. This is generally more expensive than PoE or CP biasing, which exit at a nearby surface. However, if the contribution is large enough, it may still be worth performing such long walks.

### 3.6 Committing to *Incident Illumination* sampling

It would be problematic if the two different types of biasing techniques competed for the random walk, sending it back and forth. For this reason, we set the probability of selecting either PoE or CP to zero once II is selected for the first time. This means that a random walk commits to biasing towards the light source. We always keep a minimum probability of sampling using the *Classical* technique, however.

## 4 Results

All results were rendered on a machine with eight Intel Xeon E7-8867 v3 processors running at 2.5 GHz. We used a total of 256 concurrent threads. Render times are given in thread hours.

We implemented the reference and proposed techniques in our research renderer, which uses single-wavelength spectral transport. The ray tracing backend is Embree (Wald et al., 2014). The renderer does not impose a maximum length for random walks. Terminating paths after a given number of bounces will increase performance in practice, but tends to miss important long, but high-contribution light paths that occur in the dense scattering scenes we show in our paper.

Scene	BxDF	Mean Free Path [cm]	$\alpha$
Ear	Diffuse Transmitter (textured)	[0.13, 0.09, 0.671]	[0.959, 0.764, 0.678]
Candle	Diffuse Transmitter ( $\rho = 1$ )	heterogeneous	[0.631, 0.887, 0.999]
Buddha	Smooth Dielectric ( $\eta = 1.2$ )	0.5	[0.98, 0.85, 0.2]

Table IV.1.: Material parameters for our test scenes. Colors are given in linear sRGB.

Our scenes feature a large dynamic range, and so we had to apply some amount of tone mapping to ensure acceptable printing. Still, we recommend viewing results on a computer screen. Both tone mapped and out-of-renderer images are provided in supplemental material to (Meng, Hanika, et al., 2016) as OpenEXR files.

We will now describe our test scenes, and analyze them afterwards. Material parameters for our scenes are summarized in Tab. IV.1.

## 4.1 Ear scene

The Ear scene contains three area light sources: two fill lights illuminate the scene from the front, but are not directly visible. A strong backlight is visible in frame, and causes visible sub-surface scattering in the ear lobe. The ear lobe is approximately 5 cm tall.

We show the full rendering, enlarged crops from the cheek and the ear, and RMSE in Figure IV.6. The crops were taken from roughly equal-time renders, at 9.9 thread hours each.

## 4.2 Candle scene

The only source of light in the Candle scene is a brightly emitting object approximating a flame. It causes sub-surface scattering throughout the candle, which is approximately 12 cm tall. The wax is a heterogeneous volume with spatially varying density, which results in a pattern of brighter and darker spots.

We show the full rendering, crops from the bottom part of the candle, and RMSE in Figure IV.7. The crops are roughly equal time at about 4.25 thread hours each.

## 4.3 Buddha scene

The statuette in the Buddha scene is roughly 20 cm tall. There is only one source of illumination, a strong light illuminating the scene from behind.

Again, we show the full rendering, crops from the bottom part of the statuette, and RMSE in Figure IV.8. The crops are roughly equal time at about 4.25 thread hours each.

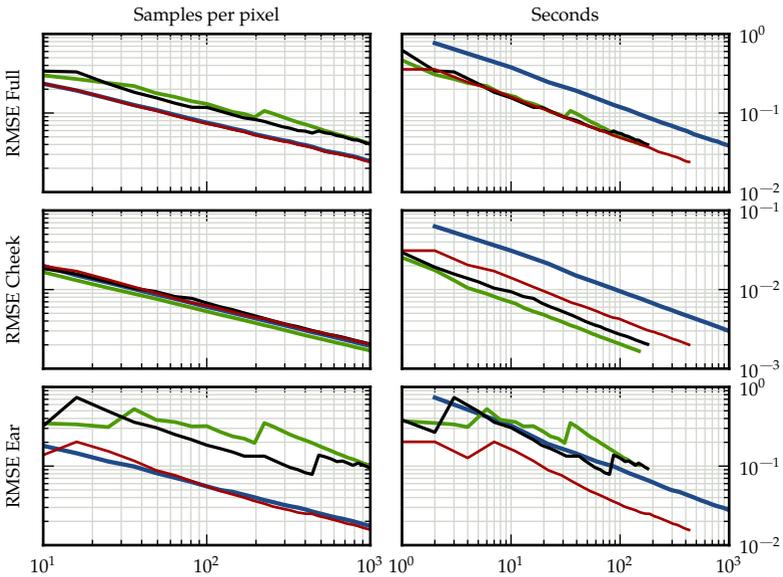
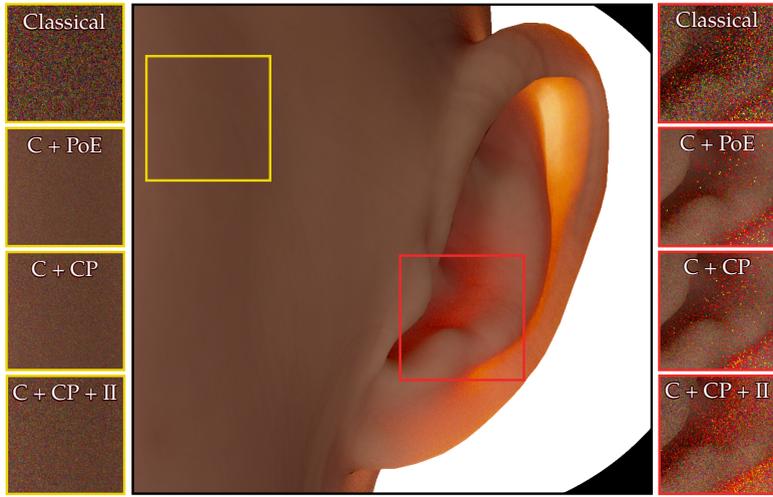


Figure IV.6.: Top: The Ear scene, which features a strong backlight behind a translucent ear, with equal-time comparisons at roughly 9.9 thread hours each. Bottom: Root mean squared error for the full Ear scene, the cheek crop, and the ear crop (Meng, Hanika, et al., 2016).

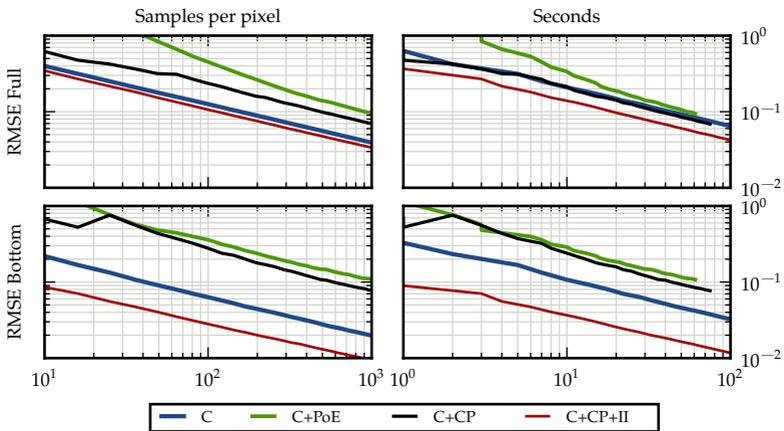
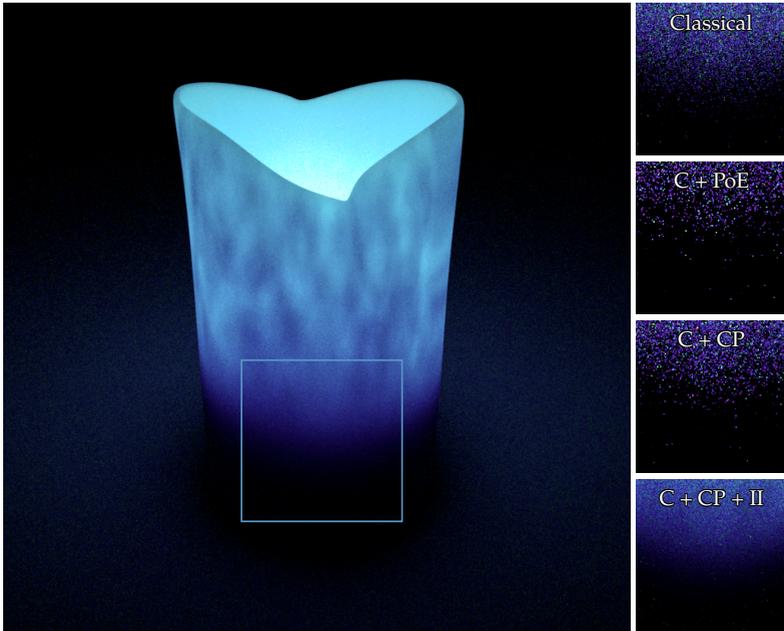


Figure IV.7.: Top: The Candle scene. The enlarged regions show roughly equal-time comparisons at about 4.25 thread hours each. Bottom: Root mean squared error for the full Candle scene and the bottom crop (Meng, Hanika, et al., 2016).

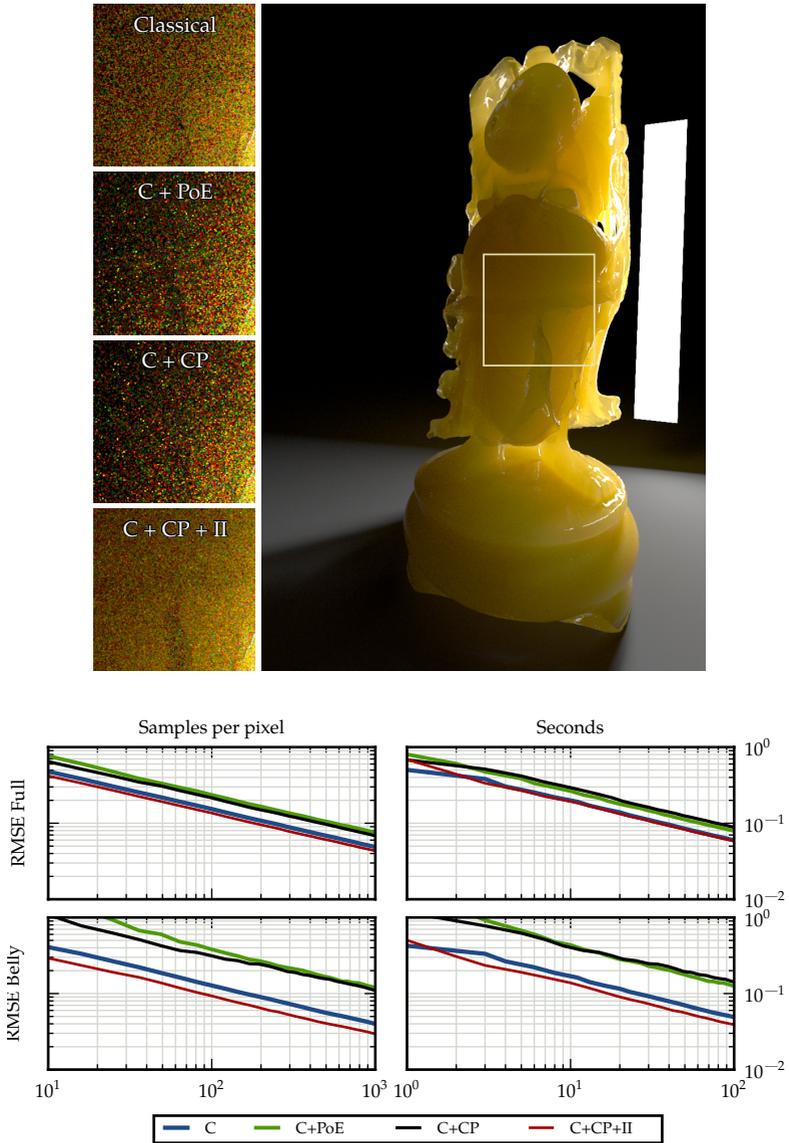


Figure IV.8: Top: The Buddha scene. The enlarged regions show roughly equal-time comparisons at about 4.25 thread hours each. Bottom: Root mean squared error for the full Buddha scene and the bottom crop (Meng, Hanika, et al., 2016).

## 4.4 Analysis

All scenes highlight a typical failure case of *Point of Entry* sampling: large contributions from random walks that cross the medium and do not return to the PoE. This is especially striking in the crops of the Candle and Buddha scenes, where PoE sampling performs far worse than *Classical* sampling. The Ear scene is the only one where PoE performs best in some parts: random walks entering at the cheek must exit there to have measurable contribution. Because of the low curvature, PoE and CP are almost identical, but the PoE can be computed more efficiently.

It is this problem that motivated our work in the first place. Our first proposed technique, *Closest Point* sampling, improves matters slightly. This technique has at least some chance of finding the light paths that PoE actively tries to avoid: If the first vertex inside the medium is closer to the back side than the front, CP will bias towards a back light. Consequently, a slight improvement can be seen in the Ear and Candle scenes. However, CP is no magic bullet. It also tends to favor turning around to exit the way it entered the medium.

Adding *Incident Illumination* sampling helps in scenarios where the illumination is not uniform, and where the light sources happen to not be in the direction of PoE or CP sampling. We find that biasing using II can speed up convergence by an order of magnitude in problematic regions when compared to *Classical* sampling. Compared to PoE biasing, the speed up can be even more dramatic, e.g. in the Candle scene where almost no random walks make it to the inside of the candle.

## 5 Discussion and Future Work

In this chapter, we demonstrated several practical improvements to the Dwivedi sampling technique as introduced by Křivánek and d'Eon (2014).

We proposed replacing the *Point of Entry* biasing technique with a more accurate *Closest Point* sampling, and showed that non-uniform illumination can be handled more robustly with *Incident Illumination* sampling. We also proposed a heuristic approach for combining these biasing strategies with *Classical* sampling using MIS.

Some open questions remain. For example, our current selection criterion only implicitly considers the strength of incident illumination by importance sampling light sources. It is conceivable that using a simple analytic model, such as a diffusion approach, to estimate incident illumination from a given slab direction, could yield further improvements. Using the recent path guiding technique by Vorba et al. (2014) for slab placement seems like an additional, promising direction of future research.

Non-constant phase functions are still difficult to handle with biased sampling. Our solution is to gracefully degrade to *Classical* sampling with increasing anisotropy. However, it would be more elegant, as suggested by Křivánek and d'Eon (2014), to use more exact analytical expressions that incorporate anisotropy directly.

Dwivedi sampling uses the assumption of a homogeneous medium. We showed in the Candle scene that it can also handle heterogeneous media. However, further applications such as clouds warrant future investigation.

# V

## A MULTI-SCALE ALGORITHM FOR RENDERING GRANULAR MATERIALS

### 1 Introduction

The very common class of *granular materials* includes sand, sugar, salt, and even snow. The defining property is the fact that many individual objects, the *grains*, form an aggregate.

In scenes with many millions of grains, it is impractical to have an artist place grains manually. For applications such as visual effects and predictive rendering, physics simulations such as (Ihmsen et al., 2012) are therefore often employed for this task. However, such simulations are costly both in their memory requirements and the processing power needed to complete them. Additionally, they produce data in baked form, and changing parameters or even individual grain geometry may require recomputation, which makes iterations on scene variants tedious. This problem becomes worse the larger scenes get. We therefore propose a procedural modelling algorithm that can generate arbitrarily sized granular aggregates. It has a low memory footprint and allows quick changes to grains and the aggregate structure. Individual grains can be complex, including specular surfaces and scattering interior media. Overall, the algorithm is a great test-bed for our research on light transport in granular materials. An example scene is shown in Figure V.1, which contains about two billion grains in total.

Apart from the geometric modelling problems, granular materials pose particular challenges for photorealistic rendering with Monte Carlo methods (Kajiya, 1986). One reason is the high geometric complexity of granular aggregates: a good way to reduce variance in Monte Carlo rendering can be bidirectional path construction, where sub-paths are created starting from the sensor and an emitter simultaneously, and the algorithm tries to connect the two. This connection, however, will usually fail in granular materials because there is no direct visibility.

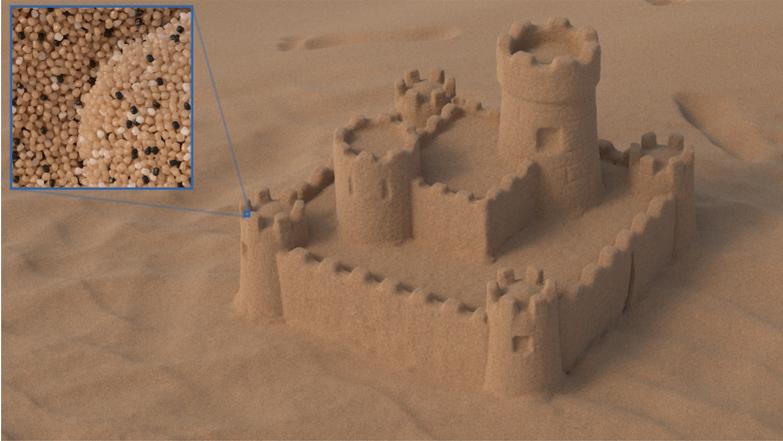


Figure V.1.: A sand castle modelled with our procedural instantiation algorithm. The castle consists of roughly 2 billion grains, and each of the grains is modelled using approximately 200,000 triangles. An inset shows a zoom-in where individual grains are visible (Meng, Papas, et al., 2015).

Another reason is rooted in the material properties of the grains themselves. In materials such as sand, snow, or sugar, the surfaces of grains are often specular or glossy, which also limits the potential for connections. Additionally, while grains often have a scattering interior, some have a very high albedo. This means that long paths that traverse many individual grains may still contribute a lot to the image, and cannot be safely terminated. To alleviate these problems, we propose replacing the grain geometry by a standard stochastic participating medium, parameterized by the single-scattering albedo  $\alpha$  and the mean free path  $\bar{l}$ , in such a way that the medium on average behaves like the grain aggregate. We leverage results on sphere packing, and exploit the structure of our procedural grain model, to derive the medium parameters from simple statistics on the aggregate. To facilitate this derivation, we employ an intermediate path construction technique, which we call *teleportation transport*, that models non-point scattering on grains, as well as transport between grains, in a stochastic fashion.

An important feature of granular materials are coherent refractive patterns and fine surface detail in close-up shots (see the inset in Figure V.1 for an example). By replacing the aggregate with a participating medium, this detail is lost. We therefore propose a multi-scale algorithm, which first uses path tracing with grain geometry, and switches to the participating medium representation automatically after a few grain interactions. This algorithm can achieve speedups of multiple orders of magnitude in some cases, making previously impractical scenes realistic to render.



Figure V.2.: Light scattering in granular materials exhibits coherency, which violates a core assumption of the radiative transfer equation (Meng, Papas, et al., 2015).

## 2 Previous work

### 2.1 Densely packed random media

The large-scale appearance of granular aggregates is reminiscent of participating materials, which scatter and absorb light. However, granular materials violate one of the basic assumptions of radiative transfer: the assumption that the distances between scattering particles are large compared to the particle size. In fact, it is known that the classical RTE only works well for very low packing rates up to roughly  $f = 0.1$  (Randrianalisoa and Baillis, 2010), where the packing rate of a material is defined using the volume of space filled with material  $\mathcal{V}_o$ , and the total volume  $\mathcal{V}$  as

$$f = \frac{\mathcal{V}_o}{\mathcal{V}}. \quad (\text{V.1})$$

Common granular materials such as sugar, sand, and snow often have higher packing rates well above 0.4 (Bubnik et al., 1998; Cuffey and Paterson, 2010; McWhorter and Sunada, 1977). As a result, successive scattering events are not independent in granular materials (Moon, Walter, et al., 2007). This can be seen in aggregates of clear granules, which act like tiny lenses, and cause coherent refraction patterns (see Figure V.2).

In poured random packings of hard spheres, where spheres are dropped into a container, packing rates of roughly 0.6 can be observed, and, if the container is then subjected to vibrations, a theoretical upper bound is  $f \approx 0.634$  (Song et al., 2008).

Since we use sphere packings throughout our techniques, they share this upper

bound. Therefore, our methods do not support densely packed aggregates of highly aspherical grains, such as rice or coins.

## 2.2 Light transport in granular aggregates

Granular aggregates can have tiny individual grains, for which wave effects are not negligible. In applications that require highly accurate results, such as nuclear reactor physics, one may therefore consider the wave nature of electromagnetic radiation explicitly (Durant et al., 2007; Foldy, 1945). Such methods, however, are extremely computationally expensive. In photorealistic image synthesis, wave effects are therefore often ignored altogether, and methods based on geometrical optics such as Monte Carlo path tracing are used instead.

Since path tracing of actual grain geometry is costly, several attempts have been made to approximate granular aggregates with homogeneous stochastic media that follow the radiative transfer equation despite the fact that it cannot represent dependent scattering in granular materials properly. These methods tend to assume that the material is statistically invariant to translations and rotations. A recent publication showed an example of how such methods can be generalized to heterogeneous granular mixtures by storing medium parameters in a spatial data structure (Müller et al., 2016).

Randrianalisoa and Baillis (2010) use a Monte Carlo simulation in the full aggregate to obtain effective parameters for the RTE. Such parameters can then be used in standard volumetric path tracing (Kajiya, 1986; Rushmeier, 1988) to render images. Our method follows the same basic idea, but avoids a costly precomputation on the full granular material.

Moon, Walter, et al. (2007) also perform a Monte Carlo simulation on the full medium, where they precompute functions that describe the exitant radiation on spherical shells of varying size due to incident radiation at the sphere center. The precomputed functions are also called *shell transfer functions*. During rendering, they always pick the largest possible precomputed shell to sample a new position and direction. This enables them to traverse the aggregate in large step, and to leave it quickly.

Singh and Kaviany (1992) perform a precomputation for a single grain that accounts for non-point scattering at grains, and use it in a modified discrete ordinates method (Chandrasekhar, 1960). However, they are restricted to dielectric spheres as their grain primitives. Our method performs a similar per-grain precomputation to obtain a non-point scattering distribution function, but allows for arbitrary geometry, surface properties, and internal scattering media.

Donovan et al. (2003) implement a procedure where free paths are sampled from a statistical distribution, and grains are then instanced at the collision site. We introduce a very similar two-step sampling framework, which uses precomputed non-point scattering distribution functions to avoid instantiating grain geometry, but use an analytic formula for the distribution of free paths.

It is possible to employ diffusion methods to approximate high-order scattering in granular materials extremely quickly (d'Eon, 2013; Meng, Papas, et al., 2015; Papas, 2015), but such methods are subject to considerable visible bias. We will not consider them in this dissertation.

We focus on granular materials in this chapter, but similar ideas have been applied to cloth rendering. In cloth, individual yarns scatter light (Marschner et al., 2003), but they also form compound shapes and exhibit large-scale scattering behaviour (Moon and Marschner, 2006; Schröder, Zhao, et al., 2012). Like Schröder, Klein, et al. (2011), we fit parameters for a stochastic medium to an aggregate of discrete scatterers, avoiding the need to consider individual particles for all bounces. This allows us to develop a more efficient light transport algorithm. We also use a procedural modelling technique to generate granular materials on the fly, an idea that has been explored for yarn structures, as well (Schröder, 2013).

### **3 A procedural model for granular materials**

Our procedural modelling algorithm for granular materials is based on two principles. Like Peytavie et al. (2009), who generate packings of irregular pebbles and rocks, we use an infinite grid of pre-packed blocks to control the topology of the granular aggregate. If a bounded aggregate is desired, we also a water-tight boundary mesh to carve shapes out of the infinite granular material.

However, we decouple topology and grain geometry. Our pre-packed blocks contain spheres, and we instance complex grain exemplar geometry into these bounding spheres during raytracing. This has multiple benefits. First, the decoupling allows for quick changes to grain geometry, grain materials, grain orientation, and mixing ratios if the aggregate is composed of multiple grain types. Second, we can exploit the knowledge that the aggregate is composed of grains whose bounding spheres do not overlap to accelerate light transport considerably. This will become clear in the later sections of this chapter, where we use results from sphere packing literature to derive participating media parameters.

#### **3.1 Packed bounding spheres**

Our algorithm employs sphere packings, where each sphere is understood as the bounding sphere of an individual grain. Grain geometry can be instantiated into a bounding sphere at run time, and so different granular materials can share the same topological structure by sharing the sphere packings.

Each sphere packing block must be infinitely tileable, so that it can be used to fill three-dimensional space seamlessly. In principle, an aperiodic set of blocks such as Wang cubes (Culik II and Kari, 1996) can be used to break up repetition artifacts, but we found in our tests that such artifacts are only visible specially constructed scenes, such as flat, axis-aligned blocks of granular material. This can be attributed to the additional per-grain randomization we apply, as well as the fact that we use irregular boundary meshes to carve shapes out of the infinite space of grains. See

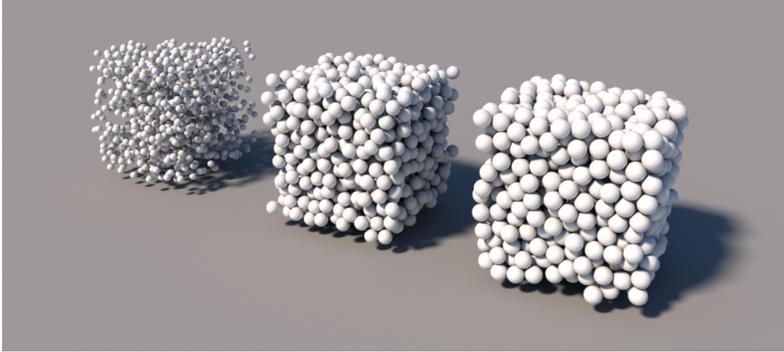


Figure V.3.: Infinitely tileable blocks of packed spheres with varying packing rate, as they are used in our algorithm. Each block contains roughly 1000 spheres.

Figure V.17 for a scene containing a cuboid aggregate that is slightly rotated off axis to hide repetition artifacts.

Instead, we use the algorithm of Skoge et al. (2006) to pack spheres into tileable blocks. This algorithm can be understood as compressing an initially loose packing until a desired packing rate is attained.

### 3.2 Ray intersection

Our method works by instantiating grain geometry at run-time, while a ray  $(\mathbf{x}, \boldsymbol{\omega})$  is being intersected with the granular aggregate.

First, three-dimensional space is conceptually divided into an infinite regular grid. Each grid cell matches the precomputed block of bounding spheres in size. It is possible to find the three-dimensional index of the cell containing the current ray position  $\mathbf{x}$  using simple arithmetic.

The block of bounding spheres is then instantiated into the current grid cell and intersected. We found it useful to accelerate the ray-intersection test using a bounding volume hierarchy (MacDonald and K. S. Booth, 1990; Wald, 2007) or even a regular grid. It is paramount to use a numerically stable ray-sphere intersection test in this stage, since individual grains tend to be tiny compared to the scene size.

If a bounding sphere intersection is found, then the three-dimensional grid cell index  $(i_x, i_y, i_z)$  together with the bounding sphere index in the cell  $i_b$  are used to create a random seed for per-grain randomization using the simple hash function (Perlin, 2002)

$$s(i_x, i_y, i_z, i_b) = \text{perm}(i_x + \text{perm}(i_y + \text{perm}(i_z + \text{perm}(i_b))))), \quad (\text{V.2})$$

where  $\text{perm}(j) = \mathcal{P}[j \bmod |\mathcal{P}|]$  accesses a permutation table of size  $|\mathcal{P}|$ .



Figure V.4.: Grain exemplars we use for instantiation. The grains all have highly complex specular surfaces with roughly 200.000 triangles each. Apart from the flour grain, they also have scattering and absorbing interiors, though this is hardly visible for the snow and sugar grains (Meng, Papas, et al., 2015).

It is important to note here that while conceptually, this algorithm can yield aggregates where no two grains are truly alike, the cell index, random seed, and random state must be stored in a finite number of bits. The aggregate will therefore at some point repeat. We use 64 bit integer types, and have not been able to find any adverse effect in simulations with unbounded materials.

The random seed is used to initialize a random number generator such as the Mersenne twister (Matsumoto and Nishimura, 1998). This generator can in turn be used to select one of multiple grain exemplars, and randomized transformation matrices, for instantiation.

The ray is now transformed into the local space of the grain exemplar, and the grain instance is intersected using a standard raytracing kernel. Due to the high geometric complexity of our grain exemplars, we again employ acceleration structures to be able to perform the ray-grain intersection efficiently.

Note that our randomization process is completely deterministic, and so is the resulting granular material.

If the ray does not intersect the grain instance, then the algorithm attempts to find the next bounding sphere intersection. If none is found, then the algorithm proceeds by testing the next grid cell using a simple three-dimensional DDA algorithm (Fujimoto et al., 1986).

To model bounded granular aggregates, the user may provide a water-tight boundary mesh. We voxelize this mesh, storing in each voxel a two-bit state variable that can assume one of the values *fully inside*, *partially inside*, and *outside*. Before the first

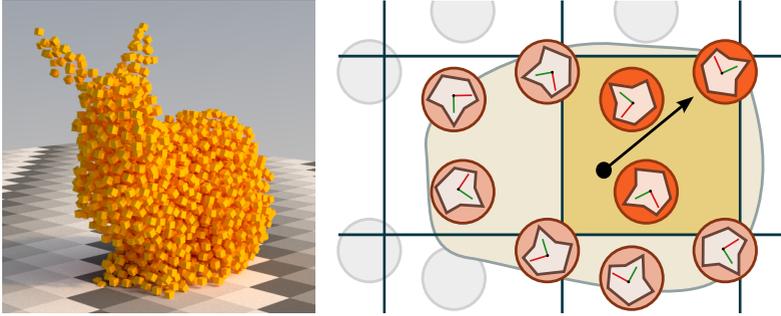


Figure V.5.: To intersect procedural granular aggregates such as this bunny of cubical grains (left), we consider all bounding spheres, from an infinitely tiled space, that reside inside a given aggregate boundary mesh. We perform intersection tests with randomly rotated, instanced grain geometry (right) inside each bounding sphere (Meng, Papas, et al., 2015).

intersection test, the voxel grid is used to skip over empty space, so that grains in bounding spheres that are in *fully outside* cells are never instantiated. For bounding spheres in *partially inside* cells, we trace a ray from the sphere center towards the boundary mesh, and only instantiate the grain if the intersected surface’s normal indicates that the bounding sphere center is inside the boundary mesh. This effectively carves the shape of the boundary mesh out of an infinite space filled with grains. We show pseudo-code for the procedural instantiation method in Algorithm 1.

## 4 Teleportation transport

In this section, we develop a path construction algorithm for granular materials. Like volumetric path tracing, it alternately samples a free path to the next collision site, and then an exitant direction *and position*. We call the non-point scattering spatial shift in the second sampling step “teleportation”, and dub the whole algorithm *teleportation transport*.

The teleportation transport model has the great advantage that it decouples inter-grain transport and intra-grain transport, and therefore, the aggregate topology and optical properties of grains.

We also exploit the fact that grains in our granular materials have non-intersecting bounding spheres, which is a direct result of our procedural model (see Section V.3). Note that one can also enforce this property in datasets created in physical simulation packages.

The two basic components of teleportation transport are a free-path distribution  $p_b(t)$ , from which we sample distances, and an angular-spatial scattering distribution, which maps incident directions and positions on a grain bounding sphere to exitant

directions and positions.

#### 4.1 Free-path distribution

The first step in every iteration of teleportation transport is to sample a free path to the next collision with a grain. Mathematically, this is done using a probability distribution  $p_b(t)$ .

Instead of running Monte Carlo simulations to obtain  $p_b(t)$ , we use a result from Dixmier (1978). He found that in packings of hard spheres, the distribution can be approximated by the simple exponential

$$p_b(t) = \mu_b e^{-\mu_b t}, \quad \text{with} \quad \mu_b = \frac{3}{4R} \frac{f}{1-f}. \quad (\text{V.3})$$

Surprisingly, this formula only depends on the radius  $R$  of the spheres and the packing rate  $f$ . The value  $\bar{l}_b = 1/\mu_b$  can be interpreted as the average distance between two bounding spheres. Various authors have confirmed that this formula works remarkably well, although it exhibits approximation error for small distances of less than 1 grain radius (Donovan et al., 2003; Levitz, 1993; Olson et al., 2006; Torquato and Lu, 1993). We confirm this in Figure V.6, but in our experiments, we found no significant visual impact of this approximation error. We therefore opted to prefer the analytic formula over a more accurate tabulated distribution.

The two parameters,  $R$  and  $f$ , are readily obtained from the tile of packed bounding

---

#### Algorithm 1 Procedural Grain Instantiation (Meng, Papas, et al., 2015)

---

```

1: function INTERSECTGRAINS(Ray r, Float tmin, Float tmax)
2:   while tmin < tmax do
3:     iv = VoxelIndex(r, tmin)
4:     if PartiallyInside(iv) then
5:       ov = VoxelOrigin(i)
6:       o = Origin(r) - ov
7:       if IntersectTile(o, Direction(r), tmin, tmax) then
8:         p = BoundingSphereCenter() + ov
9:         if SurelyInside(iv) or PointInside(p) then
10:          if IntersectInstance(r, tmin, tmax, p) then
11:            return True
12:          end if
13:        end if
14:        tmin = BoundingSphereExit()
15:      end if
16:    end if
17:    tmin = FindNextVoxelT(r, tmin)
18:  end while
19:  return False
20: end function

```

---

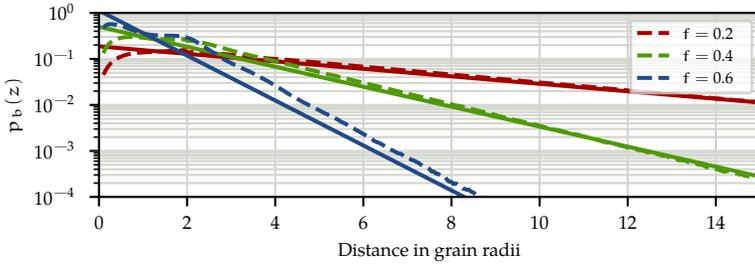


Figure V.6.: Free-flight distribution for different packing rates, computed using Monte Carlo simulations (dashed) and using the analytic model by Dixmier (1978, solid) (Meng, Papas, et al., 2015).

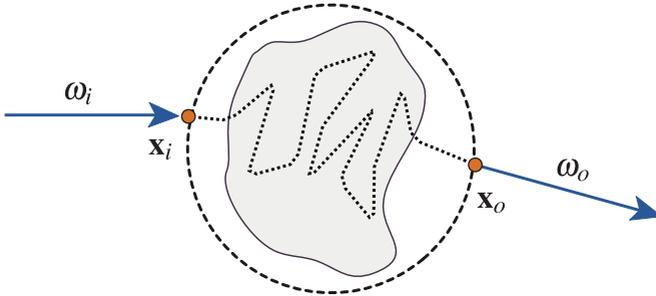


Figure V.7.: The setup we use for computing our TSDF. Photons are emitted towards a grain and its bounding sphere. The TSDF is tabulated as a function of incident direction  $\omega_i$ , incident position  $x_i$ , exitant direction  $\omega_o$ , and exitant position  $x_o$  (Meng, Papas, et al., 2015).

spheres we use in our procedural instantiation method. In fact, the algorithm we use to generate tileable sphere packings (Skoge et al., 2006) outputs the packing rate as the termination criterion. Again it is important to note that packing rate and sphere radius can also be computed efficiently for granular data sets obtained through other means.

## 4.2 Teleportation scattering distribution function

The second step in teleportation transport is to sample an exitant direction and position given that the path has just hit a grain bounding sphere.

We use a precomputed function to do this. This function, the *teleportation scattering distribution function* (TSDF), encodes both the angular and the spatial offset due to scat-

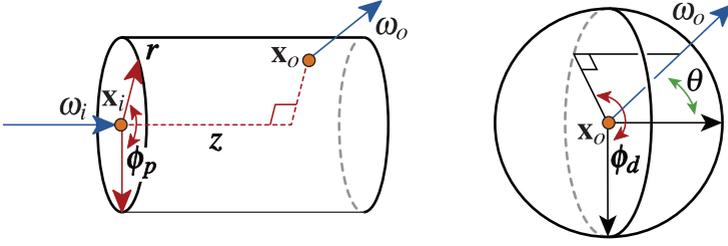


Figure V.8.: The cylindrical and spherical coordinate systems used to represent the outgoing position and direction in our TDSF (Meng, Papas, et al., 2015).

tering on a grain. The TDSF is conceptually similar to a BSSRDF (Jensen, Marschner, et al., 2001) used for sub-surface scattering, and a BFSDF (Zinke and Weber, 2007) used to model scattering in fibers.

We determine the TDSF by tabulating photons from a Monte Carlo simulation. Initially carrying unit energy, the photons are randomly spawned and shot towards a grain and its bounding sphere. We randomize both the incident direction and position. A standard volumetric path tracer is used to determine the path of each photon. Absorption is accounted for by reducing the photon energy. Whenever a photon exits the grain bounding sphere again, its energy is stored in a table.

In general, this table is eight-dimensional, mapping incident position  $\mathbf{x}_i$  and direction  $\boldsymbol{\omega}_i$  to exitant position  $\mathbf{x}_o$  and  $\boldsymbol{\omega}_o$  on the bounding sphere. We found it convenient to represent these values in cylindrical and spherical coordinate systems centered around the incident ray, where  $\mathbf{x}_o = (r, \phi_p, z)$  and  $\boldsymbol{\omega}_o = (\theta, \phi_d)$  such that  $\cos \theta = \langle \boldsymbol{\omega}_i, \boldsymbol{\omega}_o \rangle$ . Since the incident direction is randomized, the rotation around the incident ray is arbitrary, and so it suffices to consider the difference angle  $\phi_\Delta = \phi_p - \phi_d$ . In total, the TDSF becomes  $p(\theta, \phi_\Delta, r, z)$  (see Figure V.8 for an illustration).

We additionally found that for our grains, the dimensions  $\phi_\Delta$ ,  $r$ , and  $z$  tend to be uncorrelated. For this reason, instead of tabulating the full four-dimensional function, we tabulate a one-dimensional probability density  $p(\theta)$  and three two-dimensional conditional probability densities  $p(\phi_\Delta|\theta)$ ,  $p(r|\theta)$ , and  $p(z|\theta)$ :

$$p(\theta, \phi_\Delta, r, z) \approx p(\theta) \cdot p(\phi_\Delta|\theta) \cdot p(r|\theta) \cdot p(z|\theta). \quad (\text{V.4})$$

The resulting function is an approximate statistical descriptor of the grain, and can be used directly to sample exitant positions and directions due to scattering on the grain without even instantiating it. We visualize the TDSF for a dielectric sphere and a snow grain in Figure V.9.

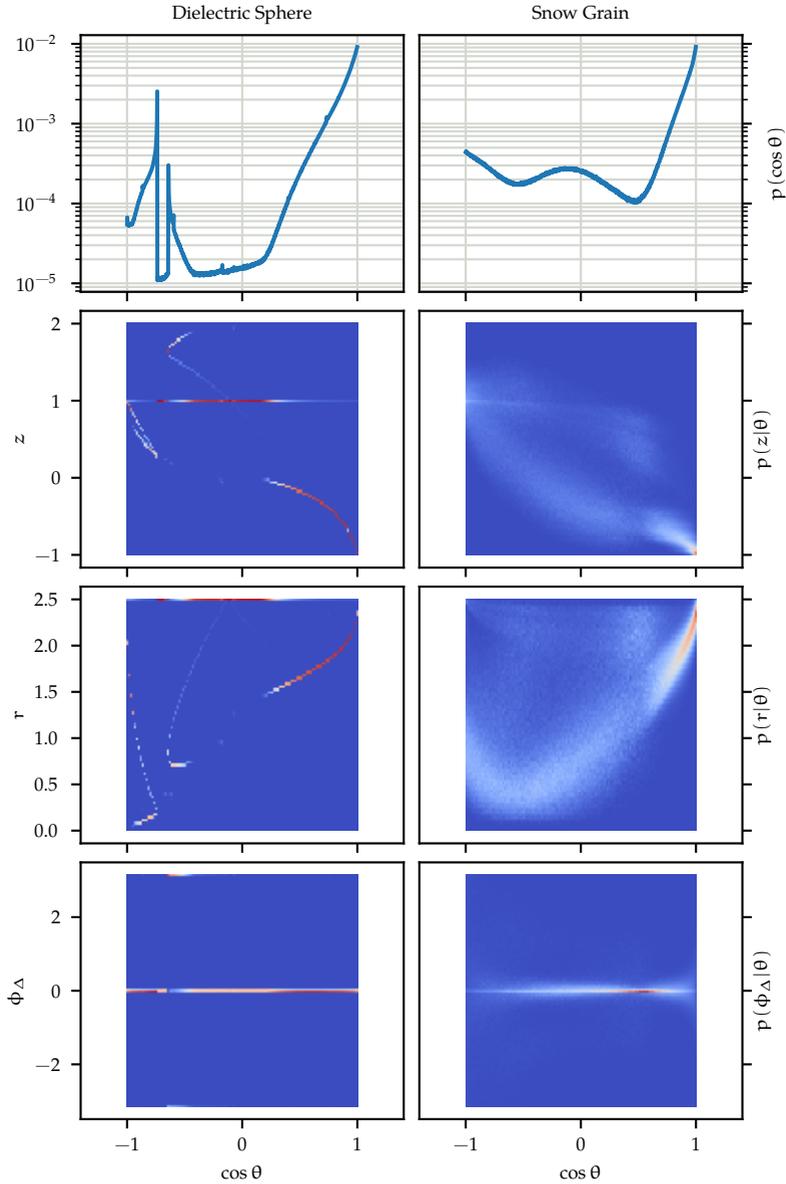


Figure V.9.: The TSDF probability density functions for a dielectric sphere with index of refraction  $\eta = 1.33$  (left) and our snow grain (right). Distances are given in units of grain radius, angles in units of radians (Meng, Papas, et al., 2015). The function  $p(\cos \theta)$  for the dielectric sphere can also be found in (Zhou et al., 2003, Figure 8).

### 4.3 Delta forward scattering

With grains that do not fill their bounding spheres completely, there is always some probability  $1 - \beta$  that a ray hitting the bounding sphere does not hit the grain inside. This results in a delta peak at  $\theta = 0$ , which is hard to properly represent in a discretized table.

We therefore separate it from the rest of the function by storing an estimate of  $\beta$ , the *hit probability* during our precomputation by counting all rays that never hit the grain instance. Together with the hit probability, we store the mean spatial offset of unscattered rays as  $\bar{l}_\delta$ .

### 4.4 Grain albedo

If  $N$  photons are used in the precomputation, then a total energy of  $Q_i = N$  is shot at the grain bounding sphere. A fraction  $(1 - \beta)Q_i$  will escape unscattered, but the part  $\beta Q_i$  experiences grain interaction and therefore possibly absorption. This means that the total exitant  $Q_o$  can be less than  $N$ . We account for this by storing the *grain albedo*

$$\alpha = \frac{Q_o}{Q_i}. \quad (\text{V.5})$$

### 4.5 Algorithm

Teleportation transport can be implemented as follows: First, sample a distance from Equation (V.3). Then, sample  $\beta$  to see if a grain interaction occurs. If this is the case, multiply the path throughput with  $\alpha$  to account for absorption. Then, sample an exitant direction using  $p(\theta)$  and  $p(\phi_\Delta|\theta)$ , and sample an exitant position using  $p(r|\theta)$  and  $p(z|\theta)$ . We call this implementation teleportation path tracing (TPT). Pseudocode for sampling the TSDF is shown in Algorithm 2.

This algorithm is very similar to volumetric path tracing, and in fact, it avoids all ray intersection tests with grain geometry. Although this is a measurable boost in efficiency, the algorithm does not enable the use of next event estimation and similar bidirectional techniques. The reason for this is the lack of a proper transmittance

---

#### Algorithm 2 Sampling $\omega_o$ and $\mathbf{x}_o$ in TPT (Meng, Papas, et al., 2015)

---

```

1: function SAMPLETSDF( $\omega_i$ )
2:    $\phi_d = \text{SampleUniform}(0, 2\pi)$ 
3:    $\theta = \text{Sample}(p_\theta)$ 
4:    $r = \text{Sample}(p_r)$ 
5:    $z = \text{Sample}(p_z)$ 
6:    $\phi_\Delta = \text{Sample}(p_{\phi_\Delta})$ 
7:    $\omega_o = R_{\omega_i}(\phi_d) (\cos(\theta)\omega_i + \sin(\theta)\omega_i^\perp)$ 
8:    $\mathbf{x}_o = R_{\omega_i}(\phi_d + \phi_\Delta) (z\omega_i + r\omega_i^\perp)$ 
9: end function

```

---

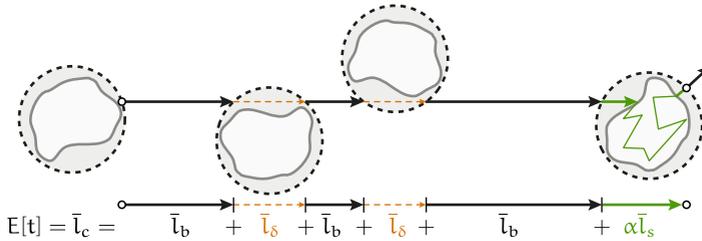


Figure V.10.: The combined expected free-flight distance between grain interactions  $\bar{l}_c$  is a potentially infinite sum  $\bar{l}_\beta$  of sphere-to-sphere distances  $\bar{l}_b$  and delta scattering distances  $\bar{l}_\delta$  due to the probability  $(1 - \beta)$  of passing through spheres unobstructed, followed by a scattering teleportation distance  $\alpha \bar{l}_s$  (Meng, Papas, et al., 2015).

function: even though a free-path distribution is used in teleportation transport, some paths may interact with many bounding spheres but never with an actual grain if  $\beta$  is low. Additionally, the teleportation step may cause paths to move more quickly, effectively decreasing the density of the medium.

Teleportation path tracing is also problematic because boundaries need special handling: a non-point scattering event may cause a path to exit the medium, or to move into a solid object. This leads to light leaking and complicates tracking the current medium. In a prototypical implementation, we find geometry intersections on the line  $\mathbf{x}_i - \mathbf{x}_o$ , and reset the ray to the position of the first intersection.

## 5 RTE parameters for granular materials

Despite the implementation problems associated with the teleportation transport model, it is a valuable tool for deriving parameters for a standard participating medium that obeys the RTE and exhibits no non-point scattering.

These parameters are the single-scattering albedo  $\alpha = \mu_s / \mu_t$ , the phase function  $\phi$ , and the mean free path  $\bar{l}$ .

For the first parameter, the single-scattering albedo  $\alpha$ , we simply re-use the grain albedo that results directly from our TSDF precomputation.

To obtain a phase function, we would like to extract the directional part of the TSDF. We could simply try to fit a parametric model such as the Henyey-Greenstein phase function (Henyey and Greenstein, 1941), for example by matching the mean cosine  $g$ . However, we found that this results in a significant loss of visual fidelity, since the directional part of the TSDF is not modeled well by a single Henyey-Greenstein lobe. We therefore decided to use the tabulated probability density  $p(\theta)$  from our TSDF data directly.

## 5.1 Combined free-flight distribution

The third parameter in a participating medium is the mean free path  $\bar{l}$ . It is simply the expected value of the exponential free-flight distribution. Research has shown that in materials such as granular aggregates, where successive scattering events are dependent, the free-flight distribution is not exponential (d'Eon, 2013; Torquato, 2001).

However, in order to exploit variance reduction techniques known in standard volumetric path tracing, we must find an exponential free-flight distribution that matches the appearance of the non-exponential material with sufficient fidelity. We therefore attempt to find the mean-free path  $\bar{l}_c = E[t]$  of a stochastic granular material as described by the teleportation transport model. This mean free path combines both normal spatial propagation and non-point scattering effects.

This derivation is one of the core results of our research. It presents a connection of non-classical light transport in granular materials and classical light transport in participating media with exponential free flights, albeit an approximate one that is only accurate on average.

We start by analyzing the different path segments that can occur on a random walk in a teleportation transport medium. Figure V.10 shows these segments for a path that traverses two grain bounding spheres unscattered, and interacts with the grain in the third bounding sphere.

Distances between bounding spheres are, on average,  $\bar{l}_b$ , as dictated by our free-flight distribution from Equation (V.6). Each bounding sphere interaction in which no scattering takes place adds another offset of  $\bar{l}_\delta$ , which we know from our TSDF precomputation. Finally, once a scattering event occurs, the teleportation will on average cause an offset of  $(1 - \alpha) \cdot 0 + \alpha \cdot \bar{l}_s$ , because a fraction of  $(1 - \alpha)$  of all rays are absorbed and thus never teleported. The value  $\bar{l}_s$  is the mean spatial offset, and it can also be obtained directly from the TSDF. We use a scalar value here instead of a three-dimensional offset vector because grains are randomly oriented and the mean teleport vector is therefore aligned with the ray direction. In Table V.1 we present the scalar statistics for the grains used in our results. The phase functions for dielectric spheres and our snow grains can be seen in Figure V.9.

Table V.1.: Grain parameters extracted from TSDF precomputations for the grains used in our results. Values are averaged over all wavelengths and rounded to two digits (Meng, Papas, et al., 2015)

	Dielectric Spheres	Flour	Pink Salt	Sand	Snow	Cane Sugar	Brown Sugar
$\bar{l}_\delta$	0.00	0.87	0.97	0.85	0.86	0.77	0.77
$\bar{l}_s$	1.44	0.18	1.00	0.50	1.40	1.17	0.88
$\alpha$	1.00	1.00	0.99	0.68	1.00	1.00	0.76
$\beta$	1.00	0.66	0.53	0.63	0.65	0.71	0.71

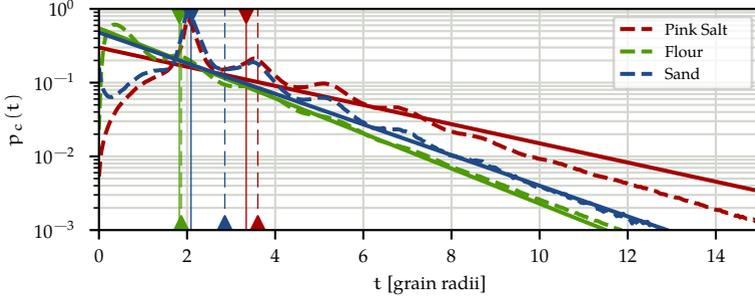


Figure V.11.: We compare our combined exponential free-flight distribution  $p_c(t)$  (solid, see Equations (V.10) and (V.12)) for three of our granular materials against Monte Carlo measurements (dashed). The packing rate is  $f = 0.63$  for all grain types. The hit probability is  $\beta = 0.53$  (Pink Salt),  $\beta = 0.66$  (Flour), and  $\beta = 0.63$  (Sand). The mean free paths are indicated by vertical lines (Meng, Papas, et al., 2015).

In total, for  $N$  unscattered events and one scattered event, the distance between the original position of the ray and the next exitant position on a bounding sphere is

$$\bar{l}_N = \alpha \bar{l}_s + N \cdot (\bar{l}_b + \bar{l}_\delta) + \bar{l}_b. \quad (V.6)$$

The probability for  $N \geq 0$  unscattered events followed by one scattered event is  $P_N = (1 - \beta)^N \beta$ , and so the mean free path can be written as

$$\begin{aligned} \bar{l}_c = E[t] &= \sum_{i=0}^{\infty} P_N \bar{l}_i \\ &= \sum_{i=0}^{\infty} (1 - \beta)^i \beta (\alpha \bar{l}_s + \bar{l}_b + i(\bar{l}_b + \bar{l}_\delta)) \\ &= \beta (\alpha \bar{l}_s + \bar{l}_b) \sum_{i=0}^{\infty} (1 - \beta)^i + \beta (\bar{l}_b + \bar{l}_\delta) \sum_{i=0}^{\infty} i(1 - \beta)^i. \end{aligned} \quad (V.7)$$

Since  $0 < \beta \leq 1$  for non-singular grains, the following equalities hold

$$\sum_{i=0}^{\infty} (1 - \beta)^i = \frac{1}{1 - \beta} \quad (V.8)$$

$$\sum_{i=0}^{\infty} i \cdot (1 - \beta)^i = \frac{1 - \beta}{\beta^2}, \quad (V.9)$$

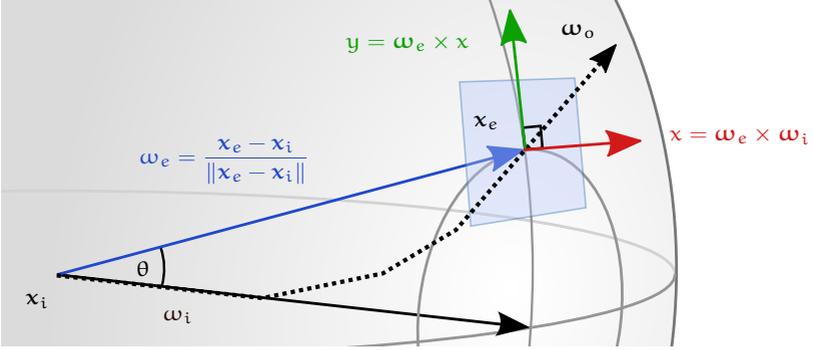


Figure V.12.: The setup we use for the comparison in Figures V.13 and V.14. We use spherical detectors of radius  $10R$  and  $20R$ , where  $R$  is the grain radius. This allows us to separately analyze short-range and long-range scattering. We used  $4 \cdot 10^8$  samples for our simulations. (Meng, Papas, et al., 2015).

and we can simplify the infinite sums to obtain the mean free path

$$\bar{l}_c = \alpha \bar{l}_s + \bar{l}_b + (\bar{l}_b + \bar{l}_\delta) \frac{1 - \beta}{\beta}. \quad (\text{V.10})$$

For spherical grains that fill their bounding sphere completely,  $\beta = 1$  and the formula collapses to

$$\bar{l}_{c,0} = \alpha \bar{l}_s + \bar{l}_b, \quad (\text{V.11})$$

which corresponds to scattering on the grain in the first bounding sphere a ray encounters. With  $\mu_c = 1/\bar{l}_c$ , the combined free-flight distribution is

$$p_c(t) = \mu_c \exp(-\mu_c t). \quad (\text{V.12})$$

We compare this analytic approximation to Monte Carlo measurements for three exemplary granular materials in Figure V.11.

For mixtures of grains, such as the sand material, we compute effective RTE parameters using a weighted average with weights corresponding to the mixing ratios.

## 5.2 Validation

To validate our derivations, we performed Monte Carlo simulations to obtain estimates of the radiance and irradiance fields on spherical shells in granular materials with explicit path tracing and our approximate volumetric representation.

Our simulation setup is as follows. Photons are spawned at random locations and with random directions inside an infinite medium. We are able to create infinite

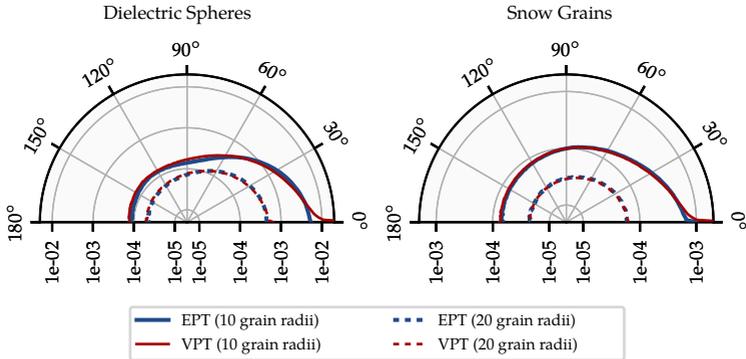


Figure V.13: Irradiance on spherical shells due to a unit beam source in an infinite granular material. The angular variable is the inclination  $\theta$  (Meng, Papas, et al., 2015).

granular aggregates because of our procedural instantiation scheme. The photons are then traced until they have experienced interactions with ten grains. For each photon, we center a spherical shell detector of a fixed radius around the photon origin. Whenever a photon passes the shell, it is accounted for in a histogram. We use shells of 10 and 20 grain radii to compare short-range and long-range effects.

Figure V.12 shows the simulation setup. A photon is traced from a beam source at location  $\mathbf{x}_i$  with direction  $\omega_i$ . The photon follows a path, which is shown as a dotted line, until it exits the detector sphere at  $\mathbf{x}_e$ . The exitant direction  $\omega_o$  is then projected onto the blue tangent plane with normal  $\omega_e$  to obtain the projected position  $(x, y)$ . Assuming invariance with respect to rotation around the initial photon direction, we bin based on the coordinate triple  $(\theta, x, y)$ , where  $\theta$  is the inclination angle.

In Figure V.13, we visualize the irradiance at our spherical shell detectors over the inclination  $\theta$ . Explicit path tracing (EPT) and our approximate volumetric method (VPT) agree well, although it is important to note that forward scattering is overestimated for short distances.

In Figure V.14, we plot the full radiance field for varying inclination angles. Again, VPT agrees well with the EPT results but overestimates forward scattering. The data displayed here is essentially the information stored in the shell transfer functions by Moon, Walter, et al. (2007). It would therefore be possible to accelerate the precomputation of shell transfer functions with our technique, or use shell transfer functions to accelerate our technique further. In fact, since our original paper (Meng, Papas, et al., 2015) was published, a combination of STF and our technique was utilized by Müller et al. (2016).

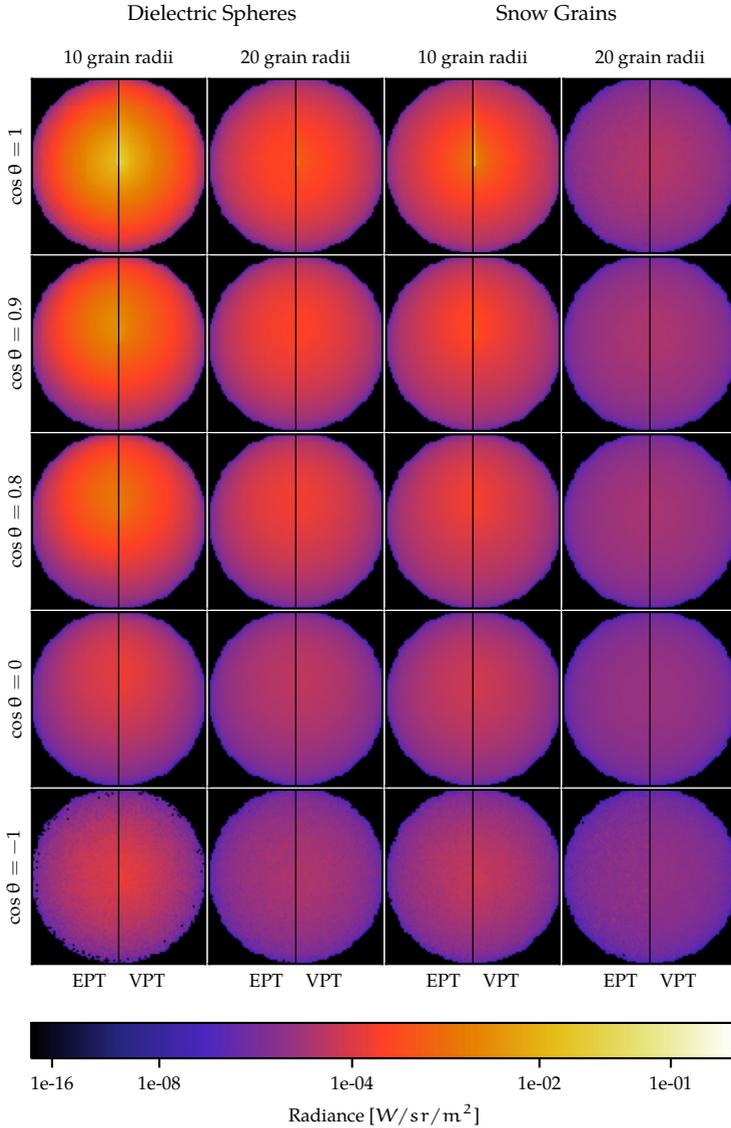


Figure V.14.: Radiance on spherical shells due to a unit beam source in infinite granular materials. We show plots for varying inclination angles  $\theta$  (Meng, Papas, et al., 2015).



Figure V.15.: Left: Explicit path tracing produces accurate low- and high-order scattering as well as fine surface detail, but takes a long time to render. Center: Volumetric path tracing is much faster and captures the overall appearance well, but cannot provide surface detail or coherent refraction patterns. Right: Our multi-stage algorithm first traces rays against grain geometry and then switches to the volumetric representation to approximate high-order scattering. (Meng, Papas, et al., 2015)

## 6 A multi-scale algorithm for rendering granular materials

We have just shown that radiance and irradiance fields are in good agreement between EPT and our VPT method, which is significantly more efficient to compute. However, it only provides good average behaviour. Fine surface detail showing individual grains cannot be captured because there is no spatially coherent scattering in our purely stochastic VPT approach. This problem can be seen in Figure V.15, where individual glass beads are visible in the EPT image (left) upon close inspection, but the sphere surface is completely smooth in the VPT image (right).

Like Li et al. (2005) and Moon, Walter, et al. (2007), we therefore use a hybrid method that transitions between rendering techniques on the fly while performing random walks. This method spawns primary rays in EPT mode, intersecting individual grains and computing full light transport within and between them. This results in a faithful representation of surface detail and low-order scattering. After a while, the path is switched to VPT mode in order to approximate high-order scattering, and no more grain geometry needs to be intersected.

This leads to a highly efficient algorithm that is able to exploit common noise reduction techniques such as next event estimation in the stochastic medium, reducing rendering times significantly, but still retains fine surface detail and coherent refraction patterns. In Figure V.15 (right), we show the application of this technique to our sphere of glass beads scene.

### 6.1 Automatic switching criterion

In order to evaluate the need to use EPT on a path, we are interested in how coherent the current beam of light still is at every grain interaction. We analyze this with a geometric criterion inspired by ray differentials (Igehy, 1999). However, instead of

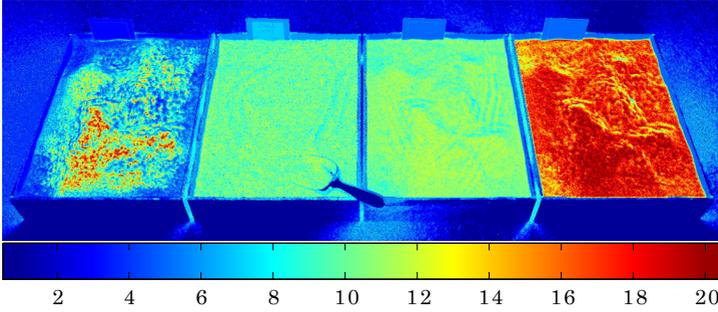


Figure V.16.: Color-coded switch depth in the spices scene (see Figure V.18). The materials are flour (left), pink salt (center left), raw sugar (center right), and raw sugar (right). The diffuse grains in the flour cause an early switch, apart from a few pixels where the rays get trapped in cavities. The raw sugar grains exhibit almost no absorption or scattering, which causes long, coherent light paths. The switch therefore occurs relatively late in this material.

transporting explicit differentials along an individual path, we opt to instead trace packets of  $N = 16$  paths simultaneously, where paths in each packet originate in the same sensor pixel.

Initially, the paths share similar directions and probably hit similar geometry in the scene. However, after a few bounces, the path packet tends to diverge. Among other reasons, this may happen because of absorption, difference in surface normals, difference in materials, and the inherently randomized nature of path tracing.

After  $k$  bounces, if  $N_k$  paths remain alive in the packet, we compute the standard deviation of their vertex positions  $(\mathbf{x}_k^1, \mathbf{x}_k^2, \dots, \mathbf{x}_k^{N_k})$ ,  $\sigma_k$ . If this standard deviation is

$$\sigma > m \cdot R \cdot \frac{N_k}{N}, \quad (\text{V.13})$$

then we deem the packet to be sufficiently incoherent for a switch to VPT. In Equation (V.13), the parameter  $m$  is a user-specified multiplier for the grain radius  $R$ , and the factor  $N_k/N$  enables smaller packets to switch more quickly.

Intuitively, this criterion rates packets as incoherent if their vertices are so far apart that the paths are likely to hit different grains. In our results, we use  $m = 4$ .

If  $\sigma_k$  exceeds the threshold, then all remaining paths switch to VPT upon hitting the next grain. This need not happen simultaneously for all paths in the packet.

If only one path remains in a packet, then the incoherency criterion cannot be evaluated anymore because the standard deviation isn't well defined. We then choose to immediately switch the path to VPT at the next grain interaction.

This automatic criterion enables us to render granular materials at vastly different

scales. In close-up shots, where individual grains are clearly visible, explicit path tracing is used for a relatively long time. However, if a granular material is seen from afar, then the technique switches to volumetric path tracing far more quickly. We demonstrate the criterion on our spices scene in Figure V.16.

## 7 Results

We implemented our rendering method as a new integrator type in Mitsuba (Jakob, 2010). All results were rendered on a massively parallel cluster where each node contained two twelve-core Intel Xeon E5-2697v2 processors running at 2.7 GHz each, and 64GB of random access memory. All render times are reported core hours, that is, the wall-clock time multiplied with the number of processor cores allocated to the job. For parallel scheduling, we rendered many 32 bit floating-point images with low sample counts and averaged them afterwards. All images are also tonemapped with an sRGB gamma curve for display.

### 7.1 Time to unit variance

To compare our techniques in a meaningful fashion, we would like to report the time required to reach some fixed quality level. In terms of Monte Carlo estimates, this quality level is usually the variance with respect to a converged reference image. Despite our access to a massively parallel compute cluster obtaining truly noise-free reference solutions from path tracing was prohibitively expensive. Additionally, our approximate techniques do not converge to the unbiased result, and so comparisons with respect to a common reference image have little informational value. This is visible in Figure V.17, where the MSE curves approach some nonzero value. We therefore use, for each technique separately, an online estimator for the residual variance after  $N$  samples. Since our compute cluster outputs  $M$  independent images of  $N$  samples per pixel each, we can compute an online estimate of variance for each of the  $M$  images, and average them to obtain the variance after  $N$  samples:

$$\widehat{\text{Var}}(N) = \frac{1}{N} \sum_{i=1}^N \widehat{\text{Var}}_i(N). \quad (\text{V.14})$$

Similarly, we obtain an estimate of the time it takes to compute  $N$  samples as

$$\widehat{t}(N) = \frac{1}{N} \sum_{i=1}^N t_i(N). \quad (\text{V.15})$$

The convergence property of Monte Carlo estimators,

$$\text{Var} = c/t \iff, \quad (\text{V.16})$$

tells us that the proportionality constant  $c$  is equal to the time required to lower

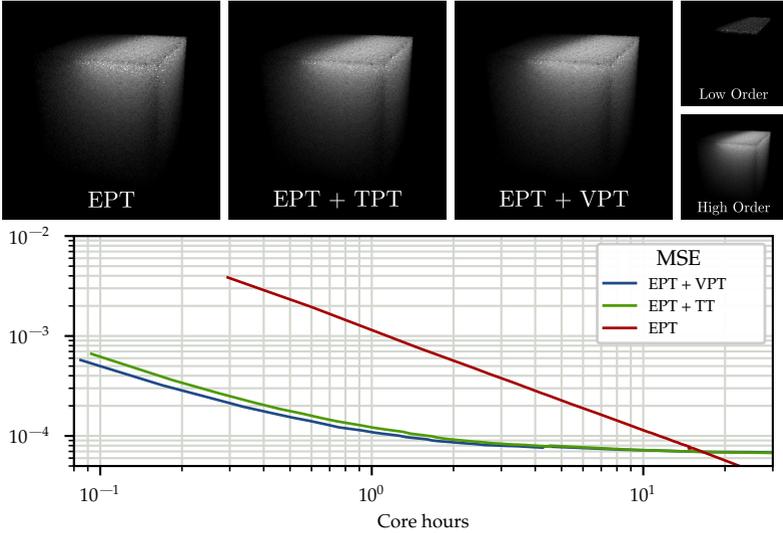


Figure V.17.: Top: Equal quality comparison of a cube of dielectric spheres, as shown by Moon, Walter, et al. (2007, Figure 9). The last column shows individual low- and high-order contributions computed with explicit path tracing and volumetric path tracing, respectively. Bottom: We show the mean squared error for explicit path tracing (EPT), EPT combined with teleportation transport (EPT+TT), and EPT combined with our volumetric approximation (EPT+VPT). Our techniques converge much more quickly, but their biased nature is clearly visible (Meng, Papas, et al., 2015).

variance to 1. We therefore call  $c$  the *time to unit variance* (TTUV), and report it as a measure of quality for each of the techniques. The TTUV can be estimated as

$$c \approx \widehat{\text{Var}}(N) \cdot \widehat{t}(N). \quad (\text{V.17})$$

## 7.2 Overview

We summarize render times for our main result scenes in Table V.2, both for explicit path tracing of grain geometry (EPT), and our hybrid method utilizing an automatic switch to the approximate volumetric path tracking model (EPT+VPT). Our main measure is time to unit variance, but we also show the speedup of the hybrid method with respect to EPT.

We consider two cases: combined render times for the full image, and high-order transport only images, where paths are not counted if they never fulfill the criterion

Table V.2.: Time to unit variance in seconds for explicit path tracing and our hybrid method. The value in parenthesis is the speedup relative to EPT alone. On the left we report times for rendering both low order and high order transport whereas on the right we report the timings and speedup only for the high order component.

Scene	Time to unit variance, combined (sec)		Time to unit variance, high order (sec)	
	EPT	EPT+VPT	EPT	EPT+VPT
Spices	1617.3	400.3 (4.00 $\times$ )	688.8	5.3 (130.8 $\times$ )
Sand Castle	30.7	14.4 (2.10 $\times$ )	11.8	0.1 (206.8 $\times$ )
Snowman large	4000.8	132.0 (30.3 $\times$ )	3548.6	8.8 (404.6 $\times$ )
Snowman small	2868.4	171.1 (16.8 $\times$ )	2338.0	6.1 (381.5 $\times$ )

for switching to VPT. The reason for this is that our method is specifically designed to optimize high-order transport, and to leave low-order transport alone as much as possible. This is what enables us to retain fine surface detail and coherent refraction patterns.

EPT+VPT results in a significant speedup overall, ranging from 2 $\times$  for the sand castle scene to over 30 $\times$  for the large snow man. If only high-order transport time is measured, then the speedup is between 130 $\times$  and 400 $\times$ .

In result images, we specify render times along with residual variance as *CPU hours high-order / CPU hours total (variance)*. For example a value of *4148 / 5646 (8.3e-5)* means that the image rendered for 5646 CPU hours in total, spending 4148 CPU hours on high-order scattering, and has a residual variance of 8.3e-5.

### 7.3 Sand Castle scene

The sand castle scene in Figure V.1 is an aggregate of roughly 2 billion individual grains, each of which contains a highly scattering participating medium, and has a specular surface. There are three types of grains with different absorption properties.

Our hybrid method manages to deliver a 2.2 $\times$  speedup compared to EPT at equal image quality.

The relatively modest overall speedup is in stark contrast to the high-order speedup of 200.6 $\times$ . The reason for this is rooted in the grain material parameters: They have an average albedo of only 0.68, and only very few paths survive long enough to be switched to VPT mode. Additionally, our implementation enforces at least one initial EPT grain interaction per path to guarantee preserved surface detail. In consequence, high-order scattering has a minor part in this scene to begin with, and optimizations designed for high-order scattering must have a minor impact, as well.

### 7.4 Spices scene

The spices scene (see Figure V.18) contains four different granular aggregates (from left to right): white flour, pink salt, brown sugar, and cane sugar.

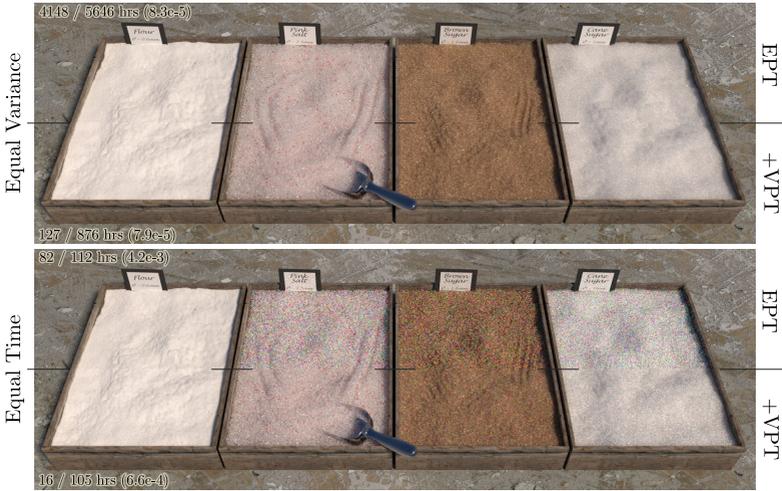


Figure V.18.: The spices scene, which contains flour (left), pink salt (center left), raw sugar (center right), and cane sugar (right). We show roughly equal variance (top) and roughly equal time (bottom) comparisons of EPT to our hybrid technique EPT+VPT (Meng, Papas, et al., 2015).

In Figure V.18 (top), we compare roughly equal variance levels. Our hybrid method manages to render the scene in 876 CPU hours, as opposed to 5646 CPU hours for explicit path tracing. In fact, it only spends roughly 14% of the render time on high-order scattering, while EPT spends 73% there. Again, this demonstrates that we optimize mainly for high-order scattering.

In Figure V.18 (bottom), we show a roughly equal time comparison. This comparison clearly demonstrates the higher variance in EPT mode as opposed to the hybrid EPT+VPT method, although this effect is again the least pronounced in flour.

It is interesting to note here that the approximative nature of our volumetric model is visible in the flour, where a subtle difference in brightness can be seen at the border between the two images.

## 7.5 Snowman scene

The snowman scene (see Figure V.19) comes in two variants: the large snowman (top) is approximately 1m tall, while the small snowman (bottom) has the same overall geometry, but is only 10cm tall. This leads to a denser look on the large snowman. The difference in scale is realized by scaling the grain geometry. This is visible in the close-ups on the right, which show a region around the left „eye” of the snowman.

The scene is lit by a sun environment map, where the sun, like in nature, is small

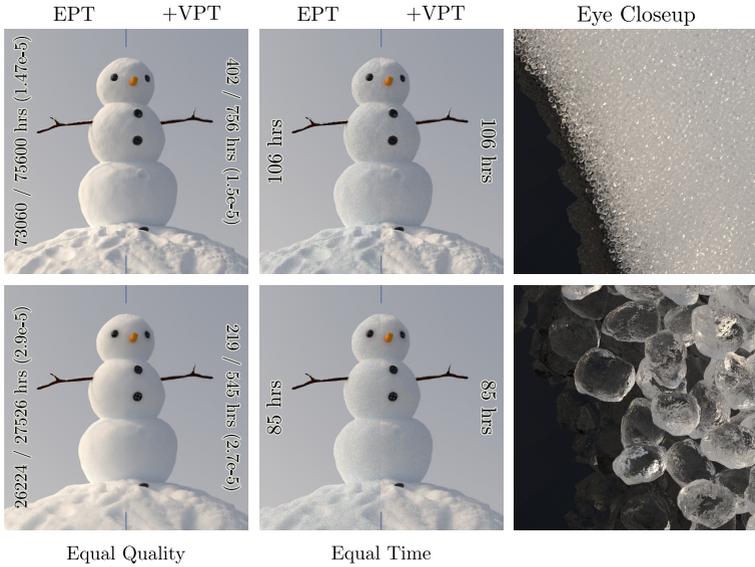


Figure V.19.: The snowman scene, showing a large (top, 1m) and a small (bottom, 10cm) figure made of ice granules. We show roughly equal quality (left) and roughly equal time (right) comparisons of EPT and our hybrid EPT+VPT method, along with close-up renderings of the snowman's eye. (Meng, Papas, et al., 2015).

and extremely bright. Since since the snow grains also exhibit almost no absorption or scattering and have highly specular surfaces, this scene presents a worst case for explicit path tracing. The reasons are twofold: first, the small size of the sun means that it is hard to find at random and second, the surface properties of the grains mean that one cannot connect a path directly to the sun.

If a path hits the sun by chance, then there is likely to be a large contribution to the image, leading to excessive noise. In Figure V.19 (center), this effect is visible: the EPT part of the image seems to be more translucent, with and is not as bright as it should be. This is because low-order reflections are still missing, simply because the path tracer has not been able to find them yet. In contrast, with our hybrid method, we can exploit variance reduction techniques to make connections to the sun far more likely.

On the equal quality renders, we give times for both high-order scattering and total transport time. Almost all time is spent in the simulation of high-order scattering, and so this scene is an ideal candidate for optimization with our technique. Indeed, we reduce the time spent in high-order scattering to approximately 50%, down from more than 95% in the EPT case.

## 8 Conclusion and future research

In this chapter, we have presented a method for rendering granular aggregates. The method is mainly geared towards optimizing high-order scattering as it occurs in high-albedo materials, where explicit path tracing will generate long paths with a high contribution, leading to excessive noise.

Our technique combines explicit path tracing with a volumetric approximation, and switches on the fly from the former to the latter. The EPT prefix is used to retain high-frequency surface detail and coherent refraction patterns, so that individual grains remain visible. The VPT postfix, in contrast, enables us to employ well known variance reduction techniques such as next event estimation by replacing complex visibility with a stochastic approximation based on the radiative transfer equation. Using this approach, we are able to reduce rendering time spent to compute high-order scattering to a level comparable to the time spent for low-order scattering. This enables rendering of scenes that were previously intractable.

### 8.1 Diffusion

In (Meng, Papas, et al., 2015), further speedups are obtained by using a diffusion method to exit the material quickly. However, this adds considerable bias especially in high albedo materials. Future research might focus on finding diffusion models that are better equipped to deal with granular materials. This has been explored to some extent by d'Eon (2013), who derived diffusion models for materials with non-exponential extinction.

### 8.2 Removing EPT

Our technique switches using an automatic criterion based on ray differentials to ensure sufficient diffusion. It always renders at least one grain interaction using explicit path tracing, however. A promising direction of future research might be to side-step EPT completely for distant views, and model surface glints using a BRDF model, similar to (Yan et al., 2014) and (Zirr and Kaplanyan, 2016).

### 8.3 Packing density

One of the main limitations of our technique is the limited packing density, which is a direct consequence of the fact that we employ results from sphere packing literature to compute RTE parameters for our volumetric approximation. To remedy

this, packings of spherocylinders were analyzed in a master's thesis (Kahl, 2016). The results indicate that at least for homogeneous packings, an analytic free-flight distribution similar to ours can be used with spherocylinders, as well. Another possible area of research are packings of ellipsoids (Donev et al., 2004), or cylinders without hemispherical caps, which could be used in packings of couins, or pills.

## 8.4 Heterogeneous mixtures

Our method assumes granular materials that are statistically invariant to translation and rotation. Müller et al. (2016) extended our method to heterogeneous mixtures by rasterizing volume properties into a three-dimensional grid. They are also able to handle animated data, but an interesting area of further research might be to exploit temporal averaging in the derivation of the RTE parameters.

## 8.5 Spectral light transport

While this dissertation is focused mainly on spectral light transport, our prototype renderer used for our hybrid EPT+VPT technique actually worked with RGB transport, and performed precomputations for the red, green, and blue channels. To use it in a spectral renderer, different approaches are possible: One approach would be to store TSDF samples in spectral bins. However, this would increase storage size, which may not be desirable. Another idea could be to perform precomputation in RGB, and use on-the-fly spectral upsampling, such as our grid-based technique, to obtain wavelength-dependent RTE parameters during rendering. It is also important to note that the aggregate topology will never depend on the wavelength, and so only the per-grain precomputation must be adapted. In any case, future research could focus on optimizing both storage and run time requirements for the spectral case.

# VI

## CONCLUSION

As we have stated in the introduction, many problems in light transport still lack a satisfactory solution. We have focused on three particular problems in this thesis: efficient and general spectral upsampling, noise reduction in dense participating media in the presence of strong backlights, and the efficient modeling and rendering of granular materials.

Our approach to spectral upsampling improves on a technique which has been in use in many rendering systems for almost twenty years, but also has severe shortcomings. Our technique allows for arbitrary input and as a consequence supports very wide gamut color spaces. It can be implemented using readily available tools, is fast, suited for Monte Carlo algorithms that often only need to determine the spectrum for a specific wavelength, and can be used as a drop-in replacement. In fact, it has been successfully used in our research rendering system for roughly two years, and has been used for several of our research projects. We therefore believe that this algorithm is generally applicable to physically-based Monte Carlo rendering systems. The full source code can be obtained as part of the supplemental material for our original publication (Meng, Simon, et al., 2015). As an area of future research, it may be interesting to look at different grid point positions, and to analyze the impact on perceived accuracy. It also is not clear that the smoothness criterion used in the precomputation is optimal. In fact, since research about the frequency content of naturally occurring reflectances is available, it may be worth trying to synthesize spectra matching a given input color given a predefined band limit.

We also presented a technique that improves the robustness of Dwivedi sampling in dense, high albedo participating media with back lights. This is a major failure case of (Křivánek and d'Eon, 2014), but it is very common in theatrically lit character portrait shots, as they appear in many movies. We have shown that such specialized techniques may improve render times in a relatively general framework, and other related techniques are likely to exist. One area of future research might be the search for even more robust selection heuristics that can discriminate better between the available techniques, or a more in-depth analysis of how multi-vertex sampling techniques such as Dwivedi sampling can be combined with bidirectional or Markov-chain methods. Additionally, using better information about incident illumination, such as a

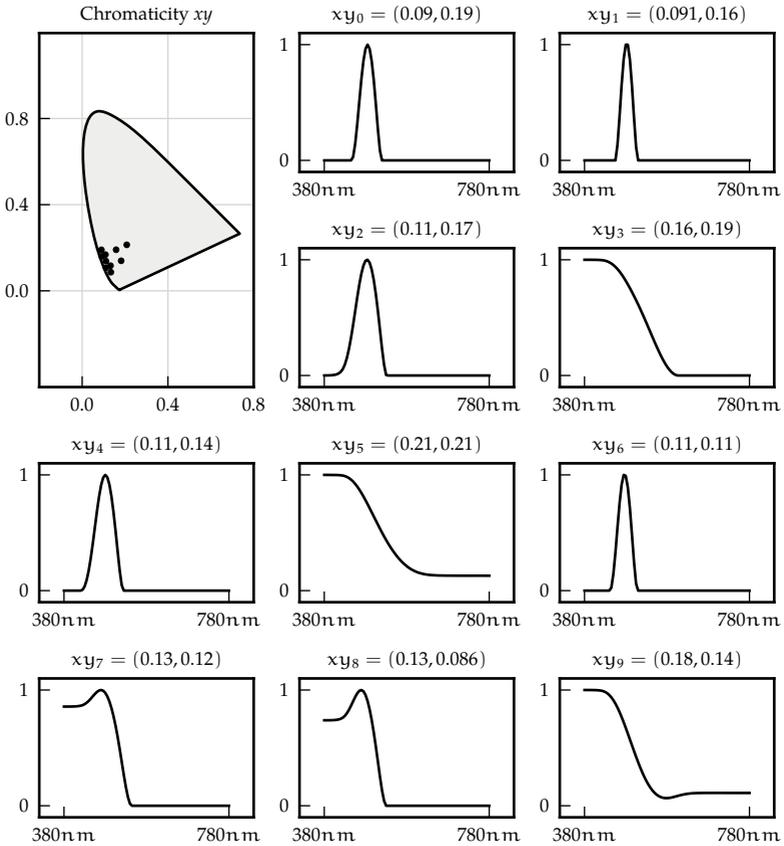
diffusion approximation or path guiding, could prove beneficial. We showed on an example scene that Dwivedi sampling can be used in heterogeneous media, but more research should be done in this area to explore it more systematically.

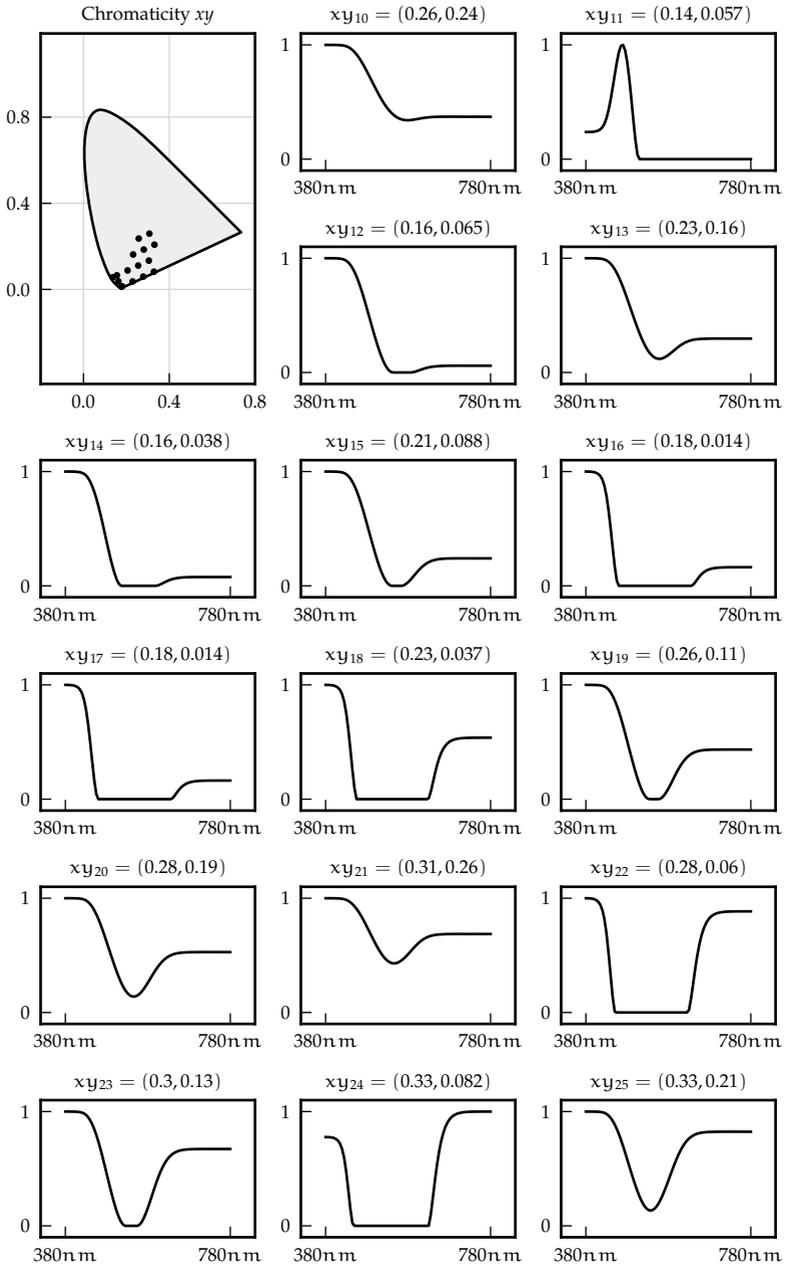
Finally, we presented our work on modeling and rendering granular materials. This topic is complex, not only because so many different granular materials exist, but also because it is not always clear if a material is granular or not. For example, snow can be modeled as an aggregate of dielectric particles, as we have done in this paper, but this is not entirely accurate. As snow is compressed, it turns into a solid block of ice, and the boundary may not be clear. Our techniques can therefore only be approximate. However, the sacrifice in accuracy comes with an immense reward: we are able to model, with little memory required, granular materials that are infinite up to integer accuracy. We can also carve granular objects out of these infinite spaces, a technique that allows for efficient modeling and prototyping. Our rendering technique uses the assumption of non-intersecting bounding sphere to find an analytic model for the parameters of a participating medium that behaves very similar to the granular aggregate, and can be used to speed up high-order scattering by multiple orders of magnitude. Several questions, however, are left unanswered in this dissertation. Our procedural model cannot animate grains in a meaningful macroscopic motion. Future research could attempt to create a model that can incorporate some amount of motion without having to use expensive physically-based rigid body simulation. In our rendering algorithm, we assume a homogeneous granular material, where mixing ratios do not vary with position. Since the publication of our original paper (Meng, Papis, et al., 2015), another publication has dealt with exactly this issue, extending our technique (Müller et al., 2016). Our technique is also limited in the maximum packing rate it can handle. This is caused by our assumption of non-overlapping bounding spheres. Some preliminary work has been carried out in order to understand the implications of non-spherical bounding volumes (Kahl, 2016), but more research is required. We believe that combining our VPT approximation with surface glint models could provide a simple and relatively cheap way to render vast background granular objects such as a beach even more efficiently.

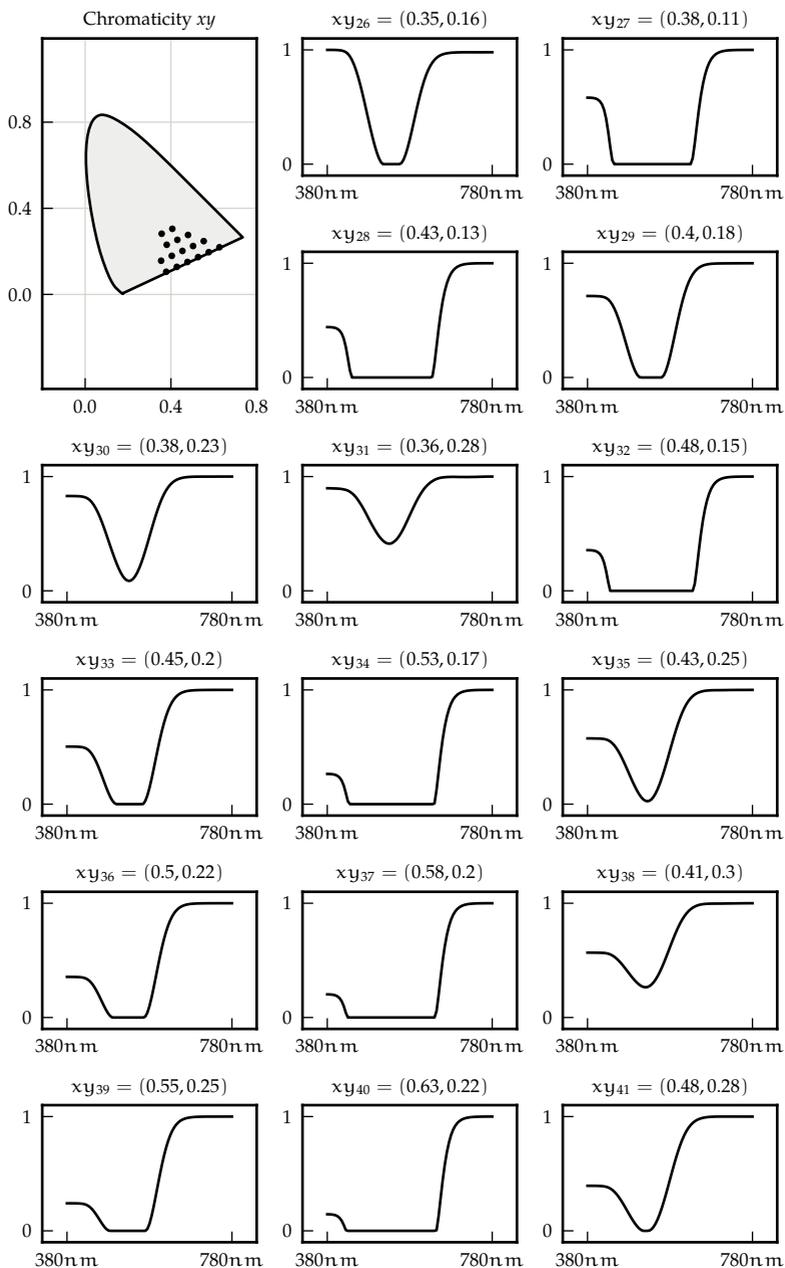
In conclusion, while we believe that we have demonstrated innovative and efficient solutions for some of the harder remaining problems of physically-based light transport, there clearly are many open questions. This means that there are still a lot of areas for future researchers to explore. Possibly one of their bigger problems will be to get a good understanding of the wealth of related work that exists, not only in the field of computer graphics, but also in the mathematics and physics communities. Computer graphics therefore requires interdisciplinary research, which can be difficult because notation and nomenclature tends to differ between communities. However, we believe that it is also an extremely rewarding experience, and hope that this work may serve as an inspiration for others to pursue similar ideas.

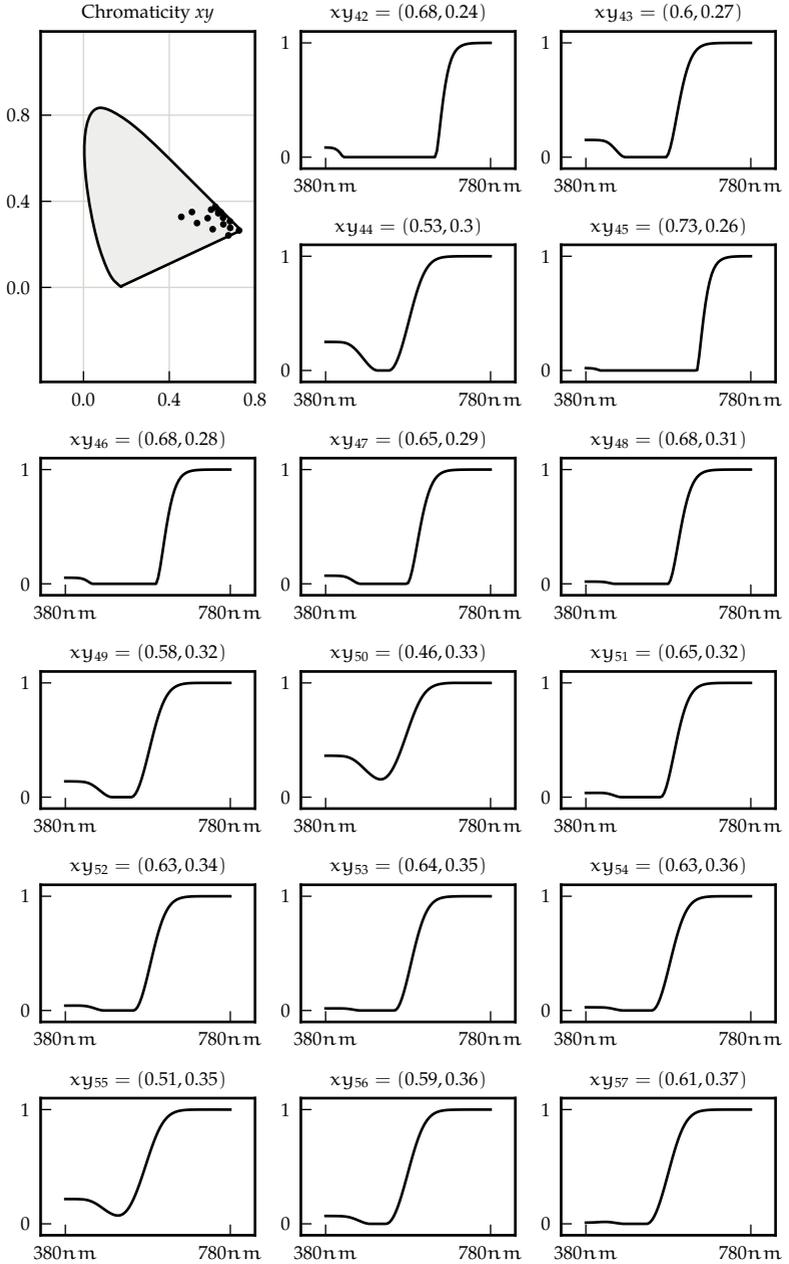
# GRID BASIS SPECTRA

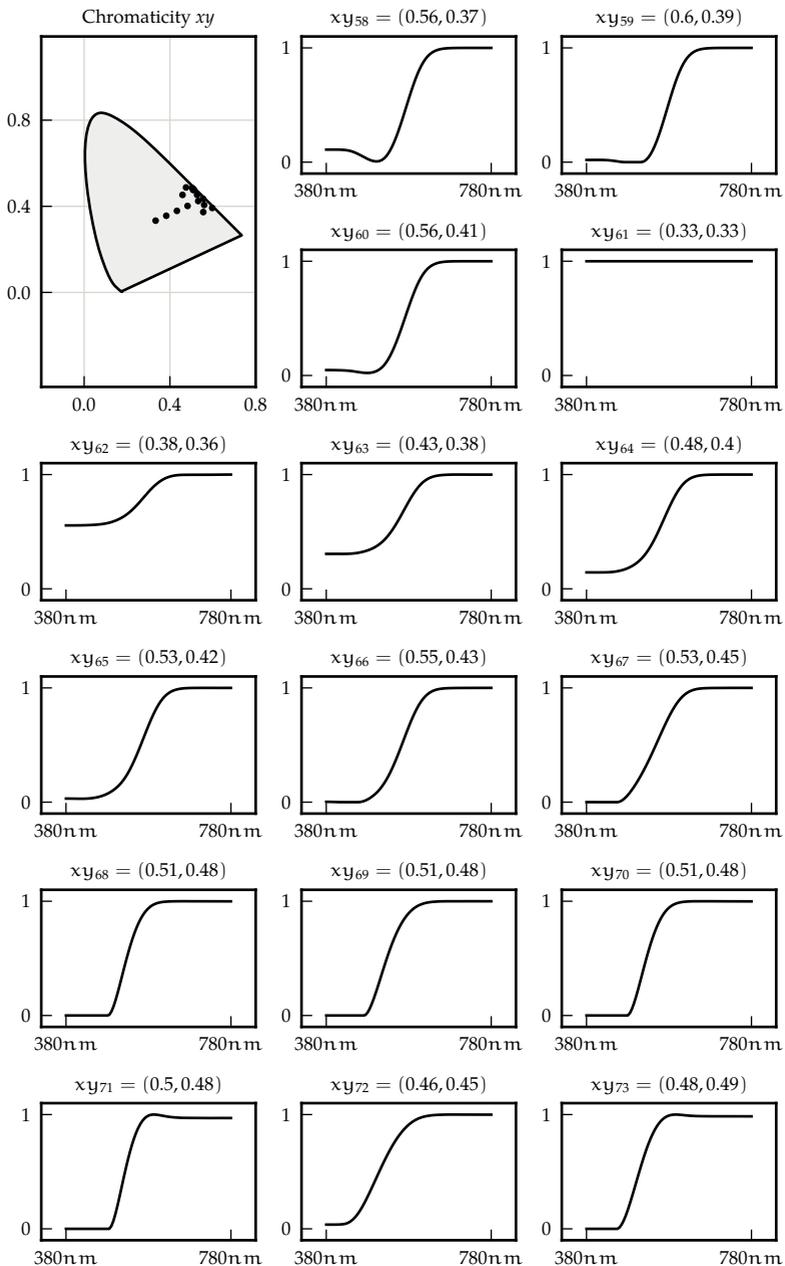
This appendix contains all 186 precomputed basis spectra used in our spectral upsampling technique from Chapter III. They are normalized to maximum 1 for visibility.

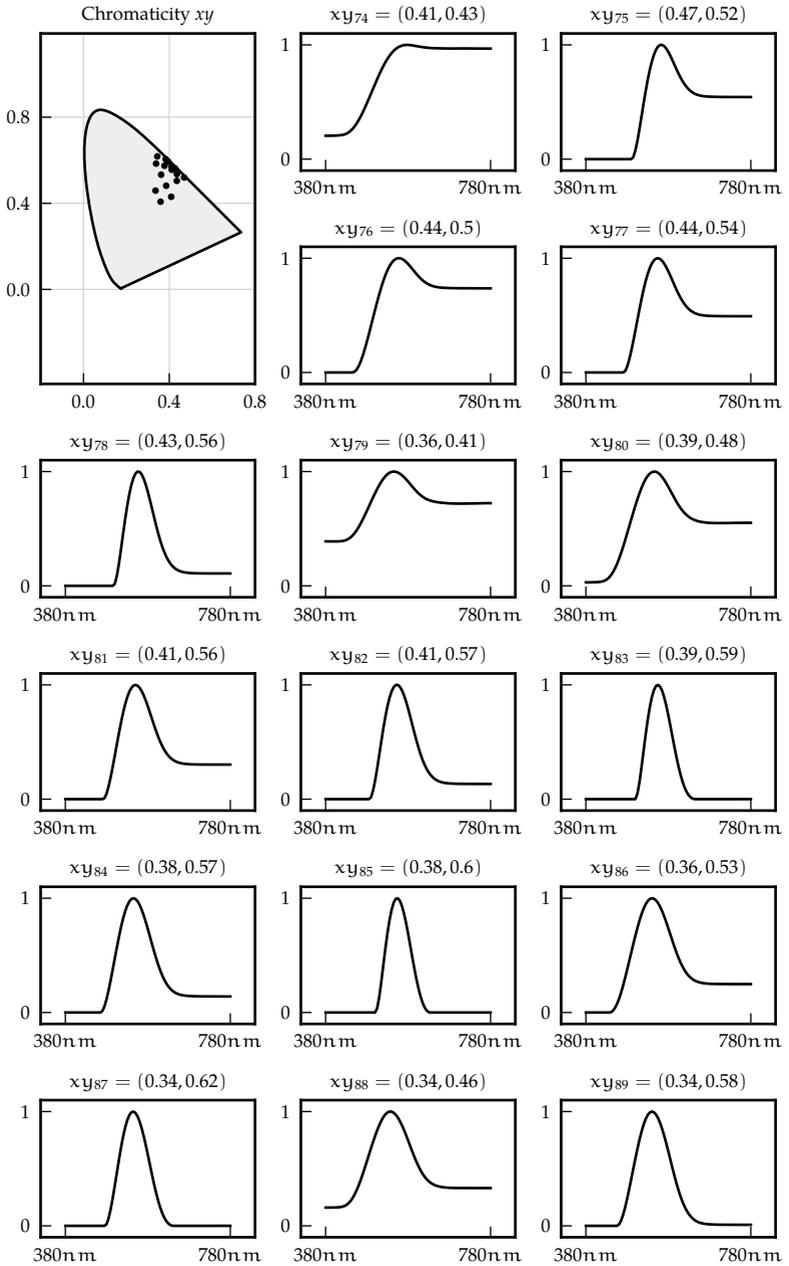


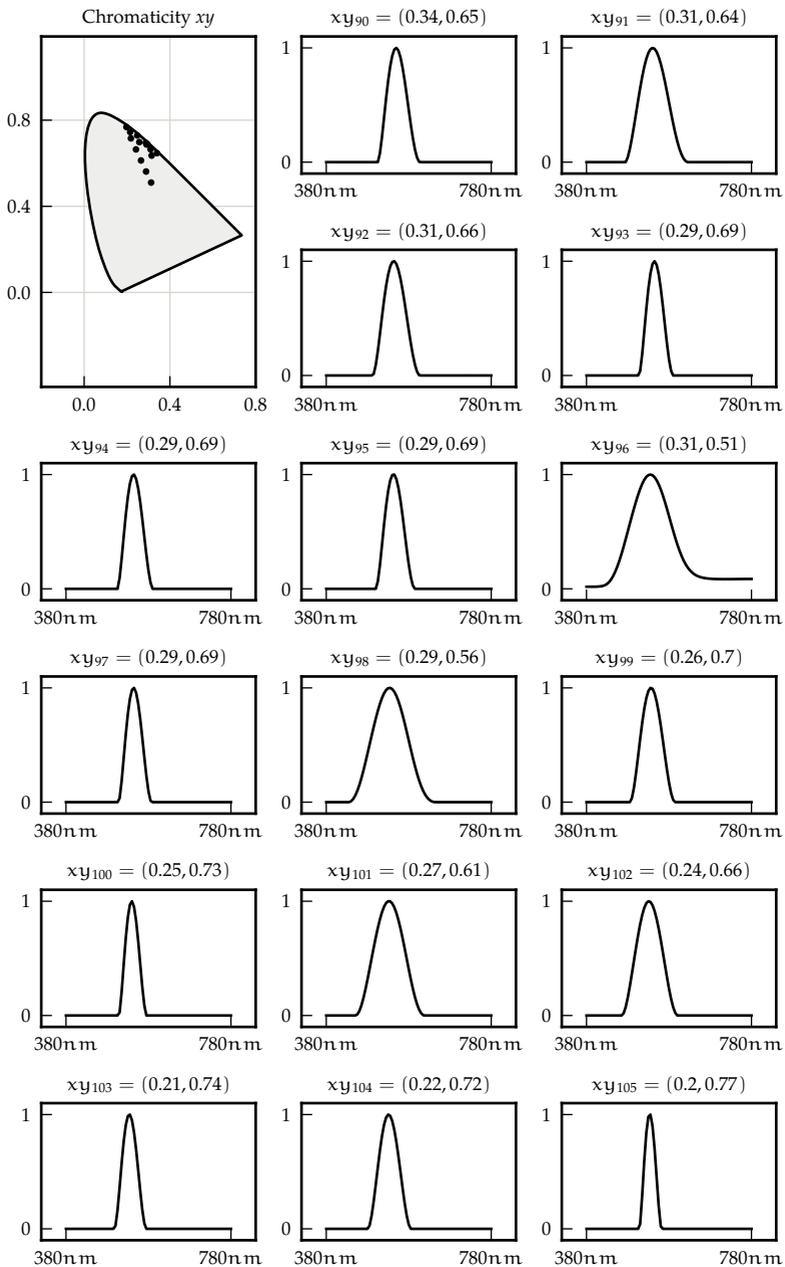




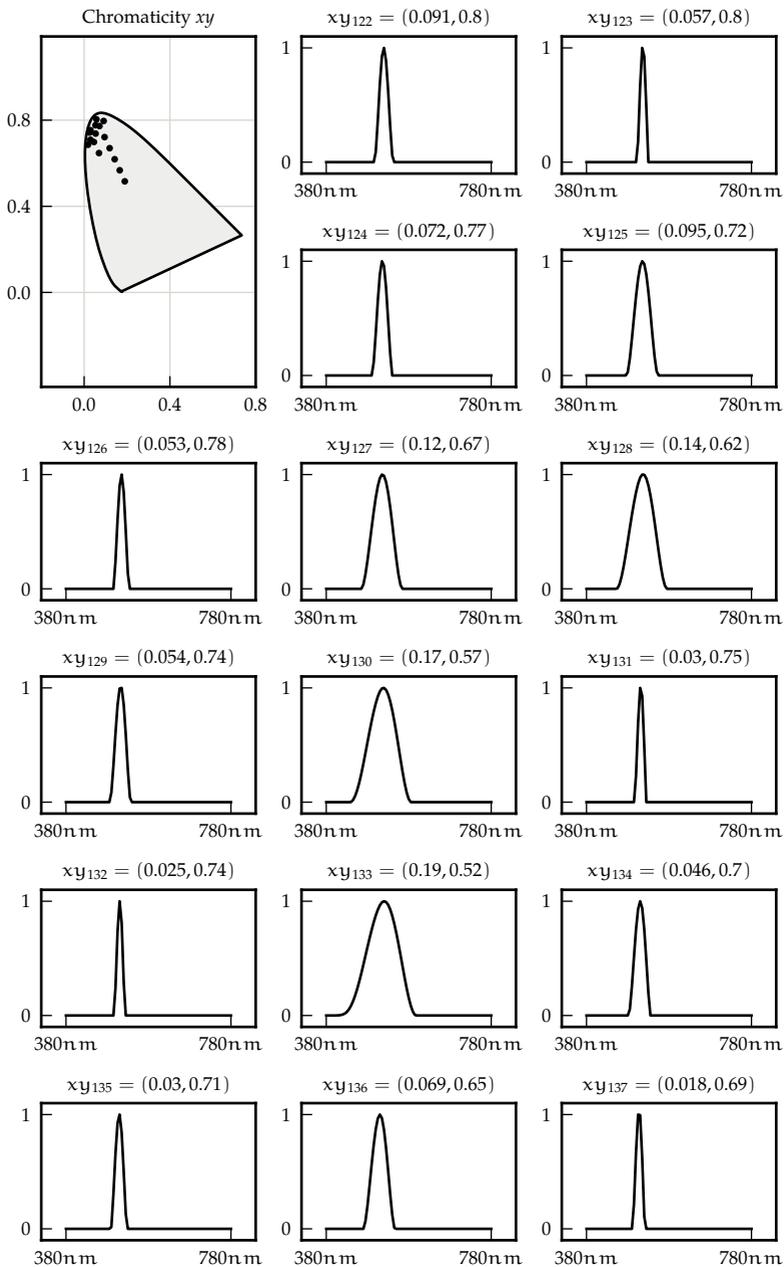


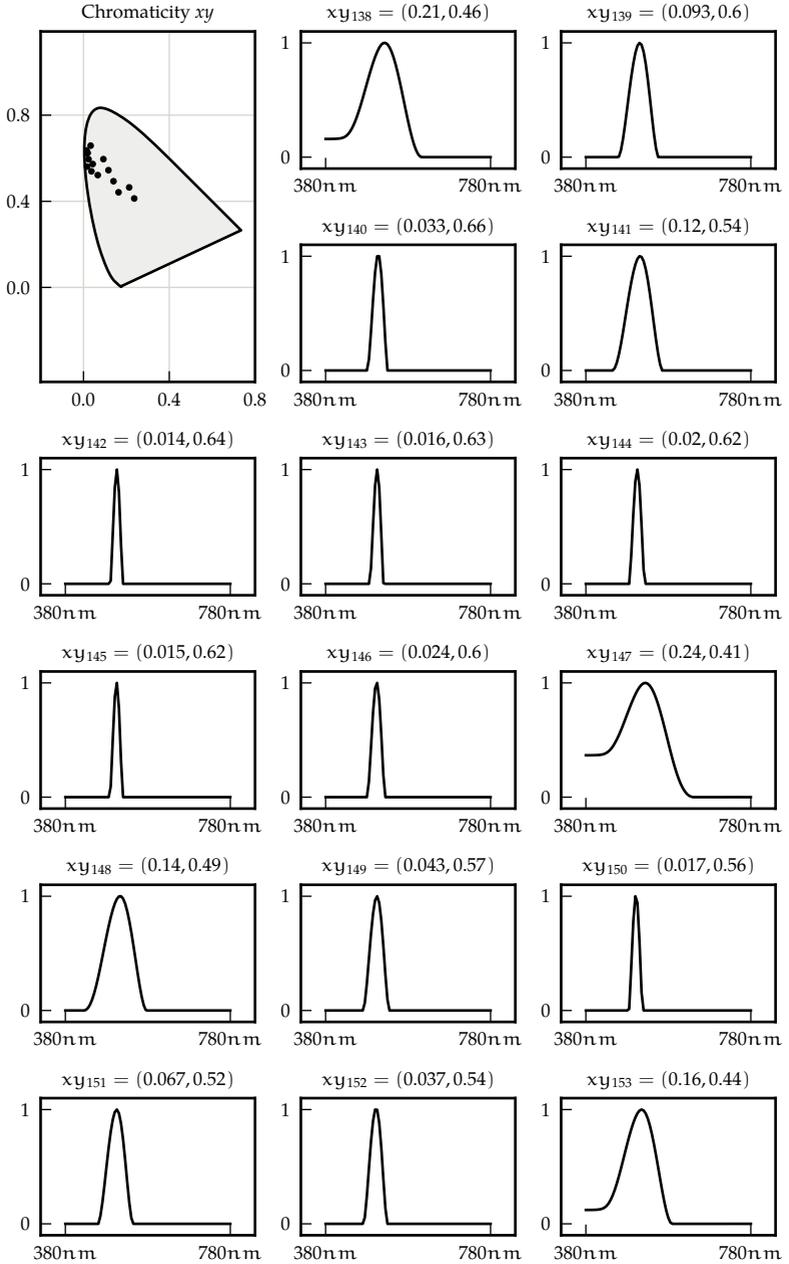


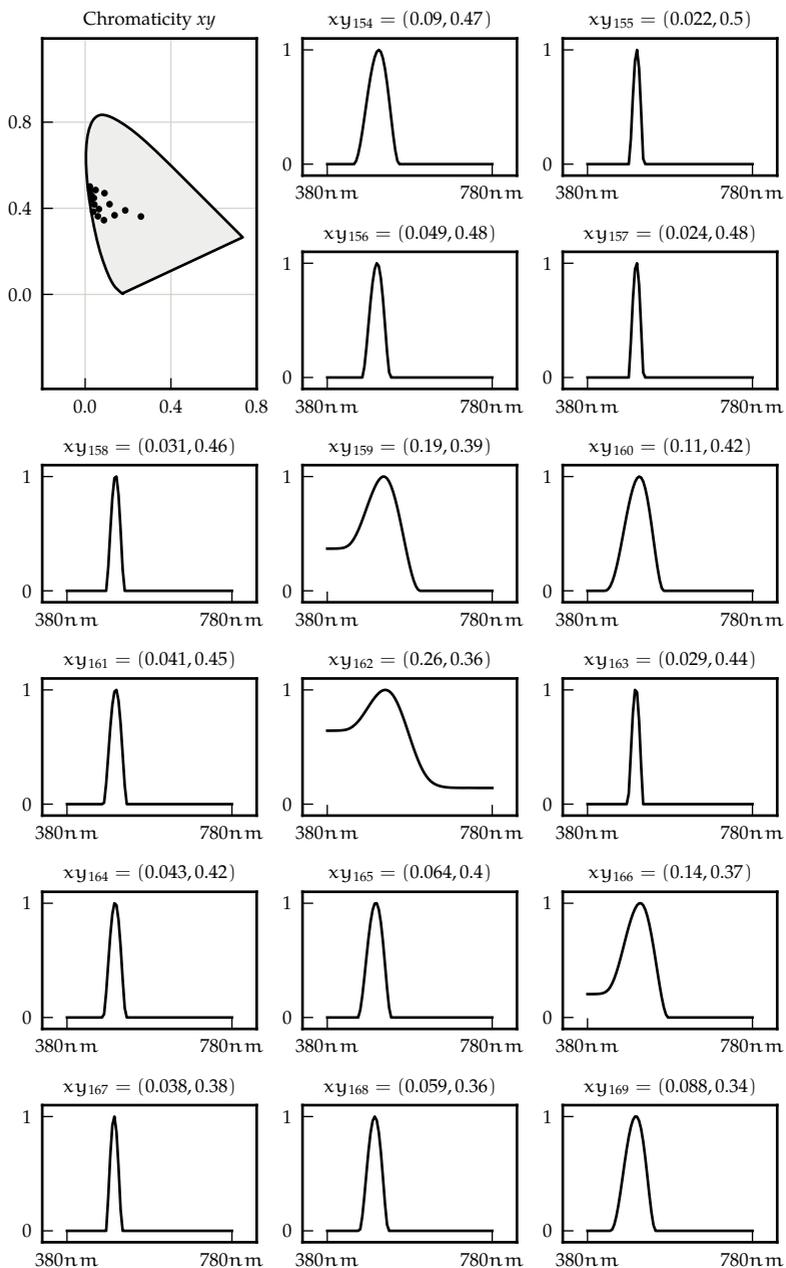


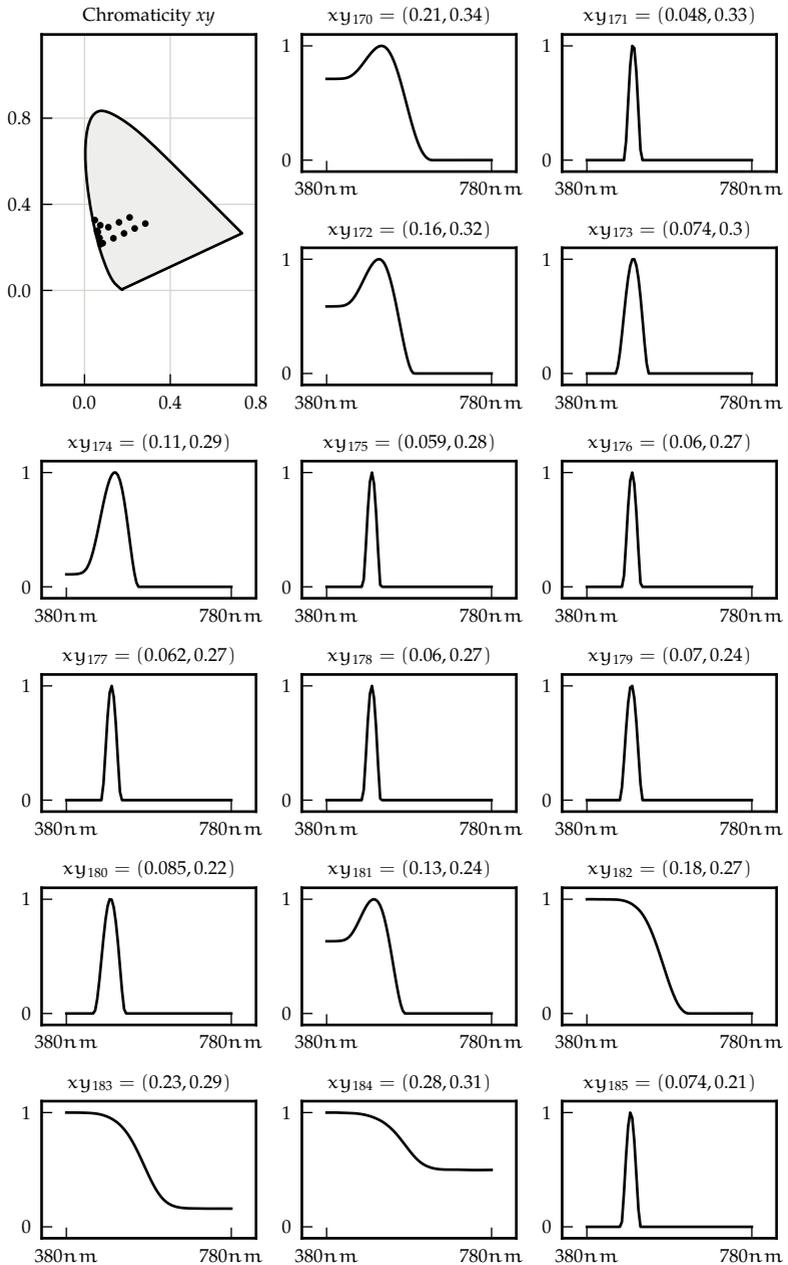












# LIST OF SYMBOLS

## Radiometric Quantities

$f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$	BxDF (p. 23)	$[1/\text{sr}]$
$\mu_a(\boldsymbol{x}, \lambda)$	Absorption Coefficient (p. 24)	$[1/\text{m}]$
$\mu_e(\boldsymbol{x}, \lambda)$	Emission Coefficient (p. 24)	$[1/\text{m}]$
$\mu_t(\boldsymbol{x}, \lambda)$	Extinction Coefficient (p. 24)	$[1/\text{m}]$
$\mu_s(\boldsymbol{x}, \lambda)$	Scattering Coefficient (p. 24)	$[1/\text{m}]$
$\rho(\boldsymbol{x})$	Density (p. 24)	$[1/\text{m}^3]$
$\Phi$	Radiant Flux (p. 20)	$[\text{J}/\text{s} = \text{W}]$
$E(\boldsymbol{x})$	Surface Irradiance (p. 21)	$[\text{W}/\text{m}^2]$
$E(\boldsymbol{x})$	Volume Irradiance (p. 21)	$[\text{W}/\text{m}^3]$
$\bar{l}(\lambda)$	Mean Free Path (p. 27)	$[\text{m}]$
$L(\boldsymbol{x}, \boldsymbol{\omega})$	Surface Radiance (p. 22)	$[\text{W}/\text{sr}/\text{m}^2]$
$L(\boldsymbol{x}, \boldsymbol{\omega})$	Volume Radiance (p. 22)	$[\text{W}/\text{sr}/\text{m}^3]$
$B(\boldsymbol{x})$	Surface Radiosity (p. 21)	$[\text{W}/\text{m}^2]$
$B(\boldsymbol{x})$	Volume Radiosity (p. 21)	$[\text{W}/\text{m}^3]$
$\sigma_a(\boldsymbol{x}, \lambda)$	Absorption Cross-Section (p. 23)	$[\text{m}^2]$
$\sigma_e(\boldsymbol{x}, \lambda)$	Emission Cross-Section (p. 24)	$[\text{m}^2]$
$\sigma_t(\boldsymbol{x}, \lambda)$	Extinction Cross-Section (p. 24)	$[\text{m}^2]$
$\sigma_s(\boldsymbol{x}, \lambda)$	Scattering Cross-Section (p. 23)	$[\text{m}^2]$
$\tau(s)$	Transmittance (p. 25)	$[1]$



# BIBLIOGRAPHY

- Adobe (2005). *Adobe RGB (1998) Color Image Encoding*. Tech. rep. Adobe Systems Incorporated (cit. on p. 37).
- Barsky, B. A., Horn, D. R., Klein, S. A., Pang, J. A., and Yu, M. (2003). "Camera models and optical systems used in computer graphics: part I, object-based techniques". In: *International Conference on Computational Science and Its Applications*. Springer, pp. 246–255 (cit. on pp. 31–32).
- BIPM (1983). "Résolution 1 – Définition du mètre". In: *Conférence Générale des Poids et Mesures*, pp. 97–98 (cit. on p. 20).
- Booth, T. E. (1987). "Generalized Zero-variance Solutions and Intelligent Random Numbers". In: *Proceedings of the 19th Conference on Winter Simulation*. WSC '87, pp. 445–451 (cit. on p. 57).
- Bubnik, Z., Kadlec, P., Urban, D., and Bruhns, M. (1998). *Sugar Technologists Manual*. Verlag Dr. Albert Bartens (cit. on p. 75).
- Case, K. M., Hoffman, F. de, and Placzek, G. (1953). *Introduction to the Theory of Neutron Diffusion*. U.S. Government Printing Office, Washington, D.C. (cit. on pp. 60–61).
- Chandrasekhar, S. (1960). *Radiative Transfer*. Dover Publications, Inc. (cit. on pp. ix, 19, 23–24, 27, 59, 76).
- Christensen, P. H. and Burley, B. (2015). *Approximate Reflectance Profiles for Efficient Subsurface Scattering*. Tech. rep. Pixar Technical Memo #15-04. Pixar Animation Studios, Walt Disney Animation Studios (cit. on p. 56).
- CIE (2004). *Colorimetry*. Tech. rep. Commission Internationale de l'Éclairage (cit. on pp. 34, 36–38).
- CIE (2008). *CIE Colorimetry - part 4: 1976 L\*a\*b\* colour space*. Tech. rep. ISO 11664-4:2008(E)/CIE S 014-4/E:2007. Joint ISO/CIE Standard (cit. on p. 51).
- Cohen, M. F. and Wallace, J. R. (1993). *Radiosity and Realistic Image Synthesis*. Morgan Kaufman (cit. on p. 22).
- Coveyou, R. R., Cain, V. R., and Yost, K. (1967). "Adjoint and importance in Monte Carlo application". In: *Nuclear Science and Engineering* 27.2, pp. 219–234 (cit. on pp. 40, 42).

- Cuffey, K. and Paterson, W. (2010). *The physics of glaciers*. Academic Press (cit. on p. 75).
- Culik II, K. and Kari, J. (1996). "An aperiodic set of Wang cubes". In: *J. UCS The Journal of Universal Computer Science*. Springer, pp. 675–686 (cit. on p. 77).
- Demtröder, W. (2015). *Experimentalphysik 1. Mechanik und Wärme*. 7th ed. Springer Spektrum (cit. on p. 38).
- d'Eon, E. (2013). "Rigorous asymptotic and moment-preserving diffusion approximations for generalized linear Boltzmann transport in arbitrary dimension". In: *Transport Theory and Statistical Physics* 42.6-7, pp. 237–297 (cit. on pp. 77, 87, 99).
- d'Eon, E. and Irving, G. (2011). "A Quantized-diffusion Model for Rendering Translucent Materials". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 30.4, 56:1–56:14 (cit. on p. 56).
- d'Eon, E., Luebke, D. P., and Enderton, E. (2007). "Efficient Rendering of Human Skin". In: *Proc. Eurographics Symposium on Rendering*, pp. 147–157 (cit. on p. 56).
- Dixmier, M. (1978). "Une Nouvelle Description des Empilements Aléatoires et des Fluides Denses". In: *Le Journal de Physique* 39, pp. 873–895 (cit. on pp. 81–82).
- Donev, A., Cisse, I., Sachs, D., Vario, E. A., Stillinger, F. H., Connelly, R., Torquato, S., and Chaikin, P. M. (2004). "Improving the Density of Jammed Disordered Packings Using Ellipsoids". In: *Science* 303.5660, pp. 990–993 (cit. on p. 100).
- Donner, C. and Jensen, H. W. (2005). "Light Diffusion in Multi-layered Translucent Materials". In: *ACM Trans. Graph.* 24.3, pp. 1032–1039 (cit. on p. 56).
- Donner, C. and Jensen, H. W. (2006). "A Spectral BSSRDF for Shading Human Skin". In: *Proc. Eurographics Symposium on Rendering*. Nicosia, Cyprus, pp. 409–417 (cit. on p. 56).
- Donovan, T., Sutton, T., and Danon, Y. (2003). "Implementation of Chord Length Sampling for Transport Through a Binary Stochastic Mixture". In: *Nuclear Mathematical and Computational Sciences: A Century in Review, A Century Anew* (cit. on pp. 76, 81).
- Durant, S., Calvo-Perez, O., Vukadinovic, N., and Greffet, J.-J. (2007). "Light scattering by a random distribution of particles embedded in absorbing media: full-wave Monte Carlo solutions of the extinction coefficient". In: *Journal of the Optical Society of America* 24.9, pp. 2953–2962 (cit. on p. 76).
- Dutré, P., Lafortune, E. P., and Willems, Y. D. (1993). "Monte Carlo light tracing with direct computation of pixel intensities". In: pp. 128–137 (cit. on p. 41).
- Dwivedi, S. R. (1982a). "A new importance biasing scheme for deep-penetration Monte Carlo". In: *Am. nucl. Energy* 9, pp. 359–368 (cit. on pp. xii, 3, 55, 57, 61).
- Dwivedi, S. R. (1982b). "Zero Variance Biasing Schemes for Monte Carlo Calculations of Neutron and Radiation Transport". In: *Nucl. Sci. Eng.* 80.1, pp. 172–178 (cit. on pp. 3, 55, 57).

- Fairman, H., Brill, M., and Hemmendinger, H. (1998). "How the CIE 1931 color-matching functions were derived from Wright-Guild data". In: *Color Research and Application* 22.1, pp. 11–23 (cit. on p. 36).
- Fernando, R. (2004). *GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics*. Pearson Higher Education (cit. on p. 29).
- Foldy, L. L. (1945). "The Multiple Scattering of Waves. I. General Theory of Isotropic Scattering by Randomly Distributed Scatterers". In: *Physical Review* 67 (3-4), pp. 107–119 (cit. on p. 76).
- Frisvad, J. R., Hachisuka, T., and Kjeldsen, T. K. (2014). "Directional dipole model for subsurface scattering". In: *ACM Trans. on Graphics* 34.1, 5:1–5:12 (cit. on p. 56).
- Fujimoto, A., Tanaka, T., and Iwata, K. (1986). "ARTSccelerated Ray-tracing System". In: *IEEE Comput. Graph. Appl.* 6.4, pp. 16–26 (cit. on p. 79).
- Georgiev, I., Krivánek, J., Davidovic, T., and Slusallek, P. (2012). "Light transport simulation with vertex connection and merging." In: *ACM Trans. Graph.* 31.6, p. 192 (cit. on p. 42).
- Glassner, A. S. (1995). *Principles of Digital Image Synthesis*. Morgan Kaufmann (cit. on p. 43).
- Hachisuka, T., Pantaleoni, J., and Jensen, H. W. (2012). "A path space extension for robust light transport simulation". In: *ACM Transactions on Graphics (TOG)* 31.6, p. 191 (cit. on p. 42).
- Haghighat, A. (2015). *Monte Carlo Methods for Particle Transport*. CRC Press Taylor & Francis Group (cit. on p. 7).
- Halton, J. H. (1964). "Algorithm 247: Radical-inverse Quasi-random Point Sequence". In: *Commun. ACM* 7.12, pp. 701–702. ISSN: 0001-0782 (cit. on p. 14).
- Hanika, J. and Dachsbacher, C. (2014). "Efficient Monte Carlo Rendering with Realistic Lenses". In: *Computer Graphics Forum (Proceedings of Eurographics)* 33.2, pp. 323–332 (cit. on p. 32).
- Hanika, J., Droske, M., and Fascione, L. (2015). "Manifold Next Event Estimation". In: *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 34.4, pp. 87–97 (cit. on p. 65).
- Heney, L. G. and Greenstein, J. L. (1941). "Diffuse radiation in the galaxy". In: *The Astrophysical Journal* 93, pp. 70–83 (cit. on pp. 66, 86).
- Holzschuch, N. (2015). "Accurate computation of single scattering in participating media with refractive boundaries". In: *Computer Graphics Forum* 34.6, pp. 48–59 (cit. on p. 65).
- Hoogenboom, J. E. (2008). "Zero-Variance Monte Carlo Schemes Revisited". In: *Nucl. Sci. Eng.* 160.1, pp. 1–22 (cit. on p. 57).
- Igehy, H. (1999). "Tracing Ray Differentials". In: *Proc. SIGGRAPH*, pp. 179–186 (cit. on p. 92).

- Ihmsen, M., Wahl, A., and Teschner, M. (2012). "High-Resolution Simulation of Granular Material with SPH." In: *VRIPHYS*, pp. 53–60 (cit. on p. 73).
- Ishimaru, A. (1978). *Wave Propagation and Scattering in Random Media. Single Scattering and Transport Theory*. Vol. 1. Academic Press (cit. on p. 19).
- ITU (2002). *Recommendation ITU-R BT.709-5: Parameter values for the HDTV standards for production and international programme exchange*. Tech. rep. International Telecommunication Union (cit. on p. 37).
- Jakob, W. (2010). *Mitsuba renderer*. <http://mitsuba-renderer.org> (cit. on p. 94).
- Jarabo, A., Marco, J., Muñoz, A., Buisan, R., Jarosz, W., and Gutierrez, D. (2014). "A Framework for Transient Rendering". In: *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 33.6 (cit. on p. 57).
- Jensen, H. W. (1996). "Global illumination using photon maps". In: *Rendering Techniques' 96*. Springer, pp. 21–30 (cit. on p. 41).
- Jensen, H. W. and Buhler, J. (2002). "A Rapid Hierarchical Rendering Technique for Translucent Materials". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 21.3, pp. 576–581 (cit. on p. 56).
- Jensen, H. W., Marschner, S. R., Levoy, M., and Hanrahan, P. (2001). "A Practical Model for Subsurface Light Transport". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: ACM, pp. 511–518 (cit. on pp. 56, 83).
- Jordan, G., Deeb, S. S., Bosten, J. M., and Mollon, J. D. (2010). "The dimensionality of color vision in carriers of anomalous trichromacy". In: *Journal of Vision* 10.8, p. 12 (cit. on p. 35).
- Kahl, D. (2016). "Modeling Granular Materials with non-spherical Objects". MA thesis. Karlsruhe Institute of Technology, Department of Informatics, Chair for Computer Graphics (cit. on pp. 100, 102).
- Kajiya, J. T. (1986). "The rendering equation". In: *Computer Graphics (Proc. SIGGRAPH)*, pp. 143–150 (cit. on pp. ix, 41, 73, 76).
- Kalos, M. H. and Whitlock, P. A. (2008). *Monte Carlo Methods*. 2nd ed. Wiley-VCH (cit. on p. 17).
- Kalos, M. (1963). "On the estimation of flux at a point by Monte Carlo". In: *Nuclear Science and Engineering* 16.1, pp. 111–117 (cit. on p. 42).
- Kelemen, C., Szirmay-Kalos, L., Antal, G., and Csonka, F. (2002). "A simple and robust mutation strategy for the metropolis light transport algorithm". In: *Computer Graphics Forum*. Vol. 21. 3. Wiley Online Library, pp. 531–540 (cit. on p. 40).
- King, A., Kulla, C., Conty, A., and Fajardo, M. (2013). "BSSRDF Importance Sampling". In: *ACM SIGGRAPH 2013 Talks*. Proc. SIGGRAPH (cit. on p. 56).
- Kraft, D. (1988). *A software package for sequential quadratic programming*. Tech. rep. DFVLR-FB 88-28. Köln, Germany (cit. on p. 47).

- Křivánek, J. and d'Eon, E. (2014). "A Zero-variance-based Sampling Scheme for Monte Carlo Subsurface Scattering". In: *ACM SIGGRAPH 2014 Talks*. Proc. SIGGRAPH. Vancouver, Canada: ACM, 66:1–66:1. ISBN: 978-1-4503-2960-6 (cit. on pp. x, xii, 3, 55, 61, 64–66, 72, 101).
- Křivánek, J., Georgiev, I., Hachisuka, T., Vévoda, P., Šik, M., Nowrouzezahrai, D., and Jarosz, W. (2014). "Unifying Points, Beams, and Paths in Volumetric Light Transport Simulation". In: *ACM Trans. Graph.* 33.4, 103:1–103:13. ISSN: 0730-0301 (cit. on p. 42).
- Lafortune, E. P. and Willems, Y. D. (1993). "Bi-directional path tracing". In: *Proceedings of Third International Conference on Computational Graphics and Visualization Techniques (Compugraphics '93)*. Alvor, Portugal, pp. 145–153 (cit. on p. 41).
- Lafortune, E. P. and Willems, Y. D. (1996). "Rendering Participating Media with Bidirectional Path Tracing". In: *Proc. Eurographics Workshop on Rendering*, pp. 91–100 (cit. on p. 41).
- Lehn, J. and Wegman, H. (2006). *Einführung in die Statistik*. 5th ed. B. G. Teubner (cit. on pp. 7–10, 12–14, 26).
- Levitz, P. (1993). "Knudsen Diffusion and Excitation Transfer in Random Porous Media". In: *Journal of Physical Chemistry* 97, pp. 3813–3818 (cit. on p. 81).
- Li, H., Pellacini, F., and Torrance, K. E. (2005). "A Hybrid Monte Carlo Method for Accurate and Efficient Subsurface Scattering". In: *Proc. Eurographics Symposium on Rendering*, pp. 283–290 (cit. on p. 92).
- MacAdam, D. L. (1935a). "Maximum Visual Efficiency of Colored Materials". In: *Journal of the Optical Society of America* 25.11, pp. 361–367 (cit. on p. 43).
- MacAdam, D. L. (1935b). "The Theory of the Maximum Visual Efficiency of Colored Materials". In: *Journal of the Optical Society of America* 25.8, pp. 249–249 (cit. on pp. 43, 47, 50–53).
- MacDonald, J. D. and Booth, K. S. (1990). "Heuristics for ray tracing using space subdivision". In: *The Visual Computer* 6.3, pp. 153–166 (cit. on p. 78).
- Maloney, L. T. (1986). "Evaluation of linear models of surface spectral reflectance with small numbers of parameters". In: *Journal of the Optical Society of America* 3.10, pp. 1673–1683 (cit. on pp. xi, 35).
- Marschner, S. R., Jensen, H. W., Cammarano, M., Worley, S., and Hanrahan, P. (2003). "Light scattering from human hair fibers". In: *ACM Transactions on Graphics (TOG)*. Vol. 22. 3. ACM, pp. 780–791 (cit. on p. 77).
- Matsumoto, M. and Nishimura, T. (1998). "Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator". In: *ACM Trans. Model. Comput. Simul.* 8.1, pp. 3–30. ISSN: 1049-3301 (cit. on pp. 14, 79).
- McCamy, C. S., Marcus, H., and Davidson, J. G. (1976). "A Color-Rendition Chart". In: *Journal of Applied Photographic Engineering* 2.3, pp. 95–99 (cit. on pp. 34, 50).

## Bibliography

- McCormick, N. J. and Kuscer, I. (1973). "Singular eigenfunction expansions in neutron transport theory". In: *Advances in Nuclear Science and Technology* 7, pp. 181–282 (cit. on pp. 58, 60).
- McWhorter, D. B. and Sunada, D. K. (1977). *Ground-water hydrology and hydraulics*. Water Resources Publications, LLC (cit. on p. 75).
- Meng, J., Hanika, J., and Dachsbacher, C. (2016). "Improving the Dwivedi Sampling Scheme". In: *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 35.4, pp. 37–44 (cit. on pp. x, xii, 3–4, 56, 58, 60, 63, 66, 68–71).
- Meng, J., Papas, M., Habel, R., Dachsbacher, C., Marschner, S., Gross, M., and Jarosz, W. (2015). "Multi-Scale Modeling and Rendering of Granular Materials". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 34.4, 49:1–49:13 (cit. on pp. xi–xiii, 4, 74–75, 77, 79–92, 95, 97–99, 102).
- Meng, J., Simon, F., Hanika, J., and Dachsbacher, C. (2015). "Physically Meaningful Rendering using Tristimulus Colours". In: *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)* 34.4, pp. 31–40 (cit. on pp. x–xi, 3–4, 44–45, 47–48, 50–53, 101).
- Metropolis, N. and Ulam, S. (1949). "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247, pp. 335–341 (cit. on p. 16).
- Moon, J. T. and Marschner, S. R. (2006). "Simulating multiple scattering in hair using a photon mapping approach". In: *ACM Trans. on Graphics (Proc. SIGGRAPH)* 25.3, pp. 1067–1074 (cit. on pp. 4, 77).
- Moon, J. T., Walter, B., and Marschner, S. R. (2007). "Rendering Discrete Random Media Using Precomputed Scattering Solutions". In: *Proc. Eurographics Symposium on Rendering*, pp. 231–242 (cit. on pp. 75–76, 90, 92, 95).
- Müller, T., Papas, M., Gross, M., Jarosz, W., and Novák, J. (2016). "Efficient Rendering of Heterogeneous Polydisperse Granular Media". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 35.6 (cit. on pp. 76, 90, 100, 102).
- Nicodemus, F. E. (1965). *Directional Reflectance and Emissivity of an Opaque Surface*. Tech. rep. AD 463473. Mountain View, California: Electronic Defense Laboratories (cit. on p. 23).
- Nicodemus, F. E., Richmond, J., Hsia, J., Ginsberg, I., and Limperis, T. (1977). *Geometrical Considerations and Nomenclature for Reflectance*. Tech. rep. NBS MN-160. Washington, D.C. 20234: National Bureau of Standards (cit. on pp. 22–23).
- Ohta, N. and Robertson, A. (2006). *Colorimetry: Fundamentals and Applications*. Wiley Publishing. ISBN: 9780470094730 (cit. on pp. 34, 50–51).
- Olson, G., Miller, D., Larsen, E., and Morel, J. (2006). "Chord length distributions in binary stochastic media in two and three dimensions". In: *Journal of Quantitative Spectroscopy & Radiative Transfer* 101, pp. 269–283 (cit. on p. 81).
- Papas, M. (2015). "A computational appearance fabrication framework and derived applications". PhD thesis. ETH Zürich (cit. on pp. 4, 77).

- Perlin, K. (2002). "Improving Noise". In: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '02. San Antonio, Texas: ACM, pp. 681–682 (cit. on p. 78).
- Peter, I. and Pietrek, G. (1998). "Importance driven construction of photon maps". In: *Rendering Techniques '98*. Springer, pp. 269–280 (cit. on p. 33).
- Peytavie, A., Galin, E., Merillou, S., and Grosjean, J. (2009). "Procedural Generation of Rock Piles Using Aperiodic Tiling". en. In: *Computer Graphics Forum (Proc. Pacific Graphics)* 28.7, pp. 1801–1810 (cit. on p. 77).
- Pitman, J. (2006). *Probability*. Springer Texts in Statistics. Springer (cit. on pp. 7, 10).
- Potmesil, M. and Chakravarty, I. (1981). "A lens and aperture camera model for synthetic image generation". In: *ACM SIGGRAPH Computer Graphics* 15.3, pp. 297–305 (cit. on p. 32).
- Randrianalisoa, J. and Baillis, D. (2010). "Radiative properties of densely packed spheres in semitransparent media: A new geometric optics approach". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 111.10, pp. 1372–1388 (cit. on pp. 75–76).
- Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo Method*. 2nd ed. John Wiley & Sons, Inc. (cit. on pp. 7, 11–12).
- Rushmeier, H. E. (1988). "Realistic Image Synthesis for Scenes with Radiatively Participating Media". PhD thesis. Ithaca, NY, USA: Cornell University (cit. on p. 76).
- Salah, B. E. A. and Teich, M. C. (2007). *Fundamentals of Photonics*. John Wiley & Sons, Inc. (cit. on pp. 19–20).
- Schrade, E., Hanika, J., and Dachsbacher, C. (2016). "Sparse high-degree polynomials for wide-angle lenses". In: *Computer Graphics Forum (Proceedings of Eurographics Symposium on Rendering)* 35.4, pp. 89–97 (cit. on p. 32).
- Schröder, K. (2013). "Visual Prototyping of Cloth". PhD thesis. Bonn, Germany: Universität Bonn (cit. on p. 77).
- Schröder, K., Klein, R., and Zinke, A. (2011). "A Volumetric Approach to Predictive Rendering of Fabrics". In: *Computer Graphics Forum (Proceedings of EGSR 2011)* 30.4, pp. 1277–1286 (cit. on p. 77).
- Schröder, K., Zhao, S., and Zinke, A. (2012). "Recent Advances in Physically-Based Appearance Modeling of Cloth". In: *ACM SIGGRAPH Asia 2012: Course Notes* (cit. on p. 77).
- Schrödinger, E. (1920). "Theorie der Pigmente von größter Leuchtkraft". In: *Annalen der Physik* 367.15, pp. 603–622 (cit. on p. 35).
- Singh, B. and Kaviany, M. (1992). "Modelling radiative heat transfer in packed beds". In: *International Journal of Heat and Mass Transfer* 35.6, pp. 1397–1405 (cit. on p. 76).

- Skoge, M., Donev, A., Stillinger, F. H., and Torquato, S. (2006). "Packing hyperspheres in high-dimensional Euclidean spaces". In: *Physical Review E* 74.4, p. 041127 (cit. on pp. 78, 82).
- Smits, B. (1999). "An RGB-to-spectrum conversion for reflectances". In: *Journal of Graphics Tools* 4.4, pp. 11–22 (cit. on pp. x–xi, 3, 43–47, 50–53).
- Song, C., Wang, P., and Makse, H. A. (2008). "A phase diagram for jammed matter". In: *Nature* 7195, pp. 629–632 (cit. on p. 75).
- Spanier, J. and Gelbard, E. M. (1969). *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley Publishing Company (cit. on pp. 7–12, 14–16, 33, 39–40).
- Stockman, A. and Sharpe, L. T. (2000). "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype". In: *Vision Research* 40.13, pp. 1711–1737. ISSN: 0042-6989 (cit. on pp. 35–36).
- Torquato, S. (2001). *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Interdisciplinary Applied Mathematics. Springer (cit. on p. 87).
- Torquato, S. and Lu, B. (1993). "Chord-length distribution function for two-phase random media". In: *Physical Review E* 47 (4), pp. 2950–2953 (cit. on p. 81).
- Veach, E. (1997). "Robust Monte Carlo Methods for Light Transport Simulation". PhD thesis. Stanford University (cit. on pp. xii, 17–19, 33).
- Veach, E. and Guibas, L. (1995). "Bidirectional estimators for light transport". In: *Photorealistic Rendering Techniques*. Springer, pp. 145–167 (cit. on p. 41).
- Veach, E. and Guibas, L. J. (1995). "Optimally combining sampling techniques for Monte Carlo rendering". In: *Proc. SIGGRAPH*, pp. 419–428 (cit. on pp. 17–18, 56).
- Veach, E. and Guibas, L. J. (1997). "Metropolis light transport". In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., pp. 65–76 (cit. on p. 40).
- Vorba, J., Karlík, O., Šik, M., Ritschel, T., and Krivánek, J. (2014). "On-line Learning of Parametric Mixture Models for Light Transport Simulation". In: *Proc. SIGGRAPH* 33.4, 101:1–101:11. ISSN: 0730-0301 (cit. on p. 72).
- Wald, I. (2007). "On fast construction of SAH-based bounding volume hierarchies". In: *Interactive Ray Tracing, 2007. RT'07. IEEE Symposium on*. IEEE, pp. 33–40 (cit. on p. 78).
- Wald, I., Woop, S., Benthin, C., Johnson, G. S., and Ernst, M. (2014). "Embree: A Kernel Framework for Efficient CPU Ray Tracing". In: *ACM Trans. on Graphics* 33.4, 143:1–143:8 (cit. on p. 67).
- Walter, B., Marschner, S., Li, H., and Torrance, K. (2007). "Microfacet Models for Refraction through Rough Surfaces". In: *Proc. Eurographics Symposium on Rendering*, pp. 195–206 (cit. on p. 65).

- Wang, Q., Xu, H., and Sun, Y. (2004). "Practical construction of reflectances for spectral rendering". In: *In Proceedings of the 22th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pp. 193–196 (cit. on p. 45).
- Wyszecki, G. and Stiles, W. (1982). *Color science: concepts and methods, quantitative data and formulae*. Wiley classics library. Wiley. ISBN: 9780471021063 (cit. on p. 36).
- Yan, L.-Q., Hašan, M., Jakob, W., Lawrence, J., Marschner, S., and Ramamoorthi, R. (2014). "Rendering Glints on High-Resolution Normal-Mapped Specular Surfaces". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 33.4, 116:1–116:9 (cit. on p. 99).
- Zhou, X., Li, S., and Stamnes, K. (2003). "Geometrical-optics code for computing the optical properties of large dielectric spheres". In: *Applied optics* 42.21, pp. 4295–4306 (cit. on p. 84).
- Zinke, A. and Weber, A. (2007). "Light Scattering from Filaments". In: *IEEE Transactions on Visualization and Computer Graphics* 13.2, pp. 342–356 (cit. on p. 83).
- Zirr, T. and Kaplanyan, A. S. (2016). "Real-time rendering of procedural multiscale materials". In: *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM, pp. 139–148 (cit. on p. 99).



# INDEX

- Adjoint radiative transfer equation,
  - 34
- Bias, 11
- Bidirectional path tracing, 41
- Bidirectional reflectance-distribution
  - function, 22
- Black body, 38
- Chebyshev inequality, 14
- Chromaticity, 37
- Color matching functions, 37, 46
- Color model, 36
- Color saturation, 35
- Color space, 36
- Color temperature, 38
- Consistency, 11
- Convergence plots, 17
- Correlation, 12
- Covariance
  - definition, 12
  - of independent random
    - variables, 13
- Cumulative distribution function, 9
- Diffusion length, 60
- Emitter, 28
- Equal-energy white, 38
- Expected value
  - definition, 10
  - of  $aX + b$ , 10
  - of  $h(X_1, \dots, X_N)$ , 10
  - of the mean squared error, 13
  - of the sample mean, 11
  - of the sample sum, 11
- Free-flight distribution, 26
- Gamut, 36
- Geometric term, 31
- Grain albedo, 85
- Granular materials, 73
- Hit probability, 85
- Importance, 33
- Importance sampling, 17
- Importon, 40
- Inversion method, 16
- Irradiance, 21
- Just noticeable difference, 51
- Law of large numbers, 14
- Light path, 32
- Light tracing, 41
- Luminance, 37
- MacAdam box spectra, 43
- Mean free path, 27
- Measurable space, 8
- Measure, 8
- Measurement contribution function,
  - 33
- Metamerism, 43

## INDEX

- Monte Carlo integration, 16
- Multiple importance sampling
  - Balance heuristic, 18
  - definition, 17
  - Implicit weights, 18
  - Multi sample model, 18
  - One sample model, 17
  - Power heuristic, 19
- Next event estimation, 42
  
- Packing rate, 75
- Participating media, 23
- Path integral, 33
- Path space, 33
- Path tracing, 41
- Phase function, 27
- Photon, 20, 40
- Photon energy, 20
- Photon mapping, 41
- Pinhole camera, 31
- Primary color, 36
- Probability density function
  - conditional, 10
  - definition, 9
  - joint, 9
  - marginal, 10
- Probability measure
  - definition, 8
  - joint, 9
- Projected solid angle, 21
- Purple line, 37
  
- Radiance, 22
- Radiant flux, 20
- Radiant intensity, 21
- Radiative transfer equation, 27
- Radiosity, 21
- Random event
  - certain, 8
  - definition, 8
  - impossible, 8
- Random experiment, 8
- Random variable
  - continuous, 9
  - definition, 9
  - discrete, 9
  - independent, 10
  - uniform, 14
- Random walk
  - Analog, 40
  - definition, 39
  - Non-analog, 40
- Rejection sampling, 14
  
- Sample mean, 11
- Sample variance, 14
- Sensors, 31
- Sets
  - $\sigma$ -algebra, 8
  - complement, 7
  - power set, 8
  - union, 7
- Singular eigenfunction expansion,
  - 58, 61
- Source term, 28
- Spectral energy density, 34
- Spectral locus, 37
- Spectral rendering, 38
- Spectral upsampling
  - definition, 37
- Spectrum
  - attenuation, 34
  - definition, 34
  - emission, 34
- Speed of light, 20
- Standard deviation
  - of the Monte Carlo estimator,
    - 16
  - definition, 12
  
- Teleportation scattering distribution
  - function, 82
- Teleportation transport, 74, 80
- Time to unit variance, 95
- Transmittance, 25
- Tristimulus rendering, 38
- Tristimulus values, 36

- Unified path sampling, 42
- Variance
  - definition, 12
  - of  $X + Y$ , 13
  - of the Monte Carlo estimator,  
16
- of the sample mean, 13
- Vertex connection and merging, 42
- Volume coefficient, 24
- Volume cross section, 23
- Volume density, 24
- White point, 37