University of California
Santa Barbara

Towards Practical Physical Optics Rendering

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Computer Science

by

Shlomi Steinberg

Committee in charge:

Professor Lingqi Yan, Chair
Professor Ömer Eğecioğlu
Professor David Weld

June 2023

The Dissertation of Shlomi Steinberg is approved.

Professor David Weld

Professor Ömer Eğecioğlu

Professor Lingqi Yan, Committee Chair

March 2023

Towards Practical Physical Optics Rendering

by

Shlomi Steinberg

Acknowledgements

I would like to express my gratitude to my advisor, Lingqi Yan, for his insight and patience through all stages of this academic endeavour. I also would like to thank other mentors and noteworthy collaborators who have supported me during my doctoral studies and shared their expertise: Ömer Eğecioğlu, Nir Davidson, and Matt Pharr. None of this would have been possible without the motivation and eternal support of my wife, Jinrui, who has been there for me through thick and thin, and provided endless inspiration.

Curriculum Vitæ

Shlomi Steinberg

Education

2023	Ph.D. in Computer Science (Expected) University of California, Santa Barbara.
2017	MSc in Mathematics and Computer Science Weizmann Institute of Science, Israel.
2015	BSc in Mathematics and Computer Science Open University, Israel.

Select Awards

2022-2023 NVIDIA PhD Fellowship

Publications

A Generalized Ray Formulation For Wave-Optics Rendering *(in review)*

Steinberg S., Ramamoorthi R., Bitterli B., D'eon E., Yan L., Pharr M.

ACM Transactions on Graphics • Jul 2023

Towards Practical Physical-Optics Rendering

Steinberg S., P. Sen, Yan L.

ACM Transactions on Graphics (Proceedings of SIGGRAPH 2022) • **Best paper honourable mention** • Jul 2022

Two-Mirror Compact System for Ideal Concentration of Diffuse Light

Steinberg S., N. Bokor, N. Davidson

Journal of the Optical Society of America A • Mar 2022

Physical Light-Matter Interaction in Hermite-Gauss Space

Steinberg S., Yan L.

ACM Transactions on Graphics (Proceedings of SIGGRAPH ASIA 2021) • Nov 2021

A Generic Framework for Physical Light Transport

Steinberg S., Yan L.

ACM Transactions on Graphics (Proceedings of SIGGRAPH 2021) • Jul 2021

Rendering of Subjective Speckle Formed by Rough Statistical Surfaces

Steinberg S., Yan L.

ACM Transactions on Graphics • Feb 2022

Accurate Rendering of Liquid-Crystals and Inhomogeneous Optically Anisotropic Media

Steinberg S.

ACM Transactions on Graphics • Apr 2020

Analytic Spectral Integration of Birefringence-Induced Iridescence

Steinberg S.

Eurographics Symposium on Rendering 2019 • Jul 2019

Efficient Distributed Execution of Multi-component Scenario-based Models

Steinberg S., Greenyer J., Gritzner D., Harel D., Katz G. and Marron A.

Model-Driven Engineering and Software Development • Jul 2018

Distributing Scenario-Based Models: A Replicate-and-Project Approach

Steinberg S., Greenyer J., Gritzner D., Harel D., Katz G. and Marron A.

MODELSWARD 2017 • Feb 2017

Please Note: Text and figures from these papers are used and appear in this dissertation

Abstract

Towards Practical Physical Optics Rendering

by

Shlomi Steinberg

Today's rendering and light transport frameworks are formulated strictly under the context of ray optics. However, many applications often call for taking the wave nature of light into consideration. This has been historically challenging, mainly because wave-optical descriptors of light are neither linear nor local, frustrating the applications of classical rendering and path-tracing techniques. Today, such path-tracing techniques power most of the complex computer-generated content in films and movies, however wave solvers struggle with scenes that would be considered exceedingly simple by the rendering community. The purpose of this work is to generalize classical path-tracing techniques to physical optics. We introduce *physical light transport* (PLT), a framework which represents the wave nature of light globally in a scene, and is consistent with Maxwell's theory of electromagnetism. We show how a wave-optical system can be sampled using *generalized rays*, i.e. local and linear classical-like point queries of the wave-optical system. These generalized rays enable the application of essentially arbitrary sampling methods, but to do so under rigorous wave optics. Therefore, our work serves as a link between computer graphics' path tracing and computational optics methods.

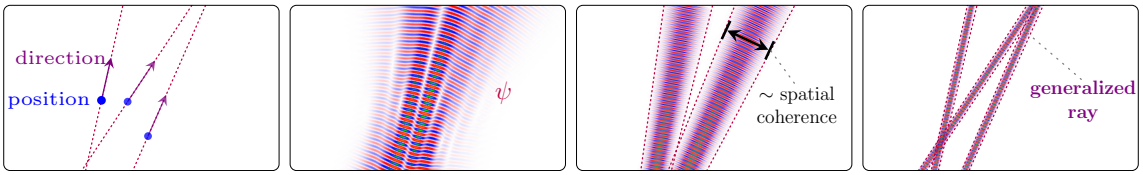
Contents

Chapter 1

Introduction

Light transport is a theoretical underpinning of rendering and computer graphics, formalising the mathematical model that drives modern rendering and image synthesis, as well as other related studies such as inverse rendering and computer vision. For many years rendering and light transport have been formulated under the context of a simplistic understanding of light: as a collection of rays. This *ray optical* view (see Fig. 1.1) is convenient, as a light ray is perfectly localized in space—it has precise position and direction of propagation—and rays add up linearly in terms of their observable radiometric properties (e.g., radiance). These traits, *locality* and *linearity*, are essential for rendering: they allow formulating a linear rendering equation; and, enable the application of powerful sampling techniques in order to sample paths through the scene, connecting a light source to the sensor. In addition, dedicated ray-tracing hardware today facilitates applying these path tracing techniques in real-time [10].

The above comes at a price: ray optics ignores the wave nature of light, meaning wave-interference effects cannot be reproduced. Such effects include the colourful glints that appear when light is scattered by scratches in metal; the colour of the wings and scales of some species of insects, snakes and fish; the grating patterns that appear on compact disks



(a) ray optics (b) wave optics (c) partially-coherent beams (d) generalized rays

Figure 1.1: **Illustration: different pictorial views of light's physics.** (a) Most often, light transport is formulated in terms of ray optics, where light is composed of particles—light rays—that can be perfectly *localized* in space: i.e. we may precisely specify the position and direction of propagation of each ray. Furthermore, these rays superpose *linearly*, i.e., when two rays arrive at a surface or a sensor, we add up their respective irradiances linearly (no interference takes place). (b) Under wave optics—a more rigorous understanding of light—the basic descriptor of light is the *wave function* ψ. Unlike a light ray under ray optics, perfect localization of light is prohibited by the uncertainty relation, and the wave function now acts as a global descriptor of light. Moreover, linearity, in terms of light's observable properties, no longer holds. The loss of locality and linearity make wave optics incompatible with path tracing algorithms. (c) By taking advantage of light's limited optical coherence (the spatial extent over which wave-interference effects may arise) light may be understood as mutually-incoherent beams. In this form, the descriptors of light regain linearity (on superposition of beams), and to some degree locality, up to the extent of light's optical coherence. However, as we describe partially-coherent light, the *sampling problem* manifests, see Fig. 1.2. (d) We solve this problem by decomposing light into coherent modes: *generalized rays* that are pair-wise mutually incoherent, thus linear; and local to a greater degree compared to other models, as their extent is only constrained by light's wavelength. We use these generalized rays to sample paths effectively.

(CD), holographs and LCD screens; goniochromatic surfaces and iridescence effects in coated or painted materials and in layered materials such as the human skin; pleochroism and stress birefringence in framed glass and molded plastics; and, the appearance of a layer of oil or metallic oxide on a surface.

A wave-optical formulation of light, in sharp contrast to the ray-optical view, admits neither a clear locality nor linearity: Under wave optics, the *wave function* replaces the ray as the descriptor of light (Fig. 1.1b). However, it is well-known that the wave function admits intrinsic *uncertainty*: its extent over spatial position and frequency (direction of propagation) may not be arbitrarily small, implying light may no longer be perfectly localized. Unlike the ray, which permits local sampling of light's behaviour, the wave

function serves as a *global* descriptor of light. Furthermore, superposition of wave functions is not linear in terms of their intensity and radiometry, but is *bilinear*: it admits an interference term. The loss of locality and linearity greatly frustrates applications of path tracing under wave optics. While path-tracing techniques power most of the complex computer-generated content in films and movies, wave solvers struggle with scenes that would be considered exceedingly simple by the rendering community.

Faithful reproduction of wave effects, as observed in the physical world, is further complicated by the fact that virtually all the light that we encounter in daily life and render with is *partially coherent*. That is, this light's underlying field contains seemingly random fluctuations and its ability to produce observable wave-interference effects is spatially and temporally limited. Therefore, even though interference happens all the time and all around us, many of these processes produce interference effects with very poor observable contrast, and are not visible; while others may produce observable interference effects, with varying contrast that depends both on the material and light properties.

Over the past decade, a significant effort has been made to reproduce and emulate the appearance of some materials that arises due to wave effects. To model such effects, electromagnetism has been extensively employed in computer graphics, but only applied locally, in order to study a particular effect. This works reasonably well on a nanoscopic scale, e.g., to reproduce thin-film interference. On the other hand, the highly chaotic nature of natural light (i.e., its limited optical coherence) renders deterministic models of light inadequate in describing the propagation of such light, and thus difficult to apply globally. But a global treatment is desired: These coherence properties of light—the light's ability to superpose and induce wave interference—are in fact a global process, as coherence arises due to propagation of light and is then altered by the interaction of light with matter and media.

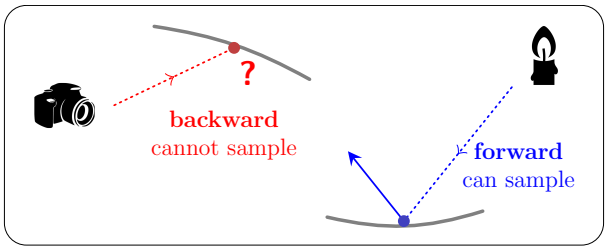


Figure 1.2: **The sampling problem.** To importance sample partially-coherent BSDFs (or light-matter interactions in general), the coherence properties of light are needed. However, these properties depend on the light source, and change on propagation and interaction with matter. Therefore, these properties are not available, and are difficult to predict or estimate, when tracing paths *backwards*, i.e. starting from a camera. For a variety of reasons, path tracers trace paths backward-only, or bidirectionally (forward and backward), hence limiting our ability to perform backward sampling is a real problem. This sampling problem applies to any framework where the descriptor of light quantifies light's optical coherence.

Physical light transport

Recognizing the crucial role that the partial coherence of light plays in driving the observable appearance of materials, in this work we introduce *physical light transport* (PLT) as a framework that enables practical wave-optical light transport. PLT generalizes classical quantities and descriptors of light into a formalism that is consistent with electromagnetism. Formally quantifying the effects of the partial-coherence of light on the optical response of materials is a core motivation for PLT.

However, the moment our descriptor of light quantifies the coherence of light and BSDFs (*bidirectional scattering distribution functions*) are functions of these coherence properties, a **sampling problem** comes into effect (see Fig. 1.2): These coherence properties depend on the source, and change on propagation and on interaction with matter. Hence, they are only known when tracing light beams from a light source, but for a variety of practical reasons, path tracing is often done *backwards*—from the sensor—or bidirectionally—backward and forward. Because BSDFs depend on the coherence properties of light, which are unknown when tracing paths backward, it is unclear how to importance sample these BSDFs: it would require predicting in some fashion what the

coherence of light would be at the interaction region, a difficult endeavour. One approach that we have preciously proposed [46] is to fix a global lower limit on the coherence of light. However that limit needs to be conservative, and while this approach works for some materials, it yields poor sampling for materials that admit strong wave-interference effects, which precisely are the materials of interest. Good sampling is critical for efficient path tracing, and solving the sampling problem is the core motivation for this work.

Evidently, if we sample partially-coherent interactions, the coherence properties of light need to be known. We choose to side-step the issue by avoiding partial coherence altogether in the formulation of path sampling. We take inspiration from the significant effort made in quantum mechanics in attempts to regain the "grainy" view of classical mechanics (with intrinsic locality and linearity). In Chapter 5, we draw upon such formulations in order to introduce a *generalized ray*: a form of a coherent state that is the closest wave-optical analogue of the classical ray. Light can always be understood as a collection of such mutually-incoherent generalized rays, thereby regaining linearity and locality (as physically allowed by the uncertainty relation). This decomposition of light into generalized rays is a form of coherent-modes decomposition [32]. In Section 5.3, we then present a simple, but powerful, two-step *sample-solve* algorithm, which is our primary contribution:

1. **Sample** — Because each generalized ray is perfectly coherent, these generalized rays can be used to sample paths in a coherence-agnostic manner, solving the sampling problem. The generalized ray plays a similar role to the classical ray under ray optics: it allows *point sampling* of light's behaviour.

2. **Solve** — Once a path connecting a light source to the sensor has been sampled, we apply PLT tools in order to *solve* for the partially-coherent transport over that path. This is possible due to the exact, formal relation between generalized rays

and optical coherence that we show.

This sample-solve approach enables efficient sampling, and, as a consequence, we show in Section 5.4 that we are able to do wave-optical rendering with complex scenes, and at interactive performance. As a matter of fact, the convergence performance of our approach is multiple orders-of-magnitude greater compared to the state-of-the-art. Moreover, as our generalized rays extend classical rays to wave optics (in the sense that locality and linearity is regained), such wave-optical rendering can be achieved with relatively minor modifications of an existing, classical path tracer: only the BSDFs need to be redesigned. More generally, sample-solve serves as a link between path tracing techniques and computational optics methods: The sample step, as formulated, applies in highly-general settings, and allows the application of a wide class of sampling techniques; while the solve step is unconstrained by sampling concerns. This link may serve to extend the applicability of some computational optics tools to complex scenes and environments, as well as to additional computer graphics applications.

In this thesis we will start (Chapter 3) with deriving the tools needed to properly quantify partially-coherent light. We will discuss optical coherence theory, sourcing, diffraction and propagation. We will then introduce our matter model (Chapter 4), discuss how partially-coherent light interacts with matter, and present the relevant light-matter interaction formulae. Finally, we turn to the practical side of things and discuss sampling the wave-optical phase space with generalized rays (Chapter 5). We will show how partially-coherent light arises as a bundle of such rays, and present our sample-solve approach that brings together wave-optical sampling and partially-coherent light transport.

Chapter 2

Background

We start with an overview of relevant topics in mathematical analysis (Section 2.2) and stochastic processes (Section 2.3), with the discussion serving as a brief review of the required background. We proceed with presenting the necessary theoretical foundations of modern electrodynamics and optical coherence theory in a comprehensive fashion, with full derivations of most relevant results and theorems.

2.1 Preliminaries

2.1.1 Notation

In this subsection we fix the notation that will be used throughout the dissertation.

Complex numbers

We denote the complex space as $\mathbb{C}$ and the imaginary unit as i. Given a complex number $z = a + \mathrm{i}b \in \mathbb{C}$, with $a, b \in \mathbb{R}$ real numbers, we define the real part and imaginary

part operators as: $\mathrm{Re}\, z = a$ and $\mathrm{Im}\, z = b$. Then, absolute value of a complex number is

$$|z| = \sqrt{\mathrm{Re}\{z\}^2 + \mathrm{Im}\{z\}^2} = \sqrt{a^2 + b^2} \ . \tag{2.1}$$

Any complex number can be written in polar form using the well-known *Euler's formula*:

$$z = re^{\mathrm{i}\theta} = r(\cos\theta + \mathrm{i}\sin\theta) \ , \tag{2.2}$$

where $r = |z|$ and $\theta = \arg(z) \in (-\pi, \pi]$ is known as the *argument* of z and is uniquely defined by z. The $\star$ superscript denotes the complex conjugate:

$$z^\star = a - \mathrm{i}b \ . \tag{2.3}$$

The complex conjugate is distributive over complex addition and multiplication, and clearly $zz^\star = |z|^2 = |z^\star|^2$.

Vectors

We denote vectors and vector-valued functions with an arrow diacritic, $\vec{\boldsymbol{k}}$. A hat diacritic over a vector denotes the unit vector, i.e. $\hat{\boldsymbol{k}} = \vec{\boldsymbol{k}}/|\vec{\boldsymbol{k}}|$. Then, the cartesian three dimensional basis is $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ and any vector $\vec{\boldsymbol{v}} \in \mathbb{C}^3$ can be written as

$$\vec{\boldsymbol{v}} = v_x\hat{\boldsymbol{x}} + v_y\hat{\boldsymbol{y}} + v_z\hat{\boldsymbol{z}} \tag{2.4}$$

where the shorthand $v_e = \hat{\boldsymbol{e}} \cdot \vec{\boldsymbol{v}}$ is used to express a component of $\vec{\boldsymbol{v}}$ is some arbitrary basis (inferred from context). The gradient, divergence, curl and Laplacian operators, in

cartesian coordinates, are:

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z} \tag{2.5}$$

$$\boldsymbol{\nabla} \cdot \vec{\boldsymbol{v}} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \tag{2.6}$$

$$\boldsymbol{\nabla} \times \vec{\boldsymbol{v}} = \hat{\boldsymbol{x}}\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) + \hat{\boldsymbol{y}}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) + \hat{\boldsymbol{z}}\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) \tag{2.7}$$

$$\nabla^2\vec{\boldsymbol{v}} = \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_z}{\partial z^2}\ , \tag{2.8}$$

respectively. A few useful vector identities are the triple product identities and some vector calculus identities that we encounter later on:

$$\vec{\boldsymbol{a}} \times \left(\vec{\boldsymbol{b}} \times \vec{\boldsymbol{c}}\right) = \vec{\boldsymbol{b}}(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{c}}) - \vec{\boldsymbol{c}}\left(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}\right) \tag{2.9}$$

$$\vec{\boldsymbol{a}} \cdot \left(\vec{\boldsymbol{b}} \times \vec{\boldsymbol{c}}\right) = \vec{\boldsymbol{b}} \cdot (\vec{\boldsymbol{c}} \times \vec{\boldsymbol{a}}) = \vec{\boldsymbol{c}} \cdot \left(\vec{\boldsymbol{a}} \times \vec{\boldsymbol{b}}\right) \tag{2.10}$$

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \vec{\boldsymbol{v}}) = 0 \tag{2.11}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = 0\ . \tag{2.12}$$

Functions

Let $\mathbb{F}$ and $\mathbb{F}'$ be fields (e.g. $\mathbb{R}$ or $\mathbb{C}$). We say that a function $f : \mathbb{F} \to \mathbb{F}'$ is absolutely integrable, or f is an L^1-function, if the following holds:

$$\int_{\mathbb{F}} \mathrm{d}x\, |f(x)| < \infty\ , \tag{2.13}$$

that is, the integral converges. Similarly, we say that f is square integrable, or f is an L^2-function, if the following holds:

$$\int_{\mathbb{F}} \mathrm{d}x\, |f(x)|^2 < \infty\ . \tag{2.14}$$

Any complex-valued function $f : \mathbb{F} \to \mathbb{C}$, can be decomposed into two real-valued functions:

$$f(w) = \phi_{\mathrm{r}}(w) + \mathrm{i}\phi_{\mathrm{i}}(w) \ , \tag{2.15}$$

such that $\phi_{\mathrm{r}} = \mathrm{Re}\{f\} : \mathbb{F} \to \mathbb{R}$ and $\phi_{\mathrm{i}} = \mathrm{Im}\{f\} : \mathbb{F} \to \mathbb{R}$. A complex-valued function $f(w)$ is then said to be *Hermitian* when $\mathrm{Re}\, f$ is an even function and $\mathrm{Im}\, f$ is an odd function, that is the following holds

$$f(w)^\star = f(-w) \ . \tag{2.16}$$

E.g., $e^{\mathrm{i}\theta}$ as a function of θ is Hermitian, hence $(e^{\mathrm{i}\theta})^\star = e^{-\mathrm{i}\theta}$.

Given an operator $\mathscr{L}$, we say that ψ is an *eigenfunction* of $\mathscr{L}$, if when acting upon ψ the operator $\mathscr{L}$ serves to only scale ψ, that is if the following holds:

$$\mathscr{L}\psi = \lambda\psi \tag{2.17}$$

in which case λ is the *eigenvalue* associated with the eigenfunction ψ. For example, as $\mathrm{d}^2/\mathrm{d}x^2 \cos x = -\cos x$, we say that the function $\cos x$ is the eigenfunction of the 2^{nd}-order derivative operator $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, with -1 being the associated eigenvalue.

A useful inequality is Schwarz's integral inequality: For any f, g L^2-functions

$$\left| \int_{\mathbb{R}} fg \right| \leq \left(\int_{\mathbb{R}} |f|^2 \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |g|^2 \right)^{\frac{1}{2}} \ , \tag{2.18}$$

which is the functional analogue of the triangle inequality.

Kronecker and Dirac delta

We denote the *Kronecker delta* as

$$\delta_{ij} \triangleq \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} . \tag{2.19}$$

E.g., for any orthonormal basis $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots\}$ it holds that $\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \delta_{ij}$. The (one-dimensional) *Dirac delta* $\delta(x)$ is an important generalized function that is defined as an impulse response, that admits the "filtering" property:

$$\forall x \neq 0 \rightarrow \delta(x) = 0 \qquad \text{and} \qquad f(x) = \int_{-\infty}^{\infty} \mathrm{d}x' \, f(x')\delta(x' - x) \, . \tag{2.20}$$

Matrices

Our notation for matrices is quite standard. The vector space $\mathbb{F}^{n \times m}$ denotes the space of $n \times m$ matrices over the field $\mathbb{F}$. We denote the identity matrix as $\boldsymbol{I}$, with the dimensions implied by context. A common source of ambiguity is the conjugate transpose operator, which we denote by the dagger, viz. $\boldsymbol{A}^\dagger$.

Additional definitions

The convolution between two functions $\boldsymbol{f}$ and $\boldsymbol{g}$ is defined as follows:

$$(f * g)(t) \triangleq \int_{-\infty}^{\infty} \mathrm{d}t' \, f(t - t')g(t') \, . \tag{2.21}$$

Commonly used in optics are rectangular and circular aperture functions:

$$\operatorname{rect}(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \qquad \operatorname{circ}(r) = \begin{cases} 1 & r \leq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{2.22}$$

2.2 Functional Analysis

2.2.1 Fourier Analysis

Fourier analysis is a crucial tool across a broad range of scientific and engineering applications, in particular in the analysis of time-harmonic fields and their statistics, which is at the centre of our discussion. The *Fourier transform* is a form of an integral transform, where a function is mapped from its original domain to another domain, or put equivalently: From one representation to another. The Fourier transform decomposes a function, or signal, into the frequencies that build up the original function. There are multiple conflicting definitions of the Fourier transform, with different disciplines favouring different normalization constants and units. We define the Fourier transform with unitary, angular frequency kernels:

$$\check{f}(\omega) = \mathscr{F}\{f\} \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}t \, f(t) e^{-\mathrm{i}\omega t} \tag{2.23}$$

$$f(t) = \mathscr{F}^{-1}\{\check{f}\} \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}\omega \, \check{f}(\omega) e^{\mathrm{i}\omega t} , \tag{2.24}$$

with natural extension to higher dimensions. This definition is only valid when the integrals converge and a sufficient conditions is that f and its Fourier transform are L^1-functions. The Fourier transform pair will generally be denoted with a wedge diacritic. The Fourier transform can also be defined *in the generalized sense* in special cases, for

example, for the Dirac delta generalized function.

Real functions

Let f be a real-valued signal and $\check{f}$ its Fourier transform. Clearly $\mathscr{F}^{-1}\{\check{f}\}^\star = \mathscr{F}^{-1}\{\check{f}\}$, and it is easy to show that

$$f(t) = \mathscr{F}^{-1}\{\check{f}\} = \sqrt{\frac{2}{\pi}}\,\mathrm{Re}\int_0^\infty \mathrm{d}\omega\, \check{f}(\omega)e^{\mathrm{i}\omega t}\ . \tag{2.25}$$

The above half-space formulation of the inverse Fourier transform suggests that given a real signal, only the non-negative frequencies of the transform pair are needed to recover the signal and the negative frequencies contain no additional information.

Theorems

We list here additional important results and theorems in Fourier analysis. Derivations can be found in introductory textbooks on Fourier analysis. We assume Fourier transformable functions f and g throughout.

Theorem 2.2.1.1 (Parseval's Theorem)**.**

$$\int_{-\infty}^{\infty} \mathrm{d}t\, f(t)g(t)^\star = \int_{-\infty}^{\infty} \mathrm{d}\omega\, \check{f}(\omega)\check{g}(\omega)^\star\ . \tag{2.26}$$

From Parseval's theorem we immediately conclude that $\int |f|^2 = \int |\check{f}|^2$, which physically can be interpreted as the claim that the total energy in a signal is invariant under the Fourier transform. A useful way to transform convolutions into multiplications is provided by the following theorem.

Theorem 2.2.1.2 (Convolution Theorem)**.**

$$\mathscr{F}\{f * g\} = \mathscr{F}\{f\}\,\mathscr{F}\{g\} \tag{2.27}$$

$$\mathscr{F}\{fg\} = \mathscr{F}\{f\} * \mathscr{F}\{g\}\ , \tag{2.28}$$

that is, a convolution in the time-domain is equivalent to multiplication in the frequency-domain, and vice versa. A direct consequence of the convolution theorem is the *cross-correlation theorem*, which makes contact with statistical analysis by associating the cross-correlation of f and g to their Fourier transforms. The cross-correlation is defined as:

$$\mathrm{Corr}\{f, g\}(t) \triangleq \int_{-\infty}^{\infty} \mathrm{d}\tau\, f(t+\tau) g(\tau)^{\star}\ . \tag{2.29}$$

Theorem 2.2.1.3 (Cross-Correlation Theorem)**.**

$$\mathrm{Corr}\{f, g\}(t) = \mathscr{F}^{-1}\{\mathscr{F}\{f\}\,\mathscr{F}\{g\}^{\star}\}\ . \tag{2.30}$$

Proof. Using the convolution theorem:

$$\begin{aligned}\mathrm{Corr}\{f, g\}(t) &= \int_{-\infty}^{\infty} \mathrm{d}\tau\, f(t+\tau) g(\tau)^{\star} = f(t) * g(-t)^{\star} \\ &= \mathscr{F}^{-1}\{\mathscr{F}\{f\}\,\mathscr{F}\{g\}^{\star}\}\end{aligned} \tag{2.31}$$

as desired. □

Fourier transforms of important functions

A few useful (temporal) Fourier transform identities are listed here. Given $a \in \mathbb{R}$, transform pairs of elementary functions are:

$$\mathscr{F}\{\delta(t-a)\} = \frac{1}{\sqrt{2\pi}} e^{\mathrm{i}a\omega} \tag{2.32}$$

$$\mathscr{F}\left\{e^{\mathrm{i}at}\right\} = \sqrt{2\pi}\delta(a+\omega) \tag{2.33}$$

$$\mathscr{F}\{\sin(at)\} = \mathrm{i}\frac{\sqrt{\pi}}{\sqrt{2}}[\delta(\omega-a) - \delta(\omega+a)] \tag{2.34}$$

$$\mathscr{F}\{\cos(at)\} = \frac{\sqrt{\pi}}{\sqrt{2}}[\delta(\omega-a) + \delta(\omega+a)] \tag{2.35}$$

$$\mathscr{F}\left\{e^{-(at)^2}\right\} = \frac{\sqrt{2}}{|a|} e^{\frac{-\omega^2}{4a^2}} \ . \tag{2.36}$$

Transform pairs of simple aperture functions:

$$\mathscr{F}\{\mathrm{rect}(t)\} = \frac{\sqrt{\pi}}{\sqrt{2}}\,\mathrm{sinc}\left(\frac{\omega}{2}\right) \tag{2.37}$$

$$\mathscr{F}\{\mathrm{circ}\,(|\vec{\boldsymbol{r}}|)\} = \frac{\sqrt{2\pi}}{|\vec{\boldsymbol{k}}|}\mathrm{J}_1\Big(|\vec{\boldsymbol{k}}|\Big) \ , \tag{2.38}$$

where J_1 is the Bessel function of first kind. The transform of the circular aperture is known as the *Airy disk*.

2.2.2 Method of Green Functions

The method of *Green functions* is a highly useful method for solving linear partial differential equations. Consider a linear differential operator $\mathscr{L}$ and an inhomogeneous partial differential equation of time and space:

$$\mathscr{L}\ \Psi(\vec{\boldsymbol{r}}, t) = -f(\vec{\boldsymbol{r}}, t) \ , \tag{2.39}$$

where f is a known *source function* and we would like to solve for $\Psi(\vec{\boldsymbol{r}}, t)$. To that end, we introduce G, the Green function associated with the operator $\mathscr{L}$, defined as the elementary impulse

$$\mathscr{L}\, G\big(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}'; t, t'\big) = -\delta^3(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}')\delta(t - t') \ . \tag{2.40}$$

If we were to find an analytic form for G, then the solution Ψ to the inhomogeneous partial differential equation immediately becomes

$$\Psi(\vec{\boldsymbol{r}}, t) = -\int_{-\infty}^{\infty} \mathrm{d}t' \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\boldsymbol{r}}'\, G\big(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}'; t, t'\big) f\big(\vec{\boldsymbol{r}}', t'\big) \ , \tag{2.41}$$

which is readily verifiable by applying $\mathscr{L}$ to both sides. That is, using the Green function the solution Ψ is expressed as a sum of point influences over different spatial and temporal points weighted by f. Note that a Green function is associated with a specific linear differential operator, and might not be unique. When $\mathscr{L}$ is translation invariant, G can be written as $G(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'; t - t')$, and the solution becomes a four-dimensional space and time convolution, viz.

$$\Psi(\vec{\boldsymbol{r}}, t) = -\iiiint \mathrm{d}t'\, \mathrm{d}^3\vec{\boldsymbol{r}}'\, G\big(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'; t - t'\big) f\big(\vec{\boldsymbol{r}}', t'\big) = -G * f \ . \tag{2.42}$$

Retarded Green function

The wave operator, $\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$, is an important linear differential operator that will be employed extensively later on when we discuss electromagnetic waves. The Green function for the wave operator admits the following analytic form (see Zangwill [61] for

derivations)

$$G_{\pm}\big(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}';t,t'\big)=\frac{\delta\big(t-t'\pm\frac{1}{c}\big|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\big|\big)}{4\pi\big|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\big|}\ , \tag{2.43}$$

which are known as the advanced (G_+) and retarded (G_-) Green functions. Physically, it is the retarded function that is more compelling as the advanced function "looks forward in time" to compute a solution. We assume the following natural boundary conditions: $\Psi\to 0$ as $\vec{\boldsymbol{r}}\to\infty$ or $t<0$. Then, we choose the Green function corresponding to the wave operator with the listed boundary conditions to be the retarded function and write it as $G_-\big(\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}';t-t'\big)$. The solution becomes

$$\begin{aligned}\Psi(\vec{\boldsymbol{r}},t)&=-\int_{-\infty}^{\infty}\mathrm{d}t'\int_{\mathbb{R}^3}\mathrm{d}^3\vec{\boldsymbol{r}}'\,G\big(\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}';t-t'\big)f\big(\vec{\boldsymbol{r}}',t'\big)\\&=\frac{1}{4\pi}\int_{\mathbb{R}^3}\mathrm{d}^3\vec{\boldsymbol{r}}'\,\frac{f\big(\vec{\boldsymbol{r}}',t-\frac{1}{c}\big|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\big|\big)}{\big|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\big|}\ .\end{aligned} \tag{2.44}$$

Free-space Green function

The Fourier domain counterpart of the wave operator is the Helmholtz operator, ∇^2+k^2. We call the associated Green function the *free-space Green function* and denote it G_0. To obtain a unique solution we impose the boundary condition known as the Sommerfeld radiation condition, viz.

$$\lim_{r\to\infty} r\left(\frac{\partial}{\partial r}-\mathrm{i}k\right)G_0=0\ , \tag{2.45}$$

where the derivative is in spherical coordinates. The Sommerfeld radiation condition implies that a point impulse solution must behave like an outgoing propagating spherical wave. Noting that the Helmholtz operator is also translation invariant, the free-space

Green function can be shown to be [61]

$$G_0\big(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\big) = \frac{1}{4\pi\big|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\big|} e^{\mathrm{i}k\left|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\right|} \ . \tag{2.46}$$

2.3 Stochastic Processes

Random variables naturally generalize to random (or stochastic) processes, where the uncertain values—the possible outcomes from an experiment—are functions instead of numbers. Consider a physical process that undergoes random fluctuations. For example, a gas-discharge lamp contains a large collection of excited atoms which randomly drop to a lower energy state and in the process emit light energy in the form of a photon. Such spontaneous emission is a stochastic process where photons with distinct energies have different probabilities—possibly varying over time—of being emitted (and those probabilities define the emission spectrum). Curious readers are encouraged to refer to more comprehensive sources: Miller [34] for an introductory textbook on random processes, and Goodman [20] for a statistical optics perspective.

We now formally define a stochastic process: Let Ω be a (finite or infinite) sample space with an accompanying probability measure. Then, a stochastic process $U(t)$ is defined as an ensemble of sample functions, called *realizations*, and can be written as $U = \{{}^{\zeta}u(t)\}$ with $\zeta \in \Omega$. Note that a stochastic process implies the existence of the underlying state space and measure, however for simplicity we omit this dependence from the notation. Each realization ${}^{\zeta}u(t)$ can be thought of as a possible outcome of a singular experiment that measures the stochastic process (e.g., the observed photons and their energy emitted over time), and is typically a function of time. Being time-dependent

processes, it is natural to define the *time-averaging operator*:

$$\left\langle {}^{\zeta}u \right\rangle_{\mathrm{t}} \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \, {}^{\zeta}u(t) \ , \tag{2.47}$$

as well as the *time cross-correlation function*:

$$\begin{aligned} \tilde{\Gamma}_{\zeta\xi}(\tau) &\triangleq \left\langle {}^{\zeta}u(t+\tau) \, {}^{\xi}u(t)^{\star} \right\rangle_{\mathrm{t}} \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \, {}^{\zeta}u(t+\tau) \, {}^{\xi}u(t)^{\star} \ , \end{aligned} \tag{2.48}$$

which describes the statistical similarity of a realization compared with a time-shifted realization. The spacial case of $\tilde{\Gamma}_{\zeta\zeta}$ is called the *time autocorrelation function*. We assume the limits above exist for any realization of a well-behaved physical process.

At times it is practical to characterize the stochastic process via its joint distributions: Let $p_U({}^{\zeta_1}u, \ldots, {}^{\zeta_n}u; t_1, \ldots, t_n)$ be the joint probability density function of $n \geq 1$ realizations taken at fixed times $t_1, \ldots, t_n$. Then, the *ensemble average* of a stochastic process can be considered as the expected outcome of the process at some instant, and is defined as the average over the constituent realizations:

$$\langle U(t) \rangle \triangleq \int_{\Omega} \mathrm{d}u \, u p_U(u; t) \ . \tag{2.49}$$

Of import is the second-order joint moment of a couple of stochastic processes U and V, i.e. the *statistical cross-correlation function*, whose definition follows

$$\begin{aligned} \Gamma_{UV}(t_1, t_2) &\triangleq \langle U(t_1) V(t_2)^{\star} \rangle \\ &= \iint \mathrm{d}u \, \mathrm{d}v \, u v^{\star} p_{UV}(u, v; t_1, t_2) \ . \end{aligned} \tag{2.50}$$

In similar fashion to the time-based case, we denote the *statistical autocorrelation function*

as Γ_{UU}. We choose not to consider higher-order moments.

Ergodicity and stationarity

A stochastic process whose characteristics do not change over time is said to be a steady-state process, or *stationary*. More formally, a stochastic process is stationary to order n if the joint probability density of order up to n is invariant under time translation. A process is *wide-sense stationary* provided $\langle U(t) \rangle$ is independent of t (and without loss of generality can be assumed to 0) and the autocorrelation function is only a function of the time difference, i.e. $\Gamma_{UU}(t_1, t_2) = \Gamma_{UU}(\tau)$ with $\tau = t_2 - t_1$ (clearly, stationarity to order 2 is a sufficient condition for wide-sense stationarity). We assume that all processes that are relevant to our discussion are stationary, at least in the wide sense.

A more restrictive class of stochastic processes is *ergodic* processes, where any realization in the ensemble of such a process fully describes the entire stochastic process. If the following holds

$$\langle U \rangle = \left\langle {}^{\zeta}u \right\rangle_{\mathrm{t}} \tag{2.51}$$

$$\Gamma_{UU}(t_1, t_2) = \Gamma_{UU}(\tau) = \tilde{\Gamma}_{\zeta\zeta}(\tau) \ , \tag{2.52}$$

for all realizations then U is an ergodic stochastic process, and we refer to Γ_{UU} as simply the autocorrelation function of the process. We immediately observe that an ergodic process must be stationary. Most stochastic processes of interest are ergodic and, unless stated otherwise, we will implicitly assume so.

As an example consider again our gas discharge lamp. The life-time of such a lamp is typically far greater than the characteristic time of emission, hence we can consider the process to be stationary: If we were to perform a couple of experiments, each starting at a different time point and each measuring an emission realizations, we'd expect the

produced sample functions to be statistically similar. Put differently, the probability of some photon energy being emitted remains constant over time. This process is also ergodic: Each realization will measure the lamp's entire spectrum of emission.

Spectral analysis

Each realization $u(t)$ of a stochastic process U is a function of time, thus we are encouraged to study the spectral decomposition $\mathscr{F}\{u\}$. Unfortunately, for many physical processes the realizations are not Fourier transformable. This holds in general for stationary processes as well, where by definition energy is distributed over all time. However, the ensemble averaged decomposition does typically exist. We assume that U is stationary, at least in the wide sense, and define the truncated, normalized Fourier transform-like of a realization to be

$$\bar{u}_T(\omega) = \frac{1}{\sqrt{2T}} \int_{-T}^{T} \mathrm{d}t \, u(t) e^{\mathrm{i}\omega t} \ . \tag{2.53}$$

By Parseval's theorem (Theorem 2.2.1.1), the total energy is preserved in transform pairs, therefore $|\bar{u}_T|^2$ quantifies the power in each spectral component of the truncated signal. At the limit, we define the *power spectral density* (PSD) of the process U:

$$S_{UU}(\omega) \triangleq \lim_{T\to\infty} \left\langle |\bar{u}_T(\omega)|^2 \right\rangle \ , \tag{2.54}$$

which has units of power per unit frequency. The definition above is trivially extended to become *cross power spectral density* of a couple of stochastic processes U and V:

$$S_{UV}(\omega) \triangleq \lim_{T\to\infty} \left\langle \bar{u}_T(\omega) \bar{v}_T(\omega)^\star \right\rangle \ , \tag{2.55}$$

which is an important quantity in the theory of optical coherence. For ergodic processes, the statistical cross- and autocorrelation functions can be interchanged with their time-based counterparts.

We now ready to present the celebrated *Wiener–Khinchin theorem*:

Theorem 2.3.0.1 (Wiener–Khinchin Theorem)**.** *The spectral decomposition of the autocorrelation function of a stationary process is its power spectral density.*

Proof. With the definition of the PSD (Eq. (2.54)) as the starting point, note that integration and ensemble averaging commute and by using the properties of a stationary process we deduce:

$$\begin{aligned} S_{UU}(\omega) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \int_{-T}^{T} \mathrm{d}t' \left\langle u(t)u(t')^{\star} \right\rangle e^{\mathrm{i}\omega(t-t')} \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \int_{-T-t}^{T-t} \mathrm{d}\tau\, \Gamma_{UU}(\tau) e^{\mathrm{i}\omega\tau} \\ &= \lim_{T\to\infty} \int_{-T}^{T} \mathrm{d}\tau\, \Gamma_{UU}(\tau) e^{\mathrm{i}\omega\tau} = \sqrt{2\pi}\,\mathscr{F}\{\Gamma_{UU}\} \ . \end{aligned} \tag{2.56}$$

where we performed the variable substitution $\tau = t' - t$. □

That is, the PSD and the autocorrelation function are intrinsically linked as Fourier transform pairs for any stationary stochastic process.

It can be shown that even when the limit in Eq. (2.54) does not converge, a spectral decomposition still exists in form of some power spectrum distribution. Furthermore, the theorem can also be generalized to a wider class of stochastic processes. We won't prove either of those claims, however, as the above formulation is sufficient for our use. When we consider a spectral decomposition of the autocorrelation we assume its existence is given by the Wiener–Khinchin theorem.

2.4 Electromagnetic Waves in Vacuum

Light energy is propagated by electromagnetic waves. The intrinsic quantities that give rise to those waves are the electric, $\vec{\boldsymbol{E}}$, and magnetic, $\vec{\boldsymbol{B}}$, fields whose behaviour is governed by Maxwell's famous set of equations. Those equations are the fundamental principles that describe the interaction of those fields with charge and current, as well as the time-dependent interaction between the fields themselves. It is that interaction between the electric and magnetic fields that results in self-supporting fields that oscillate in unison and give rise to time-harmonic electromagnetic waves. We provide a very brief discussion about the relevant classical electrodynamics. We use the Gaussian-cgs units in this write-up.

We start with the general Maxwell equations:

$$\boldsymbol{\nabla}\cdot\vec{\boldsymbol{E}} = 4\pi\rho \qquad\qquad \boldsymbol{\nabla}\cdot\vec{\boldsymbol{B}} = 0\ , \tag{2.57}$$

$$\boldsymbol{\nabla}\times\vec{\boldsymbol{E}} = -\frac{1}{c}\frac{\partial\vec{\boldsymbol{B}}}{\partial t} \qquad\qquad \boldsymbol{\nabla}\times\vec{\boldsymbol{B}} = \frac{4\pi}{c}\vec{\boldsymbol{j}} + \frac{1}{c}\frac{\partial\vec{\boldsymbol{E}}}{\partial t}\ , \tag{2.58}$$

where $\rho, \vec{\boldsymbol{j}}$ are the charge and current distributions, respectively, and c is the speed of light. Directly from the two curl equations we note that

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\vec{\boldsymbol{E}} = -\frac{1}{c}\frac{\partial}{\partial t}\Big(\boldsymbol{\nabla}\times\vec{\boldsymbol{B}}\Big) = -\frac{1}{c^2}\frac{\partial^2\vec{\boldsymbol{E}}}{\partial t^2} - \frac{4\pi}{c^2}\frac{\partial\vec{\boldsymbol{j}}}{\partial t}\ , \tag{2.59}$$

and by using the vector triple product identity (2.9), we deduce that the electric and magnetic fields satisfy the inhomogeneous wave equations:

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\boldsymbol{E}} = 4\pi\left(\frac{1}{c}\frac{\partial\vec{\boldsymbol{j}}}{\partial t} + \boldsymbol{\nabla}\rho\right)\ , \tag{2.60}$$

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\boldsymbol{B}} = -\frac{4\pi}{c}\boldsymbol{\nabla}\times\vec{\boldsymbol{j}}\ . \tag{2.61}$$

The operator on the left-hand side in Eqs. (2.60) and (2.61) is known as the *d'Alembert operator*, or the wave operator, and we denote it via

$$\triangle_{\mathrm{m}} = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \ . \tag{2.62}$$

Away from sources, that is where no charge or current is present ($\rho \equiv 0$ and $\vec{\boldsymbol{j}} \equiv 0$), the wave equations for the electric and magnetic fields reduce to their homogeneous counterparts:

$$\triangle_{\mathrm{m}} \vec{\boldsymbol{E}} = 0 \qquad\qquad \triangle_{\mathrm{m}} \vec{\boldsymbol{B}} = 0 \ . \tag{2.63}$$

While Eqs. (2.60) and (2.61) govern the sourcing of electromagnetic waves, the homogeneous partial differential equations above govern the propagation of those waves. The propagation of electromagnetic waves in space and matter, on the other hand, is the foundation of light transport and the homogeneous wave equations above are the driving principle.

Transverse electromagnetic waves

An important class of solutions to the free-space wave Equations (2.63) are *transverse electromagnetic* waves (TEM). This simple kind of waves are commonly used in practical applications (like computer rendering), provide some physical insight and are useful to study propagation in matter and polarization, which will be discussed later.

TEM waves are waves where the electric and magnetic fields are perpendicular to the direction of propagation. Let $\vec{\boldsymbol{k}}$ be the wave's direction of propagation, i.e. the *wavevector*, and without loss of generality we set $\vec{\boldsymbol{k}} = k\hat{\boldsymbol{z}}$, with $k = |\vec{\boldsymbol{k}}|$ being the *wavenumber*. The spatial frequency is then $2\pi k$ and the wavelength is related to the

wavenumber via $\lambda = \frac{2\pi}{k}$. Then, we are looking for fields $\vec{\boldsymbol{E}}(z,t)$, $\vec{\boldsymbol{B}}(z,t)$ that satisfy the homogeneous wave equations as well as Maxwell's source-free equations. From Eqs. (2.57) and (2.58) we immediately deduce that

$$\frac{\partial E_z}{\partial z} = 0 \qquad \frac{\partial E_z}{\partial t} = c\hat{\boldsymbol{z}} \cdot \boldsymbol{\nabla} \times \vec{\boldsymbol{B}} = c\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) = 0 \ , \tag{2.64}$$

and similarly for B_z, which implies that E_z, B_z are constants and we set them to 0. The general solution is known as *d'Alembert formula* and is a superposition of two plane-waves propagating in opposite directions, $\boldsymbol{f}_\perp(z-ct)$ and $\boldsymbol{g}_\perp(z+ct)$ such that $\boldsymbol{f}_\perp \cdot \hat{\boldsymbol{z}} = \boldsymbol{g}_\perp \cdot \hat{\boldsymbol{z}} = 0$:

$$\vec{\boldsymbol{E}} = \boldsymbol{f}_\perp(z-ct) + \boldsymbol{g}_\perp(z+ct) \ . \tag{2.65}$$

Further, using Eq. (2.58) we write

$$\begin{aligned} \frac{\partial \vec{\boldsymbol{B}}}{\partial t} = -c\boldsymbol{\nabla} \times \vec{\boldsymbol{E}} &= -c\hat{\boldsymbol{z}} \times \frac{\partial}{\partial z}(f_\perp(z-ct) + \boldsymbol{g}_\perp(z+ct)) \\ &= \hat{\boldsymbol{z}} \times \frac{\partial}{\partial t}(f_\perp(z-ct) - \boldsymbol{g}_\perp(z+ct)) \ . \end{aligned} \tag{2.66}$$

And thus

$$\vec{\boldsymbol{B}} = \hat{\boldsymbol{z}} \times \boldsymbol{f}_\perp(z-ct) - \hat{\boldsymbol{z}} \times \boldsymbol{g}_\perp(z+ct) \ , \tag{2.67}$$

up to a constant which we can neglect. Noting again that $\vec{\boldsymbol{k}} \parallel \hat{\boldsymbol{z}}$ we deduce that indeed $\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{k}} = \vec{\boldsymbol{B}} \cdot \vec{\boldsymbol{k}} = 0$, and hence Eqs. (2.65) and (2.67) describe a TEM wave. However, observe that it does not hold in general that $\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{B}} = 0$ (such waves always admit some standing wave characteristic).

The most simple form of a TEM wave is the monochromatic plane-wave. Those kind of waves have very simple geometric properties, which is why they are ubiquitously used

in computer rendering. Let $\boldsymbol{f}_\perp \equiv 0$, then the solution to the homogeneous wave equations becomes:

$$\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}, t) = \mathcal{E}_\perp e^{\mathrm{i}(\vec{\boldsymbol{k}}\cdot\vec{\boldsymbol{r}} - \omega t)} \ , \tag{2.68}$$

$$\vec{\boldsymbol{B}}(\vec{\boldsymbol{r}}, t) = -(\hat{\boldsymbol{k}} \times \mathcal{E}_\perp) e^{\mathrm{i}(\vec{\boldsymbol{k}}\cdot\vec{\boldsymbol{r}} - \omega t)} \ , \tag{2.69}$$

with $\omega = c|\vec{\boldsymbol{k}}| = ck$ being the *angular frequency*. The quantity $\mathcal{E}_\perp$ is generally a complex-valued (temporally and spatially-invariant) vector that describes the direction of the electric field $\vec{\boldsymbol{E}}$ in space, i.e. its polarization, and we discuss polarization later. Note a physical characteristic of plane-waves: $|\vec{\boldsymbol{E}}| = |\vec{\boldsymbol{B}}|$. The complex fields $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ above are the analytic signal representation of the wave, and it is the real parts that describe the ("physical") time-varying electric and magnetic disturbances. Indeed, the energy density of any electromagnetic wave is:

$$u_{\mathrm{EM}} = \frac{1}{8\pi}\left[\left(\mathrm{Re}\,\vec{\boldsymbol{E}}\right)^2 + \left(\mathrm{Re}\,\vec{\boldsymbol{B}}\right)^2\right] \ , \tag{2.70}$$

that is, the complex parts of $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ contribute no energy. The complex representation is used solely for convenient and ease of computation. Note that for plane-waves it always holds that $\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{B}} = 0$ hence the vectors $\{\vec{\boldsymbol{k}}, \vec{\boldsymbol{E}}, \vec{\boldsymbol{B}}\}$ give rise to an orthogonal triad, with $\vec{\boldsymbol{B}}$ being uniquely defined by the former two vectors. This means we do not need to concern ourselves with the $\vec{\boldsymbol{B}}$ field. The *Poynting vector*—the energy current density and direction of the electromagnetic flux—is defined as

$$\vec{\boldsymbol{S}} = \frac{c}{4\pi}\left(\vec{\boldsymbol{E}} \times \vec{\boldsymbol{B}}\right) \tag{2.71}$$

for any electromagnetic wave. In the case of plane-waves, $\vec{\boldsymbol{S}}$ is clearly aligned with $\vec{\boldsymbol{k}}$ and

the time-averaged energy density and Poynting vector are easily calculated:

$$\langle u_{\mathrm{EM}} \rangle_{\mathrm{t}} = \frac{1}{8\pi} |\mathcal{E}_{\perp}|^2 \ , \tag{2.72}$$

$$\left\langle \vec{\boldsymbol{S}} \right\rangle_{\mathrm{t}} = \langle u_{\mathrm{EM}} \rangle_{\mathrm{t}} \, c\hat{\boldsymbol{k}} \ . \tag{2.73}$$

As our sensors of interest (e.g., eye, camera) observe electromagnetic radiation over a period long compared to the angular frequency ω, it is the time-averaged quantities that are of more significance to us than the instantaneous values. The observed intensity of the wave is then simply the magnitude of the (real part of) time-averaged Poynting vector:

$$I(\vec{\boldsymbol{r}}) = \left| \mathrm{Re} \left\langle \vec{\boldsymbol{S}}(\vec{\boldsymbol{r}}) \right\rangle_{\mathrm{t}} \right| \ . \tag{2.74}$$

2.5 Optical Coherence Theory

So far we have been concerned with deterministic fields, waves and wave packets. Real-life radiation from physical sources, on the other hand, often calls for a statistical treatment. The optical coherence theory, which can be considered as the study of observable optical quantities, has been pioneered by Emil Wolf, who also remained a dominant figure and contributor to the field. This section loosely follows Wolf [59]. Additional resources are Born and Wolf [8], Goodman [20], Mandel and Wolf [32].

Temporal coherence

To simplify the discussion we start by ignoring polarization. Polarization does play an important role in optical coherence and will be revisited later in Section 2.6. Let $\psi(\vec{\boldsymbol{r}}, t)$ be a scalar wave packet (i.e., a *wave function*). Consider an experiment involving

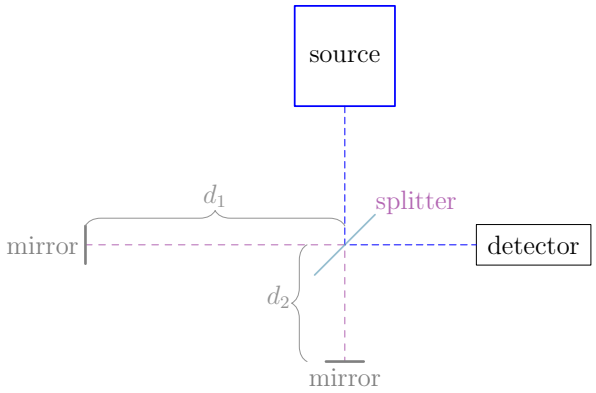


Figure 2.1: Michelson interferometer: Light is emitted by a source and is then split into two beams. Each beam travels until it is reflected by a mirror, then the beams are recombined and are observed by the detector. The distance difference between the travelled beams, $d = d_2 - d_1$, is controllable by moving the mirrors. This distance difference induces a time-delay of $\tau = \frac{d}{c}$ between the beams which enables us to measure the light's ability to interfere with a time-delayed version of itself, i.e. its temporal coherence.

the Michelson interferometer (Fig. 2.1), a device where light is split into two paths, then recombined inducing a time-delay between the recombined beams. The recombined beam is then observed by a detector. The distance difference between the paths is denoted as d resulting in a $\tau = \frac{d}{c}$ time-delay between the recombined beams. The amplitude of the field at the observation point at some time t, denoted $w(\tau)$, is then the superposition of the recombined beams:

$$w(\tau) = \alpha_1 \psi(t) + \alpha_2 \psi(t + \tau) \tag{2.75}$$

where $\alpha_{1,2}$ are real constants that describe the field strength that remains in each beam after interacting with the Michelson interferometer. If the splitting and recombining machinery introduces no phase-shifts then we can assume that $\alpha_{1,2}$ are real values. Further, if we assume that the energy lost in the process is negligible, then $\alpha_1^1 + \alpha_2^2 = 1$. At optical frequencies, the fluctuations of the field ψ are not directly observable, we therefore treat ψ as a *wave ensemble*—a (stationary, in the wide sense, and ergodic) stochastic process.

Denote $\psi_1 = \psi(t)$, $\psi_2 = \psi(t+\tau)$, then the observed intensity of w by the detector is the time-averaged quantity (up to a constant):

$$\begin{aligned} I(w(\tau)) &= \langle w(\tau) w(\tau)^\star \rangle_{\mathrm{t}} \\ &= \alpha_1^2 \langle \psi_1 \psi_1^\star \rangle_{\mathrm{t}} + \alpha_2^2 \langle \psi_2 \psi_2^\star \rangle_{\mathrm{t}} + 2\,\mathrm{Re}\{\alpha_1 \alpha_2 \langle \psi_1 \psi_2^\star \rangle_{\mathrm{t}}\} \\ &= I(\psi) + 2\alpha_1 \alpha_2 \,\mathrm{Re}\,\tilde{\Gamma}_{\psi\psi}(\tau) \ , \end{aligned} \tag{2.76}$$

where we assumed that $I(\psi(t)) = I(\psi(t+\tau))$, that is energy-loss due to propagation inside the device is negligible. If we were to further assume that $\alpha = \alpha_1 = \alpha_2 = 1/\sqrt{2}$, meaning that the energy is split evenly between the beams, then $I(w(\tau)) = I(\psi) + \mathrm{Re}\,\tilde{\Gamma}_{\psi\psi}(\tau)$. $\tilde{\Gamma}_{\psi\psi}$ is the time autocorrelation function of the analytic signal ψ. In the context of optical coherence theory it is known as the *temporal coherence function*, simply denoted $\Gamma(\tau)$ henceforth, and is an important quantity of second-order coherence. We define

$$\gamma_{\psi\psi}(\tau) \triangleq \frac{\Gamma_{\psi\psi}(\tau)}{\Gamma_{\psi\psi}(0)} \ . \tag{2.77}$$

Notice that $|\gamma_{\psi\psi}| \leq 1$, and we can rewrite Eq. (2.76) in the form of the *temporal interference law for stationary fields*:

$$I(\psi_1 + \psi_2) = I(\psi_1) + I(\psi_2) + 2\sqrt{I(\psi_1)}\sqrt{I(\psi_2)}\,\mathrm{Re}\,\gamma_{\psi\psi}(\tau) \tag{2.78}$$

for any stationary ψ and any τ, the time-delay between ψ_1 and ψ_2. When $I(w) = 0$ total *destructive interference* has occurred, while when $I(w) = 2I(\psi)$ total *constructive interference* has occurred. The significance of the Michelson interferometer is its ability to measure the *temporal coherence* of the observed light: By slowly changing the distance difference between the paths, and thus the time-delay τ, we measure the ability of the

light beam to interfere with a time-delayed version of itself. This physical property is quantified by the *degree of temporal coherence* $\gamma_{\psi\psi}$.

Spatial coherence

We now consider the spatial extent of a light source. Consider Young's famous double slit experiment (Fig. 2.2) where light passes through two thin slits situated in close proximity to each other and then is observed on a screen. Assume the following geometry: An opaque screen, with two slits notched in it at points $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$, and a screen with an observation point $\vec{\boldsymbol{p}}$. Both screens are parallel to the xy-axis. A light source is placed such that the screen containing the slits is between the source and the observation screen, and the centre of the light source, the middle point between the slits and $\vec{\boldsymbol{p}}$ all lie on the z-axis. Denote the distance between the slits $d = |\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2|$. As before, let ψ be the amplitude of the emitted light and the amplitude at the point of interest $\vec{\boldsymbol{p}}$ can be written as the superposition

$$\psi(\vec{\boldsymbol{p}}, t) = K_1 \psi\left(\vec{\boldsymbol{r}}_1, t - \frac{|\vec{\boldsymbol{p}} - \vec{\boldsymbol{r}}_1|}{c}\right) + K_2 \psi\left(\vec{\boldsymbol{r}}_2, t - \frac{|\vec{\boldsymbol{p}} - \vec{\boldsymbol{r}}_2|}{c}\right) \tag{2.79}$$

(up to a constant), with $K_{1,2}$ are some (complex-valued) constants that arise due to the diffraction. We are interested in the observed (time-averaged) intensity $I(\vec{\boldsymbol{p}})$. Then, denoting $t' = t - |\vec{\boldsymbol{p}} - \vec{\boldsymbol{r}}_1|/c$ and $t'' = t - |\vec{\boldsymbol{p}} - \vec{\boldsymbol{r}}_2|/c$ we get

$$\begin{aligned} I(\psi(\vec{\boldsymbol{p}}, t)) =& |K_1|^2 I(\psi(\vec{\boldsymbol{r}}_1, t')) + |K_2|^2 I(\psi(\vec{\boldsymbol{r}}_2, t'')) \\ &+ 2\,\mathrm{Re}\left\{ K_1 K_2^\star \left\langle \psi(\vec{\boldsymbol{r}}_1, t') \psi(\vec{\boldsymbol{r}}_2, t'')^\star \right\rangle_{\mathrm{t}} \right\} . \end{aligned} \tag{2.80}$$

Denote $\tau = t'' - t'$ as the time difference, and the generalization of the temporal coherence function, known as the *mutual coherence function* for stationary processes, immediately

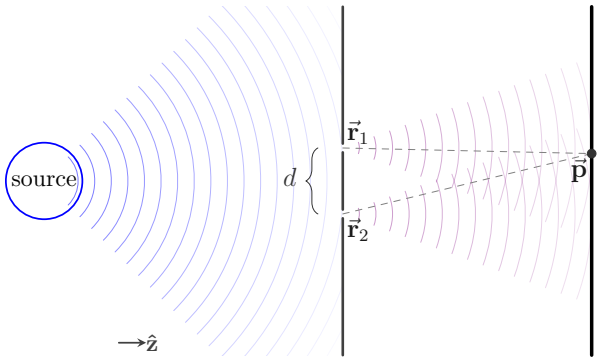


Figure 2.2: Young's double slit experiment: Light is incident upon an opaque screen (center) with two thin slits notched in it. The slits are positioned at points $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$. The light diffracts through the slits and produces an interference pattern that is observed on a screen (right). The intensity of the interference pattern at an observation point $\vec{\boldsymbol{p}}$ depends both on the temporal coherence properties of the light, due to time difference between the light arriving from $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$; as well as the spatial coherence properties that come into effect due to distance between the slits $d = |\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2|$.

follows

$$\Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) \triangleq \langle \psi(\vec{\boldsymbol{r}}_1, t)\psi(\vec{\boldsymbol{r}}_2, t+\tau)^\star \rangle_{\mathrm{t}} \ . \tag{2.81}$$

A basic property of the mutual coherence function that is of interest is the relation of its complex conjugate to itself: $\Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau)^\star = \Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_2, \vec{\boldsymbol{r}}_1, -\tau)$, which is readily verifiable by using the stationarity of u.

In an identical fashion to the temporal degree of coherence, we can also define the *degree of coherence* as the normalized autocorrelation:

$$\gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) \triangleq \frac{\Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau)}{\sqrt{\Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_1, 0)}\sqrt{\Gamma_{\psi\psi}(\vec{\boldsymbol{r}}_2, \vec{\boldsymbol{r}}_2, 0)}} \ , \tag{2.82}$$

and it is trivially verifiable that $|\gamma| \leq 1$ (by Schwarz's inequality, Eq. (2.18)). Using the

degree of coherence we can rewrite Eq. (2.80) as the *interference law for stationary fields*

$$I(\psi(\vec{\boldsymbol{r}}_1,t)+\psi(\vec{\boldsymbol{r}}_2,t+\tau)) = I(\psi(\vec{\boldsymbol{r}}_1,t)) + I(\psi(\vec{\boldsymbol{r}}_2,t)) + 2\sqrt{I(\psi(\vec{\boldsymbol{r}}_1,t))}\sqrt{I(\psi(\vec{\boldsymbol{r}}_2,t))}\,\mathrm{Re}\,\gamma_{\psi\psi}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\tau)\ . \tag{2.83}$$

The mutual coherence function is the fundamental quantity on the theory of optical coherence. As with the temporal coherence function, we drop the subscripts and denote it $\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\tau)$. Furthermore, when $\tau \equiv 0$, i.e. we ignore temporal contributions, we slightly abuse notation to write $\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2) \triangleq \Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,0)$, which is the "equal-time" mutual coherence.

Often we model a light source as a collection of a great many independent elementary radiators, an accurate model for spontaneous emission sources. Let ${}^{j}u(\vec{\boldsymbol{r}},t)$ be a singular wavelet emitted by a radiator. The wave ensemble is then a collection of all such emitted wavelets. The instantaneous intensity is then:

$$I(u(\vec{\boldsymbol{p}},t)) = \langle u(\vec{\boldsymbol{p}},t)u(\vec{\boldsymbol{p}},t)^{\star}\rangle\ . \tag{2.84}$$

Somewhat informally, we can note that when the light source is small, the distance difference between the elementary radiators that give rise to the contributing phasors is small as well, and thus most of the arriving wavelets will be in-phase and interfere. With a greater source, the contributing phasors will be out-of-phase, until the distribution of the phases is uniform and no interference takes place anymore (more accurately, constructive and destructive interference takes place with equal probability). For ergodic process, the time and ensemble averages are equal.

2.5.1 Coherence in the Spectral Domain

The space-time formulation of the mutual coherence presented in the previous subsection is useful when the light is monochromatic, or satisfies some monochromaticity conditions. When studying the interaction of light and matter, it is more natural to study each frequency component independently in-place of the time-dependant field. In addition to rendering some of the mathematics more tractable, an apparent advantage is that by treating each spectral component individually, the coherence time is effectively infinite and we can ignore temporal effects.

First, we formally present the space-frequency formulation of coherence: In Section 2.3 we discussed the power spectral density of a stationary process and its relation to the autocorrelation. This relation is described by the Wiener–Khinchin theorem (Theorem 2.3.0.1) which states that the transform pair of the autocorrelation function is the cross power spectral density, known as the *cross-spectral density* under the context of optical coherence theory. Viz.

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \triangleq \mathscr{F}\left\{\Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, t)\right\} . \tag{2.85}$$

Just as the mutual coherence Γ is the central quantity in the space-time formulation of coherence, the cross-spectral density $\mathcal{C}$ is the central quantity in the space-frequency formulation of coherence. To show that different spectral components are uncorrelated, at least for ergodic processes, consider a truncated Fourier transform of the signal (an unnormalized version of Eq. (2.53)):

$$\bar{\psi}_T(\vec{\boldsymbol{r}}, \omega) = \int_{-T}^{T} \mathrm{d}t\, \psi(\vec{\boldsymbol{r}}, t) e^{\mathrm{i}\omega t} . \tag{2.86}$$

Then, consider the following ensemble average:

$$\lim_{T\to\infty}\left\langle \bar{\psi}_T(\vec{\boldsymbol{r}},\omega)\bar{\psi}_T\big(\vec{\boldsymbol{r}}',\omega'\big)^\star\right\rangle = \lim_{T\to\infty}\int_{-T}^{T}\mathrm{d}t\, e^{\mathrm{i}(\omega-\omega')t} \times\int_{-T+t}^{T+t}\mathrm{d}\tau\left\langle \psi(\vec{\boldsymbol{r}},t)\psi\big(\vec{\boldsymbol{r}}',t+\tau\big)^\star\right\rangle e^{-\mathrm{i}\omega'\tau}\ . \tag{2.87}$$

Due to stationarity the integrand of the second integral is independent of t. Then, at the limit the second integral is the (inverse) Fourier transform of $\Gamma(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\tau)$, i.e. the cross-spectral density, while the first integral is calculated trivially using the Fourier identity (2.33), resulting in the expression

$$\lim_{T\to\infty}\left\langle \bar{u}_T(\vec{\boldsymbol{r}},\omega)\bar{u}_T\big(\vec{\boldsymbol{r}}',\omega'\big)^\star\right\rangle = 2\pi\delta(\omega-\omega')\,\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',-\omega'\right)\ . \tag{2.88}$$

As expected, for stationary process, which are not Fourier transformable, the limit does not strictly exist for $\omega=\omega'$. However, whenever $\omega\neq\omega'$ the Dirac delta vanishes, the limit exists and the ensemble average is 0, indicating that the spectral components of distinct frequencies are uncorrelated.

In similar fashion to the mutual coherence function, it is at times more convenient to work with the normalized version of the cross-spectral density, which known as the *spectral degree of coherence*:

$$\mathfrak{w}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega)\triangleq\frac{\mathcal{C}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right)}{\sqrt{\mathcal{C}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_1,\omega\right)}\sqrt{\mathcal{C}\left(\vec{\boldsymbol{r}}_2,\vec{\boldsymbol{r}}_2,\omega\right)}}\ . \tag{2.89}$$

The observed intensity of a wave ensemble at a given point is simply its self coherence, i.e. $I=\Gamma(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}})$. Note that the same information is contained in the cross-spectral density

if we were to integrate over the relevant frequencies:

$$\Gamma(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, 0) = \mathscr{F}^{-1}\{\mathcal{C}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega\right)\}\Big|_{t=0} = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty \mathrm{d}\omega\, \mathcal{C}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega\right) \tag{2.90}$$

Integrating over the non-negative frequencies only is sufficient as we pertain to real underlying signals. It makes sense then to call the integrand $\mathcal{C}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega\right)$ the *spectral density*.

Coherent-modes representation of the spectrum

An important result presented by Wolf [58] is the decomposition of the cross-spectral density. Let $D \subset \mathbb{R}^3$ be a closed set, large enough to be the spatial domain of interest. Note that for any ω the following holds

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \mathcal{C}\left(\vec{\boldsymbol{r}}_2, \vec{\boldsymbol{r}}_1, \omega\right)^\star \tag{2.91}$$

$$\int_D \int_D \mathrm{d}\vec{\boldsymbol{r}}\, \mathrm{d}\vec{\boldsymbol{r}}' \left|\mathcal{C}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \omega\right)\right|^2 < \infty\ , \tag{2.92}$$

where the integral inequality holds because the cross-spectral density is point-wise finite (for a physical process) and D is a finite domain. The above are the conditions of the Hilbert–Schmidt kernel, therefore $\mathcal{C}$ with domain $D \times D$ is a kernel function. It is also a positive semi-definite kernel (See Wolf [58] for a proof), therefore Mercer's condition is satisfied as well. Then, by Mercer's theorem, there exists a decomposition:

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \sum_{n=1}^{\infty} \alpha_n(\omega) \phi_n(\vec{\boldsymbol{r}}_1) \phi_n(\vec{\boldsymbol{r}}_2)^\star\ , \tag{2.93}$$

where $\alpha_n \geq 0$ are eigenvalues and ϕ_n are eigenfunctions of the Hilbert–Schmidt integral operator associated with the kernel function. The set of eigenfunctions form an orthonormal basis, viz. $\int_D \phi_n \phi_m = \delta_{nm}$, and is a countable set at most. Such decomposition is

called the *coherent-modes* decomposition, where $\mathcal{C}^{(n)}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega) = \alpha_n(\omega)\phi_n(\vec{\boldsymbol{r}}_1)\phi_n(\vec{\boldsymbol{r}}_2)^\star$ are the coherent-modes. Notice now that by taking the inverse Fourier transform of each side in Eq. (2.85), we get:

$$\Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, t) = \mathscr{F}^{-1}\{\mathcal{C}\} = \sum_n \mathscr{F}^{-1}\{\alpha_n(\omega)\phi_n(\vec{\boldsymbol{r}}_1)\phi_n(\vec{\boldsymbol{r}}_2)^\star\} \ , \tag{2.94}$$

and we write $\Gamma^{(n)} = \mathscr{F}^{-1}\Big\{\mathcal{C}^{(n)}\Big\}$, the *space-time coherent-modes*.

Using the above decomposition we can also construct an ensemble of space-frequency realizations faithful to the original signal, despite the fact that a spectral decomposition in the sense of a Fourier transform does not generally exist for a realization of a stationary process. To demonstrate that, consider the ensemble of elementary monochromatic wavelets $\{v(\vec{\boldsymbol{r}}, \omega)e^{-\mathrm{i}\omega t}\}$, such that

$$v(\vec{\boldsymbol{r}}, \omega) = \sum_{n=1}^{\infty} a_n(\omega)\phi_n(\vec{\boldsymbol{r}}) \ , \tag{2.95}$$

where the coefficients $a_n(\omega)$ are some frequency-dependent random variables that fulfil:

$$\mathrm{E}[a_n(\omega)a_m(\omega)^\star] = \delta_{nm}\alpha_n(\omega) \qquad \text{and} \qquad \sum_n |a_n(\omega)| < \infty \ . \tag{2.96}$$

We define now the *space-frequency ensemble average* over a frequency as

$$\begin{aligned} \langle v(\vec{\boldsymbol{r}}_1, \omega)v(\vec{\boldsymbol{r}}_2, \omega)^\star\rangle_\omega &\triangleq \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{E}[a_n(\omega)a_m(\omega)^\star]\phi_n(\vec{\boldsymbol{r}}_1)\phi_m(\vec{\boldsymbol{r}}_2)^\star \\ &= \mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \ , \end{aligned} \tag{2.97}$$

indicating that the space-frequency ensemble average over the ensemble of monochromatic realizations, as defined, indeed reproduces the cross-spectral density. In addition,

using Eq. (2.90) we immediately observe that the observed intensity at a point becomes:

$$\pi I(\vec{\boldsymbol{r}}) = \int_0^\infty \mathrm{d}\omega \, \langle v(\vec{\boldsymbol{r}},\omega) v(\vec{\boldsymbol{r}},\omega)^\star \rangle_\omega = \int_0^\infty \mathrm{d}\omega \, \mathcal{C}\,(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}},\omega) \ . \tag{2.98}$$

Hence, the wave ensemble $\{v(\vec{\boldsymbol{r}},\omega)e^{-\mathrm{i}\omega t}\}$ is a faithful decomposition into monochromatic wavelets.

The *spectral interference law* for stationary wave ensembles can now be formulated: Suppose $\{v(\vec{\boldsymbol{r}},\omega)e^{-\mathrm{i}\omega t}\}$ is a decomposition of some space-time signal into elementary monochromatic wavelets. Recall Young's double slit experiment. We assume that light arrives at an observation point $\vec{\boldsymbol{p}}$ from a couple of thin slits, $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$, and we would like to compute the spectral density of $v(\vec{\boldsymbol{p}},\omega)$. It takes light $r_{1,2} = |\vec{\boldsymbol{r}}_{1,2} - \vec{\boldsymbol{p}}|$ time to travel from each slit, respectively, to the observation point, therefore the light (of frequency ω) that arrives at $\vec{\boldsymbol{p}}$ can be formulated as the superposition

$$v(\vec{\boldsymbol{p}},\omega)e^{-\mathrm{i}\omega t} = v(\vec{\boldsymbol{q}}_1,\omega)e^{-\mathrm{i}\omega\frac{r_1}{c}}e^{-\mathrm{i}\omega t} + v(\vec{\boldsymbol{q}}_2,\omega)e^{-\mathrm{i}\omega\frac{r_2}{c}}e^{-\mathrm{i}\omega t} \ . \tag{2.99}$$

Denoting $\mathfrak{s} = \mathcal{C}(\vec{\boldsymbol{p}},\vec{\boldsymbol{p}},\omega)$ and $\mathfrak{s}'_{1,2} = \mathcal{C}(\vec{\boldsymbol{r}}_{1,2},\vec{\boldsymbol{r}}_{1,2},\omega)$, the spectral density at the observation point is then

$$\mathfrak{s} = \mathfrak{s}'_1 + \mathfrak{s}'_2 + 2\sqrt{\mathfrak{s}'_1}\sqrt{\mathfrak{s}'_2}\,\mathrm{Re}\left[\mathfrak{w}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega)e^{-\mathrm{i}\omega\frac{r_1-r_2}{c}}\right] \ , \tag{2.100}$$

which is the spectral interference law. While the interference law (Eq. (2.83)) describes the intensity of a superposition of waves of any state of coherence, the spectral interference law describes the spectral density of the superposition of monochromatic wavelets of any state of coherence.

Separability of temporal and spatial coherence

So far we have shown that under the space-frequency formulation, distinct frequencies can be treated independently while ignoring temporal effects. Clearly temporal effects do not just disappear, and we will show now that temporal coherence gets reintroduced when taking the intensity in the space-frequency domain. Consider the spectral interference law (Eq. (2.100)), applying Eq. (2.98) we integrate both sides to calculate the intensity:

$$I = I_1' + I_2' + \frac{2}{\pi}\,\mathrm{Re}\int_0^\infty \mathrm{d}\omega\,\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) e^{-\mathrm{i}\omega\frac{r_1 - r_2}{c}} \tag{2.101}$$

with $I = I(\vec{\boldsymbol{p}})$ and $I'_{1,2} = I(\vec{\boldsymbol{r}}_{1,2})$. Carrying out the integration of the last term we recognize the inverse Fourier transform and we get

$$\frac{2}{\pi}\,\mathrm{Re}\int_0^\infty \mathrm{d}\omega\,\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) e^{-\mathrm{i}\omega\frac{r_1 - r_2}{c}} = 2\,\mathrm{Re}\,\Gamma\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \frac{r_1 - r_2}{c}\right) \tag{2.102}$$

which is identical to the interference law with the appropriate time-delay. Hence, the temporal coherence phenomena get reintroduced when we observe the ensemble, i.e. calculate the intensity.

It is of theoretical importance to discuss when spatial and temporal effects are orthogonal and can be separated from each other. Write the mutual coherence function as a product of spatial coherence and temporal coherence functions, viz.

$$\Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) = \Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2)\tilde{\Gamma}(\tau) \tag{2.103}$$

then, by taking the Fourier transform of each side, we get

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2)\,\mathscr{F}\left\{\tilde{\Gamma}(\tau)\right\} \tag{2.104}$$

the last term, $\mathscr{F}\left\{\tilde{\Gamma}(\tau)\right\}$, is the space-invariant power spectral density. That is, the mutual coherence function is separable into spatial and temporal components if and only if the spectral profile of the wave ensemble is constant over space. The interaction of light with matter, be it propagation through matter or scattering, is typically frequency dependent and admits a dispersive relation. Therefore, we can conclude that for our domain of interest, spatial and temporal coherence effects can not be separated.

2.5.2 The Van Cittert–Zernike Theorem

The thought experiment we have carried out earlier when considering a natural source with a positive spatial extent as a collection of elementary radiators gives insight into how "incoherent" sources can give rise to wave ensembles which exhibit spatial coherence over extended regions of space. If we were to measure the amplitudes of the wavelets emitted by the source at close proximity to the source, we would find that the independent elementary radiators give rise to highly irregular phasors with uncorrelated phases. On the other hand, far away from the source the incoming wave-trains would share some semblance, and indeed very far from the source the source can approximated as a "coherent" point source. This phenomenon frequently arises in astronomy: Despite the massive size of radiating stars, due to the immense distance from the source, light that is observed on earth can exhibit a significant degree of coherence. This observation suggests that spatial coherence is "produced" (in a sense) via light propagation.

Let the light source be composed of a continuous distribution of independent radiators. Each radiator positioned at $\vec{\boldsymbol{s}}$ gives rise to the scalar wavelet ${}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{r}}, t)$. For simplicity assume that the light is perfectly monochromatic with angular frequency ω, therefore we can ignore temporal coherence effects. The amplitude at point $\vec{\boldsymbol{r}}$ is then the contributions from the entire ensemble, and can be formulated as a spatial integral over the volume of

the source:

$$\psi(\vec{\boldsymbol{r}},t)=\int_{V(S)}\mathrm{d}^3\vec{\boldsymbol{s}}\,{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{r}},t)\ . \tag{2.105}$$

We are interested in the mutual coherence of ψ at any two points $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$, i.e. $\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2)$. An expression for mutual coherence function under general conditions is given by the *Van Cittert–Zernike theorem*, the central theorem of optical coherence theory:

Theorem 2.5.2.1 (Van Cittert–Zernike Theorem)**.** *Assume a light source as described above. Given a pair of points $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$ the (equal-time) mutual coherence is as follows*

$$\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2)=\int_{V(S)}\mathrm{d}^3\boldsymbol{s}\,\frac{e^{-\mathrm{i}\omega\frac{r_{\boldsymbol{s}}-r'_{\boldsymbol{s}}}{c}}}{r_{\boldsymbol{s}}r'_{\boldsymbol{s}}}I(\vec{\boldsymbol{s}})\ ,$$

where for brevity we denoted $r_{\boldsymbol{s}}=|\vec{\boldsymbol{r}}_1-\vec{\boldsymbol{s}}|$ and $r'_{\boldsymbol{s}}=|\vec{\boldsymbol{r}}_2-\vec{\boldsymbol{s}}|$.

Proof. The amplitude of each ensemble member can be written as the time-delayed version of the emitted wavelet with the appropriate phase progression:

$${}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{r}},t)={}^{\boldsymbol{s}}\psi\left(\vec{\boldsymbol{s}},t-\frac{|\vec{\boldsymbol{r}}-\vec{\boldsymbol{s}}|}{c}\right)\frac{1}{|\vec{\boldsymbol{r}}-\vec{\boldsymbol{s}}|}e^{-\mathrm{i}\omega\frac{|\vec{\boldsymbol{r}}-\vec{\boldsymbol{s}}|}{c}}\ . \tag{2.106}$$

We assume simple spherical waves and hence the amplitude decays as the inverse of the distance from the source. We also make the natural assumption that the radiators are zero-meaned, i.e. $\langle{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{r}},t)\rangle_{\mathrm{t}}=0$. Then, an almost incoherent ensemble that corresponds to a propagating wave can be written as [20]

$$\begin{aligned}\langle{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t)\,{}^{\boldsymbol{q}}\psi(\vec{\boldsymbol{q}},t+\tau)^{\star}\rangle_{\mathrm{t}}&=\begin{cases}\langle{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t)\,{}^{\boldsymbol{q}}\psi(\vec{\boldsymbol{q}},t+\tau)^{\star}\rangle_{\mathrm{t}} & ,\boldsymbol{s}=\boldsymbol{q}\\ \langle{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t)\rangle_{\mathrm{t}}\,\langle{}^{\boldsymbol{q}}\psi(\vec{\boldsymbol{q}},t+\tau)\rangle_{\mathrm{t}}^{\star} & ,\boldsymbol{s}\neq\boldsymbol{q}\end{cases}\\ &=\delta^3(\vec{\boldsymbol{s}}-\vec{\boldsymbol{q}})\,\langle{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t)\,{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t+\tau)^{\star}\rangle_{\mathrm{t}}\ ,\end{aligned} \tag{2.107}$$

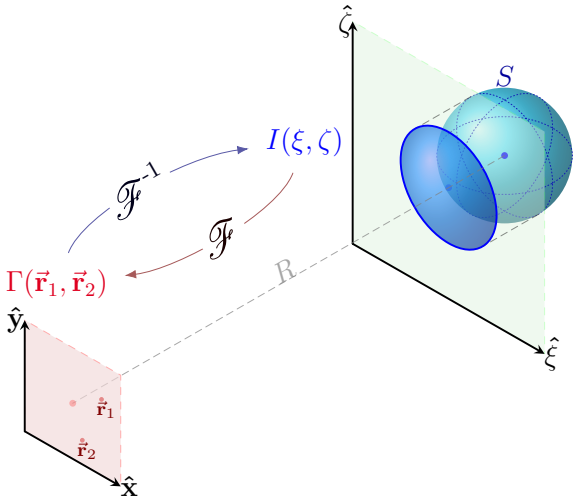


Figure 2.3: A light source S is imaged by a far-away observer. The light source gives rise to an intensity distribution on the source plane (right), denoted $I(\xi,\zeta)$, and we are interested in the mutual coherence, Γ, of the light that is observed on the image plane (left). The Far-Field Van Cittert–Zernike Theorem relates the two quantities as Fourier transform pairs.

up to a constant, for any $\vec{\boldsymbol{s}}$, $\vec{\boldsymbol{q}}$ and τ. By using the equations above we get the following expression for the mutual coherence:

$$\begin{aligned}\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2) &= \langle \psi(\vec{\boldsymbol{r}}_1,t)\psi(\vec{\boldsymbol{r}}_2,t)^\star\rangle_{\mathrm{t}} \\ &= \int_{V(S)} \mathrm{d}^3\boldsymbol{s} \int_{V(S)} \mathrm{d}^3\boldsymbol{q}\, \langle {}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{r}}_1,t)\,{}^{\boldsymbol{q}}\psi(\vec{\boldsymbol{r}}_2,t)^\star\rangle_{\mathrm{t}} \\ &= \int_{V(S)} \mathrm{d}^3\boldsymbol{s}\, \frac{e^{-\mathrm{i}\omega\frac{r_{\boldsymbol{s}}-r'_{\boldsymbol{s}}}{c}}}{r_{\boldsymbol{s}}r'_{\boldsymbol{s}}} \left\langle {}^{\boldsymbol{s}}\psi\left(\vec{\boldsymbol{s}},t-\frac{r_{\boldsymbol{s}}}{c}\right) {}^{\boldsymbol{s}}\psi\left(\vec{\boldsymbol{s}},t-\frac{r'_{\boldsymbol{s}}}{c}\right)^\star\right\rangle_{\mathrm{t}} . \end{aligned} \tag{2.108}$$

As mentioned, we assume full temporal coherence, therefore $I(\vec{\boldsymbol{s}}) = \langle {}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t)\,{}^{\boldsymbol{s}}\psi(\vec{\boldsymbol{s}},t+\tau)\rangle_{\mathrm{t}}$ is the intensity of a singular radiator. Hence:

$$\Gamma(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2) = \int_{V(S)} \mathrm{d}^3\boldsymbol{s}\, \frac{e^{-\mathrm{i}\omega\frac{r_{\boldsymbol{s}}-r'_{\boldsymbol{s}}}{c}}}{r_{\boldsymbol{s}}r'_{\boldsymbol{s}}} I(\vec{\boldsymbol{s}}) , \tag{2.109}$$

as desired. □

Let R_1, R_2 be the distances from the source to $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$, respectively. If we are far from

the source, where $R_1 \approx r_{\boldsymbol{s}}$ and $R_2 \approx r'_{\boldsymbol{s}}$ are a reasonable approximation for any $\vec{\boldsymbol{s}} \in V(S)$, then the Van Cittert–Zernike theorem simplifies to:

$$\Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) \approx \sqrt{2\pi}\frac{1}{R_1 R_2}\int_{V(S)} \mathrm{d}^3\boldsymbol{s}\, I(\vec{\boldsymbol{s}}) e^{-\mathrm{i}k(r_{\boldsymbol{s}} - r'_{\boldsymbol{s}})} \ , \tag{2.110}$$

where $\omega = kc$ as usual. Assume that we image the light arriving from the source, and would like to study its coherence properties. The imaging effectively transforms the source into a two-dimensional plane, which without loss of generality we set to coincide with the xy-plane. Our points of interest are now placed on the "image" plane, i.e. the imaging sensor, which we assume to be co-planar to the source and located at $z = R \approx R_{1,2}$. See Fig. 2.3 for a depiction of the considered geometry.

Theorem 2.5.2.2 (Far-Field Van Cittert–Zernike Theorem)**.** *Given the planar geometry described above and $\vec{\boldsymbol{r}}_1$, $\vec{\boldsymbol{r}}_2$ on the "image" plane, we denote $I(\xi, \zeta)$ as the radiance that leaves the imaged source at planar coordinates (ξ, ζ) and is observed by the sensor. The implied convention is that $I(\xi, \zeta) \equiv 0$ outside the light source. Then,*

$$\Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) = 2\pi e^{-\mathrm{i}k\vartheta}\, \mathscr{F}\{I\}(k\Delta_x, k\Delta_y) \ ,$$

where $\vec{\boldsymbol{r}}_{1,2} = (x_{1,2}, y_{1,2})$ and we denote the shorthands: $\Delta_x = \frac{x_2 - x_1}{R}$, $\Delta_y = \frac{y_2 - y_1}{R}$ and $\vartheta = \frac{1}{2R}\big(|\vec{\boldsymbol{r}}_1|^2 - |\vec{\boldsymbol{r}}_2|^2\big)$.

Proof. Start with Eq. (2.110). Given (ξ, ζ), coordinates on the source plane, we adopt the following approximation:

$$r_{\boldsymbol{s}} = \sqrt{R^2 + (\xi - x_1)^2 + (\zeta - y_1)^2} \approx R + \frac{(\xi - x_1)^2 + (\zeta - y_1)^2}{2R} \ ,$$

where we power expanded the square root and neglected the $\mathcal{O}\big(\frac{1}{R^3}\big)$ terms. A similar

expression for $r'_{\boldsymbol{s}}$ follows. Define $\Delta_x = \frac{x_2 - x_1}{R}$ and $\Delta_y = \frac{y_2 - y_1}{R}$ then the distance difference becomes:

$$\begin{aligned} r_{\boldsymbol{s}} - r'_{\boldsymbol{s}} &= \frac{(\xi - x_1)^2 + (\zeta - y_1)^2}{2R} - \frac{(\xi - x_2)^2 + (\zeta - y_2)^2}{2R} \\ &= \vartheta + \xi\Delta_x + \zeta\Delta_y \ , \end{aligned} \tag{2.111}$$

with

$$\vartheta = \frac{x_1^2 + y_1^2 - x_2^2 - y_2^2}{2R} = \frac{1}{2R}\big(|\vec{\boldsymbol{r}}_1|^{\,2} - |\vec{\boldsymbol{r}}_2|^{\,2}\big) \ , \tag{2.112}$$

And we conclude that the mutual coherence is:

$$\begin{aligned} \Gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) &\approx e^{-\mathrm{i}k\vartheta} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}\xi\, \mathrm{d}\zeta\, I(\xi,\zeta) e^{-\mathrm{i}k(\xi\Delta_x + \zeta\Delta_y)} \\ &= 2\pi e^{-\mathrm{i}k\vartheta}\, \mathscr{F}\{I\}(k\Delta_x, k\Delta_y) \ . \end{aligned} \tag{2.113}$$

The R^{-2} factor from Eq. (2.110) was eliminated due to $I(\xi,\zeta)$ having units of radiance (flux per solid angle per area). □

The far-field Van Cittert–Zernike theorem is of great practical import: When the source is far, compared to the size of the source and the distance between the points of interest, it relates the mutual coherence to the intensity distribution across the source as Fourier transform pairs.

Before we conclude, it is worthwhile to clarify some terminology: We sometimes refer to a light source as "incoherent" or "coherent", however strictly speaking only wave ensembles at particular regions in time and space can be described as coherent or incoherent. An "incoherent" light source, which are essentially all natural sources, are

incoherent in the sense that they are composed of uncorrelated radiators. Nonetheless, radiation from such sources can still be coherent. The spatial region over which a wave ensemble remains somewhat coherent can be quantified, and we will do so next.

Spatial Coherence Area

Using the Van Cittert–Zernike theorem we can now present a well-known and practically very important lemma that readily allows us to estimate the spatial region over which light from natural sources remains coherent.

We define the *spatial coherence area* as

$$A_{\mathrm{C}}(\vec{\boldsymbol{r}}) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}x\,\mathrm{d}y\,|\gamma(x,y)|^2 \ . \tag{2.114}$$

The degree of coherence $\gamma(x,y)$ is the intensity normalized version of the planar form of the mutual coherence, similar to the far-field Van Cittert–Zernike theorem (Theorem 2.5.2.2).

Lemma 2.5.2.3. *Given a monochromatic disk-shaped planar source with radius ζ and with a uniform radiant flux per unit area (i.e. constant radiosity). The spatial coherence area A_{C} satisfies the following proportionality relation:*

$$A_{\mathrm{C}}(\vec{\boldsymbol{r}}) \propto \frac{\lambda^2}{\Omega_S} \ ,$$

where λ is the wavelength and Ω_S is the solid angle subtended by the source at $\vec{\boldsymbol{r}}$.

Proof. The intensity that reaches a point $\vec{\boldsymbol{r}}$ is $I(\vec{\boldsymbol{r}}) = \pi\zeta^2 I_0 R^{-2}$, with R being the distance to the source. Away from the source the spatial coherence area can be computed using

the far-field Van Cittert–Zernike theorem:

$$\begin{aligned} A_{\mathrm{C}}(\vec{\boldsymbol{r}}) &= \iint \mathrm{d}x\,\mathrm{d}y \left| \frac{\Gamma(x,y)}{I(\vec{\boldsymbol{r}})} \right|^2 \\ &\approx \iint \mathrm{d}x\,\mathrm{d}y \left| \frac{I_0 R^{-2} \iint_{\mathrm{disk}} \mathrm{d}\xi\,\mathrm{d}\zeta\, e^{-\mathrm{i}k\frac{\xi x+\zeta y}{R}}}{\pi\zeta^2 I_0 R^{-2}} \right|^2 \\ &= \frac{1}{\pi^2\zeta^4} \iint \mathrm{d}x\,\mathrm{d}y \left| \frac{2\pi\zeta \mathrm{J}_1\Big(k\zeta R^{-1}\sqrt{x^2+y^2}\Big)}{kR^{-1}\sqrt{x^2+y^2}} \right|^2 \\ &= \frac{8\pi R^2}{k^2\zeta^2} \int_0^\infty \mathrm{d}\rho\,\rho \left| \frac{\mathrm{J}_1(k\zeta R^{-1}\rho)}{\rho} \right|^2 , \end{aligned} \tag{2.115}$$

where we used the circular aperture Fourier transform identity (2.38), and transformed from cartesian coordinates to polar coordinates via $\mathrm{d}x\,\mathrm{d}y = \mathrm{d}\theta\,\rho\,\mathrm{d}\rho$. The Bessel function of the first kind, $\mathrm{J}_1(a)$ admits its first zero at $a_0 \approx 3.8317$ [1]. It makes practical sense to limit the integration till a_0, and thereby we get the expression

$$\begin{aligned} A_{\mathrm{C}}(\vec{\boldsymbol{r}}) &\approx \frac{8\pi R^2}{k^2\zeta^2} \int_0^{a_0} \frac{\mathrm{d}\rho'}{\rho'} |\mathrm{J}_1(\rho')|^2 \\ &= \frac{4\pi R^2}{k^2\zeta^2}\big(1 - \mathrm{J}_0(a_0)^2 - \mathrm{J}_1(a_0)^2\big) \approx \frac{4R^2\lambda^2}{5\pi\zeta^2} \propto \frac{\lambda^2}{\Omega_S} , \end{aligned} \tag{2.116}$$

where we used the following Bessel integral identity [8]: $\int_0^a \rho^{-1}\mathrm{J}_1(\rho)^2 = \frac{1}{2}\big(1 - \mathrm{J}_0(a)^2 - \mathrm{J}_1(a)^2\big)$, and the approximation $1 - \mathrm{J}_0(a_0)^2 \approx \frac{4}{5}$. Also, $k = 2\pi\lambda^{-1}$, as usual. □

Lemma 2.5.2.3 presents an order-of-magnitude relation: The spatial coherence area induced by a perfectly circular light source is proportional to the inverse solid angle subtended by the source. It is also implied that longer wavelength light will generally remain coherent over greater spatial regions compared to light with shorter wavelengths. This approximative relation is conserved, in practice, for sources of other shapes.

It is of interest to briefly examine the behaviour of the degree of coherence, $|\gamma|$, when

$\rho > a_0$ in the integral in Eq. (2.115). For $a \gg 1$ the following asymptotic expansion of the Bessel function of the first kind is known [1]:

$$\mathrm{J}_1(a) \approx \sqrt{\frac{2}{\pi a}}\left[\cos\left(a - \frac{3\pi}{4}\right) + \mathcal{O}\big(|\rho|^{-1}\big)\right] . \tag{2.117}$$

That is, when the distance between the points of interest $\vec{\boldsymbol{r}}$ and $\vec{\boldsymbol{r}}'$ is exactly $\rho_0 = R\frac{\lambda a_0}{2\pi\zeta}$ the degree of coherence is zero, however as the distance increases past ρ_0 the degree of coherence oscillates, but nonetheless remains small. Indeed, $|\gamma|$ as a function of distance behaves like an Airy disk.

2.5.3 Propagation of Coherence Functions

From our discussion of the Van Cittert–Zernike theorem, we know that spatial coherence is generated by propagation. In this subsection we examine the propagation operators of the space-time formulation as well as the space-frequency formulation. As before, let ψ be a wave ensemble. Representing a wave, ψ obeys the wave operator:

$$\nabla^2\psi(\vec{\boldsymbol{r}}, t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi(\vec{\boldsymbol{r}}, t) , \tag{2.118}$$

multiply both sides by $\psi(\vec{\boldsymbol{r}}', t')^\star$ and take the time-average:

$$\nabla^2_{\vec{\boldsymbol{r}}}\left\langle\psi(\vec{\boldsymbol{r}}, t)\psi(\vec{\boldsymbol{r}}', t')^\star\right\rangle_{\mathrm{t}} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left\langle\psi(\vec{\boldsymbol{r}}, t)\psi(\vec{\boldsymbol{r}}', t')^\star\right\rangle_{\mathrm{t}} , \tag{2.119}$$

where the derivatives are with respect to the unprimed coordinates, $\vec{\boldsymbol{r}}$ and t. The above results in

$$\nabla^2_{\vec{\boldsymbol{r}}}\Gamma\big(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', t' - t\big) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Gamma\big(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', t' - t\big) . \tag{2.120}$$

That is, the mutual coherence function propagates in accordance to the wave operator.

The mutual coherence function can be written as the inverse Fourier transform of the cross-spectral density, viz.

$$\Gamma\big(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\tau\big)=\mathscr{F}^{-1}\big\{\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\omega\right)\big\}\ . \tag{2.121}$$

By applying the wave operator (Eq. (2.62)) to each side of the equation above, and formally interchanging the order of integration and differentiation, we get

$$\begin{aligned}\triangle_{\mathrm{m}}\,\Gamma\big(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\tau\big)&=\triangle_{\mathrm{m}}\,\mathscr{F}^{-1}\big\{\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\omega\right)\big\}\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\triangle_{\mathrm{m}}\left[\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\omega\right)e^{-\mathrm{i}\omega\tau}\right]\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\left[\nabla^2+\frac{\omega^2}{c^2}\right]\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\omega\right)e^{-\mathrm{i}\omega\tau}\ ,\end{aligned} \tag{2.122}$$

which can only vanish for all time differences, τ, if the integrand is identically 0. The same conclusion can also be trivially reached by considering the spectral decomposition of ψ into elementary monochromatic wavelets, $\{v(\vec{\boldsymbol{r}},\omega)e^{-\mathrm{i}\omega t}\}$, as discussed in Subsection 2.5.1. Applying the wave operator to each wavelet in the ensemble and taking the space-frequency ensemble average immediately shows that indeed the cross-spectral density obeys the Helmholtz equation.

Hence, we can conclude:

$$\triangle_{\mathrm{m}}\,\Gamma\big(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\tau\big)=0\ , \tag{2.123}$$

$$\left[\nabla^2+\frac{\omega^2}{c^2}\right]\mathcal{C}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}}',\omega\right)=0\ . \tag{2.124}$$

That is, the propagation of the mutual coherence is dictated by the wave operator. And, as expected, the cross-spectral density—being the Fourier transform pair of the mutual

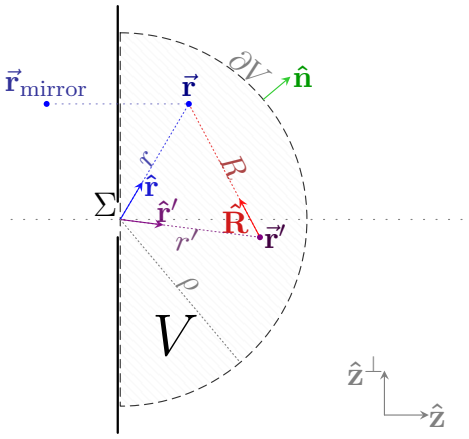


Figure 2.4: Propagation of cross-spectral density.

coherence—propagates in accordance to the Helmholtz equation, which is generally easier to solve.

Propagation as a diffraction problem

As the mutual coherence and the cross-spectral density functions propagate as waves, we can formulate the propagation of those coherence functions as a diffraction problem. We focus on the cross-spectral density. The cross-spectral density propagates according to the Helmholtz equation, which calls for a Green function treatment. Recall the free-space Green function G_0 (Eq. (2.46)):

$$\left[\nabla^2 + k^2\right] G_0\left(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\right) = -\delta^3(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}') \ . \tag{2.125}$$

Consider a product of two free-space Green functions that are acted upon by a couple of Helmholtz operators:

$$\begin{aligned}\left[\nabla_1'^2 + k^2\right]\left[\nabla_2'^2 + k^2\right] G_0\left(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_1'\right)^{\star} G_0\left(\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}_2'\right) \\ = \delta^3(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_1')\delta^3(\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}_2') \ .\end{aligned} \tag{2.126}$$

We denote $\nabla^2_{1,2}$, $\nabla'^2_{1,2}$ as the laplacians with respect to $\vec{\boldsymbol{r}}_{1,2}$ and $\vec{\boldsymbol{r}}'_{1,2}$, respectively. Multiply both sides by $\mathcal{C}$ and integrate twice over the space $V = \{z > 0\} \cap B_\rho$, where B_ρ is a ball of some radius ρ centred at the origin (see Fig. 2.4 for a visualization of the geometry):

$$\begin{aligned}\mathcal{C}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega) = &\int_V \mathrm{d}^3\vec{\boldsymbol{r}}'_1 \left[\nabla'^2_1 + k^2\right] G_0\big(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}'_1\big)^\star \\ &\times \int_V \mathrm{d}^3\vec{\boldsymbol{r}}'_2 \left[\nabla'^2_2 + k^2\right] G_0\big(\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}'_2\big)\, \mathcal{C}\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right) . \end{aligned} \tag{2.127}$$

Consider the second integral on the right-hand side: By noting that $k^2\,\mathcal{C} = -\nabla'^2_2\,\mathcal{C}$ (directly from Eq. (2.124), the Helmholtz equation) and applying Green's second identity we get

$$\begin{aligned}\int_V \mathrm{d}^3\vec{\boldsymbol{r}}'_2 \left[\nabla'^2_2 + k^2\right] G_0\,\mathcal{C} &= \int_V \mathrm{d}^3\vec{\boldsymbol{r}}'_2 \left[\mathcal{C}\,\nabla'^2_2 G_0 - G_0 \nabla'^2_2\,\mathcal{C}\right] \\ &= \oint_{\partial V} \mathrm{d}^2\vec{\boldsymbol{r}}'_2\, \hat{\boldsymbol{n}} \cdot \left[\mathcal{C}\,\boldsymbol{\nabla}'_2 G_0 - G_0 \boldsymbol{\nabla}'_2\,\mathcal{C}\right] , \end{aligned} \tag{2.128}$$

where $\boldsymbol{\nabla}'_2$ is the gradient with respect to $\vec{\boldsymbol{r}}'_2$ and ∂V is th boundary of V with $\hat{\boldsymbol{n}}$ being its outward normal at a point. Denote the shorthand $\vec{\boldsymbol{R}}_2 = \vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}'_2$. The gradient of the free-space Green function follows immediately from rewriting Eq. (2.46) in spherical coordinates:

$$\boldsymbol{\nabla}' G_0\big(\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}'_2\big) = \hat{\boldsymbol{R}}_2 \frac{\partial}{\partial R_2} G_0(R_2) = \hat{\boldsymbol{R}}_2 \frac{\mathrm{i}kR_2 - 1}{4\pi R_2^2} e^{\mathrm{i}kR_2} , \tag{2.129}$$

where $R_2 = |\vec{\boldsymbol{R}}_2|$ and $\hat{\boldsymbol{R}}_2 = \vec{\boldsymbol{R}}_2 / R_2$ is the radius basis vector in spherical coordinates (see Fig. 2.4). The Sommerfeld radiation condition (Eq. (2.45)) applies to outgoing spherical wavelets, and therefore it applies to the cross-spectral density as well. Therefore $\mathcal{C} \to 0$ as $\vec{\boldsymbol{r}}'_2 \to \infty$ at least faster than $1/R_2$, and thus the integrand on the right-hand side in Eq. (2.128) decays to 0 faster than R_2^{-2}. We conclude that the integrand vanishes

everywhere at the limit $\vec{\boldsymbol{r}}_2' \to \infty$. Hence, by taking the limit $\rho \to \infty$ (recall ρ is the radius of the half-sphere ∂V) and substituting the expressions for G_0 and its gradient into Eq. (2.128), we get

$$\begin{aligned}\lim_{\rho\to\infty}\int_V \mathrm{d}^3\vec{\boldsymbol{r}}_2' \left[\nabla_2'^2 + k^2\right] G_0\, \mathcal{C} \\ = -\int_{\hat{\boldsymbol{z}}^\perp} \mathrm{d}^2\vec{\boldsymbol{r}}_2' \, \frac{e^{\mathrm{i}kR_2}}{4\pi R_2}\hat{\boldsymbol{z}} \cdot \left[\hat{\boldsymbol{R}}_2 \frac{\mathrm{i}kR_2 - 1}{R_2}\,\mathcal{C} - \boldsymbol{\nabla}_2'\,\mathcal{C}\right] ,\end{aligned} \tag{2.130}$$

where $\hat{\boldsymbol{z}}^\perp$ is the orthogonal complement of the z-axis, i.e. the xy-plane. The integral above is a superposition of spherical wavelets that propagate from the plane $\hat{\boldsymbol{z}}^\perp$. Indeed, it is a mathematical formulation of Huygens' principle and the integral itself is nothing more than the Kirchhoff diffraction integral. Proceeding with the same approach we also apply Kirchhoff's approximation and assume that the electric field (and in turn, the cross-spectral density) vanishes everywhere on $\hat{\boldsymbol{z}}^\perp$, except an area $\Sigma \subset \hat{\boldsymbol{z}}^\perp$ which can be considered either as an aperture or a light source. Kirchhoff's boundary conditions produce an overdetermined problem by also assuming that the derivative of $\mathcal{C}$ vanishes on $\hat{\boldsymbol{z}}^\perp$. This is not necessary, and we follow Zangwill [61] by replacing the free-space Green function with a mirror-image variant:

$$\tilde{G}_0\big(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}'\big) \triangleq G_0\big(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\big) - G_0\big(\vec{\boldsymbol{r}}_{\text{mirror}} - \vec{\boldsymbol{r}}'\big) , \tag{2.131}$$

where in the second Green function $\vec{\boldsymbol{r}}_{\text{mirror}} = \hat{\boldsymbol{x}} r_x + \hat{\boldsymbol{y}} r_y - \hat{\boldsymbol{z}} r_z$, i.e. $\vec{\boldsymbol{r}}$ mirrored with respect to the $\hat{\boldsymbol{z}}^\perp$ plane. Clearly, $\tilde{G}_0$ vanishes on $\vec{\boldsymbol{r}} \in \hat{\boldsymbol{z}}^\perp$, which dispenses with the $\boldsymbol{\nabla}_2'\,\mathcal{C}$ term—the Neumann boundary condition—in Eqs. (2.128) and (2.130). When acted upon by the Helmholtz operator $\tilde{G}_0$ produces two Dirac deltas, however one of them lies outside of V and does not contribute to Eq. (2.127). Hence, by replacing G_0 with $\tilde{G}_0$ in Eq. (2.128)

we are left with a mathematically well-defined problem and Eq. (2.130) reduces to

$$\lim_{\rho\to\infty}\int_V \mathrm{d}^3\vec{\boldsymbol{r}}_2'\left[\nabla_2'^2+k^2\right]\tilde{G}_0\,\mathcal{C}=-\int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}_2'\,\mathcal{C}\,\frac{\partial}{\partial z}\left[\frac{e^{\mathrm{i}kR_2}}{2\pi R_2}\right]\,, \tag{2.132}$$

which is the Rayleigh-Sommerfeld diffraction integral of the first kind [8]. The integral above is a solution of the Helmholtz equation in the $z>0$ half-space expressed in terms of the boundary values on Σ.

In an entirely analogous manner the same process is readily applied to the first integral (over $\vec{\boldsymbol{r}}_1'\in V$) in Eq. (2.127). Note that $\frac{\partial}{\partial z}=\frac{\partial}{\partial R_j}\frac{\partial}{\partial z}R_j=\frac{z}{R_j}\frac{\partial}{\partial R_j}$ and $\frac{\partial}{\partial R_j}\frac{e^{\mathrm{i}kR_j}}{R_j}=\frac{e^{\mathrm{i}kR_j}}{R_j}[\mathrm{i}k-1/R_j]$, for $j\in\{1,2\}$. At which point we finally end up with the following expression

$$\begin{aligned}\mathcal{C}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right)=&\frac{1}{4\pi^2}\int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}_1'\int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}_2'\,\mathcal{C}\left(\vec{\boldsymbol{r}}_1',\vec{\boldsymbol{r}}_2',\omega\right)\\&\times\frac{z_1z_2}{R_1^2R_2^2}e^{\mathrm{i}k(R_1-R_2)}\left(-\mathrm{i}k-\frac{1}{R_1}\right)\left(\mathrm{i}k-\frac{1}{R_2}\right)\,,\end{aligned} \tag{2.133}$$

where z_1 and z_2 are the z-components of $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$, respectively. Denote $\hat{\boldsymbol{r}}_j$ as the unit vector in direction of $\vec{\boldsymbol{r}}_j$ and r_j the distance to $\vec{\boldsymbol{r}}_j$, and similarly with primed variables (see Fig. 2.4). Then, in the Fresnel-Fraunhofer diffraction region—far from the source with respect to the characteristic size of the source—we can approximate $R_1\approx r_1$ and $R_2\approx r_2$ (for $\vec{\boldsymbol{r}}_j'\in\Sigma$). Furthermore, for visible light clearly $kr_{1,2}\gg 1$ in that region, and thus the fractions $1/R_j$ are negligible. In the high-frequency exponent a better approximation is required and we power expand $R_{1,2}$ (in similar manner to the proof of Theorem 2.5.2.2):

$$R_j=\left|\vec{\boldsymbol{r}}_j-\vec{\boldsymbol{r}}_j'\right|=r_j+\frac{r_j'^2-2\vec{\boldsymbol{r}}_j\cdot\vec{\boldsymbol{r}}_j'}{2r_j}+\mathcal{O}\left(\frac{1}{r_j^2}\right)\approx r_j-\hat{\boldsymbol{r}}_j\cdot\vec{\boldsymbol{r}}_j'\,. \tag{2.134}$$

Applying all of the above, Eq. (2.133) simplifies to

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \approx \frac{\cos\theta_1 \cos\theta_2}{\lambda^2 r_1 r_2} e^{\mathrm{i}k(r_1 - r_2)} \times \int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}'_1 \, e^{-\mathrm{i}k\hat{\boldsymbol{r}}_1 \cdot \vec{\boldsymbol{r}}'_1} \int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}'_2 \, \mathcal{C}\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right) e^{\mathrm{i}k\hat{\boldsymbol{r}}_2 \cdot \vec{\boldsymbol{r}}'_2} \ , \tag{2.135}$$

where $\lambda = \frac{2\pi}{k}$ is the wavelength and θ_j is the inclination angle, i.e. $\cos\theta_j = \hat{\boldsymbol{r}}_j \cdot \hat{\boldsymbol{z}}$. The above formula is the (scalar) propagation integral for the cross-spectral density, and a essentially is a pair of 2D spatial Fourier transforms.

2.6 Polarization

The electric and magnetic fields supporting propagating electromagnetic waves generally admit transverse components. The wave's *polarization* describes the geometry of the transverse oscillations of the supporting fields. Assume that a wave packet's direction of propagation is concentrated along the $+\hat{\boldsymbol{z}}$ direction. Denote $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$ as the orthonormal basis of the orthogonal complement $\hat{\boldsymbol{z}}^\perp$ of the propagation direction. Then, we concentrate on studying the polarization of the (time-dependent) electric field $\vec{\boldsymbol{E}}$. We treat the electric field as a stationary, at least in the wide sense, stochastic process and let $E_{1,2} = \hat{\boldsymbol{e}}_{1,2} \cdot \vec{\boldsymbol{E}}$ be the mutually-orthogonal transverse components of the electric field. Define the *polarization matrix* as the cross-correlations between the transverse components:

$$\boldsymbol{J} \triangleq \frac{c}{8\pi} \begin{bmatrix} \langle E_1 E_1^\star \rangle_{\mathrm{t}} & \langle E_1 E_2^\star \rangle_{\mathrm{t}} \\ \langle E_2 E_1^\star \rangle_{\mathrm{t}} & \langle E_2 E_2^\star \rangle_{\mathrm{t}} \end{bmatrix} \ . \tag{2.136}$$

The polarization matrix was introduced by Born and Wolf in older literature as the "coherency matrix", before the advent of optical coherence theory, however the term is

a misnomer: The polarization matrix elements are the (equal-time) cross-correlations of the transverse fields at a single spatial position, and the matrix describes the polarization characteristics of the wave packet at that position. Coherence studies the cross-correlation between points in multiple space and time points. Denote the matrix elements J_{xy}, with $x, y \in \{1, 2\}$, and we first briefly consider the mathematical properties of the polarization matrix. The matrix is clearly Hermitian, i.e. $\boldsymbol{J}^\dagger = \boldsymbol{J}$. By applying Schwarz's inequality (Eq. (2.18)) we deduce that

$$|J_{12}| = |J_{21}| \leq \sqrt{|J_{11}|}\sqrt{|J_{22}|} \ , \tag{2.137}$$

and it follows that $\boldsymbol{J}$ is also positive semi-definite, i.e. $\boldsymbol{J} \succeq 0$, and thus $\det \boldsymbol{J} \geq 0$ and all the eigenvalues are real and non-negative. The diagonal elements of the equal-time polarization matrix are always real, and are the time-averaged energy density carried by the electric field. For the idealized case of a plane-wave, that is exactly the observed intensity (Eq. (2.74)) of each orthogonal component of the wave, and thus the intensity of the wave is the sum of the diagonal elements, i.e.

$$I = \operatorname{tr} \boldsymbol{J} \ , \tag{2.138}$$

Also note that the determinant and the trace of $\boldsymbol{J}$ are invariant under the choice of the basis $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$: Indeed, for any rotation matrix $\boldsymbol{R}$ the rotated transverse field components become

$$\begin{bmatrix} E_1' \\ E_2' \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} E_1 \cos\theta - E_2 \sin\theta \\ E_1 \sin\theta + E_2 \cos\theta \end{bmatrix} \ , \tag{2.139}$$

where θ is any (counter-clockwise) rotation angle. With a little algebra we can verify that if the polarization matrix of the new decomposition is $\boldsymbol{J}'$, then $\det \boldsymbol{J} = \det \boldsymbol{J}'$ and $\operatorname{tr} \boldsymbol{J} = \operatorname{tr} \boldsymbol{J}'$, as we would physically expect. It is also noteworthy that the polarization matrix of a superposition of statistically independent waves is simply the sum of the polarization matrices of each wave. This result follows immediately from the fact that the time-average of independent process vanishes.

We are now ready to define the *degree of correlation* of the orthogonal components, which is the normalized off-diagonal element

$$j_{12} \triangleq \frac{1}{\sqrt{J_{11}}\sqrt{J_{22}}} J_{12} \tag{2.140}$$

at a spatial position of interest. Clearly, $|j_{12}| \leq 1$ (when one of the diagonal elements vanishes, we can check that $|j_{12}| = 1$ at the limit).

Unpolarized light We define *unpolarized light* such that its polarization matrix fulfils the following: 1. under any choice of basis, the transverse field components are uncorrelated; and 2. the observed intensity along any transverse direction is identical. Or, equivalently, $|j_{12}| = 0$ and $J_{11} = J_{22}$. It is easy to check that light that is unpolarized remains unpolarized regardless of how we decompose the transverse field oscillations. Given the conditions above, we can write the polarization matrix of an unpolarized wave as

$$\boldsymbol{J}_{\mathrm{u}} = \frac{1}{2} \begin{bmatrix} I_{\mathrm{u}} & \\ & I_{\mathrm{u}} \end{bmatrix} , \tag{2.141}$$

where I_{u} is the intensity of the wave (for a plane wave).

Polarized light Given $|j_{12}| = 1$, the transverse oscillations are fully correlated and we call such light completely *polarized light*. It trivially follows that

$$\det \boldsymbol{J} = J_{11}J_{22} - J_{12}(J_{12})^{\star} = J_{11}J_{22} - \left|\sqrt{J_{11}}\sqrt{J_{22}}j_{12}\right|^2 = 0 \ , \tag{2.142}$$

and, as the determinant is basis invariant, we deduce that completely polarized light, like unpolarized light, remains completely polarized under rotation of the transverse frame. The polarization matrix in this case can be written in the following form:

$$\boldsymbol{J}_{\mathrm{p}} = \frac{c}{8\pi}\begin{bmatrix} J_{11} & \sqrt{J_{11}}\sqrt{J_{22}}e^{\mathrm{i}\varphi} \\ \sqrt{J_{11}}\sqrt{J_{22}}e^{-\mathrm{i}\varphi} & J_{22} \end{bmatrix} \ , \tag{2.143}$$

where φ is any real number.

Corollary 2.6.0.0.1. *A completely polarized polarization matrix has eigenvalues* 0 *and* $J_{11} + J_{22}$*, and its Cholesky decomposition is* $\boldsymbol{L}\boldsymbol{L}^{\dagger}$ *with*

$$\boldsymbol{L} = \begin{bmatrix} \sqrt{J_{11}} & 0 \\ \sqrt{J_{22}}e^{-\mathrm{i}\varphi} & 0 \end{bmatrix} \ . \tag{2.144}$$

Observe from Eq. (2.143) that the phase difference between E_1 and E_2, φ, must be independent of time, as otherwise E_1 and E_2 are not fully correlated. When $\varphi = m\pi$, for some integer m, then the transverse oscillations are fully in-phase or delayed by π (which is equivalent to fully in-phase with a negated E_2), and we deduce that such a wave is completely *linearly polarized*. A necessary and sufficient condition for completely linearly polarized light is then $\det \boldsymbol{J} = 0$ and $\operatorname{Im} j_{12} = 0$. When $\operatorname{Im} j_{12} \neq 0$ the wave will generally be a (correlated) superposition of linearly polarized light and non-linearly polarized light. Under the context of a monochromatic plane-wave, a typical treatment of the subject of

optical polarization further classifies the states of polarization into circular polarization ($\varphi = \pm \mathrm{i}$), and elliptical polarization (the general complete polarization state, with linear and circular being special cases).

The *major axis* of the polarization matrix is defined as the transverse direction in which most of the energy of the wave packet is concentrated. The following lemma provides an easy formula to locate that direction.

Lemma 2.6.0.1. *For any polarization matrix, the angle between the major axis and* $\hat{\boldsymbol{e}}_1$ *is* ψ*, where*

$$\tan(2\psi) = 2\,\mathrm{Re}\,\frac{J_{12}}{J_{11} - J_{22}}\ .$$

Proof. See Born and Wolf [8]. □

Suppose we were to align $\hat{\boldsymbol{e}}_1$ with the major axis of a polarization matrix by applying a (counter-clockwise) rotation through an angle ψ to the transverse frame. Then, J_{11} would be maximized in that frame. Under that frame, the quantity $\frac{J_{11}-J_{22}}{J_{11}+j_{22}}$ is called the *degree of linear polarization*. For example, if the polarization matrix were to be completely linearly polarized then J_{22} would vanish, meaning that the entirety of the energy is concentrated in the direction of the major axis and the degree of linear polarization is 1. On the other hand, polarization matrices of unpolarized light or of completely polarized light with $\mathrm{Re}\, j_{12} = 0$ admit $J_{11} = J_{22}$ under any frame, as the energy is equally distributed in any direction and the degree of linear polarization is 0.

As a side note, the geometric shape traced by transverse field lines of polychromatic light is called a "Lissajous" figure, which closes upon itself with a regular period when the frequencies composing the wave packet are commensurable, that is when $\omega_{j+1}/\omega_j \in \mathbb{Q}$ for all (finite) frequencies ω_j.

Decomposition of light of any state of polarization

Consider again a polychromatic wave packet propagating in direction $\hat{\boldsymbol{z}}$. Its polarization matrix $\boldsymbol{J}$, under the most general conditions, is as presented in Eq. (2.136). We will now show that it is possible to decompose the wave packet into a superposition of unpolarized light and completely polarized light.

Lemma 2.6.0.2 (Canonical Decomposition of the Polarization Matrix)**.** *Any polarization matrix $\boldsymbol{J}$ can always be decomposed into*

$$\boldsymbol{J} = \boldsymbol{J}_{\mathrm{u}} + \boldsymbol{J}_{\mathrm{p}} \ ,$$

where each matrix retains the mathematical properties of the polarization matrix, namely Hermiticity and positive semi-definiteness, and the following holds:

- $\boldsymbol{J}_{\mathrm{u}} = \frac{I_{\mathrm{u}}}{2}\boldsymbol{I}$*, that is $\boldsymbol{J}_{\mathrm{u}}$ describes the completely unpolarized part of the light.*
- $\det \boldsymbol{J}_{\mathrm{p}} = 0$*, i.e. $\boldsymbol{J}_{\mathrm{p}}$ is a completely polarized polarization matrix.*

Furthermore, this decomposition is unique.

Proof. Let J_{xy}, with $x, y \in \{1, 2\}$, be the elements of the matrix $\boldsymbol{J}$ and $j_{12} = (J_{11}J_{22})^{\frac{1}{2}}J_{12}$ be the degree of correlation, as before. We are looking for a decomposition

$$\boldsymbol{J} = \frac{I_{\mathrm{u}}}{2}\boldsymbol{I} + \boldsymbol{J}_{\mathrm{p}} \tag{2.145}$$

such that $\det \boldsymbol{J}_{\mathrm{p}} = 0$, which implies that $|\boldsymbol{J} - \frac{1}{2}I_{\mathrm{u}}\boldsymbol{I}| = 0$, i.e. $\frac{1}{2}I_{\mathrm{u}}$ is an eigenvalue of $\boldsymbol{J}$. $\boldsymbol{J}$ is Hermitian and positive semi-definite, therefore it has real non-negative eigenvalues. Note further that if both the eigenvalues vanish, $\boldsymbol{J}$ must be a singular matrix and $\det \boldsymbol{J} =$

0, i.e. it is completely polarized, then $\boldsymbol{J}_{\mathrm{u}} = 0$. Solving for the eigenvalues we get

$$I_{\mathrm{u}} = J_{11} + J_{22} \pm \sqrt{\left(J_{11} - J_{22}\right)^2 + 4J_{11}J_{22}|j_{12}|^2} \ . \tag{2.146}$$

The completely polarized fraction of the energy would then be

$$\boldsymbol{J}_{\mathrm{p}} = \boldsymbol{J} - \boldsymbol{J}_{\mathrm{u}} = \begin{bmatrix} J'_{11} & J_{12} \\ J^{\star}_{12} & J'_{22} \end{bmatrix} \ , \tag{2.147}$$

where $J'_{xx} = J_{xx} - \frac{1}{2}I_{\mathrm{u}}$. Note that J'_{xx} is negative when we select the greater eigenvalue (with the positive sign), violating the positive semi-definiteness requirement. Therefore, I_{u} must be the smaller eigenvalue, and $\boldsymbol{J}_{\mathrm{u}}$ follows immediately.

Clearly, as we always select the smaller eigenvalue, the decomposition is unique. □

Corollary 2.6.0.2.1. *Let the following be a decomposition of a completely polarized matrix $\boldsymbol{J}_{\mathrm{p}}$ into a couple of polarization matrices:*

$$\boldsymbol{J}_{\mathrm{p}} = \boldsymbol{J}'_{\mathrm{p}} + \boldsymbol{J}''_{\mathrm{p}} \ .$$

We state, without proving, that the only possible such decomposition is a trivial decomposition, i.e. $\boldsymbol{J}'_{\mathrm{p}} = \alpha\boldsymbol{J}_{\mathrm{p}}$ and $\boldsymbol{J}''_{\mathrm{p}} = (1-\alpha)\boldsymbol{J}_{\mathrm{p}}$ for some $0 \leq \alpha \leq 1$. It is easy to check that a completely polarized matrix admits no decomposition with a partially polarized or unpolarized matrix. Of more consequence is the conclusion that the phase-shifts, i.e. φ in Eq. (2.143)*, of each of the decomposing matrices is identical to the phase-shift of $\boldsymbol{J}$.*

That is, a completely polarized wave ensemble is fully correlated, and therefore can not be decomposed into different states of polarization.

To conclude, we define the important *degree of polarization* as the ratio between the

polarized fraction of the energy and the total energy carried by the wave:

$$\mathcal{P} \triangleq \frac{\operatorname{tr} \boldsymbol{J}_{\mathrm{p}}}{\operatorname{tr} \boldsymbol{J}} = \frac{I_{\mathrm{p}}}{I} = \sqrt{1 - \frac{4 \det \boldsymbol{J}}{(\operatorname{tr} \boldsymbol{J})^2}} \ , \tag{2.148}$$

where the identity $I_{\mathrm{p}} = (\operatorname{tr} \boldsymbol{J})^2 - 4 \det \boldsymbol{J}$ follows directly from Eq. (2.146). When $\mathcal{P} = 1$ the light is completely polarized, while $\mathcal{P} = 0$ indicates unpolarized light. When $0 < \mathcal{P} < 1$ we say the light is *partially polarized*.

2.6.1 Unified Theory of Polarization and Coherence

We now make contact with optical coherence theory presented in Section 2.5. The discussion in this section outlines a unified theory of optical coherence that takes polarization into account and presents a few interesting results. Our discussion is based on Wolf [59].

The polarization matrix presented in the previous subsection can be generalized into the mutual coherence matrix that also accounts for coherence phenomena, viz.

$$\begin{aligned} \boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) &\triangleq \frac{c}{8\pi} \begin{bmatrix} \Gamma_{11}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) & \Gamma_{12}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) \\ \Gamma_{21}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) & \Gamma_{22}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \tau) \end{bmatrix} \\ &= \frac{c}{8\pi} \begin{bmatrix} \langle E_1(\vec{\boldsymbol{r}}_1, t) E_1(\vec{\boldsymbol{r}}_2, t+\tau)^\star \rangle_{\mathrm{t}} & \langle E_1(\vec{\boldsymbol{r}}_1, t) E_2(\vec{\boldsymbol{r}}_2, t+\tau)^\star \rangle_{\mathrm{t}} \\ \langle E_2(\vec{\boldsymbol{r}}_1, t) E_1(\vec{\boldsymbol{r}}_2, t+\tau)^\star \rangle_{\mathrm{t}} & \langle E_2(\vec{\boldsymbol{r}}_1, t) E_2(\vec{\boldsymbol{r}}_2, t+\tau)^\star \rangle_{\mathrm{t}} \end{bmatrix} \ , \end{aligned} \tag{2.149}$$

where Γ_{xy} is the mutual coherence function relating the orthogonal field components. The above is the space-time formulation of optical coherence, and the space-frequency

readily follows with the cross-spectral density matrix:

$$\begin{aligned}\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right) &\triangleq \frac{c}{8\pi}\begin{bmatrix}\mathcal{C}_{11}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right) & \mathcal{C}_{12}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right)\\ \mathcal{C}_{21}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right) & \mathcal{C}_{22}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right)\end{bmatrix}\\ &= \frac{c}{8\pi}\begin{bmatrix}\langle E_1(\vec{\boldsymbol{r}}_1)E_1(\vec{\boldsymbol{r}}_2)^\star\rangle_\omega & \langle E_1(\vec{\boldsymbol{r}}_1)E_2(\vec{\boldsymbol{r}}_2)^\star\rangle_\omega\\ \langle E_2(\vec{\boldsymbol{r}}_1)E_1(\vec{\boldsymbol{r}}_2)^\star\rangle_\omega & \langle E_2(\vec{\boldsymbol{r}}_1)E_2(\vec{\boldsymbol{r}}_2)^\star\rangle_\omega\end{bmatrix}\,,\end{aligned} \tag{2.150}$$

where the ensemble averages are over same-frequency wavelets (see Eq. (2.97)). The relation between the matrix-versions of the mutual coherence and cross-spectral density is preserved, with them being Fourier transform pairs, viz:

$$\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right) = \mathscr{F}\left\{\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\tau)\right\}\,, \tag{2.151}$$

When $\vec{\boldsymbol{r}}_1 = \vec{\boldsymbol{r}}_2$, both matrices, $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\Gamma}$, clearly are Hermitian and positive semi-definite. Therefore, the cross-spectral density and mutual coherence matrices evaluated at a point, i.e. $\boldsymbol{\mathcal{C}}(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}},\omega)$ and $\boldsymbol{\Gamma}(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}},\tau)$, are polarization matrices. Accordingly, the rest of our analysis of polarization matrices applies.

The mutual coherence and cross-spectral density matrices are the akin to their scalar counterparts discussed in Section 2.5. Under matrix formulation, the energy carried by a wave becomes $I(\vec{\boldsymbol{r}}) = \operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}})$, where we again abuse notation and drop the time-delay for equal-time mutual coherence, and similarly the spectral density becomes $\mathfrak{s}(\vec{\boldsymbol{r}},\omega) \triangleq \operatorname{tr}\boldsymbol{\mathcal{C}}(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}},\omega)$. Then, in an analogous way, we define the matrix degree of coherence (matrix analogue of Eq. (2.82)) and the matrix spectral degree of coherence

(matrix analogue of Eq. (2.89)) as

$$\boldsymbol{\gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\tau) \triangleq \frac{\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\tau)}{\sqrt{\operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_1)}\sqrt{\operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_2,\vec{\boldsymbol{r}}_2)}} \;, \tag{2.152}$$

$$\boldsymbol{\varpi}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega) \triangleq \frac{\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right)}{\sqrt{\operatorname{tr}\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_1,\omega\right)}\sqrt{\operatorname{tr}\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_2,\vec{\boldsymbol{r}}_2,\omega\right)}} \;, \tag{2.153}$$

where the positive semi-definiteness of the matrices implies that $|\boldsymbol{\gamma}| \leq 1$ and likewise $|\boldsymbol{\varpi}| \leq 1$.

The interference laws also mostly retain their scalar form. Let $\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}},t)$ be the electric field of a wave ensemble. We are interested in the observed intensity of a superposition between the energy arriving from different points, $\vec{\boldsymbol{E}}'(\vec{\boldsymbol{r}},t) = \vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_1,t_1) + \vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_2,t_2)$. The expression for the intensity is the matrix form of the interference law for stationary fields (analogue of Eq. (2.83)):

$$\begin{aligned} I'(\vec{\boldsymbol{r}},t) = {} & \operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_1) + \operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_2,\vec{\boldsymbol{r}}_2) \\ & + 2\sqrt{\operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_1)}\sqrt{\operatorname{tr}\boldsymbol{\Gamma}(\vec{\boldsymbol{r}}_2,\vec{\boldsymbol{r}}_2)}\operatorname{Re}\operatorname{tr}\boldsymbol{\gamma}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,t_2-t_1) \;. \end{aligned} \tag{2.154}$$

In the spectral domain, the matrix form of the spectral interference law (Eq. (2.100)) takes on the familiar form:

$$\begin{aligned} \mathfrak{s}(\vec{\boldsymbol{r}},\omega) = \operatorname{tr}\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}},\vec{\boldsymbol{r}},\omega\right) = {} & \mathfrak{s}(\vec{\boldsymbol{r}}_1,\omega) + \mathfrak{s}(\vec{\boldsymbol{r}}_2,\omega) \\ & + 2\sqrt{\mathfrak{s}(\vec{\boldsymbol{r}}_1,\omega)}\sqrt{\mathfrak{s}(\vec{\boldsymbol{r}}_2,\omega)}\operatorname{Re}\left[\operatorname{tr}\boldsymbol{\varpi}(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega)e^{-\mathrm{i}\omega\frac{r_1-r_2}{c}}\right] \;, \end{aligned} \tag{2.155}$$

where r_1 and r_2 are the distances from $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$, respectively, to the observation point $\vec{\boldsymbol{r}}$. The derivation is identical to the derivations that preceded Eq. (2.100).

The matrix-based space-time and the space-frequency formalisms of optical coherence serve to merge polarization and coherence into a single entity: The diagonal elements ex-

press the coherence properties, while the off-diagonal elements describe the polarization characteristics. This is not done solely in the name of convenience—both polarization and coherence affect observable properties of electromagnetic waves and guide their interaction with matter. Indeed, polarization and coherence might appear as unrelated aspects at first, but nonetheless they are very closely intertwined in describing the behaviour of a wave ensemble. This surprisingly tight relationship is demonstrated by the observation that the coherence properties of the wave induce polarization changes on propagation [27].

2.6.2 Diffraction of Polarized Light

We now formulate the diffraction problem for electromagnetic waves, when polarization is taken into account. Under vector diffraction theory we treat incoming light as an electromagnetic incident plane wave, with its electric and magnetic components known on some aperture Σ, and we seek a unique solution that satisfies Maxwell's equations (see Fig. 2.4). As before, the geometry is an aperture that is located on $z = 0$ and we solve for $\vec{\boldsymbol{r}} = (x, y, z \geq 0)$. We follow Zangwill [61] by looking for solutions formulated as a superposition of monochromatic spherical waves, each varying as $e^{-\mathrm{i}\omega t}$, with $\omega = ck$ being the angular frequency of interest. That approach gives rise to *Smythe's diffraction formula*:

$$\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}) = 2\boldsymbol{\nabla} \times \int_{z'=0} \mathrm{d}^2\vec{\boldsymbol{r}}'_{\perp} G_0\big(\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'_{\perp}\big)\hat{\boldsymbol{z}} \times \vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_{\perp}\big) \ , \tag{2.156}$$

$$\vec{\boldsymbol{B}}(\vec{\boldsymbol{r}}) = -\frac{\mathrm{i}}{k_0}\boldsymbol{\nabla} \times \vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}) \ , \tag{2.157}$$

with G_0 being the free-space Green function. The above expression for the magnetic field arises immediately from Maxwell equations, Eq. (2.58). It can be verified that if $\vec{\boldsymbol{E}}$

satisfies Maxwell's divergence equation, $\boldsymbol{\nabla}\cdot\vec{\boldsymbol{E}}=0$ (Eq. (2.57)), on the aperture, then $\boldsymbol{\nabla}\cdot\vec{\boldsymbol{E}}=0$ on $z>0$ as well. Smythe's formula is a manifestation of Huygens' principle, with every point on the aperture being regarded a source of spherical wavelets. With a little algebra we can rewrite the expression for the electric field as

$$\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}})=-2\int_{z'=0}\mathrm{d}^2\vec{\boldsymbol{r}}'_{\perp}\Big[\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_{\perp}\big)\Big]\times\boldsymbol{\nabla}G_0\ . \tag{2.158}$$

We proceed in a fashion similar to the scalar diffraction case. At the Fresnel-Fraunhofer diffraction region, far from the aperture with respect to the wavelength and characteristic size of the aperture, we can simplify considerably. The gradient of the free-space Green function was computed previously (Eq. (2.129)):

$$\boldsymbol{\nabla}G_0\big(\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'_{\perp}\big)=\hat{\boldsymbol{R}}\frac{\mathrm{i}kR-1}{4\pi R^2}e^{\mathrm{i}kR}\ , \tag{2.159}$$

with $\vec{\boldsymbol{R}}=\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'_{\perp}$ and $R=|\vec{\boldsymbol{R}}|$, $\hat{\boldsymbol{R}}=\frac{\vec{\boldsymbol{R}}}{R}$, as usual. Using the power expansion $R\approx r-\hat{\boldsymbol{r}}\cdot\vec{\boldsymbol{r}}'_{\perp}$ (Eq. (2.134)), as well as the fact that $kR\ggg 1$, we get

$$\boldsymbol{\nabla}G_0\big(\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'_{\perp}\big)\approx ik\hat{\boldsymbol{r}}\frac{e^{\mathrm{i}kr}e^{-\mathrm{i}k\hat{\boldsymbol{r}}\cdot\vec{\boldsymbol{r}}'_{\perp}}}{4\pi r}\ . \tag{2.160}$$

As with scalar diffraction, we apply the Kirchhoff approximation, that is $\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}=0$ everywhere on $z=0$ except the aperture $\Sigma\subset\{z=0\}$. This is a good approximation at the high-frequency limit with wavelengths small compared with the aperture size. Finally, by substituting Eq. (2.160) into Eq. (2.158) the far-field Smythe's diffraction formula under the Kirchhoff approximation becomes

$$\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}})=\mathrm{i}k\frac{e^{\mathrm{i}kr}}{2\pi r}\hat{\boldsymbol{r}}\times\int_{\Sigma}\mathrm{d}^2\vec{\boldsymbol{r}}'_{\perp}\Big[\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_{\perp}\big)\Big]e^{-\mathrm{i}k\hat{\boldsymbol{r}}\cdot\vec{\boldsymbol{r}}'_{\perp}}\ . \tag{2.161}$$

2.7 Phase-Space Optics

In this section we discuss Hamiltonian optics, starting with the ray optical picture, and then progressing to wave optics.

Under the ray picture of light, light consists of "luminous" corpuscles. As such a particle evolves by propagating and interacting with optical systems, it traces a "light ray". A particle's position and *momentum*, i.e. the particle's direction of propagation, serve as a complete description of the light ray at a particular instant. Therefore, it is convenient to study the dynamics of ray optics in *phase space*: a $2n$-dimensional space that consists of n independent position coordinates, as well as n momenta coordinates (often referred to as the *canonically conjugate* variables). A ray, at a particular instant, corresponds to a point in phase-space. This phase-space pictorial view of light is adopted, sometimes implicitly, by rendering theory: we perform point queries in phase space by tracing rays from a particular point, in a particular direction.

The concept of "*locality*" then becomes central to our discussion: Ray optics permits a precise, simultaneous knowledge of position and momentum. This perfect localization is what enables us to make use of spatial-subdivision acceleration structures for ray tracing, even achieving real-time performance.

On the other hand, under wave optics such locality is not possible. In wave optics, the basic descriptor of light is the wave function, which is the spatial function of the complex excitations of the underlying electric field, and the momentum space becomes the Fourier space. It is well known that a function and its Fourier conjugate (the Fourier transformed function) cannot both have finite support, leading to an *uncertainty relation*: position and momentum may not be both specified with perfect precision. Therefore, in sharp contrast to ray optics, where the descriptor of light—a ray or a collection of rays—is local, the wave function and its Fourier conjugate serve as a global description of light.

This loss of locality in wave optics nullifies our ability to perform simple point queries in phase space, and indeed this inherent uncertainty is a major difficulty in devising a formalism of wave-optics rendering.

A rich history of research focuses on attempts to restore, to a degree, that "grainy" phase-space view of ray optics. Most notably, the *Wigner distribution function* [55] (also known as the Wigner-Ville distribution in mathematics) is a complete descriptor of light that simultaneously provides information about both the spatial and angular spectrum properties of the wave function. Thereby, the Wigner distribution function serves to define the dynamics of wave optics in a phase space. For a more comprehensive discussion on Wigner optics, as well as the role the Wigner distribution function plays in wave and quantum optics, the curious reader is referred to Testorf et al. [48], Torre [51].

2.7.1 Ray Optics

Hamiltonian optics are developed from Fermat's principle—the principle of extremal action, which in the optical context means the extremal optical path. Specifically, the path taken by a light ray from point $\vec{\boldsymbol{q}}_1$ to point $\vec{\boldsymbol{q}}_2$ fulfils

$$\delta \int_{\vec{\boldsymbol{q}}_1}^{\vec{\boldsymbol{q}}_2} \mathrm{d}s \, \eta\big(\vec{\boldsymbol{q}}'\big) = 0 \ , \tag{2.162}$$

with η being the refractive index of the medium and s the arc length. That is, the path where the optical path length (path length times refractive index) is an extremum or is stationary, therefore the ray path must follow the refractive-index gradient:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\eta(\vec{\boldsymbol{q}})\frac{\mathrm{d}\vec{\boldsymbol{q}}}{\mathrm{d}s}\right] = \boldsymbol{\nabla}\eta(\vec{\boldsymbol{q}}) \ . \tag{2.163}$$

The above is reminiscent of Newton's second law, hence a ray behaves as a classical point particle, with the refractive-index of the medium serving as the mass of the particle. A force $\boldsymbol{\nabla}\eta$ acts upon this particle, thereby light bends—traces an Eikonal—as it propagates through a refractive-index graded medium.

From Eq. (2.163) we recognize the light particle's momentum as

$$\vec{\boldsymbol{p}} \triangleq \eta(\vec{\boldsymbol{q}})\frac{\mathrm{d}\vec{\boldsymbol{q}}}{\mathrm{d}s} \ . \tag{2.164}$$

The momenta $\vec{\boldsymbol{p}}$ are the *canonically conjugate* variables to the position variables $\vec{\boldsymbol{q}}$, and are the optical direction cosines (ray direction scaled by the refractive index). We denote the vector

$$\vec{\boldsymbol{u}}(s) \triangleq \begin{pmatrix} \vec{\boldsymbol{q}} \\ \vec{\boldsymbol{p}} \end{pmatrix} \tag{2.165}$$

as a *ray*. The ray $\vec{\boldsymbol{u}}$ lives in *phase-space*: a vector space defined as the cartesian product of the position and momentum space. The dynamics of that ray, as it evolves w.r.t. s, are quantified by the *Hamiltonian*

$$H(\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}} \ ; \ s) = -\sqrt{\eta^2(\vec{\boldsymbol{q}}) - p^2} \ , \tag{2.166}$$

and Hamiltonian's equations

$$\frac{\mathrm{d}q_\beta}{\mathrm{d}s} = \frac{\partial H}{\partial p_\beta} \quad \text{and} \quad \frac{\mathrm{d}p_\beta}{\mathrm{d}s} = -\frac{\partial H}{\partial q_\beta} \ , \tag{2.167}$$

with $\beta \in \{x, y, z\}$.

The above can be recast in operator notation into

$$\frac{\mathrm{d}}{\mathrm{d}s}\vec{\boldsymbol{u}} = \boldsymbol{\mathcal{H}}\vec{\boldsymbol{u}} \ , \tag{2.168}$$

$$\text{with} \quad \boldsymbol{\mathcal{H}} \triangleq \sum_{\beta\in\{x,y,z\}} \left[\frac{\partial H}{\partial p_\beta}\frac{\partial}{\partial q_\beta} - \frac{\partial H}{\partial q_\beta}\frac{\partial}{\partial p_\beta}\right] . \tag{2.169}$$

Eq. (2.168) is the *ray equation*, an operator-valued differential first-order equation, and $\boldsymbol{\mathcal{H}}$ is the Lie operator associated with the optical Hamiltonian H. The ray equation yields the closely-related Eikonal equation, as well as the Snell's law of refraction and the law of reflection at an interface between two media.

The solution to the ray equation, representing the evolution of the ray from s_0 to some s, can be written via the *ray-transfer operator* $\boldsymbol{\mathcal{T}}$:

$$\vec{\boldsymbol{u}}(s) = \boldsymbol{\mathcal{T}}\vec{\boldsymbol{u}}(s_0) \ , \qquad \text{with} \quad \boldsymbol{\mathcal{T}} \triangleq \mathrm{e}^{(s-s_0)\boldsymbol{\mathcal{H}}} \ . \tag{2.170}$$

The exponential map above maps the Lie algebra $\{\boldsymbol{\mathcal{H}}\}$ into the corresponding symplectic Lie group of ray-transfer operators.

Linear optical systems and quadratic Hamiltonian When the light rays propagate roughly in the same direction, say the z-axis, we take a paraxial view: The ray evolution variable s is replaced with z, and $\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}}$ become 2-dimensional vectors that live on the xy-plane at a particular instant $z = z'$ of a ray's evolution. Paraxiality implies $p_x^2 + p_y^2 \ll \eta^2$, hence the optical Hamiltonian H (Eq. (2.166)) can be written in the quadratic approximation:

$$H(\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}} \ ; \ z) = \tfrac{1}{2\eta(\vec{\boldsymbol{q}})}p^2 - \eta(\vec{\boldsymbol{q}}) \ . \tag{2.171}$$

An interesting special case of paraxial optical systems are simple systems, where $\boldsymbol{\mathcal{H}}$

does not dependant on z. Such systems are known as linear optical systems, or "ABCD" systems. The latter refers to the fact that $\boldsymbol{\mathcal{T}}$ can be written in the following block-structural form:

$$\begin{pmatrix} \vec{\boldsymbol{q}}(z) \\ \vec{\boldsymbol{p}}(z) \end{pmatrix} = \boldsymbol{\mathcal{T}} \begin{pmatrix} \vec{\boldsymbol{q}}(z_0) \\ \vec{\boldsymbol{p}}(z_0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \vec{\boldsymbol{q}}(z_0) \\ \vec{\boldsymbol{p}}(z_0) \end{pmatrix} , \tag{2.172}$$

with $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ being 2×2 real matrices, and for non-absorbing systems $|\boldsymbol{\mathcal{T}}| = 1$.

ABCD systems are of particular interest, as they include propagation, and reflection and refraction of light at simple interfaces, as well as curved interfaces (like lenses). For example, propagation through a medium with constant refractive-index, or focusing by a thin lens, admit the following ray-transfer matrices

$$\boldsymbol{\mathcal{T}}_{\text{propagation}} = \begin{pmatrix} 1 & \frac{d}{\eta} \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\mathcal{T}}_{\text{thinlens}} = \begin{pmatrix} 1 & 0 \\ \frac{1}{f} & 1 \end{pmatrix} , \tag{2.173}$$

respectively. We use scalars for the ABCD elements of the matrices above to indicate that these systems are rotationally-invariant. d is the (scaled) distance of propagation, and f is the focal length of the lens.

An Ensemble of Rays and Liouville's Theorem

The discussion above centred upon the dynamics of a single ray. We now extend the discussion to a statistical ensemble of rays. Let $\rho(\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}}\ ;\ z)$ be the *ray density* function, which is a probability density function quantifying the statistical distribution of rays over phase space. Given an arbitrary function of position and momentum $f(\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}})$, the average

value of f over the entire statistical ensemble of rays is

$$\langle f \rangle = \int \mathrm{d}^2\vec{\boldsymbol{q}}\,\mathrm{d}^2\vec{\boldsymbol{p}}\, f(\vec{\boldsymbol{q}},\vec{\boldsymbol{p}})\rho(\vec{\boldsymbol{q}},\vec{\boldsymbol{p}}\ ;\ z)\ , \tag{2.174}$$

with the integration over the entire phase space of the system at instant z. The function f can be understood as an "observable", for example, the response of a camera sensor to light, or the reflectivity of a surface.

It can be shown that the dynamics of ρ are

$$\frac{\partial}{\partial z}\rho = -\boldsymbol{\mathcal{H}}\rho \tag{2.175}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\rho = 0\ , \tag{2.176}$$

which in Hamiltonian dynamics are referred to as *Liouville's equation* and *Liouville theorem*, respectively. The above illustrates important physics: Eq. (2.175) means that the ray density evolves (up to a sign) just as a singular ray. As a mental model, the "optical flow" of light rays in phase-space can be thought of as the motion of an incompressible fluid. The total quantity of that fluid is the optical flux, while the phase-space volume occupied by that fluid is known as the *Étendue*. Liouville theorem (Eq. (2.176)) implies that the ray density in phase-space is a conserved quantity (ignoring absorption), both locally and globally:

- That "optical fluid" being incompressible means that Étendue is conserved, i.e. the optical fluid may move around in phase-space, but the volume it occupies is unchanged, leading to **global** conservation of density.
- Given any distinguished ray $(\vec{\boldsymbol{q}},\vec{\boldsymbol{p}})$, the density $\rho(\vec{\boldsymbol{q}},\vec{\boldsymbol{p}}\ ;\ z)$ in an infinitesimal volume around that ray can be understood as a *property of that ray*, and propagates along the ray's trajectory, i.e. remains conserved **locally** around that ray as the system

evolves.

If, at some particular instant of evolution z, we "scoop" some of the "optical fluid" out of the system, then the Étendue may decrease. Étendue may only increase if we add additional fluid into the system (i.e., inject optical flux).

We may relate the above to classical radiometry: the well-known radiometric radiance is defined as

$$L = \eta^2 \frac{\partial \Phi}{\partial G} \ , \tag{2.177}$$

that is, the (differential) total quantity of fluid—the optical flux Φ—over the (differential) volume this fluid occupies—the Étendue G. The well-known conservation of basic radiance, viz. L/η^2, in simple, non-absorbing optical systems, is then an immediate consequence of Liouville theorem.

A more comprehensive formulation of Hamiltonian optics can be found in the textbooks: Buchdahl [9] and Bass et al. [5].

2.7.2 Wave Optics

It is possible to recover aspects of ray optics as a limiting case of wave optics (specifically, the Helmholtz equation reduces to the Eikonal equation at the limit $\hbar \to 0$). However, wave optics cannot be formulated from ray optics purely via mathematical analysis. Instead, wave optics is typically brought forth from classical ray optics in a manner analogous to how quantum mechanics arises from classical Hamiltonian mechanics, and we briefly retrace these steps: A heuristic approach known as "quantization" (or "wavization"), first proposed separately by both Dirac and Heisenberg in different variations, is designed to obtain a quantum theory from a classical theory, hence quantization is a mapping between the theories. Quantization works by replacing the classical functions

$f(q,p)$ on phase space with operators ("observables") $\hat{\boldsymbol{f}}$, acting upon wave functions, as well as replacing the classical dynamic laws (e.g., Eq. (2.167)) with quantum dynamics.

Our starting point is a general classical non-relativistic particle with Hamiltonian

$$H(\vec{\boldsymbol{q}},\vec{\boldsymbol{p}}) = \frac{p^2}{2m} + V(\vec{\boldsymbol{q}}) \ , \tag{2.178}$$

where m is mass and V is the potential function. The canonical position and momenta variables (Eq. (2.164)) are mapped to their operator counterparts

$$q \mapsto \hat{\boldsymbol{q}} \triangleq q \qquad \text{and} \qquad p \mapsto \hat{\boldsymbol{p}} \triangleq -\mathrm{i}\hbar\frac{\partial}{\partial q} \ . \tag{2.179}$$

The Hamiltonian operator then becomes

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\boldsymbol{q}},\hat{\boldsymbol{p}}) = -\frac{\hbar^2}{2m}\nabla^2_{\vec{\boldsymbol{q}}} + V(\vec{\boldsymbol{q}}) \ , \tag{2.180}$$

with $\nabla^2_{\vec{\boldsymbol{q}}}$ being the Laplacian w.r.t. the spatial variable $\vec{\boldsymbol{q}}$.

Switching to the optical context, we note that the classical quadratic optical Hamiltonian, Eq. (2.171), is of the form of Eq. (2.178), with $m = \eta$ and $V = -\eta$. Then, applying the mapping in Equation (2.179),

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\boldsymbol{q}},\hat{\boldsymbol{p}}) = -\frac{\hbar^2}{2\eta(\vec{\boldsymbol{q}})}\nabla^2_{\vec{\boldsymbol{q}}} - \eta(\vec{\boldsymbol{q}}) \tag{2.181}$$

becomes the quadratic wave-optical Hamiltonian operator.

We denote $\psi(\vec{\boldsymbol{q}}\ ;\ t)$ as the *wave function*, with $\vec{\boldsymbol{q}}$ being spatial position and t time. The wave function is a complex function that serves as a descriptor of light under the wave-optical context (it may be understood as the excitations of the underlying electric

field). The evolution of the wave function is dictated by a time-evolution operator

$$\psi(\vec{\boldsymbol{q}}\ ;\ t) = \hat{\boldsymbol{\mathcal{U}}}(t, t_0)\psi(\vec{\boldsymbol{q}}\ ;\ t_0)\ . \tag{2.182}$$

The time-evolution operator must fulfil the *Schrödinger wave equation*:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\hat{\boldsymbol{\mathcal{U}}} = \hat{\boldsymbol{\mathcal{H}}}\hat{\boldsymbol{\mathcal{U}}}\ , \tag{2.183}$$

which takes a form reminiscent of its classical counterpart (Eq. (2.168)), with time t now playing the role of the classical paraxial system evolution variable z. The Helmholtz equation of classical wave optics can be derived from the wave equation above.

We may identify the momentum space as the Fourier-conjugate of the position space: recognizing the eigenfunctions of $\hat{\boldsymbol{p}}$ as $\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{q}}\cdot\vec{\boldsymbol{k}}}$, we may write

$$\psi(\vec{\boldsymbol{q}}) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int \mathrm{d}^3\vec{\boldsymbol{k}}\,\tilde{\psi}\Big(\vec{\boldsymbol{k}}\Big)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{q}}\cdot\vec{\boldsymbol{k}}}\ , \tag{2.184}$$

with the appropriate normalization constant added. The above is simply the inverse Fourier transform of $\tilde{\psi}(\vec{\boldsymbol{k}})$. Therefore, in wave optics it is often convenient to introduce the frequency operator

$$\hat{\boldsymbol{k}} \triangleq \frac{1}{\hbar}\hat{\boldsymbol{p}}\ , \tag{2.185}$$

and we refer to $\tilde{\psi}(\vec{\boldsymbol{k}})$ as the Fourier-conjugate of the wave function, with $\vec{\boldsymbol{k}} = \frac{1}{\hbar}\vec{\boldsymbol{p}}$ being the *wavevector*, i.e. $|\vec{\boldsymbol{k}}| = \eta\frac{2\pi}{\lambda}$ is the *wavenumber*, where λ is the wavelength.

Uncertainty relation Without loss of generality, assume that the signals $\psi, \tilde{\psi}$ are centred (zero mean). The variances of these signals, along a particular axis, say x, are

$$\sigma_{qx}^2 \triangleq \int \mathrm{d}^3\vec{\boldsymbol{q}}\, q_x^2 |\psi(\vec{\boldsymbol{q}})|^2 \,, \tag{2.186}$$

$$\sigma_{kx}^2 \triangleq \int \mathrm{d}^3\vec{\boldsymbol{k}}\, k_x^2 \left|\tilde{\psi}\left(\vec{\boldsymbol{k}}\right)\right|^2 \,. \tag{2.187}$$

Then, the Fourier relation between position and momentum, outlined by Eq. (2.184), and a bit of analysis, gives rise to the important *uncertainty relation*:

$$\sigma_{qx}^2 \sigma_{kx}^2 \geq \frac{1}{2} \,. \tag{2.188}$$

The uncertainty relation implies that the wave function and its conjugate cannot both be precisely localized in space.

ABCD optical systems and line-spread kernels Let $\psi(\vec{\boldsymbol{q}}\ ;\ z)$ be some wave function, under the paraxial approximation. Under the special case where the Hamiltonian $\hat{\boldsymbol{\mathcal{H}}}$ is not z-dependant (i.e., linear optical systems), the solution to the evolution of the system (Eq. (2.182))

$$\hat{\boldsymbol{\mathcal{U}}}(z, z_0) = \exp\left(-\mathrm{i}\frac{z - z_0}{\hbar}\hat{\boldsymbol{\mathcal{H}}}\right) \,, \tag{2.189}$$

can be rewritten (due to the linearity of the above) via a *line-spread function* acting on the wave function, viz.

$$\psi(\vec{\boldsymbol{q}}\ ;\ z) = \int \mathrm{d}^2\vec{\boldsymbol{q}}'\, g\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{q}}'\right)\psi(\vec{\boldsymbol{q}}'\ ;\ z_0) \,. \tag{2.190}$$

Note that time t is replaced by z as the system's evolution variable, under the paraxial setting. For Hamiltonians that admit only quadratic monomials in $\hat{\boldsymbol{q}}, \hat{\boldsymbol{p}}$, it can be shown that [51]

$$g(q, q') = \sqrt{\frac{-\mathrm{i}}{2\pi\hbar B}} \exp\left[\mathrm{i}\frac{1}{2\hbar B}\Big(Dq^2 + A{q'}^2 - 2qq'\Big)\right], \tag{2.191}$$

where separation into dimensions is implied. It can be shown that the conjugate $\tilde{\psi}$ transforms in similar manner to ψ, but the k-space (frequency space) $ABCD$ parameters relate to the q-space (position space) via

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_k = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix}_q \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{2.192}$$

The above is the kernel of a linear canonical transform, which generalizes Fresnel transforms and fractional Fourier transforms. Therefore, any diffraction problem that can be solved via Fourier optics tools or the Huygens-Fresnel principle is, in fact, an ABCD system. The parametrization of the ABCD variables, from the ray-transfer matrix (Eq. (2.172)) to the wave-optical line-spread function (Eq. (2.191)), changes as $B \mapsto \hbar B$ and $C \mapsto \frac{1}{\hbar} C$, due to the remapping from p-space to k-space (Eq. (2.185)).

Summary As we transition from ray optics to wave optics, the position-momentum identification of a ray $\vec{\boldsymbol{u}} = \left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{p}}\right)$ is replaced by the optical wave function and its conjugate $\left(\psi, \tilde{\psi}\right)$ as the descriptor of light. However, while under ray optics the precise and simultaneous specification of a ray's position and momentum is possible, the uncertainty relation implies that in wave optics, such a local specification is not possible. Therefore, while a ray is a *local* descriptor of light, the wave function and its conjugate serve as a *global* descriptor. Indeed, a (non-zero) wave function ψ or its conjugate $\tilde{\psi}$ will always

admit infinite support.

The dynamics of the relevant systems are governed by the ray equation or wave equation. The special case of ABCD systems are of special interest for us, as these include the majority of the light transport (not accounting for interaction with materials) around a typical scene (perfect reflections and refractions, including curved surfaces, like lenses, and propagation in media with constant or slowly-varying refractive-index).

In attempt to regain a classical-like view of a wave optical system, where position-momentum pairs can be locally sampled, we will next introduce a wave-optical phase space. Crucially, we will show that point-queries in that wave-optical phase space evolve in identical fashion to their classical counterparts, under interaction by ABCD optical systems.

2.7.3 The Wigner Distribution and the Wave-Optical Phase Space

The *Wigner distribution function* (WDF) [55] is defined as

$$\begin{aligned}\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right) &\triangleq \frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{\boldsymbol{q}}'\,\psi\left(\vec{\boldsymbol{q}}-\frac{\vec{\boldsymbol{q}}'}{2}\right)\psi^\star\left(\vec{\boldsymbol{q}}+\frac{\vec{\boldsymbol{q}}'}{2}\right)\mathrm{e}^{-\mathrm{i}\vec{\boldsymbol{q}}'\cdot\vec{\boldsymbol{k}}} \\ &\triangleq \frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{\boldsymbol{k}}'\,\tilde{\psi}\left(\vec{\boldsymbol{k}}-\frac{\vec{\boldsymbol{k}}'}{2}\right)\tilde{\psi}^\star\left(\vec{\boldsymbol{k}}+\frac{\vec{\boldsymbol{k}}'}{2}\right)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{q}}\cdot\vec{\boldsymbol{k}}'}\ ,\end{aligned} \tag{2.193}$$

with both definitions equivalent. The WDF belongs to the wider *Cohen's class* [14] of bilinear signal representations. Being a joint representation of the wave function both in q-space and k-space, the WDF gives rise to a wave-optical phase space. In this subsection, we will analyze the relevant properties of the WDF, and in-turn the wave-optical dynamics in this induced phase space. It is possible to recover the wave function, up to a phase term, from the WDF via an inverse transform.

When ψ is understood as a stochastic process—a statistical ensemble of waves—then a definition of the WDF in terms of the ensemble average is possible:

$$\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right) \triangleq \frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{\boldsymbol{q}}'\,\mathcal{C}\left(\vec{\boldsymbol{q}}-\frac{\vec{\boldsymbol{q}}'}{2},\vec{\boldsymbol{q}}+\frac{\vec{\boldsymbol{q}}'}{2}\right)\mathrm{e}^{-\mathrm{i}\vec{\boldsymbol{q}}'\cdot\vec{\boldsymbol{k}}}\,, \tag{2.194}$$

where $\mathcal{C}$ is the *cross-spectral density* of light: the space-frequency formulation of optical coherence. Clearly, as the WDF and the cross-spectral density function are Fourier pairs, they contain the same information, and one can be recovered unequivocally from the other. For completeness, we explicitly note the inverse transform:

$$\mathcal{C}\left(\vec{\boldsymbol{q}}-\tfrac{1}{2}\vec{\boldsymbol{x}},\vec{\boldsymbol{q}}+\tfrac{1}{2}\vec{\boldsymbol{x}}\right) = \int \mathrm{d}^3\vec{\boldsymbol{k}}'\,\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}'\right)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{k}}'\vec{\boldsymbol{x}}}\,, \tag{2.195}$$

or, equivalently, if we define $\vec{\boldsymbol{q}}_{1,2} = \vec{\boldsymbol{q}} \mp \frac{1}{2}\vec{\boldsymbol{x}}$:

$$\mathcal{C}\left(\vec{\boldsymbol{q}}_1,\vec{\boldsymbol{q}}_2\right) = \int \mathrm{d}^3\vec{\boldsymbol{k}}'\,\mathscr{W}\left(\frac{\vec{\boldsymbol{q}}_1+\vec{\boldsymbol{q}}_2}{2},\vec{\boldsymbol{k}}'\right)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{k}}'(\vec{\boldsymbol{q}}_2-\vec{\boldsymbol{q}}_1)}\,. \tag{2.196}$$

It should be stressed that Eq. (2.193) and Eq. (2.194) are employed under different contexts: the former when we deal with a deterministic wave function, while the latter when the underlying field is modelled as a stochastic process. For more information about optical coherence theory, see Mandel and Wolf [32], Wolf [59].

Given an arbitrary observable $\hat{\boldsymbol{f}}(\hat{\boldsymbol{q}},\hat{\boldsymbol{k}})$, its expectation value is

$$\left\langle \hat{\boldsymbol{f}} \right\rangle_{\psi} = \left\langle \psi \middle| \hat{\boldsymbol{f}} \middle| \psi \right\rangle = \int \mathrm{d}^3\vec{\boldsymbol{q}}\,\psi^{\star}(\vec{\boldsymbol{q}})\hat{\boldsymbol{f}}\left(\hat{\boldsymbol{q}},\hat{\boldsymbol{k}}\right)\psi(\vec{\boldsymbol{q}})\,. \tag{2.197}$$

It is possible to map the observable $\hat{\boldsymbol{f}}$ to its corresponding "classical" phase-space function $f(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}})$ via the *Wigner-Weyl transform* [13]. Given such a pair, $\hat{\boldsymbol{f}}$ and f, the expectation

value of the observable, i.e. Eq. (2.197), can be recast as

$$\left\langle \hat{\boldsymbol{f}} \right\rangle_{\psi} = \int \mathrm{d}^3\vec{\boldsymbol{q}}\, \mathrm{d}^3\vec{\boldsymbol{k}}\, f\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) , \tag{2.198}$$

which takes a similar form to the expectation of an observable w.r.t. the classical ray density ρ (Eq. (2.174)). Note that Eq. (2.197) is formulated in terms of operators, while Eq. (2.198) is written in terms of c-functions, typically yielding a simpler expression that is more amenable to analytic tools. The WDF then serves a role similar to the classical ray density ρ: it allows us to "ask wave-optical questions", but in a manner resembling classical phase-space queries.

The properties of Wigner distribution function The WDF fulfils most of the postulates expected from a phase-space density function.

(I) Realness — $\mathscr{W} \in \mathbb{R}$.

(II) Marginals — the position and momentum densities are the corresponding marginals of the WDF:

$$|\psi(\vec{\boldsymbol{q}})|^2 = \int \mathrm{d}^3\vec{\boldsymbol{k}}\, \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \tag{2.199}$$

$$\left|\tilde{\psi}\left(\vec{\boldsymbol{k}}\right)\right|^2 = \int \mathrm{d}^3\vec{\boldsymbol{q}}\, \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) . \tag{2.200}$$

(III) Unit measure — if the wave function is normalized, viz. $\int \mathrm{d}^3\vec{\boldsymbol{q}}\, |\psi(\vec{\boldsymbol{q}})|^2 = 1$ then the WDF integrates to one over the entire phase space:

$$\int \mathrm{d}^3\vec{\boldsymbol{q}}\, \mathrm{d}^3\vec{\boldsymbol{k}}\, \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) = 1 . \tag{2.201}$$

The converse holds as well. In general, the WDF can be normalized as $\int \mathrm{d}\vec{\boldsymbol{q}}\, \mathrm{d}\vec{\boldsymbol{k}}\, \mathscr{W} = 0$

if and only if $\psi \equiv 0$.

(IV) Galilei invariance — the WDF is invariant under Galilean transformations:

$$\psi'(\vec{\boldsymbol{q}}) = \psi\big(\vec{\boldsymbol{q}} + \vec{\boldsymbol{q}}'\big) \qquad \Longrightarrow \qquad \mathscr{W}'\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) = \mathscr{W}\left(\vec{\boldsymbol{q}} + \vec{\boldsymbol{q}}', \vec{\boldsymbol{k}}\right) \tag{2.202}$$

$$\tilde{\psi}'\left(\vec{\boldsymbol{k}}\right) = \tilde{\psi}\left(\vec{\boldsymbol{k}} + \vec{\boldsymbol{k}}'\right) \qquad \Longrightarrow \qquad \mathscr{W}'\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) = \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}} + \vec{\boldsymbol{k}}'\right) . \tag{2.203}$$

(V) Support — Given convex $S_q, S_k \subseteq \mathbb{R}^3$ such that

$$\forall \vec{\boldsymbol{q}} \notin S_q,\ \psi(\vec{\boldsymbol{q}}) = 0 \qquad \text{and} \qquad \forall \vec{\boldsymbol{k}} \notin S_k,\ \tilde{\psi}\left(\vec{\boldsymbol{k}}\right) = 0 ,$$

the WDF vanishes outside these volumes as well:

$$\mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \neq 0 \qquad \text{only if} \quad \left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \in S_q \times S_k . \tag{2.204}$$

That is, the support of the WDF in q-space and k-space is the support of ψ and $\tilde{\psi}$, respectively.

(VI) Liouville transformation laws — under the paraxial approximation, given a quadratic Hamiltonian (with only quadratic monomials in $\hat{\boldsymbol{q}}, \hat{\boldsymbol{p}}$), the WDF obeys:

$$\frac{\partial}{\partial z}\mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\ ;\ z\right) = -\boldsymbol{\mathcal{H}}\mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\ ;\ z\right) \tag{2.205}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\ ;\ z\right) = 0 , \tag{2.206}$$

i.e. the Liouville's equation and Liouville theorem of Hamiltonian mechanics, viz. Eqs. (2.175) and (2.176), and note that $\boldsymbol{\mathcal{H}}$ above is the classical Hamiltonian of ray optics (Eq. (2.169)). Also note that, as before, under the paraxial setting z replaces t as the system evolution variable, and the q and k-spaces are now 2-dimensional, meaning

the phase space becomes 4-dimensional.

(VII) Superposition — Given wave functions ψ_1 and ψ_2, the WDF of the superposition $\psi = \psi_1 + \psi_2$ is

$$\mathscr{W} = \mathscr{W}_1 + \mathscr{W}_2 + 2\,\mathrm{Re}\,\mathscr{W}_{12} \ , \tag{2.207}$$

where $\mathscr{W}_{12}$ is the *cross-term*:

$$\mathscr{W}_{12}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \triangleq \frac{1}{(2\pi)^3} \int \mathrm{d}^3\vec{\boldsymbol{q}}'\, \psi_1\left(\vec{\boldsymbol{q}} - \frac{\vec{\boldsymbol{q}}'}{2}\right) \psi_2^\star\left(\vec{\boldsymbol{q}} + \frac{\vec{\boldsymbol{q}}'}{2}\right) \mathrm{e}^{-\mathrm{i}\vec{\boldsymbol{q}}'\cdot\vec{\boldsymbol{k}}} \ . \tag{2.208}$$

The above highlights the bilinearity of the WDF.

Moments Important information about the underlying wave functions, and the optical beams these wave functions encode, can be gleaned from the WDF moments. The total energy contained in the beam is

$$E \triangleq \int \mathrm{d}^3\vec{\boldsymbol{q}}\, |\psi(\vec{\boldsymbol{q}})|^2 = \int \mathrm{d}^3\vec{\boldsymbol{q}}\, \mathrm{d}^3\vec{\boldsymbol{k}}\, \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \ . \tag{2.209}$$

Clearly, when we understand the WDF strictly as a (quasi-)probability density function, then we only consider $E = 1$. First-order moments (mean) are

$$\begin{pmatrix} \bar{\boldsymbol{q}} \\ \bar{\boldsymbol{k}} \end{pmatrix} \triangleq \frac{1}{E} \int \mathrm{d}^3\vec{\boldsymbol{q}}\, \mathrm{d}^3\vec{\boldsymbol{k}} \begin{pmatrix} \vec{\boldsymbol{q}} \\ \vec{\boldsymbol{k}} \end{pmatrix} \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}\right) \ . \tag{2.210}$$

Second-order moments give information about the gyration of beam energy about the mean, in position and frequency spaces. The second-order moments are grouped into the

real, symmetric *moments matrix* of the WDF:

$$\begin{aligned}\boldsymbol{M} &\triangleq \begin{pmatrix} m_{xx} & m_{xy} & m_{xz} & m_{x\tilde{x}} & m_{x\tilde{y}} & m_{x\tilde{z}} \\ m_{yx} & m_{yy} & m_{yz} & m_{y\tilde{x}} & m_{y\tilde{y}} & m_{y\tilde{z}} \\ & & \cdots & \cdots & & \\ m_{\tilde{z}x} & m_{\tilde{z}y} & m_{\tilde{z}z} & m_{\tilde{z}\tilde{x}} & m_{\tilde{z}\tilde{y}} & m_{\tilde{z}\tilde{z}} \end{pmatrix} \\ &\triangleq \frac{1}{E}\int \mathrm{d}^3\vec{\boldsymbol{q}}\,\mathrm{d}^3\vec{\boldsymbol{k}} \left[\begin{pmatrix}\vec{\boldsymbol{q}}\\ \vec{\boldsymbol{k}}\end{pmatrix} - \begin{pmatrix}\bar{\boldsymbol{q}}\\ \bar{\boldsymbol{k}}\end{pmatrix}\right]\left[\begin{pmatrix}\vec{\boldsymbol{q}}\\ \vec{\boldsymbol{k}}\end{pmatrix} - \begin{pmatrix}\bar{\boldsymbol{q}}\\ \bar{\boldsymbol{k}}\end{pmatrix}\right]^{\mathsf{T}} \mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right) , \end{aligned} \tag{2.211}$$

where subscripts of the matrix elements $m_{\xi\zeta}$ are unaccented or accented with a tilde to indicate q-space or k-space axes, respectively. The second-order moments on the main diagonal of $\boldsymbol{M}$ provide information about the width of the beam in phase space, i.e. both in position and frequency spaces. Our interest lies primarily in these main diagonal moments. Mixed moments are used in the optical literature to characterize beam twist, curvature as well as beam quality. Furthermore, mixed moments quantify the longitudinal components of the orbital angular momentum.

Transformation of the WDF The properties above suggest that the WDF can, to a degree, be understood as the classical phase-space density function ρ. A point-query of the wave-optical phase space, viz. $\mathscr{W}(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}})$, then can be understood as a "ray", and we write $\mathscr{W}(\vec{\boldsymbol{u}})$, with $\vec{\boldsymbol{u}} = \left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right)$ resembling its ray optical analogoue (i.e. Eq. (2.165)). Property (VI) then implies that the WDF transforms in a manner similar to a classical ray under interaction with an ABCD optical system:

$$\mathscr{W}\left(\vec{\boldsymbol{u}}\ ;\ z\right) = \mathscr{W}\left(\boldsymbol{\mathcal{T}}^{-1}\vec{\boldsymbol{u}}\ ;\ z_0\right) , \tag{2.212}$$

where the matrix $\boldsymbol{\mathcal{T}}$ is the appropriate ABCD ray-transfer matrix (Eq. (2.172)), though note that it should be transformed to q-k representation of the phase space, from the q-p representation of the ray optical phase space, as discussed in Subsection 2.7.2.

Eq. (2.212) means that a point-query $\vec{\boldsymbol{u}}$ in the wave-optical phase space (a "ray") transforms just as its ray-optical analogue, under quadratic Hamiltonian wave optics. Of particular interest is the fact that the WDF moments matrix (Eq. (2.211)) also transforms via the ray-transfer matrix, as:

$$\boldsymbol{M}(z) = \boldsymbol{\mathcal{T}}\boldsymbol{M}(z_0)\boldsymbol{\mathcal{T}}^{\mathsf{T}} \ , \tag{2.213}$$

on interaction with an ABCD optical systems.

The negativity of the WDF One postulate of a probability density function not fulfilled by the WDF is non-negativity. The WDF may take negative values, a consequence of the uncertainty relation: The quantization process employed to promote the symplectic ray optics to metaplectic wave optics serves to quantize phase space into cells (the volume of which is dictated by the uncertainty relation, Eq. (2.188)). These cells are not discrete cells with "sharp" boundaries, but overlap and interact with each other, therefore points within a phase-space cell do not constitute mutually-exclusive probability events (violating the σ-additivity of a probability measure), hence the WDF is only a quasi-probabilistic density function.

It can be shown that *anisotropic Gaussian Schell-model* (AGSM) beams are the only class of wave functions that admit non-negative Wigner distribution functions. Furthermore, AGSM beams have the most compact support in phase space (occupy the least phase space volume) relative to any other wave function. Thus, a Gaussian beam can be understood as an elementary construct that is the closest analogue of the classical

ray: serving as a form of a "generalized ray" with non-singular extent in position and momentum spaces. It should be noted that a superposition of a pair of AGSM beams does not, in general, yield a non-negative WDF: the bilinearity of the WDF gives rise to cross-terms on superposition (Property (VII) above), and it is these interference terms that are the source of negative values.

See Bastiaans [6], Testorf et al. [48] for additional discussion and applications of the WDF in optics.

Chapter 3

Physical Light Transport

3.1 Overview

In this chapter we derive and formulate our descriptors of light. We build upon optical coherence theory to describe the statistical, *observable* properties of light, as well as its ability to produce wave-interference effects on superposition.

To date, light transport formulated under the context of classical radiometry remains the de facto formalism in computer graphics. Its strength can be found in its simplicity and intuitive geometrical properties. Nonetheless, in the quest for physical realism the computer graphics community has oft turned to the electromagnetic theory in the aim of reproducing "wave optics" effects, diffraction and wave-interference, which can not be described radiometrically. Global application of electromagnetism—beyond small, localized instances—remains elusive: Finding a solution to Maxwell's differential equations typically involves considerable analytic difficulty, which is exacerbated by the fact that natural light is composed of very many distinct waves. Thus, applications of the classical electromagnetic theory in computer graphics remain confined to reproducing optical phenomena that arise on microscopical spatial scales. Optical coherence theory, like tra-

ditional light transport, concerns itself with the observable (time-averaged) properties of light, however it fully accounts for the electromagnetic nature of light and is able to reproduce the phenomena that arise due to that nature.

Our aim is to bridge the theoretical gap between a physical formulation of light and the practical radiometric description employed in computer graphics. Informally, our formalism can be thought of as (rigorously and accurately) fusing the traditional radiance with two-point information describing the coherence properties of light. And indeed, we will show that some of the intuitive concepts of radiance can still be recovered, and, additionally, at the short-wavelength limit our formalism reduces to "geometric optics". Our incentive for this approach is the crucial feature that arises: this formalism can describe an entire spectrum between classical radiometry and the physical optics that is consistent with electromagnetism. That is, these "radiance-carrying" cross spectral density functions can serve as theoretical drop-in replacements for classical radiance in computer graphics algorithms, effectively injecting all the information needed to quantify diffractive effects. Then, substituting the classical propagation with physical diffraction in the algorithm can be done selectively as it makes practical sense.

Generalized radiance Similar notions have been studied before: In optical literature, the seminal paper by Walther [53] proposed using the related Wigner distribution function (WDF) as a form of a *generalized radiance*, generalizing some aspects of the radiometric concept of radiance to physical optics. A vast body of work that followed attempted to connect the theory of partial coherence with classical radiometry, and the WDF has also seen use in computer graphics (see Section 3.2). The WDF is a powerful space-frequency signal analysis tool that is defined as the Fourier transform of the autocorrelation function of a signal (e.g., the cross-spectral density function, Eq. (2.85)) around a point. Attractive properties arise: the WDF is real-valued and, being a directional quantity, conserved

along geometric rays. It is not non-negative, however, and it has been shown that no function that is linearly related to the cross-spectral density may satisfy all the physical requirements of classical radiance [18].

The generalized radiance is connected to the wave-optical phase-space, in the sense it generalized the classical ray density of Hamiltonian optics to wave optics. We will briefly discuss the generalized radiance in Section 5.1. For now, it is worth nothing that the non-negativity of the WDF prohibits its (direct) interpretation as an energy density. Therefore, recovering intensity from the WDF requires additional work, i.e. a Fourier transform or spatial integration, and similarly additional analytic work is required to compute the WDF of model light sources. We do not aim to come up with a "true" generalized radiance. Instead, we devise a physical optics generalization of the light transport principles used in contemporary computer graphics, formulated using a form of the cross-spectral density function (introduced in Section 3.4), giving rise to a general purpose framework of partially-coherent light transport.

3.2 Related Work

Of particular relevance is the work by Oh et al. [37], which shares some of the motivational basis with our work: They propose generalizing the classical light transport used in computer graphics with a more physical formalism by using the Wigner distribution function (WDF) as a "generalized radiance" that retains some of the intuition of ray-based optics. However, no model that deals with transport of natural light—partially-coherent light—is formulated, nor are the relevant rendering equations derived.

Some work consider the limited coherence of light *locally*, but is not able to propagate optical coherence around a scene. For example, in an attempt to quantify the limited spatial coherence of radiation a Gaussian footprint is used to modulate coherent con-

tributions for the purpose of rendering diffractive scratches [54], measuring diffraction grating patterns using Jones calculus [50] and the rendering of biological diffractive surfaces [16]. A more sophisticated approach is to use overlapping kernels, where each kernel roughly models the spatial coherence footprint, and contributions from distinct kernels are added incoherently. Such Gaussian kernels were used to accurately model diffractions that arise from explicitly defined surface micro-geometry [17, 60], and simple square kernels for the construction of high-resolution BRDFs [30]. Two-point information has been used by Steinberg and Yan [45], in the form of mutual intensities related by the mutual coherence function, for the purpose of drawing a subjective speckle pattern that arises on scatter from a statistical surface. Such an approach is accurate and able to describe well arbitrary shapes of the mutual coherence function, however as this is a deterministic formalism, integration is slow. Gkioulekas et al. [19], Kotwal et al. [28] interferometrically decompose light transport in an imaged scene by using light with controlled coherence properties.

Fully coherent light transport, described electromagnetically, has been employed to render optical speckle fields [3, 4]. A large body of work has applied the electromagnetic theory locally, for the study of a particular optical effect. Stam [41] presented the first work that considered the electromagnetic nature of light. Additional works studied diffraction-aware BSDFs [15, 49, 52], iridescence in pearlescent materials [21], thin-film interference [7, 24] and real-time rendering of birefringent materials [42]. Sadeghi et al. [38] introduce a framework for simulating the scattering that arises by non-spherical particles, with application to rendering rainbows. Some work employ various forms of integral or differential solvers to devise a solution to Maxwell's equations for some special case. Steinberg [43] renders liquid-crystal micrographs via an approximative analytic solution to a light transport differential system. Musbach et al. [36] use the finite-difference time-domain (FDTD) method for numerically solving the propagation of electromagnetic

radiation in a scene. The FDTD method was also adapted the purpose of nanofabrication of materials with pigment-free structural colours Auzinger et al. [2].

The shortcomings of all work listed above is their inability to accurately propagate the coherence information of light, i.e., they deal with local phenomena, and ignore the light transport aspects of rendering.

3.3 Theoretical Foundations

3.3.1 Traditional Light Transport

Traditional light transport in computer graphics draws upon radiometry to describe the propagation and (equilibrium) distribution of power. The basic quantity is the *radiance*, defined as $L = \frac{\partial^2 \Phi}{\partial \Omega \partial A \cos\theta}$ (Φ is radiant flux, Ω is the solid angle subtended by the collector, A is the cross-sectional area of the source and $\cos\theta$ is the obliquity factor), and the driving equation is the iconic *rendering equation* [23], to which we refer as the *light transport equation* (LTE):

$$\begin{aligned} L_o(\vec{\boldsymbol{p}}, \Omega_o, \omega, t) &= L_e(\vec{\boldsymbol{p}}, \Omega_o, \omega, t) \\ &\quad + \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, f_r(\vec{\boldsymbol{p}}, \Omega', \Omega_o, \omega, t) L_i(\vec{\boldsymbol{p}}, \Omega', \omega, t) |\Omega' \cdot \hat{\boldsymbol{n}}| \ , \end{aligned} \tag{3.1}$$

where L_o is the exitant spectral radiance, L_e is the emitted radiance, L_i is the incident spectral radiance, $\mathscr{S}^2$ is the unit sphere centred at the point of interest $\vec{\boldsymbol{p}}$, ω is the radiation angular frequency, t is time, f_r is the bidirectional scattering distribution function (BSDF), $\Omega \equiv (\theta, \phi)$ is the spherical angle with $\mathrm{d}^2\Omega = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$ being the differential solid angle, and Ω_o, Ω' are the exitant and incident angles, respectively.

Such a radiometric description of light (sometimes referred to as "geometric optics") admits nice properties:

1. The LTE is linear;

2. Radiance is linear under superposition of light, which is no longer the case when wave-interference effects are taken into account (that is, the intensity and radiance of the superposed radiation is no longer a linear combination of the constituents' intensity and radiance);

3. Radiance is conserved along geometric rays in an optical system (in absence of transmission losses).

Furthermore, the LTE remains a good approximation for physical light transport when the electromagnetic nature of light plays no major role. Hence, the LTE has been a cornerstone of modern computer graphics for decades and those light transport principles were employed by a vast body work.

Limitations Nevertheless, as the electromagnetic nature of light is ignored when formulating light transport under the constraints of the LTE, all effects that arise from interference and diffraction are neglected. Unfortunately, those limitations are inherit to the formalism: It is well-known that this radiometric formulation is inconsistent with Maxwellian electromagnetism [57] and, as we shall formally demonstrate in Section 3.5, perfectly incoherent light does not propagate and does not give rise to (far-field) radiation fields! As a matter of fact, classical radiance, being a time-averaged quantity that discards phase information, is not a fundamental property of an electromagnetic wave (that is, a property that can not be derived from more basic properties), or an ensemble of such waves, and does not provide sufficient information to formulate the propagation of the underlying electromagnetic field. Furthermore, there is no way to glean the required information from a radiometric, single-point description of light.

3.4 The Spectral-Density Transport Equation

The cross-spectral density function encodes the spectral power carried by the electromagnetic radiation as well as spatial correlation information. To gain some insight into the cross-spectral density function, consider a narrow beam of light emitted or scattered by a surface patch $\delta\vec{\boldsymbol{p}} \subset \mathbb{R}^3$ (the notation $\delta\vec{\boldsymbol{p}}$ is used to describe a small two-dimensional patch or some small volume around the point $\vec{\boldsymbol{p}}$) and then propagated to a point $\vec{\boldsymbol{r}} \in \mathbb{R}^3$. Assuming the power distribution across the beam's cross section changes slowly (i.e., a quasi-homogeneous source), the cross-spectral density could be decomposed as follows

$$\mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \approx I(\vec{\boldsymbol{r}}, \omega)\gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) \tag{3.2}$$

for points $\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2$ close to $\vec{\boldsymbol{r}}$, i.e. $\vec{\boldsymbol{r}}_{1,2} \in \delta\vec{\boldsymbol{r}}$. I is the spectral power carried by the beam and γ is a unit-less complex-valued spatial correlation function, called the *spatial degree-of-coherence*, which fulfils: $|\gamma| \leq 1$, $\gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) = \gamma(\vec{\boldsymbol{r}}_2, \vec{\boldsymbol{r}}_1)^\star$ and $\gamma(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2) = 1$ when $\vec{\boldsymbol{r}}_1 = \vec{\boldsymbol{r}}_2$. It is important to note that this decomposition only serves an illustrative purpose, and in general such a decomposition is not possible. We do not use this decomposition for our formal results but it is a useful way to intuitively picture the cross-spectral density.

The cross-spectral density function is the ensemble average of the mutual intensity of equal-frequency constituents of the coherent-modes decomposition of a wave ensemble (Eq. (2.85)). Hence, akin to time-averaged intensity of an electromagnetic wave (Eq. (2.74)), the cross-spectral density quantifies directional power flow—the electromagnetic flux Φ impinging on a differential surface area $\mathrm{d}A_r$—and has units of spectral power per area, i.e. *spectral irradiance*. The differential surface area is trivially related to a differential solid angle, subtended by $\delta\vec{\boldsymbol{r}}$ at the source $\vec{\boldsymbol{p}}$, via $\mathrm{d}A_r = r^2\,\mathrm{d}\Omega$ (with

$r = |\vec{\boldsymbol{r}}|$, as usual), thus

$$I = \frac{\partial \Phi}{\partial A_r} = \frac{1}{r^2} \frac{\partial \Phi}{\partial \Omega} \tag{3.3}$$

Differentiating by the projected area of the source $\delta\vec{\boldsymbol{p}}$, denoted A_s, we get

$$r^2 \frac{\partial I}{\partial A_s \cos\theta} = \frac{\partial^2 \Phi}{\partial \Omega \partial A_s \cos\theta} \tag{3.4}$$

where θ is the inclination angle. The right-hand-side is immediately recognizable as *spectral radiance* (spectral power per solid angle per projected source area). Thus, we define the scalar and matrix *radiance cross-spectral density* (RCSD) functions as

$$\mathcal{L}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \triangleq r^2 \frac{\partial \mathcal{C}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right)}{\partial A_s \cos\theta} \ , \qquad \text{and} \qquad \boldsymbol{\mathcal{L}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) \triangleq r^2 \frac{\partial \boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right)}{\partial A_s \cos\theta} \ , \tag{3.5}$$

respectively. Substitute the decomposition outlined in Eq. (3.2) into the definition of the RCSD, yielding $\mathcal{L} = L\gamma$ where L is radiance. Indeed, the RCSD quantifies the power, in-form of spectral radiance, carried by a wave ensemble as well as the spectral spatial correlation. While that formal decomposition generally does not exist, we will see later (when discussion sourcing and propagation of the RCSD) that some of the convenient characteristics of the classical radiometric radiance are preserved: $\mathcal{L}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega\right)$ is a non-negative real quantity that is conserved in free-space propagation. The RCSD should then be thought of as a function augmenting classical radiance with a two-point spatial coherence information. This two-point information is what enables this formalism to be fully consistent with electromagnetism.

Consider a surface patch $\delta\vec{\boldsymbol{p}}$ with normal $\hat{\boldsymbol{n}}$. The irradiance, in terms of a cross-spectral density matrix, impinging upon that surface patch can be written as the super-

position of the incident RCSD matrices, viz.

$$\boldsymbol{\mathcal{C}}\Big|_{\vec{\boldsymbol{p}}} = \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, |\Omega' \cdot \hat{\boldsymbol{n}}| \, \boldsymbol{\mathcal{L}}^{(i)}_{\vec{\boldsymbol{r}}' \to \delta\vec{\boldsymbol{p}}} \ , \tag{3.6}$$

where $\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\boldsymbol{r}}' \to \delta\vec{\boldsymbol{p}}}$ is the RCSD of the incident radiation that arrived from a point $\vec{\boldsymbol{r}}'$ (that lies in direction Ω' from $\vec{\boldsymbol{p}}$). We are interested in the RCSD of the radiation that scatters and propagates from that surface patch. We will discuss propagation and diffraction in detail in the sections that follow, but for our current purpose it is sufficient to introduce $\mathscr{D}_{\delta\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}}$: an abstract diffraction operator—the "wave optics" generalization of the classical BSDF—that diffracts and propagates the cross-spectral density on the surface patch $\delta\vec{\boldsymbol{p}}$ to a region in space $\delta\vec{\boldsymbol{r}}$. Then,

$$\begin{aligned} \boldsymbol{\mathcal{L}}^{(o)}_{\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} &= \mathscr{D}_{\delta\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} \left\{ \boldsymbol{\mathcal{C}}\Big|_{\vec{\boldsymbol{p}}} \right\} \\ &= \mathscr{D}_{\delta\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} \left\{ \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, |\Omega' \cdot \hat{\boldsymbol{n}}| \, \boldsymbol{\mathcal{L}}^{(i)}_{\vec{\boldsymbol{r}}' \to \delta\vec{\boldsymbol{p}}} \right\} , \end{aligned} \tag{3.7}$$

with $\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}}$ being the quantity of interest: the RCSD of the radiation scattered from $\vec{\boldsymbol{p}}$ and propagated to region $\delta\vec{\boldsymbol{r}}$. Making the assumption that $\mathscr{D}$ is a linear operator and formally interchangeable with the integral in Eq. (3.7), we are now ready to formulate the *spectral-density transport equation* (SDTE):

Definition 3.4.1 (Spectral-Density Transport Equation)**.** *The SDTE governs the transport of light at the boundary between two media:*

$$\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} = \boldsymbol{\mathcal{L}}^{(e)}_{\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} + \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, |\Omega' \cdot \hat{\boldsymbol{n}}| \mathscr{D}_{\delta\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}} \left\{ \boldsymbol{\mathcal{L}}^{(i)}_{\vec{\boldsymbol{r}}' \to \delta\vec{\boldsymbol{p}}} \right\} , \tag{3.8}$$

where $\boldsymbol{\mathcal{L}}^{(e)}_{\vec{\boldsymbol{p}} \to \delta\vec{\boldsymbol{r}}}$ is the RCSD of the radiation emitted from $\delta\vec{\boldsymbol{p}}$.

Clear similarities can be drawn between the classical LTE (Eq. (3.1)) and the SDTE,

both describe the local, directional behaviour of light at an interface between media. We summarise the key differences:

1. The SDTE deals with functions and not numeric values. For this reason our formalism is described in terms of operators that act upon those RCSD functions.

2. While the LTE propagates values of radiance point-to-point, an inherent fact of our formalism is that an RCSD function describes information about the radiation within a small region in space, and propagation is between surface patches or regions.

3. The BSDF is replaced with a more general diffraction operator.

The superposition of RCSDs functions, as it appears in Eq. (3.6) and in the SDTE, should be understood as a summation of contributions from statistically independent sources, and for each source we integrate over the solid angle subtended by that source. Statistically independent sources can be different primary light sources, or secondary sources (e.g., a scattering surface) that are far, with respect to the spatial coherence of the light, from each other. For natural light, where spatial coherence is measured in micrometres, "far" typically refers to any supra-microscopic scales. We will formally show in Section 3.6 that the RCSD of a superposition of radiation produced by sources that are statistically independent is a linear combination of the individual RCSDs.

We make the claim, but do not prove it, that the SDTE generalizes the classical LTE in the short-wavelength limit: That is, the SDTE reduces to the LTE at the limit $k \to \infty$ (with k being the wavenumber). It is well-known that this limit can also be considered as the "geometric optics" limit. Intuitively this can be understood by observing that at the short-wavelength limit any aperture or surface detail becomes infinitely larger than the wavelength, effectively discarding diffraction effects and, thus, light propagation adheres to geometric rays.

3.5 Sourcing from Natural Light Sources

In this section we analyze and define what constitutes a natural light source as well as the radiation produced by such a source. A variety of artificial and natural light sources produce essentially all the light that we observe daily: Blackbody radiators, like the stars, the sun and the filaments of incandescent bulbs, convert internal thermal energy into electromagnetic energy; gas-discharge lamps (e.g., fluorescent lamps, sodium-vapour lamps, neon lights and HID lamps) emit light via gases excited by an electric current; light emitting diodes (LEDs) produce light by passing electrons through a p-n junction. The emission characteristics of those sources depend on body temperature, gases used, substrate material and phosphors (if any), and while the means used to generate light vary greatly, all those types of sources share the same underlying physical process: *spontaneous emission*. Spontaneous emission is a quantum mechanical process where a particle—an electron, atom or a molecule—in an excited state spontaneously undergoes transition to a lower energy state and emits the energy difference in form of a photon, in a random direction and potentially after a short delay.

Those light sources, where light production is driven by spontaneous emission, are readily modelled as a large collection of *elementary radiators*, which are those electrons, atoms or molecules that are excited by internal thermal energy, electric current, a chemical reaction, or another mechanism. These elementary radiators randomly and isotropically emit photons that give rise to the rather chaotic polychromatic electromagnetic radiation that we term as *natural light*—a core concept in our theory. We define a large collection of elementary radiators to constitute an idealised model of a source producing natural light if the following holds:

1. *Radiation from elementary radiators does not interact with the light source.* In practice, photons that are produced by radiators inside the source do interact with

other particles, such as electrons. However, such a photon will ultimately escape the source, and thus we assume that those interactions can be neglected when time-averaging.

2. *Each elementary radiator gives rise to isotropic, randomly polarized radiation* (when time-averaged).

In addition, in order to simplify analytic derivations, we also assume that (i) the source is homogeneous, i.e. all regions of the light source radiate with identical power and spectrum; and (ii) that the source is a perfect sphere. This implies that if different types of radiators are present, e.g. different gases, then these are all spatially uniformly and independently distributed within the source. A sphere is chosen for its simplicity and generality. Any geometry, as well as sources with non-uniform radiation properties work in practice: we approximate the source as a collection of disjoint spheres and source from each sphere independently.

3.5.1 Model Radiation from Natural Sources

Having defined the model of a natural light source, we now turn our attention to the emitted radiation. See Fig. 3.1 for an illustration of the geometry. Each of the elementary radiators produces electromagnetic radiation and the electromagnetic fields inside the source are a superposition of all the fields produced by all the radiators. Our goal is to find an expression for the RCSD $\boldsymbol{\mathcal{L}}$ of the fields that have radiated away from the source. However, we are only interested in the time-harmonic far-field radiation fields—that is, fields that vary as $1/r$, with r being distance from the source (and thus the intensity decays quadratically with distance). Therefore, we have to solve a far-field diffraction problem, and to that end employ the Rayleigh-Sommerfeld (scalar) diffraction

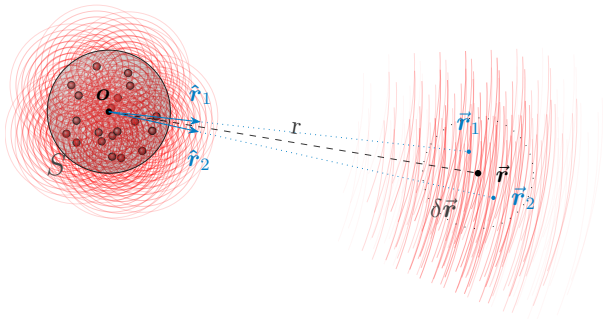


Figure 3.1: A spherical natural light source of volume S is centred at the origin $\boldsymbol{o}$. The source consists of very many *elementary radiators* (illustrated as red balls) that radiate isotropically (wavefronts illustrated as red curves). This gives rise to the highly chaotic radiation that we call *natural light*. To quantify the statistical properties of such light that has propagated a distance r to a small region $\delta\vec{\boldsymbol{r}}$ far from the source, we examine the statistical similarity of the wave ensemble at some points $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$ (in directions $\hat{\boldsymbol{r}}_1$, $\hat{\boldsymbol{r}}_2$ from the center of the source) in that region.

integral of the first kind [8]:

$$u(\vec{\boldsymbol{r}}) = \frac{1}{2\pi}\int_{\Sigma} \mathrm{d}^2\vec{\boldsymbol{r}}'\, u\big(\vec{\boldsymbol{r}}'\big)\boldsymbol{\nabla}_{\hat{\boldsymbol{r}}}\left[\frac{e^{\mathrm{i}k\left|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\right|}}{\left|\vec{\boldsymbol{r}}-\vec{\boldsymbol{r}}'\right|}\right], \tag{3.9}$$

where Σ is a diffracting aperture (centred at the origin), as before k is the wavenumber and $\boldsymbol{\nabla}_{\hat{\boldsymbol{r}}}$ is the directional derivative in direction $\hat{\boldsymbol{r}}$, i.e. the direction of the point-of-interest $\vec{\boldsymbol{r}}$ from the source. A scalar diffraction formalism is sufficient because elementary radiators radiate with random polarization. Applying Eq. (3.9) to the definitions of the cross-spectral density (Eq. (2.85)) and the RCSD (Eq. (3.5)), generalizing to a three-dimensional source (which should be understood in the sense of Huygens' principle: each elementary radiator is source to spherical wavelets) and simplifying gives us the scalar propagation equation for the RCSD (at normal inclination), which is immediately

recognizable as a double (spatial) Fourier transform:

$$\boldsymbol{\mathcal{L}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \frac{1}{\lambda^2}\boldsymbol{I} e^{\mathrm{i}k(r_1 - r_2)} \times \frac{\partial}{\partial A_s} \int_S \mathrm{d}^3\vec{\boldsymbol{r}}'_1 \, e^{-\mathrm{i}k\hat{\boldsymbol{r}}_1 \cdot \vec{\boldsymbol{r}}'_1} \int_S \mathrm{d}^3\vec{\boldsymbol{r}}'_2 \, \mathcal{C}'\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right) e^{\mathrm{i}k\hat{\boldsymbol{r}}_2 \cdot \vec{\boldsymbol{r}}'_2} \ , \tag{3.10}$$

where $\lambda = \frac{2\pi}{k}$ is the wavelength, S is the volume of the source, A_s is the source area and $\boldsymbol{I}$ is the identity matrix, which arises as the source radiates with random polarization. The RCSD above is defined for $\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \in \delta\vec{\boldsymbol{r}}$ for some point $\vec{\boldsymbol{r}}$ away from the source.

Consider the quantity $\mathcal{C}'\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right)$ that appears in Eq. (3.10). This is the cross-spectral density between the fields inside the source. The number density of the elementary radiators is typically large (e.g., roughly $10^{24}\mathrm{m}^{-3}$ for xenon in a high-pressure gas discharge lamp [35]), hence by the time radiation emitted from an elementary radiator propagates even a single wavelength, it will have passed many other independently-radiating elementary radiators. Therefore, it is reasonable to expect the cross-spectral density inside the source to drop very rapidly to 0 as the distance between the points $\vec{\boldsymbol{r}}'_1$, $\vec{\boldsymbol{r}}'_2$ increases. However, observe that if we were to set $\mathcal{C}'\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right) = 0$ for $\vec{\boldsymbol{r}}'_2 \neq \vec{\boldsymbol{r}}'_2$ inside the source, then Eq. (3.10) would yield $\boldsymbol{\mathcal{L}} \equiv 0$ everywhere outside the source. With the important theoretical conclusion being that *incoherent radiation does not propagate*, and all the light we observe exhibits some degree of optical coherence. Experiments performed by Carminati and Greffet [11] show that the near-field spatial coherence of natural light is roughly half a wavelength. Therefore, we choose to model the spatial correlation between the emitted fields inside the source as a Gaussian with a deviation of half a wavelength:

$$\mathcal{C}'\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2, \omega\right) = \Lambda(\omega) e^{-\frac{1}{2}\left[\left(\frac{1}{2}\lambda\right)^{-1}\left|\vec{\boldsymbol{r}}'_2 - \vec{\boldsymbol{r}}'_1\right|\right]^2} = \Lambda(\omega) e^{-\frac{k^2\left|\vec{\boldsymbol{r}}'_2 - \vec{\boldsymbol{r}}'_1\right|^2}{2\pi^2}} \ , \tag{3.11}$$

where $\Lambda(\omega)$ is the power spectral density of the light source, i.e. the spectrum of the emitted light. The exact shape and length of the correlation inside the source is of little import, because those characteristics only serve to scale the total emitted power, but do not affect the coherence properties of the emitted light.

Let ρ be the radius of the light source, and we assume that $\rho \gg \lambda$, i.e. the source extent is much greater than the wavelength. Then, the Gaussian in Eq. (3.11) effectively constraints the integration volume of the inner integral in Eq. (3.10) to a volume far smaller than S. This readily allows us to derive an accurate closed-form form expression for the sourced (and propagated) RCSD of the electromagnetic radiation emitted by a spherical natural light source (we fold a few constants into Λ for brevity and convenience):

Definition 3.5.2 (Incoherent Sourcing Operator)**.** *Given Λ, the emitted power spectral density, and ρ, the natural source radius, the sourced RCSD is produced by the* incoherent sourcing operator *as follows:*

$$\boldsymbol{\mathcal{L}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \mathscr{S}\{\Lambda, \rho\} \triangleq \boldsymbol{I} \frac{\Lambda(\omega)}{\sqrt{\rho}} e^{\mathrm{i}k(r_1 - r_2)} \frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k |\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2|\right]}{|\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2|^{\frac{3}{2}}} \ , \tag{3.12}$$

where $\mathrm{J}_{\frac{3}{2}}$ is the Bessel function of the first kind of order $\frac{3}{2}$. The quantities $r_{1,2}$ and $\hat{\boldsymbol{r}}_{1,2}$ are with respect to the sourcing point.

Note that the shape of the cross-spectral density function is influenced neither by the power spectral density nor by the characteristics of the spatial correlation of the fields inside the source, instead it is dictated solely by the spatial distribution of the elementary radiators.

Looking at the limit $\vec{\boldsymbol{r}}_2 \to \vec{\boldsymbol{r}}_1$, we observe that the radiance carried by the RCSD

formalised by Definition 3.5.2 is independent of the distance r, viz.

$$\lim_{\vec{\boldsymbol{r}}_2 \to \vec{\boldsymbol{r}}} \boldsymbol{\mathcal{L}}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}_2, \omega\right) = \lim_{\vec{\boldsymbol{r}}_2 \to \vec{\boldsymbol{r}}} \mathscr{S}\{\Lambda, \rho\} \propto \boldsymbol{I} \rho k^{\frac{3}{2}} \Lambda(\omega) \ , \tag{3.13}$$

as expected. Also, by rewriting $|\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2| \approx \sin\alpha$, where α is the separation angle between $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$, we conclude that the absolute value of the spatial degree-of-coherence of such radiation sourced form a natural light source is then only a function of the separation angle between the points and the wavenumber k.

Analysis The cross-spectral density expression above predicts that light remains spatially coherent over a length inversely proportional to the solid angle subtended by the source, which is consistent with well-known theory and the Van Cittert–Zernike theorem [8]. If we were to plug a wavelength of $\lambda = 0.50\,\mu\text{m}$, the solar radius $\rho \approx 696\,000.00\,\text{km}$ and the distance to earth's surface from the sun, $r \approx 1.00\,\text{au}$, into Definition 3.5.2, and then solve for the first zero of the Bessel function (≈ 4.49341), this would yield a coherence radius of $|\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2| \approx 77.00\,\mu\text{m}$, showing excellent agreement between our model and experimental measurements performed by Mashaal et al. [33] of sunlight's coherence on earth. Likewise, we also found very good agreement with measurements of the coherence of a white LED [25].

Our derivations employ far-field diffraction theory (Eq. (3.9)) that makes the paraxial approximation. To study the validity domain of this approximation, Charnotskii [12] employ the Van Cittert–Zernike theorem to compare the spatial coherence of spherical incoherent sources calculated numerically with and without the paraxial (far-field) approximation. For very small ρ (source radii), roughly $\rho \leq \frac{\lambda}{2}$, the radiation remains coherent across all space. This fits our theory well: Such small sources will consist of few elementary radiators that all fall within the modelled spatial coherence distance in-

side the source (Eq. (3.11)), giving rise to rather coherent radiation. However, with just slightly larger sources, $\rho \geq 2\lambda$, and distances from the source of only $r \geq \frac{3}{2}\rho$, the far-field approximation begins to coincide virtually identically with the non-paraxial result. This suggests that our derivations are applicable to essentially any practical source size and distance from the source.

3.6 Propagation and Diffraction

When electromagnetic waves encounter an obstacle, a brief chaotic moment ensues during which the electromagnetic fields reorganize. After that short instant of disarray, electric and magnetic fields resume their mutual embrace, and carry energy away from the obstacle in form of newly-formed time-harmonic electromagnetic waves. This is known as *diffraction*, and a diffraction problem can be formally stated as: Given a known distribution of the electromagnetic field in a confined region of space, a solution is sought that extends that field to other regions while satisfying Maxwell's equations. That is, physical propagation of light and any scattering and interaction of electromagnetic waves with matter and media are in-fact diffraction problems (note that this is a broader definition than typically followed in computer graphics). Indeed, in Section 3.5 we solved a simple scalar diffraction problem in order to derive an expression for the cross-spectral density of the light sourced and propagated from a natural light source. In this section we consider the problem in its utmost generality and derive the propagation and diffraction formula for the cross-spectral density.

After the brief instant of chaos that follows the disruption of waves by an obstacle, far-field radiation fields materialize. Seeking an analytic expressions for these fields we start with Smythe's diffraction formula, which is a diffraction integral formulated under vectorized diffraction theory and in the far-field takes the following form (see Subsec-

tion 2.6.2 for derivation):

$$\begin{aligned}\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}) &= \mathrm{i}k\frac{e^{\mathrm{i}kr}}{2\pi r}\hat{\boldsymbol{r}} \times \int_{\Sigma} \mathrm{d}^2\vec{\boldsymbol{r}}'_{\perp}\Big[\hat{\boldsymbol{n}} \times \vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_{\perp}\big)\Big]e^{-\mathrm{i}k\hat{\boldsymbol{r}}\cdot\vec{\boldsymbol{r}}'_{\perp}} \\ &= \mathrm{i}k\frac{e^{\mathrm{i}kr}}{2\pi r}\hat{\boldsymbol{r}} \times \mathscr{F}_{\Sigma}\Big\{\hat{\boldsymbol{n}} \times \vec{\boldsymbol{E}}\Big\}(\hat{\boldsymbol{r}}) \ , \end{aligned} \tag{3.14}$$

where Σ is an "aperture"—a small planar region in space upon which the electric field $\vec{\boldsymbol{E}}$ is known, $\hat{\boldsymbol{n}}$ is the aperture normal and the (spatial) Fourier transform is restricted to that aperture. We fix our coordinate system such that Σ is centred at the origin. Because we employ vector diffraction theory, we need to choose the transverse orthonormal bases for our fields. For the incident field, i.e. the known field that impinges upon the aperture, we assume decomposition under some given basis $\{\hat{\boldsymbol{e}}'_1, \hat{\boldsymbol{e}}'_2\}$, with the implied assumption being that the incident radiation is concentrated in some direction $\hat{\boldsymbol{k}}'$ and thus $\hat{\boldsymbol{e}}'_1 \cdot \hat{\boldsymbol{k}}' = \hat{\boldsymbol{e}}'_2 \cdot \hat{\boldsymbol{k}}' = 0$ and naturally $\hat{\boldsymbol{e}}'_1 \cdot \hat{\boldsymbol{e}}'_2 = 0$. Likewise, the arbitrary orthonormal basis under which the propagated light is decomposed is denoted as $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$, such that $\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{r}} = \hat{\boldsymbol{e}}_2 \cdot \hat{\boldsymbol{r}} = 0$ as well as $\hat{\boldsymbol{e}}_1 \cdot \hat{\boldsymbol{e}}_2 = 0$. See Fig. 3.2 for an illustration. By making the Born approximation [61], the planar aperture can be immediately extended to an arbitrary small volume. Physically, this should be understood in the sense that an incident field induces electromagnetic excitations in that volume, and these excitations give rise to the superposition of plane-wave modes described by the Fourier transform in Eq. (3.14). This is a very good approximation when Σ is small and not a good conductor.

Let the incident field's cross-spectral density matrix be denoted as $\boldsymbol{\mathcal{C}}'$. We are interested in deriving an expression for the cross-spectral density matrix $\boldsymbol{\mathcal{C}}$ of the radiation that diffracted and propagated to $\vec{\boldsymbol{r}}$. An element $\mathcal{C}_{\xi\zeta}$ (indexed by $\xi, \zeta \in \{1, 2\}$) of that

cross-spectral density matrix (Eq. (2.150)) can be written as

$$\begin{aligned}\mathcal{C}_{\xi\zeta}\left(\vec{\boldsymbol{r}}_1,\vec{\boldsymbol{r}}_2,\omega\right) &= \left\langle \left[\hat{\boldsymbol{e}}_\xi\cdot\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_1)\right]\left[\hat{\boldsymbol{e}}_\zeta\cdot\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_2)\right]^\star\right\rangle_\omega \\ &= \frac{k^2e^{\mathrm{i}k(r_1-r_2)}}{4\pi^2r_1r_2}\left\langle\left[\hat{\boldsymbol{e}}_x\cdot\left(\hat{\boldsymbol{r}}_1\times\int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}'_\perp\Big(\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_\perp\big)\Big)e^{-\mathrm{i}k\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{r}}'_\perp}\right)\right]\right. \\ &\quad\left.\times\left[\hat{\boldsymbol{e}}_y\cdot\left(\hat{\boldsymbol{r}}_2\times\int_\Sigma \mathrm{d}^2\vec{\boldsymbol{r}}'_\perp\Big(\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}\big(\vec{\boldsymbol{r}}'_\perp\big)\Big)e^{-\mathrm{i}k\hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{r}}'_\perp}\right)\right]^\star\right\rangle_\omega ,\end{aligned} \tag{3.15}$$

where we applied the far-field Smythe's diffraction formula to each of the field components. As usual, $r_j = |\vec{\boldsymbol{r}}_j|$ and $\hat{\boldsymbol{r}}_j = \vec{\boldsymbol{r}}_j/r_j$. Then, the vector expressions, inside the ensemble average, are rewritten as follows

$$\begin{aligned}\hat{\boldsymbol{e}}_j\cdot\left[\hat{\boldsymbol{r}}_l\times\Big(\hat{\boldsymbol{z}}\times\vec{\boldsymbol{E}}\Big)\right] &= \hat{\boldsymbol{e}}_j\cdot\left[\hat{\boldsymbol{z}}\Big(\hat{\boldsymbol{r}}_l\cdot\vec{\boldsymbol{E}}\Big)-(\hat{\boldsymbol{r}}_l\cdot\hat{\boldsymbol{z}})\vec{\boldsymbol{E}}\right] \\ &= (\hat{\boldsymbol{e}}_j\cdot\hat{\boldsymbol{z}})\Big(\hat{\boldsymbol{r}}_l\cdot\vec{\boldsymbol{E}}\Big)-(\hat{\boldsymbol{r}}_l\cdot\hat{\boldsymbol{z}})\Big(\hat{\boldsymbol{e}}_j\cdot\vec{\boldsymbol{E}}\Big) \\ &= (\hat{\boldsymbol{e}}_j\cdot\hat{\boldsymbol{z}})\hat{\boldsymbol{r}}_l\cdot\left(\vec{\boldsymbol{e}}'_1E_1+\vec{\boldsymbol{e}}'_2E_2\right)-(\hat{\boldsymbol{r}}_l\cdot\hat{\boldsymbol{z}})\hat{\boldsymbol{e}}_j\cdot\left(\vec{\boldsymbol{e}}'_1E_1+\vec{\boldsymbol{e}}'_2E_2\right),\end{aligned} \tag{3.16}$$

where we applied the triple product vector identities (Eqs. (2.9) and (2.10)). Denote the common factors

$$\begin{aligned}{}_j\mathfrak{h}_{\xi,\zeta} &= (\hat{\boldsymbol{e}}_j\cdot\hat{\boldsymbol{z}})\big(\hat{\boldsymbol{r}}_\xi\cdot\hat{\boldsymbol{e}}'_\zeta\big)-(\hat{\boldsymbol{r}}_\xi\cdot\hat{\boldsymbol{z}})\big(\hat{\boldsymbol{e}}_j\cdot\hat{\boldsymbol{e}}'_\zeta\big) \\ &= (\hat{\boldsymbol{e}}_j\times\hat{\boldsymbol{r}}_\xi)\cdot\big(\hat{\boldsymbol{z}}\times\hat{\boldsymbol{e}}'_\zeta\big),\end{aligned} \tag{3.17}$$

with $j,\xi,\zeta\in\{1,2\}$ being indices. Then, by substituting Eqs. (3.16) and (3.17) into Eq. (3.15), formally interchanging the orders of ensemble averaging and integration and applying a series of elementary algebraic manipulations, the propagated cross-spectral

density finally becomes

$$\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2, \omega\right) = \frac{1}{\lambda^2} \frac{e^{\mathrm{i}k(r_1 - r_2)}}{r_1 r_2} \boldsymbol{H}_1 \boldsymbol{\Psi} \boldsymbol{H}_2^T \ , \tag{3.18}$$

with $\boldsymbol{H}$ being the following shorthand

$$\boldsymbol{H}_\xi \triangleq \begin{bmatrix} {}_1\mathfrak{h}_{\xi,1} & {}_1\mathfrak{h}_{\xi,2} \\ {}_2\mathfrak{h}_{\xi,1} & {}_2\mathfrak{h}_{\xi,2} \end{bmatrix} \ , \tag{3.19}$$

and with the matrix $\boldsymbol{\Psi}$ being the central quantity of interest: the (double) two-dimensional Fourier transform of the cross-spectral density on the aperture, viz.

$$\begin{aligned} \boldsymbol{\Psi} &\triangleq \int_\Sigma \mathrm{d}^2 \vec{\boldsymbol{r}}'_\perp \int_\Sigma \mathrm{d}^2 \vec{\boldsymbol{r}}''_\perp \, \boldsymbol{\mathcal{C}}'\left(\vec{\boldsymbol{r}}'_\perp, \vec{\boldsymbol{r}}''_\perp, \omega\right) e^{-\mathrm{i}k\left(\hat{\boldsymbol{r}}_1 \cdot \vec{\boldsymbol{r}}'_\perp - \hat{\boldsymbol{r}}_2 \cdot \vec{\boldsymbol{r}}''_\perp\right)} \\ &= (2\pi)^2 \, \mathscr{F}\left\{ \mathbb{1}_\Sigma\left(\vec{\boldsymbol{r}}'_\perp\right) \mathscr{F}\left\{ \mathbb{1}_\Sigma\left(\vec{\boldsymbol{r}}''_\perp\right) \boldsymbol{\mathcal{C}}'\left(\vec{\boldsymbol{r}}'_\perp, \vec{\boldsymbol{r}}''_\perp, \omega\right) \right\}(-k\hat{\boldsymbol{r}}_2) \right\}(k\hat{\boldsymbol{r}}_1) \ . \end{aligned} \tag{3.20}$$

The $\mathbb{1}_\Sigma$ denotes the characteristic function of the aperture and the Fourier transforms are with respect to the variables $\vec{\boldsymbol{r}}'_\perp$ and $\vec{\boldsymbol{r}}''_\perp$.

We may now formally define the diffraction operator:

Definition 3.6.1 (Diffraction Operator)**.** *Let $\boldsymbol{\mathcal{C}}'$ be the cross-spectral density function of the radiation that arises in a small region $\Sigma \subset \mathbb{R}^3$. The RCSD of the diffracted and propagated radiation is:*

$$\mathscr{D}\{\boldsymbol{\mathcal{C}}', \Sigma\} \triangleq \frac{e^{\mathrm{i}k(r_1 - r_2)}}{\lambda^2 A_\Sigma} \boldsymbol{H}_1 \boldsymbol{\Psi}(\hat{\boldsymbol{r}}_1, \hat{\boldsymbol{r}}_2, \omega) \boldsymbol{H}_2^T \ , \tag{3.21}$$

where A_Σ is the cross-sectional area of Σ. The quantities $r_{1,2}$ and $\hat{\boldsymbol{r}}_{1,2}$ are with respect to the point subtending the RCSD.

We briefly verify that the propagated cross-spectral density matrix, $\boldsymbol{\mathcal{C}}$, is a polariza-

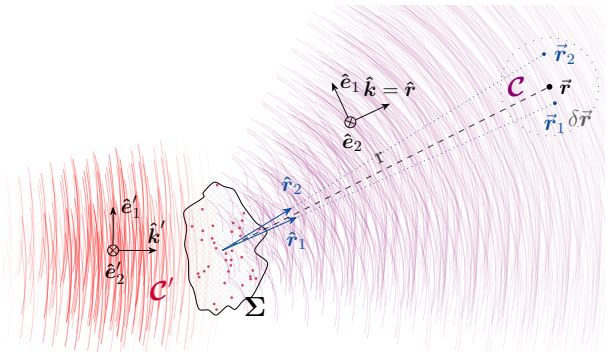


Figure 3.2: Diffraction of natural light by an obstacle Σ. The obstacle may be any matter or medium or, equivalently, an opening in a conductive screen (i.e. an aperture). An incident field (wavefronts illustrated as red lines), with its coherence properties quantified by $\boldsymbol{\mathcal{C}}'$, induces electromagnetic excitations in Σ (small purple balls), giving rise to *diffracted* radiation (wavefronts illustrated as purple lines). The solution to the diffraction problem is then $\boldsymbol{\mathcal{C}}$, which quantifies the diffracted wave ensemble that has propagated to some point $\vec{\boldsymbol{r}}$. The wavevectors $\hat{\boldsymbol{k}}$ and the orthonormal bases $\hat{\boldsymbol{e}}$ are illustrated as well. Note, as the diffracted wave ensemble propagates the wavefronts become more similar, illustrating that light gains coherence on propagation.

tion matrix (see Section 2.6) when evaluated at a single point. Indeed, when $\vec{\boldsymbol{r}}_1 = \vec{\boldsymbol{r}}_2$, it is easy to see that $\boldsymbol{H}_1 = \boldsymbol{H}_2$ and thus $(\boldsymbol{H}_1\boldsymbol{\Psi}\boldsymbol{H}_2^T)^\dagger = \boldsymbol{H}_1\Psi^\dagger\boldsymbol{H}_2^T$. As $\boldsymbol{\Psi} = \boldsymbol{\Psi}^\dagger$, due to $\boldsymbol{\Psi}$ being the Fourier transform of a Hermitian matrix, we deduce that $\boldsymbol{\mathcal{C}}(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega)$ is Hermitian. Further, $\boldsymbol{\Psi}$ is positive semi-definite, thus we can write $\boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{A}^\dagger$ for some matrix $\boldsymbol{A}$ and then $\boldsymbol{H}_1\boldsymbol{\Psi}\boldsymbol{H}_2^T = (\boldsymbol{H}_1\boldsymbol{A})(\boldsymbol{H}_1\boldsymbol{A})^\dagger$, therefore $\boldsymbol{\mathcal{C}}(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega) \succeq 0$. We conclude that if $\boldsymbol{\mathcal{C}}'(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega)$ is a polarization matrix, then $\boldsymbol{\mathcal{C}}(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}, \omega)$ is a polarization matrix.

Definition 3.6.1 is an accurate far-field diffraction and propagation formula for the cross-spectral density matrix of light of any state of coherence, polarization and spectrum, diffracted by a small region in space, under the context of full electromagnetism. To our knowledge, this expression has not been previously derived.

Note that under the highly general diffraction formulation that was discussed in this section, the radiance carried to the far-field by the diffracted RCSD conveniently does not depend on r. Furthermore, $\boldsymbol{\Psi}$, and in turn the absolute value of the spatial degree-of-coherence, is only a function of the directions $\hat{\boldsymbol{r}}_{1,2}$, in contrast to the points $\vec{\boldsymbol{r}}_{1,2}$, indicating

again that spatial coherence stems from propagation.

Superposition of light of any state of coherence Assume that light from two different sources (primary or secondary) is superposed, with the sources being Σ_1 and Σ_2, assumed to be disjoint. We study the diffracted RCSD by applying the diffraction formula (Definition 3.6.1) to the diffracting region $\Sigma = \Sigma_1 \cup \Sigma_2$. The matrix $\boldsymbol{\Psi}$ (Eq. (3.20)) then can be written as:

$$\begin{aligned}\boldsymbol{\Psi} &= \int_{\Sigma}\int_{\Sigma} \boldsymbol{\mathcal{C}}'\left(\vec{\boldsymbol{r}}', \vec{\boldsymbol{r}}'', \omega\right) e^{-\mathrm{i}k\left(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{r}}' - \hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{r}}''_{\perp}\right)} \\ &= \sum_{\xi,\zeta}\int_{\Sigma_\xi}\int_{\Sigma_\zeta} \boldsymbol{\mathcal{C}}'\left(\vec{\boldsymbol{r}}', \vec{\boldsymbol{r}}'', \omega\right) e^{-\mathrm{i}k\left(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{r}}' - \hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{r}}''_{\perp}\right)} \ , \end{aligned} \tag{3.22}$$

with $\xi, \zeta \in \{1, 2\}$ being indices. The above is trivially extended to any number of sources. Clearly, if the distance between a pair of sources Σ_ξ and Σ_ζ is large, with respect to the distances over which light remains spatially coherent, then $\boldsymbol{\mathcal{C}}'\left(\vec{\boldsymbol{r}}', \vec{\boldsymbol{r}}'', \omega\right) \approx 0$ whenever $\xi \neq \zeta$. In that case, we say that the the sources are statistically independent, and the sum in Eq. (3.22) reduces to a sum over $\xi = \zeta$ only. Hence, the RCSD of the superposed radiation is simply the linear combination of the RCSDs of the light from each source.

Scalar and vector diffraction treatment We choose to employ a rigorous vector formalism, both for our discussion of diffraction and our entire framework. This is justified by our motivation to develop a general-purpose theory applicable to a wide variety of applications. A quick comparison of our derived diffraction formulae for the scalar and matrix RCSDs, reveals that Definition 3.6.1 is equivalent to its scalar counterpart, Eq. (3.10) (where in general we need to also account for inclination), only at normal incidence and when $\hat{\boldsymbol{r}} \approx \hat{\boldsymbol{n}}$. In which case it can be shown that indeed $\boldsymbol{H}_{1,2} \approx \boldsymbol{I}$. However, away from normal incidence and exitance, the vectorized theory provides a more

accurate formalism. Under this vectorized formalism the polarization state of the wave ensemble becomes a first-class citizen, which is important as polarization plays a crucial role in many light-matter interactions and applications.

Helmholtz Reciprocity and the Short-Wavelength Limit A defining property of the classical BSDF f_r is reciprocity, which can be formally stated as $f_r(\Omega_i, \Omega_o, \omega) = f_r(\Omega_o, \Omega_i, \omega)$. Assume that a single source gives rise to the irradiance impinging on an aperture Σ from direction Ω_1, and using Definition 3.4.1 the cross-spectral density can be written as (assuming quasi-homogeneous source, which serves a decent illustrative approximation in the far-field)

$$\boldsymbol{\mathcal{C}}\Big|_{\Sigma} = \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, |\Omega' \cdot \hat{\boldsymbol{n}}| \, \boldsymbol{\mathcal{L}}^{(i)}_{\Omega'} = |\Omega_1||\Omega_1 \cdot \hat{\boldsymbol{n}}| \, \boldsymbol{\mathcal{L}}^{(i)}_{\Omega_1} \ , \tag{3.23}$$

where $\boldsymbol{\mathcal{L}}^{(i)}_{\Omega_1}$ is the RCSD incident from the source and $|\Omega_1|$ is the solid angle subtended by the source. Using the diffraction operator, the spectral radiance quantified by the RCSD of the radiation diffracted and scattered in direction Ω_2 can be expressed as follows:

$$\begin{aligned} \boldsymbol{\mathcal{L}}^{(s)}_{\Omega_1}\Big|_{\hat{\boldsymbol{r}}_1=\hat{\boldsymbol{r}}_2=\Omega_2} &= \mathscr{D}\Big\{\boldsymbol{\mathcal{C}}\Big|_{\Sigma}, \Sigma\Big\}\Big|_{\hat{\boldsymbol{r}}_1=\hat{\boldsymbol{r}}_2=\Omega_2} \\ &= \frac{|\Omega_1||\Omega_1 \cdot \hat{\boldsymbol{n}}| e^{\mathrm{i}k(r_1-r_2)}}{\lambda^2 A_\Sigma |\Omega_2 \cdot \hat{\boldsymbol{n}}|} \boldsymbol{H}_{\Omega_1} \boldsymbol{\Psi}_{\Omega_1} \boldsymbol{H}^T_{\Omega_1} \ , \end{aligned} \tag{3.24}$$

which the subscript denoting the direction of the source. Note that the matrices $\boldsymbol{H}$ in the expression above are equal when $\hat{\boldsymbol{r}}_1 = \hat{\boldsymbol{r}}_2$. The reciprocal problem is the setting with a source radiating from Ω_2 and observing the diffracted radiation in direction Ω_1, which analogously is

$$\boldsymbol{\mathcal{L}}^{(s)}_{\Omega_2}\Big|_{\hat{\boldsymbol{r}}_1=\hat{\boldsymbol{r}}_2=\Omega_1} = \frac{|\Omega_2||\Omega_2 \cdot \hat{\boldsymbol{n}}| e^{\mathrm{i}k(r_1-r_2)}}{\lambda^2 A_\Sigma |\Omega_1 \cdot \hat{\boldsymbol{n}}|} \boldsymbol{H}_{\Omega_2} \boldsymbol{\Psi}_{\Omega_2} \boldsymbol{H}^T_{\Omega_2} \ . \tag{3.25}$$

We choose our bases such that $\hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{e}}'_1$ and both are perpendicular to the plane spanned by Ω_1 and Ω_2, and this fixes the second base components (up to a sign). In this case, with some vector algebra we deduce that $\boldsymbol{H}_{\Omega_1} = |\Omega_2 \cdot \hat{\boldsymbol{n}}|\, \boldsymbol{I}$ and $\boldsymbol{H}_{\Omega_2} = |\Omega_1 \cdot \hat{\boldsymbol{n}}|\, \boldsymbol{I}$. Hence, if $|\Omega_1| = |\Omega_2|$, then the "geometric" terms in the expressions Eqs. (3.24) and (3.25) (all terms except $\boldsymbol{\Psi}$) become equal. Nonetheless, clearly $\boldsymbol{\Psi}$ is not reciprocal in general. We conclude: diffraction of partially-coherent light is reciprocal only up to these "geometric" terms.

3.7 Example Application and Validation

In this section we present one example application.

3.7.1 Rendering Diffraction on Reflection from a Surface

Consider a surface described explicitly by some height function, $h : \mathbb{R}^2 \to \mathbb{R}$, where the height deviations are from the mean surface plane, the xy-plane. The surface normal is denoted $\hat{\boldsymbol{n}} = \hat{\boldsymbol{z}}$. We denote points on the mean surface plane as $\vec{\boldsymbol{x}}_\perp$, and the corresponding surface position is $\vec{\boldsymbol{x}} = \vec{\boldsymbol{x}}_\perp + h(\vec{\boldsymbol{x}}_\perp)\hat{\boldsymbol{n}}$. Assume that a natural spherical light source of radius ρ and with emitted power spectral density Λ is positioned at point $\vec{\boldsymbol{s}}$. Radiation from the source is incident to a surface patch $\delta\vec{\boldsymbol{x}}$ at incidence direction Ω_i (relative to the normal $\hat{\boldsymbol{n}}$), and we are interested in measuring the observed radiance of light reflecting (or scattering) from the surface patch and arriving at an imaging device at position $\vec{\boldsymbol{e}}$.

Applying the SDTE (Definition 3.4.1), we are interested in the quantity

$$L(\vec{\boldsymbol{e}}) = \mathscr{L}\left\{ \int_{\mathscr{S}^2} \mathrm{d}^2\Omega' \, |\Omega' \cdot \hat{\boldsymbol{n}}| \mathscr{D}_{\delta\vec{\boldsymbol{x}},\Omega'\to\Omega_o}\{\mathscr{S}_{\mathrm{inc}}\{\Lambda,\rho\}\}, \vec{\boldsymbol{e}} \right\}, \tag{3.26}$$

where Ω_o is the direction of reflection of the surface patch and $\mathscr{D}$ is an abstract surface diffraction operator. Our focus is on deriving an analytic expression for $\mathscr{D}$ acting upon the sourced RCSD generated by $\mathscr{S}_{\text{inc}}$. When the characteristic size of the surface patch is large compared to the height fluctuations described by the function h, the surface can be considered as an aperture. We assume that the reflected field strength ratio is constant across the surface patch, hence we use the Church reflection matrix $\boldsymbol{Q}$ [29], in similar manner to the Harvey-Shack surface scatter theory. Having $\boldsymbol{Q}$ be independent of the position on the surface patch simplifies the numeric calculations to complex scalars (instead of complex 2×2 matrices), but otherwise is done purely for simplicity and convenience.

Without loss of generality we center the surface patch at the origin, viz. $\vec{\boldsymbol{x}} = 0$. Then, $\vec{\boldsymbol{s}}$ is the vector from the patch to the source, $s = |\vec{\boldsymbol{s}}|$ is the mean distance from the surface patch to the source and $\hat{\boldsymbol{s}} = \vec{\boldsymbol{s}}/s$ is the direction to the source from the surface patch. And similar notation is employed for $\vec{\boldsymbol{e}}$. The sourced RCSD incident upon the surface patch is generated via Definition 3.5.2. We denote $\vec{\boldsymbol{x}}'$ and $\vec{\boldsymbol{x}}''$ as a pair of points on the surface and their projections onto the mean surface plane as $\vec{\boldsymbol{x}}'_\perp$, $\vec{\boldsymbol{x}}''_\perp$, likewise, we denote $\vec{\boldsymbol{s}}'$ as a point on the cross-section of the light source facing the surface patch. Making the typical far-field paraxial (small-angle) approximation, the distance between a point $\vec{\boldsymbol{x}}'$ on the surface to $\vec{\boldsymbol{s}}'$ on the source is

$$\left|\vec{\boldsymbol{x}}' - \vec{\boldsymbol{s}}'\right| = s + \frac{\left|\vec{\boldsymbol{x}}'\right| - 2\vec{\boldsymbol{s}}' \cdot \vec{\boldsymbol{x}}'}{2s'} + \mathcal{O}\left(\frac{1}{s'^2}\right) \approx s' - \hat{\boldsymbol{s}}' \cdot \vec{\boldsymbol{x}}' \ , \tag{3.27}$$

where the power expansion was simplified using $|\vec{\boldsymbol{x}}'| \ll s$. And similarly for the distance between a point on the surface and the imaging device

$$\left|\vec{\boldsymbol{x}}' - \vec{\boldsymbol{e}}\right| \approx e - \hat{\boldsymbol{e}} \cdot \vec{\boldsymbol{x}}' \ . \tag{3.28}$$

Then, the RCSD that enters the "aperture" is denoted $\boldsymbol{\mathcal{L}}'$ and is the reflected and propagated sourced RCSD, viz.

$$\boldsymbol{\mathcal{L}}'\left(\vec{\boldsymbol{x}}'_{\perp},\vec{\boldsymbol{x}}''_{\perp},\omega\right)=\boldsymbol{Q}\boldsymbol{C}e^{\mathrm{i}k(\hat{\boldsymbol{s}}'+\hat{\boldsymbol{e}})\cdot\left(\vec{\boldsymbol{x}}'-\vec{\boldsymbol{x}}''\right)}\frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k|\hat{\boldsymbol{x}}'-\hat{\boldsymbol{x}}''|\right]}{s'^{2}|\hat{\boldsymbol{x}}'-\hat{\boldsymbol{x}}''|^{\frac{3}{2}}}\boldsymbol{Q}^{\dagger}\ , \tag{3.29}$$

where we applied Eqs. (3.27) and (3.28) and, for brevity, we denote $\boldsymbol{C}$ as the shorthand for the (position-independent) constants in Definition 3.5.2. The familiar wave propagation term, $\exp\left[\mathrm{i}k(\hat{\boldsymbol{s}}'+\hat{\boldsymbol{e}})\cdot\vec{\boldsymbol{x}}'\right]$, that appears in Eq. (3.29), can be regraded as the propagator that propagates the sourced RCSD to the "aperture"-plane (after reflection) taking the height fluctuations into account. We refer to it as the *frequency transmission function* [8] that arises due to residual surface roughness. While the frequency transmission function describes the height fluctuations of the surface geometry, the scalar fraction that appears in Eq. (3.29) describes the coherence properties (second-order statistics) of the light incident upon the surface patch. We denote it as the *impulse response function*, $\mathcal{P}$, induced by the wave ensemble. Making the approximation

$$|\hat{\boldsymbol{x}}'-\hat{\boldsymbol{x}}''|\approx\frac{1}{s'}|\hat{\boldsymbol{s}}'\times\left(\vec{\boldsymbol{x}}'_{\perp}-\vec{\boldsymbol{x}}''_{\perp}\right)| \tag{3.30}$$

(for some $\vec{\boldsymbol{s}}'$) which is an excellent approximation in the far-field, the impulse response function for our incoherent spherical source can be written as

$$\mathcal{P}\left(\vec{\boldsymbol{x}}'_{\perp}-\vec{\boldsymbol{x}}''_{\perp},\omega\right)\triangleq\frac{s'^{\frac{3}{2}}\mathrm{J}_{\frac{3}{2}}\left[\frac{\rho k}{s'}\left|\hat{\boldsymbol{s}}'\times\left(\vec{\boldsymbol{x}}'_{\perp}-\vec{\boldsymbol{x}}''_{\perp}\right)\right|\right]}{\left|\hat{\boldsymbol{s}}'\times\left(\vec{\boldsymbol{x}}'_{\perp}-\vec{\boldsymbol{x}}''_{\perp}\right)\right|^{\frac{3}{2}}}\ . \tag{3.31}$$

We now apply the diffraction operator, Definition 3.6.1, to Eq. (3.29) and substitute

the resulting expression into Eq. (3.26) yields

$$L(\vec{\boldsymbol{e}}) = \frac{\Theta}{\sqrt{\rho}} \frac{1}{A_\Sigma |\hat{\boldsymbol{e}} \cdot \hat{\boldsymbol{n}}|} \int_0^\infty \frac{\mathrm{d}\omega}{\lambda^2} \operatorname{tr} \left[\boldsymbol{H}_1 \int_{\mathscr{S}^2_+} \mathrm{d}^2 \hat{\boldsymbol{s}}' \, \hat{\boldsymbol{s}}' \cdot \hat{\boldsymbol{n}} \, \mathscr{F}^2_\Sigma \{ \boldsymbol{\mathcal{L}}' \} \boldsymbol{H}_2^T \right] , \tag{3.32}$$

where $\mathscr{S}^2_+$ is the upper unit hemisphere in direction $\hat{\boldsymbol{n}}$ centred on $\vec{\boldsymbol{x}}$, $\Sigma = (\delta \vec{\boldsymbol{x}})_\perp$ is the area on the mean surface plane that subtends $\delta \vec{\boldsymbol{x}}$ and Θ is the solid angle subtended by the source

$$\Theta \triangleq 2\pi \left(1 - \sqrt{1 - \frac{\rho^2}{s^2}} \right) . \tag{3.33}$$

The double two-dimensional Fourier transform of the sourced and reflected RCSD in the expression above is then over the surface geometry. As a double Fourier transform is cumbersome to numerically compute, we proceed by simplifying the expression:

$$\begin{aligned} \mathscr{F}^2_\Sigma \{ \boldsymbol{\mathcal{L}}' \} &= \int_\Sigma \mathrm{d}^2 \vec{\boldsymbol{x}}'_\perp \int_\Sigma \mathrm{d}^2 \vec{\boldsymbol{x}}''_\perp \, \boldsymbol{\mathcal{L}}' \left(\vec{\boldsymbol{x}}', \vec{\boldsymbol{x}}'', \omega \right) e^{-\mathrm{i}k \left(\hat{\boldsymbol{e}} \cdot \vec{\boldsymbol{x}}'_\perp - \hat{\boldsymbol{e}} \cdot \vec{\boldsymbol{x}}''_\perp \right)} \\ &= \boldsymbol{Q} \boldsymbol{C} \boldsymbol{Q}^\dagger \int_\Sigma \mathrm{d}^2 \vec{\boldsymbol{x}}'_\perp \, e^{-\mathrm{i}k \hat{\boldsymbol{e}} \cdot \vec{\boldsymbol{x}}'_\perp} \\ &\qquad \times (2\pi) \Big[\mathscr{F}_\Sigma \Big\{ e^{-\mathrm{i}k (\hat{\mathbf{s}}' + \hat{\mathbf{e}}) \cdot \left(\vec{\boldsymbol{x}}' - \vec{\boldsymbol{x}}'' \right)} \Big\} * \mathscr{F}_\Sigma \big\{ \mathcal{P} \big(\vec{\boldsymbol{x}}'_\perp - \vec{\boldsymbol{x}}''_\perp, \omega \big) \big\} \Big] , \end{aligned} \tag{3.34}$$

where we applied the convolution theorem (Theorem 2.2.1.2). The $*$ denotes the two-dimensional convolution over the mean surface plane. Both the Fourier transforms in the expression above are over Σ and with respect to the double-primed integration variable $\vec{\boldsymbol{x}}''_\perp$. The convolution is evaluated at frequency $(-k\hat{\boldsymbol{e}})$. Note that by the Fourier transform shift identity and the variable change $\vec{\boldsymbol{y}}'_\perp = \vec{\boldsymbol{x}}'_\perp - \vec{\boldsymbol{x}}''_\perp$ the following holds

$$\mathscr{F} \left\{ \mathcal{P} (\vec{\boldsymbol{x}}'_\perp - \vec{\boldsymbol{x}}''_\perp, \omega) \right\} \left(\vec{\boldsymbol{\xi}} \right) = -e^{-\mathrm{i} \vec{\boldsymbol{\xi}} \cdot \vec{\boldsymbol{x}}'_\perp} \, \mathscr{F} \left\{ \mathcal{P} (\vec{\boldsymbol{y}}'_\perp, \omega) \right\} \left(-\vec{\boldsymbol{\xi}} \right) , \tag{3.35}$$

making the Fourier transform of the impulse response function independent of $\vec{\boldsymbol{x}}'_\perp$. Then, formally interchanging the orders of convolution and Fourier transform in Eq. (3.34) leads to

$$\begin{aligned}\mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\} = -\boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q}^\dagger \int_{\mathbb{R}^2} \mathrm{d}^2\vec{\boldsymbol{\xi}}\, \mathscr{F}\left\{\mathcal{P}(\vec{\boldsymbol{x}}''_\perp,\omega)\right\}\left(-\vec{\boldsymbol{\xi}}\right) \int_\Sigma \mathrm{d}^2\vec{\boldsymbol{x}}'_\perp\, e^{-\mathrm{i}\vec{\boldsymbol{\xi}}\cdot\vec{\boldsymbol{x}}'_\perp} \\ \times\, e^{-\mathrm{i}k\hat{\boldsymbol{e}}\cdot\vec{\boldsymbol{x}}'_\perp} \int_\Sigma \mathrm{d}^2\vec{\boldsymbol{x}}''_\perp\, e^{\mathrm{i}\left(k\hat{\boldsymbol{e}}+\vec{\boldsymbol{\xi}}\right)\cdot\vec{\boldsymbol{x}}''_\perp} e^{-\mathrm{i}k(\hat{\boldsymbol{s}}'+\hat{\boldsymbol{e}})\cdot\left(\vec{\boldsymbol{x}}'-\vec{\boldsymbol{x}}''\right)}\ ,\end{aligned} \tag{3.36}$$

with $\mathbb{R}^2$ being the mean surface plane. By rearranging the terms and performing the variable change $\vec{\boldsymbol{\zeta}} = -\vec{\boldsymbol{\xi}}$ in the expression above we can rewrite it as a convolution of independent expressions. Thus, the double Fourier transform reduces to the more computationally-tractable expression

$$\mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\} = \boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q}^\dagger\left[\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\}\right](k\hat{\boldsymbol{e}})\ , \tag{3.37}$$

with $\mathcal{T}$ being the Fourier transform of the frequency transmission function, viz.

$$\mathcal{T}\left(\vec{\boldsymbol{\xi}}\right) \triangleq \mathscr{F}_\Sigma\left\{e^{\mathrm{i}k(\hat{\boldsymbol{s}}'+\hat{\boldsymbol{e}})\cdot\left[\vec{\boldsymbol{x}}'_\perp+\hat{\boldsymbol{n}}h(\vec{\boldsymbol{x}}'_\perp)\right]}\right\}(\vec{\boldsymbol{\xi}}) \tag{3.38}$$

(where we write $\vec{\boldsymbol{x}}' = \vec{\boldsymbol{x}}'_\perp + \hat{\boldsymbol{n}}h(\vec{\boldsymbol{x}}'_\perp)$ to make the height function explicit) and $\mathcal{T}^2(\vec{\boldsymbol{\xi}}) = \mathcal{T}(\vec{\boldsymbol{\xi}})\mathcal{T}(\vec{\boldsymbol{\xi}})^\star$ is the *frequency response function*. Eq. (3.37) is simply the two-dimensional convolution of the Fourier transform of the impulse response function with the frequency response function. Note that the impulse response function $\mathcal{P}$ is a symmetric real function, therefore its Fourier transform is real. Likewise, the frequency response function $\mathcal{T}^2$ is evidently real. Therefore, the double Fourier transform expression above is real as well, as expected seeing as we are computing the observed radiance.

Finally, substituting Eq. (3.37) into Eq. (3.32) and applying the paraxial approxi-

mation by settings $\vec{\boldsymbol{s}}' = \vec{\boldsymbol{s}}$, i.e. far-field with $s \gg \rho$, yields the simplified expression for the observed radiance of partially-coherent light reflected from an explicitly described surface:

$$L(\vec{\boldsymbol{e}}) = \frac{\Theta}{\sqrt{\rho}} \frac{1}{A_\Sigma |\hat{\mathbf{e}} \cdot \hat{\mathbf{n}}|} \int_0^\infty \frac{\mathrm{d}\omega}{\lambda^2} \Lambda(\omega) \times \operatorname{tr}\left[\boldsymbol{H}_1 \boldsymbol{Q} (\boldsymbol{H}_2 \boldsymbol{Q})^\dagger\right] \left(\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\}\right)(k\hat{\mathbf{e}}) \ . \tag{3.39}$$

Numeric evaluation The evaluation of Eq. (3.39) will typically be done numerically. To that end it is more suitable to rewrite the convolution as multiplication in Fourier space, viz.

$$\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} = \mathscr{F}\left\{\mathscr{F}^{-1}\{\mathcal{T}^2\} \cdot \mathcal{P}\right\} \ . \tag{3.40}$$

Then, $\mathscr{F}^{-1}\{\mathcal{T}^2\}$ is evaluated via a pair of fast Fourier transforms (FFT). As we only need a single frequency of the last transform, the Goertzel algorithm (generalized to arbitrary frequencies [47]) is used as it is faster, compared to an FFT, and provides a sinc interpolated result. Overall, this implies a time-complexity of $\mathcal{O}(n \log n)$.

Discussion The derivations in this section have been focused on an incoherent spherical source. A different light source would give rise to a different impulse response function. However, as we evaluate the impulse response function numerically, the same approach remains applicable for any incident RCSD function.

Observe that for at the perfectly coherent limit, the impulse response function becomes a constant, viz. $\mathcal{P}(\vec{\boldsymbol{y}}, \omega) \propto k^2$. Then, its Fourier transform is $\mathscr{F}\{\mathcal{P}\} \propto 2\pi k^2 \delta(\vec{\boldsymbol{\xi}})$, i.e. proportional to a Dirac delta. Using the fact that the Dirac delta is the identity under convolution algebra, we get $\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} \propto 2\pi k^2 \mathcal{T}^2$. Hence, the observed intensity

is proportional to the complex magnitude squared (i.e. power) of the Fourier transform of the amplitude distribution over the surface, as described by Eq. (3.14). That is, the problem reduces to a simple diffraction problem where we ignore coherence, as expected.

In this section we have derived an expression for the RCSD of light reflected of a surface under the context of a vectorized diffraction theory. Our primary result is showing that the reflected RCSD can be formulated as the convolution between a function describing the coherence properties of the incident wave ensemble and a function describing the surface-induced perturbations: That is, a convolution between the Fourier transform of the impulse response function $\mathcal{P}$, which is independent of the surface properties, and the frequency response function $\mathcal{T}^2$, which is the complex magnitude squared of the Fourier transforms of the surface's frequency transmission function and is independent of the properties of the incident wave ensemble. Some parallels can be drawn between this conclusion and known results in optics regarding transmission of mutual intensities through an optical system [8]. See Figs. 3.3 to 3.6 for rendered BRDFs visualizing the scattering of some specific surfaces. Likewise, see Fig. 3.7 for a comparison with Yan et al. [60].

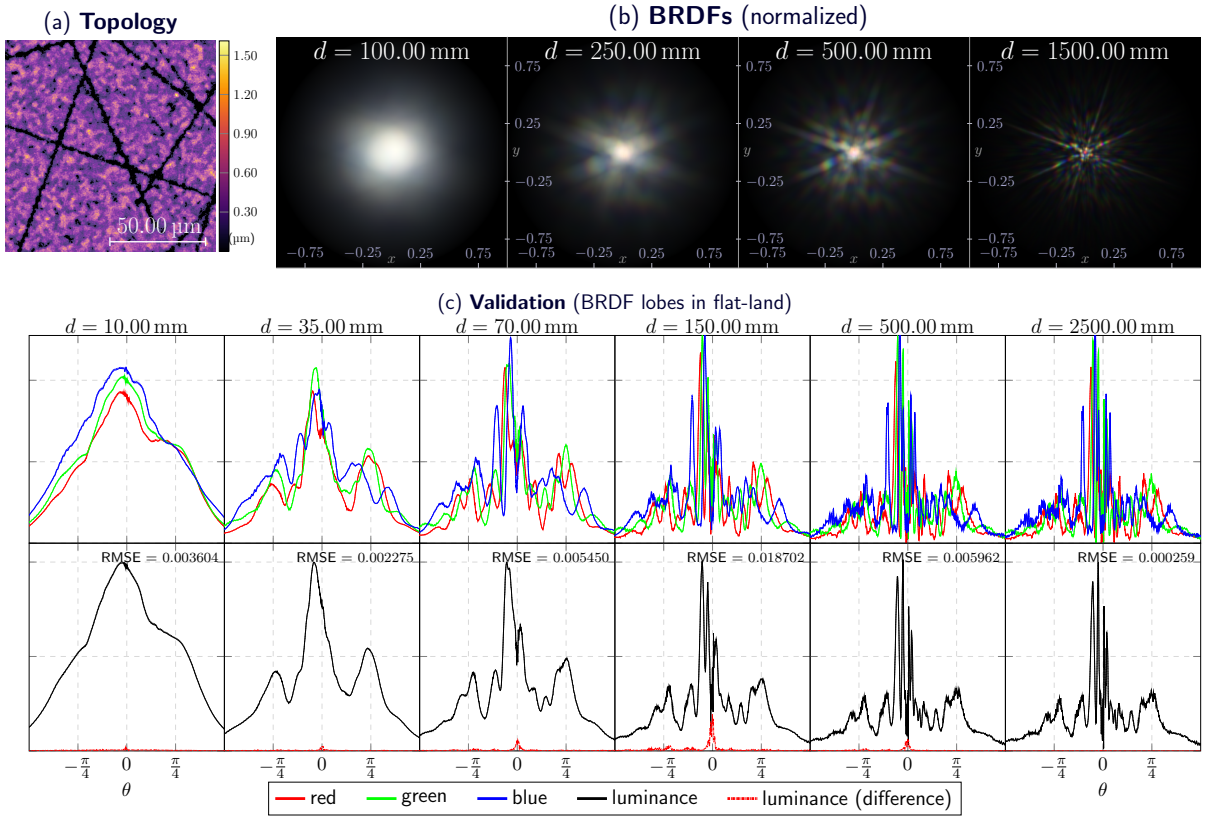


Figure 3.3: (b) BRDF visualizations (normalized intensity) of the diffraction effects that arise when natural, partially-coherent light scatters of a surface were rendered using our method. A uniformly scratched aluminum surface (topography visualized in figure (a)) is illuminated by a D65 light source of radius 8.00 mm that is positioned directly above the surface. By varying the distance of the source from the surface we change the coherence properties of the incident wave ensemble, resulting in starkly different scattering behaviour. (c) Numeric validation of our method (solid plots) against the ground truth (dashed plots) in flat-land for the same scratched surface. The radius of the light source is 1.00 mm and the ground truth is generated using 49 087.00 elementary radiators. We compare (first row) each primary colour component as well as the (second row) total luminance. Absolute luminance difference is plotted as well (second row, dash-dotted red plot).

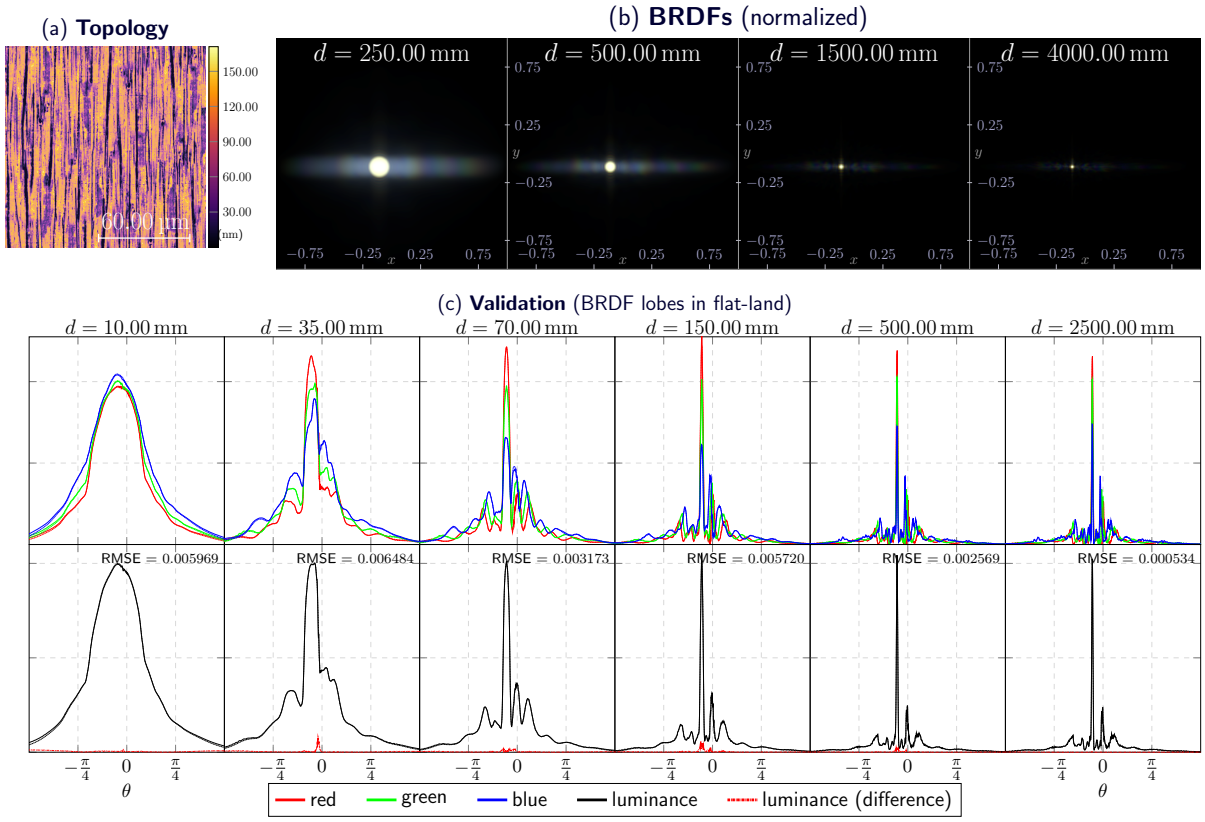


Figure 3.4: (a) An aluminum surface that has undergone electric polish resulting in low intrinsic roughness and highly anisotropic details. (b) BRDF visualizations (normalized intensity) under partially-coherent illumination produced by a fluorescent F1 model light source. The anisotropic surface features give rise to a strongly directional scattering behaviour. (c) Validation in flat-land. For validation the radius of the light source is set to 1.00 mm and the ground truth is generated using 49 087.00 elementary radiators.

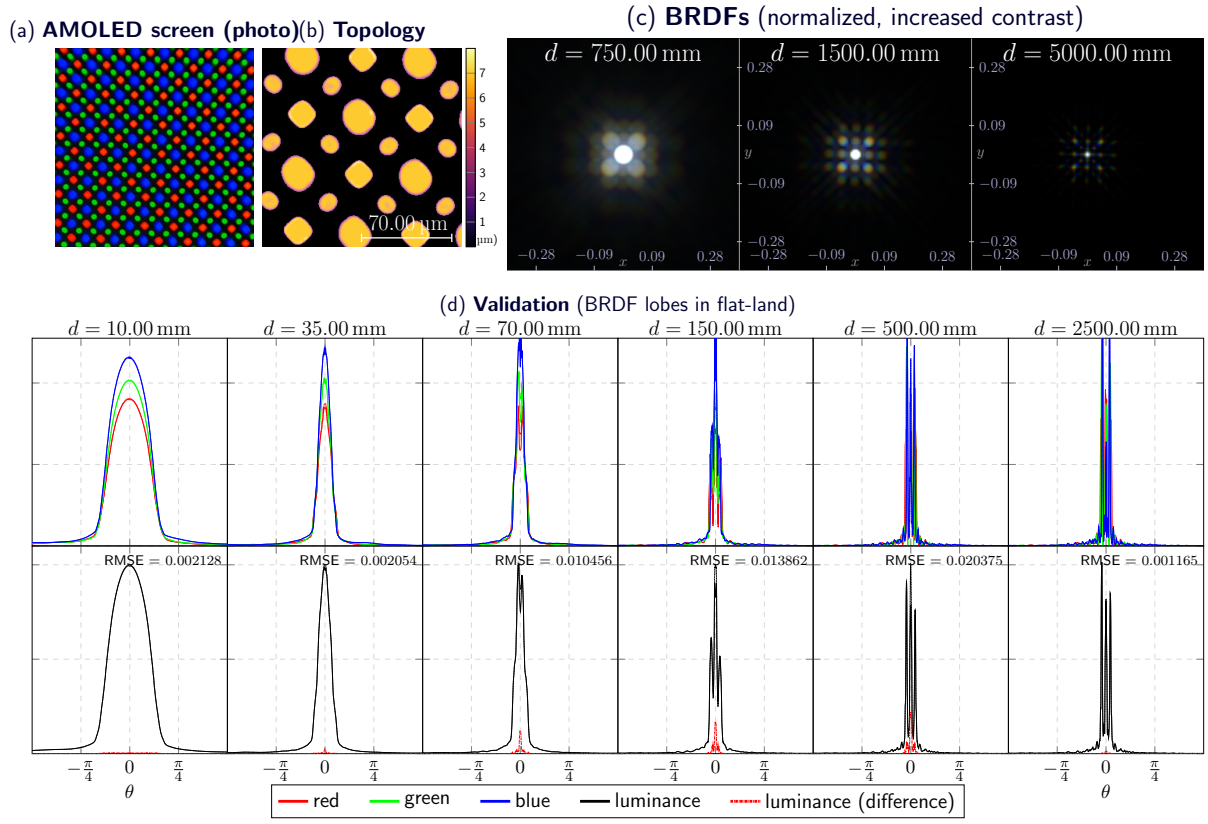


Figure 3.5: (a) A microscopy photograph of an AMOLED screen with a diamond pitch pixel arrangement (used in some hand-held devices), with its (b) surface topology visualized. That topology admits an interesting 8-way symmetry in the induced diffraction patterns, which can be seen in the (c) rendered BRDFs (zoomed-in and with artificially increased contrast for visualization purposes). We used an F1 fluorescent light source with a radius of 8.00 mm positioned directly above the surface for the rendered BRDFs. The distances between the constructive-interference peaks is consistent with Bragg's diffraction law. (d) Validation in flat-land. For validation the radius of the light source is set to 1.00 mm and the ground truth is generated using 49 087.00 elementary radiators.

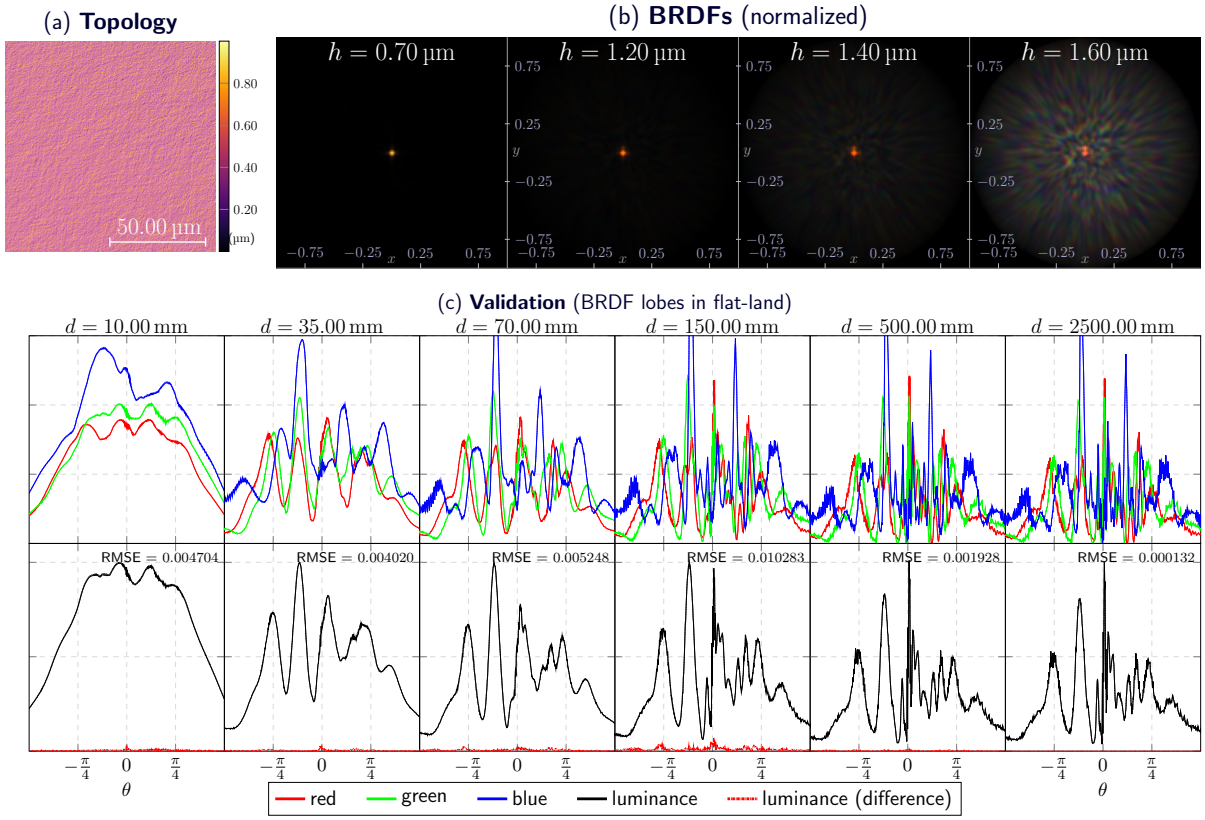


Figure 3.6: (a) A sandpaper polished gold surface with isotropic statistics. (b) BRDF visualizations (normalized intensity) generated when the surface is illuminated by a D65 light source. We vary the height-scale of the surface, h, to compare the BRDFs between different levels of roughness: For a very smooth surface (on the left), the only visible details are the reflection of the light source, however, as we increase the roughness (left to right), surface diffractions effects appear isotropically. (c) Validation in flat-land. For validation the radius of the light source is set to 1.00 mm and the ground truth is generated using 49 087.00 elementary radiators.

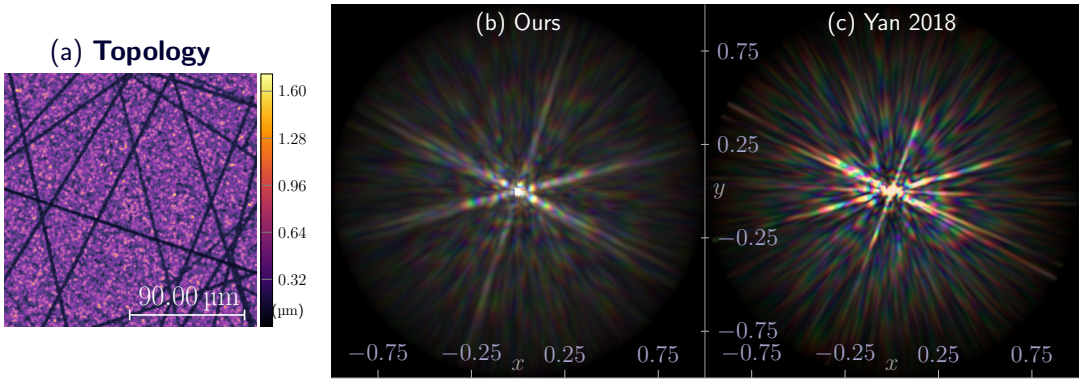


Figure 3.7: Comparison of (b) our method against (c) Yan et al. [60]. Both methods were employed on the (a) same surface topology. We used the implementation provided by Yan et al. [60] to render the BRDF produced by their method. Some differences in colour temperature and white point are expected, however the produced patterns are similar. Both BRDFs were rendered with 32 spectral samples. The methods admit comparable rendering times.

Chapter 4

Light-Matter Interactions

In this chapter we focus on quantifying matter's scattering characteristics, as well as deriving computationally tractable formulae for the interaction of partially-coherent light with matter. While in Section 3.7 a simple example of light-matter interaction was discussed, in this chapter general theorems and formulae are introduced.

4.1 The Jones Calculus and Generalized Mueller Calculus

Given a matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ with rows $\vec{\boldsymbol{a}}_j$, let the *stack operator* acting on $\boldsymbol{A}$ be denoted as $\boldsymbol{A}^{\mathscr{S}}$. That is, $\boldsymbol{A}^{\mathscr{S}} = [\vec{\boldsymbol{a}}_1^{\mathsf{T}}, \vec{\boldsymbol{a}}_2^{\mathsf{T}}, \ldots, \vec{\boldsymbol{a}}_n^{\mathsf{T}}]^{\mathsf{T}}$ is the vectorized from of the matrix $\boldsymbol{A}$: the column vector that arises by stacking the rows of $\boldsymbol{A}$.

We mention a well-known property of the Kronecker product:

Lemma 4.1.0.1. $\boldsymbol{A}\boldsymbol{D}\boldsymbol{B}^{\mathsf{T}} = \boldsymbol{C}$ *if and only if* $(\boldsymbol{B}\otimes\boldsymbol{A})\boldsymbol{D}^{\mathscr{S}} = \boldsymbol{C}^{\mathscr{S}}$.

Proof. See [31]. □

Jones calculus The Jones calculus [39] is a simple and useful formal method, where the transverse components of an electromagnetic field are written as a 2-element complex vector—the *Jones vector*—and the actions of linear optical elements are quantified via 2×2 matrices—the *Jones matrices*—that act upon these vectors. As before, let $\vec{\boldsymbol{E}}(\vec{\boldsymbol{k}}, t)$ quantify the spatiotemporal electric disturbances of a plane wave, with wavevector $\vec{\boldsymbol{k}}$. Under a given transverse basis $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$, the transverse components of this plane wave are written as the Jones vector $\vec{\boldsymbol{j}}(\vec{\boldsymbol{k}}, t) = [\hat{\boldsymbol{x}} \cdot \vec{\boldsymbol{E}},\ \hat{\boldsymbol{y}} \cdot \vec{\boldsymbol{E}}]^{\intercal}$. Let that plane wave be incident upon a particle, whose scattering characteristics are quantified by the Jones matrix $\boldsymbol{J}$. The scattered plane wave is then $\vec{\boldsymbol{j}}' = \boldsymbol{J}\vec{\boldsymbol{j}}$. Jones matrices can be complex, quantifying the phase-shifts induced by conductive particles. This calculus deals directly with the underlying electric field, and may only quantify interactions with perfectly-coherent, perfectly-polarized light.

Generalized Mueller calculus In order to derive a coherence-aware linear calculus, Korotkova [26] generalize the classical *Mueller-Stokes calculus* into a two-point formalism.

Recall that the *cross-spectral density matrix* (defined in Eq. (2.150)) is a matrix composed of the different cross-spectral densities $\mathcal{C}_{\alpha\beta}$, with $\alpha\beta \in \{x, y\}$, relating the mutual coherence between different transverse components, and may be written as the outer product:

$$\boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2\ ;\ \omega\right) \triangleq \left\langle \vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_1)\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_2)^{\dagger} \right\rangle_{\omega} = \begin{pmatrix} \mathcal{C}_{xx} & \mathcal{C}_{xy} \\ \mathcal{C}_{yx} & \mathcal{C}_{yy} \end{pmatrix}. \tag{4.1}$$

And note again that we drop the dependence on time from the notation, as the ensemble-average effectively discards time. The *generalized Stokes parameters* (gSP) vectors describes the same information as the CSD matrix above, but in a different (sometimes more convenient) analytic form. The gSP, as well as its radiance-carrying counterpart,

can be defined as:

$$\overset{\leftrightarrow}{\boldsymbol{\mathcal{S}}} \triangleq \boldsymbol{A}_{\text{mül}}\, \boldsymbol{\mathcal{C}}^{\mathscr{S}} \ , \tag{4.2}$$

$$\overset{\leftrightarrow}{\boldsymbol{\mathcal{L}}} \triangleq \boldsymbol{A}_{\text{mül}}\, \boldsymbol{\mathcal{L}}^{\mathscr{S}} \ , \tag{4.3}$$

where the $\boldsymbol{A}_{\text{mül}}$, defined by Savenkov [39, Eq. 3.12], is

$$\boldsymbol{A}_{\text{mül}} \triangleq \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -\mathrm{i} & \mathrm{i} & 0 \end{pmatrix} . \tag{4.4}$$

Let the local, spatially-varying scattering characteristics of a medium be quantified by the Jones matrix $\boldsymbol{J}(\vec{\boldsymbol{p}})$, with $\vec{\boldsymbol{p}}$ being a point in the medium. Apply $\boldsymbol{J}$ to the electric field $\vec{\boldsymbol{E}}$, and substitute into Eq. (4.1), yielding the CSD matrix of the radiation after interaction, viz.

$$\begin{aligned} \boldsymbol{\mathcal{C}}_{(\mathrm{o})}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \ ; \ \omega\right) =& \left\langle \left[\boldsymbol{J}(\vec{\boldsymbol{r}}_1)\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_1)\right]\left[\boldsymbol{J}(\vec{\boldsymbol{r}}_2)\vec{\boldsymbol{E}}(\vec{\boldsymbol{r}}_2)\right]^{\dagger} \right\rangle_{\omega} \\ =& \boldsymbol{J}(\vec{\boldsymbol{r}}_1)\, \boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \ ; \ \omega\right) \boldsymbol{J}(\vec{\boldsymbol{r}}_2)^{\dagger} \ . \end{aligned} \tag{4.5}$$

The above relates the CSD matrices of the incident radiation and radiation immediately after interaction (but before propagation), $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{C}}'$, respectively. Using Lemma 4.1.0.1 we deduce

$$\boldsymbol{\mathcal{C}}_{(\mathrm{o})}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \ ; \ \omega\right)^{\mathscr{S}} = \left[\boldsymbol{J}(\vec{\boldsymbol{r}}_2)^{\star} \otimes \boldsymbol{J}(\vec{\boldsymbol{r}}_1)\right] \boldsymbol{\mathcal{C}}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \ ; \ \omega\right)^{\mathscr{S}} \ , \tag{4.6}$$

and, finally, by the definition of generalized Stokes parameters, Eq. (4.2):

$$\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{o})} = \boldsymbol{A}_{\mathrm{m\ddot{u}l}}[\boldsymbol{J}(\vec{\boldsymbol{r}}_2)^{\star} \otimes \boldsymbol{J}(\vec{\boldsymbol{r}}_1)]\boldsymbol{A}_{\mathrm{m\ddot{u}l}}^{-1} \overleftrightarrow{\boldsymbol{\mathcal{S}}} \,, \tag{4.7}$$

which is a relation between the generalized Stokes parameters of the radiation before and after interaction. The quantity

$$\boldsymbol{\mathcal{M}}(\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2) \triangleq \boldsymbol{A}_{\mathrm{m\ddot{u}l}}[\boldsymbol{J}(\vec{\boldsymbol{p}}_2)^{\star} \otimes \boldsymbol{J}(\vec{\boldsymbol{p}}_1)]\boldsymbol{A}_{\mathrm{m\ddot{u}l}}^{-1} \tag{4.8}$$

is the *generalized Mueller matrix*. That is, $\boldsymbol{\mathcal{M}}(\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2)$, acting upon the generalized Stokes parameters, quantifies the same action as the pair of Jones matrices $\boldsymbol{J}(\vec{\boldsymbol{p}}_1)$ and $\boldsymbol{J}(\vec{\boldsymbol{p}}_2)$ acting upon the CSD matrix. Hence, $\boldsymbol{\mathcal{M}}$ is the equivalent, matricial form of Eq. (4.5), giving rise to the generalized Mueller calculus [26].

Discussion Note that the scattered CSD matrix $\boldsymbol{\mathcal{C}}'$ and generalized Stokes parameters vector $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{o})}$, which arise in Eqs. (4.6) and (4.7), are local quantities before propagation. These quantities carry units of intensity (spectral irradiance). Both the CSD matrix and generalized Stokes parameters describe the same information. As will become clear later, our motivation to employ a formalism based on the generalized Stokes parameters and generalized Mueller calculus, instead of working with CSD matrices directly, is due to:

1. Eq. (4.5) frustrates analytic progress because we may not take advantage of the properties of locally-stationary matter (see Eq. (4.15) later). This is because these matrices, $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{J}$, generally do not commute.

2. The classical Mueller matrix $\bar{\boldsymbol{m}}$ that arises in Eq. (4.15) is a simple, classical quantity, that is easy to understand and has seen wide usage in computer graphics.

4.2 The Light-Matter Interaction Theorem

Propagation Let $\mathbf{P} \subset \mathbb{R}^3$ be a scattering region (centred at the origin). Let $\vec{\boldsymbol{E}}$ be the monochromatic electric field that is incident upon the aperture, and let $\vec{\boldsymbol{E}}'$ be the scattered field. Using the far-field Smythe diffraction formula (Eq. (3.14)), we have previously derived a propagation formula for the CSD matrix, viz. Definition 3.6.1. We may trivially generalize that formula to diffraction by a volumetric region (under the Born first-order approximation) by extending the aperture to an arbitrary three-dimensional volume, $\mathbf{P}$. The far-field propagation formula for the CSD matrix becomes:

$$\begin{aligned}\boldsymbol{\mathcal{C}}_{(\mathrm{o})}\left(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \,;\, \omega\right) =& \frac{\cos^2\vartheta_o \mathrm{e}^{\mathrm{i}k\hat{\boldsymbol{r}}\cdot(\vec{\boldsymbol{r}}_1-\vec{\boldsymbol{r}}_2)}}{\lambda^2 r^2}\int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_1 \int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_2 \\ &\times \mathrm{e}^{-\mathrm{i}k(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{p}}_1-\hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{p}}_2)}\boldsymbol{H}_1\,\boldsymbol{\mathcal{C}}_P\left(\vec{\boldsymbol{r}}'_1, \vec{\boldsymbol{r}}'_2 \,;\, \omega\right)\boldsymbol{H}_2^{\mathsf{T}}\,, \end{aligned} \tag{4.9}$$

which is a double three-dimensional Fourier transform with $\boldsymbol{\mathcal{C}}_P$ being the CSD matrix of the radiation at the region $\mathbf{P}$ and ϑ_o the inclination angle. $\hat{\boldsymbol{r}}$ and r are the mean propagation direction and distance. In the far field, $\hat{\boldsymbol{r}} \approx \hat{\boldsymbol{r}}_1 \approx \hat{\boldsymbol{r}}_2$ and likewise $r \approx r_1 \approx r_2$. The transformation matrices $\boldsymbol{H}_{1,2}$ transform the incident CSD matrix from its chosen transverse basis to the transverse basis of the scattered radiation, viz. (simplified version of Eq. (3.19))

$$\boldsymbol{H}_{1,2} \triangleq \begin{pmatrix} \hat{\boldsymbol{x}}'_\perp \cdot \hat{\boldsymbol{x}} & \hat{\boldsymbol{x}}'_\perp \cdot \hat{\boldsymbol{y}} \\ \hat{\boldsymbol{y}}'_\perp \cdot \hat{\boldsymbol{x}} & \hat{\boldsymbol{y}}'_\perp \cdot \hat{\boldsymbol{y}} \end{pmatrix}\,, \tag{4.10}$$

where $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ and $\{\hat{\boldsymbol{x}}', \hat{\boldsymbol{y}}'\}$ are the given transverse bases of the incident and scattered radiation, respectively, and $\hat{\boldsymbol{x}}'_\perp, \hat{\boldsymbol{y}}'_\perp$ are the normalized projections of the scattered transverse basis onto the incident transverse plane. The above is a simple rotation matrix with angle φ, which is the angle between $\hat{\boldsymbol{x}}'_\perp$ and $\hat{\boldsymbol{x}}$ (in the singular case where $\hat{\boldsymbol{x}}'$ is

perpendicular to the incident transverse plane, we take the angle between $\hat{\boldsymbol{y}}'_{\perp}$ and $\hat{\boldsymbol{y}}$). The corresponding Mueller matrix arises immediately:

$$\boldsymbol{T} \triangleq \boldsymbol{A}_{\text{mül}}[\boldsymbol{H}_1 \otimes \boldsymbol{H}_2]\boldsymbol{A}^{-1}_{\text{mül}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\varphi & -\sin 2\varphi & 0 \\ 0 & \sin 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{4.11}$$

We may now write the general diffraction formulae Eq. (4.9) in an equivalent Stokes parameters form:

$$\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\text{o})} = \frac{\cos^2\vartheta_o \mathrm{e}^{\mathrm{i}k\hat{\boldsymbol{r}}\cdot(\vec{\boldsymbol{r}}_1-\vec{\boldsymbol{r}}_2)}}{\lambda^2 r^2} \int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_1 \int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_2 \, \mathrm{e}^{-\mathrm{i}k(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{p}}_1-\hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{p}}_2)} \boldsymbol{T} \overleftrightarrow{\boldsymbol{\mathcal{S}}}_P \, . \tag{4.12}$$

with $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\text{o})}$ and $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_P$ being the Stokes parameters vectors after propagation and at the scattering volume, respectively. Using Eq. (4.3), we may also write an equivalent propagation formula for the radiance-carrying gSP $\overleftrightarrow{\boldsymbol{\mathcal{L}}}$

$$\begin{aligned} \overleftrightarrow{\boldsymbol{\mathcal{L}}}_{(\text{o})} =& \frac{\cos^2\vartheta_o \mathrm{e}^{\mathrm{i}k\hat{\boldsymbol{r}}\cdot(\vec{\boldsymbol{r}}_1-\vec{\boldsymbol{r}}_2)}}{\lambda^2 A} \\ & \times \int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_1 \int_{\mathbf{P}} \mathrm{d}^3\vec{\boldsymbol{p}}_2 \, \mathrm{e}^{-\mathrm{i}k(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{p}}_1-\hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{p}}_2)} \boldsymbol{T} \overleftrightarrow{\boldsymbol{\mathcal{S}}}_P \, , \end{aligned} \tag{4.13}$$

where A is the projected area of $\mathbf{P}$ in direction $\hat{\boldsymbol{r}}$.

4.2.1 Interaction With Matter

Locally-stationary matter Let the scattering characteristics of the scattering region $\mathbf{P}$ be quantified by a Jones matrix $\boldsymbol{J}(\vec{\boldsymbol{p}})$, which we now be regarded as a spatial stochastic process, describing the spatially-varying polarimetric scattering potential of the medium. The respective Mueller matrix $\boldsymbol{\mathcal{M}}$ is defined via Eq. (4.8). An autocorrelation function

of that process can defined as the generalized Mueller matrix

$$\boldsymbol{\rho}_{JJ}(\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2) \triangleq \langle \boldsymbol{\mathcal{M}}(\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2) \rangle \ , \tag{4.14}$$

where the operator $\langle \cdot \rangle$ averages over the statistical ensemble of all matter realizations that conform to the statistics of the process $\boldsymbol{J}$ (not to be confused with $\langle \cdot \rangle_\omega$). The quantity $\langle \boldsymbol{\mathcal{M}} \rangle$ qualifies as an autocorrelation function, as it quantifies the ensemble-averaged mutual scattering characteristics at a pair of points.

We now restrict the stochastic process $\boldsymbol{J}$ to the class of locally-stationary processes (a less restrictive class than weak stationarity) [40]. This gives rise to *locally-stationary matter* [44]. We define the single-point function $\bar{\boldsymbol{m}}(\vec{\boldsymbol{r}}) \triangleq \langle \boldsymbol{\mathcal{M}}(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}) \rangle$, which should be understood as the ensemble-averaged, local Mueller matrix that describes the deterministic features of the matter (e.g., density and polarimetric properties of scattering particles in a medium). This quantity is *local*, in the sense that it isn't averaged over the scattering region, and *deterministic*, in the sense that it is ensemble-averaged over all matter realizations. By definition, the autocorrelation function of a weakly-stationary process takes the following form of a product of real, classical Mueller matrices:

$$\boldsymbol{\rho}_{JJ}(\vec{\boldsymbol{p}}_1, \vec{\boldsymbol{p}}_2) = \bar{\boldsymbol{m}}\left(\tfrac{\vec{\boldsymbol{p}}_1 + \vec{\boldsymbol{p}}_2}{2}\right) \boldsymbol{R}_{JJ}(\vec{\boldsymbol{p}}_1 - \vec{\boldsymbol{p}}_2) \ . \tag{4.15}$$

The Mueller matrix $\boldsymbol{R}_{JJ}$ is the *stationary autocorrelation* (a function of the difference vector $\vec{\boldsymbol{p}}_1 - \vec{\boldsymbol{p}}_2$ only) of the locally-stationary process, which describes statistical perturbations across the matter. These quantities are typically wavelength dependent.

Interaction formulae The light-matter interaction is then quantified by the matter's stochastic generalized Mueller matrix $\boldsymbol{\rho}_{JJ}$. Set

$$\overleftrightarrow{\boldsymbol{\mathcal{S}}}_P(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2 \,;\, \omega) \triangleq \boldsymbol{\rho}_{JJ}(\vec{\boldsymbol{r}}_1, \vec{\boldsymbol{r}}_2 \,;\, \omega) \overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2 \,;\, \omega) \;, \tag{4.16}$$

where $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}$ and $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_P$ are the incident (radiance-carrying) gSP before interaction with the matter and immediately after interaction, but before propagation, respectively. We also made the quasi-homogeneous assumption, where $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}$ and $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_P$ are functions of the difference $\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2$ only. The propagation is then done via the propagation formula, Eq. (4.13). Substituting Eq. (4.15) into Eq. (4.13) yields

$$\begin{aligned}\overleftrightarrow{\boldsymbol{\mathcal{L}}}_{(\mathrm{o})} =& \frac{\cos^2\vartheta_o \mathrm{e}^{\mathrm{i}k\hat{\boldsymbol{r}}\cdot(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2)}}{\lambda^2 A} \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\boldsymbol{p}}_1 \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\boldsymbol{p}}_2 \, \mathrm{e}^{-\mathrm{i}k(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{p}}_1 - \hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{p}}_2)} \\ & \times \bar{\boldsymbol{m}}\left(\tfrac{\vec{\boldsymbol{p}}_1 + \vec{\boldsymbol{p}}_2}{2}\right) \boldsymbol{R}_{JJ}(\vec{\boldsymbol{p}}_1 - \vec{\boldsymbol{p}}_2) \boldsymbol{T} \overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})} \;,\end{aligned} \tag{4.17}$$

where we denote $A = A_P / A_{(\mathrm{i})}$. Note, we are now integrating over $\mathbb{R}^3 \times \mathbb{R}^3$, the restriction to the scattering region is assumed to be performed by $\bar{\boldsymbol{m}}$, i.e. set $\bar{\boldsymbol{m}}(\vec{\boldsymbol{p}}) \equiv 0$ when $\vec{\boldsymbol{p}} \notin \mathbf{P}$. Perform the variable changes $\vec{\boldsymbol{\xi}} = \vec{\boldsymbol{p}}_1 - \vec{\boldsymbol{p}}_2$ and $\vec{\boldsymbol{\zeta}} = \frac{1}{2}(\vec{\boldsymbol{p}}_1 + \vec{\boldsymbol{p}}_2)$. Then, the quantities that appear in the integrand above can be rewritten as follows:

$$\mathrm{e}^{-\mathrm{i}k(\hat{\boldsymbol{r}}_1\cdot\vec{\boldsymbol{p}}_1 - \hat{\boldsymbol{r}}_2\cdot\vec{\boldsymbol{p}}_2)} = \mathrm{e}^{-\mathrm{i}k\frac{1}{2}\vec{\boldsymbol{\xi}}\cdot(\hat{\boldsymbol{r}}_1 + \hat{\boldsymbol{r}}_2)} \mathrm{e}^{-\mathrm{i}k\vec{\boldsymbol{\zeta}}\cdot(\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2)} \;, \tag{4.18}$$

$$\bar{\boldsymbol{m}}\left(\tfrac{\vec{\boldsymbol{p}}_1 + \vec{\boldsymbol{p}}_2}{2}\right) = \bar{\boldsymbol{m}}\left(\vec{\boldsymbol{\zeta}}\right) \;, \tag{4.19}$$

$$\boldsymbol{R}_{JJ}(\vec{\boldsymbol{p}}_1 - \vec{\boldsymbol{p}}_2) = \boldsymbol{R}_{JJ}\left(\vec{\boldsymbol{\xi}}\right) \;, \tag{4.20}$$

and recall that $\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}$ is a function of $\vec{\boldsymbol{\xi}}$ only. This serves to decouple the double integral in Eq. (4.17) into separate spatial integrals for each integration variable, $\vec{\boldsymbol{\zeta}}$ and $\vec{\boldsymbol{\xi}}$.

The first integral becomes the following Fourier transform, which can be understood

as the *angular spectrum function*:

$$\begin{aligned}\widetilde{\boldsymbol{M}}(k\hat{\boldsymbol{r}}_1 - k\hat{\boldsymbol{r}}_2) \triangleq& \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\boldsymbol{\zeta}}\, \mathrm{e}^{-\mathrm{i}k\vec{\boldsymbol{\zeta}}\cdot(\hat{\boldsymbol{r}}_1-\hat{\boldsymbol{r}}_2)}\bar{\boldsymbol{m}}\Big(\vec{\boldsymbol{\zeta}}\Big)\\ =&(2\pi)^{\frac{3}{2}}\,\mathscr{F}\{\bar{\boldsymbol{m}}\}(k\hat{\boldsymbol{r}}_1 - k\hat{\boldsymbol{r}}_2)\ .\end{aligned} \tag{4.21}$$

As $\hat{\boldsymbol{r}}_1 \approx \hat{\boldsymbol{r}}_2$ and by the far-field assumption $\bar{\boldsymbol{m}}$ changes slowly as a function of $\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}}$, $\widetilde{\boldsymbol{M}}(0)$ is a very good approximation to that integral. The second integral is a Fourier transform of the product of the matter's stationary autocorrelation function with the incident gSP, viz.

$$\vec{\boldsymbol{f}}\big(k\tfrac{\hat{\boldsymbol{r}}_1+\hat{\boldsymbol{r}}_2}{2}\big) \triangleq \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\boldsymbol{\xi}}\, \mathrm{e}^{-\mathrm{i}\frac{k}{2}\vec{\boldsymbol{\xi}}\cdot(\hat{\boldsymbol{r}}_1+\hat{\boldsymbol{r}}_2)}\boldsymbol{R}_{JJ}\Big(\vec{\boldsymbol{\xi}}\Big)\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}\Big(\vec{\boldsymbol{\xi}}\Big)\ . \tag{4.22}$$

Define $\mathcal{P} \triangleq \mathscr{F}\{\boldsymbol{R}_{JJ}\}$, the Fourier transform of the stationary autocorrelation function, i.e. the *stationary power spectral density*. We also define the classical *pBSDF* (polarimetric BSDF) as the averaged (over the entire scattering region) Mueller matrix, viz.

$$\boldsymbol{M} \triangleq \frac{1}{|P|}\int_P \bar{\boldsymbol{m}} = \frac{1}{|P|}\widetilde{\boldsymbol{M}}(0)\ . \tag{4.23}$$

Applying the convolution theorem (Theorem 2.2.1.2) and the approximation $\frac{1}{2}(\hat{\boldsymbol{r}}_1+\hat{\boldsymbol{r}}_2) \approx \hat{\boldsymbol{r}}$ to Eq. (4.22), and simplifying, yields:

Theorem 4.2.1.1 (Interaction with Locally-Stationary Matter Theorem)**.** *The scattered radiance-carrying gSP that arise on scattering by locally-stationary matter is*

$$\overleftrightarrow{\boldsymbol{\mathcal{L}}}_{(\mathrm{o})} = \frac{\cos\vartheta_o \mathrm{e}^{\mathrm{i}k\hat{\boldsymbol{r}}\cdot(\vec{\boldsymbol{r}}_1-\vec{\boldsymbol{r}}_2)}}{\lambda^2(2\pi)^{\frac{3}{2}}}\boldsymbol{M}\boldsymbol{T}\Big(\mathcal{P} * \mathscr{F}\Big\{\overleftrightarrow{\boldsymbol{\mathcal{S}}}_{(\mathrm{i})}\Big\}\Big)\Big(\vec{\boldsymbol{h}}\Big)\vec{\boldsymbol{S}}^{(\mathrm{i})}\ , \tag{4.24}$$

with the shorthand $\vec{\boldsymbol{h}} \triangleq k(\hat{\boldsymbol{r}} + \hat{\boldsymbol{s}})$.

The most important result in this chapter is the introduction of locally-stationary matter, which allows the formulation of tractable interaction formulae. The stationary power spectral density (PSD) function $\mathcal{P}$ and the classical pBSDF fully quantify the matter's scattering properties. As the PSD $\mathcal{P}$ is real, the convolution that appears in Theorem 4.2.1.1 is real as well, and this convolution quantifies the wave-interference effects. The classical Mueller matrix $\boldsymbol{M}$ acts upon the classical incident Stokes parameters vector (incident irradiance). The resulting $\overleftrightarrow{\boldsymbol{\mathcal{L}}}_{(\mathrm{o})}$ is a (radiance-carrying) classical Stokes parameters vector, that is constant up to the propagator term.

4.2.2 Example Materials

We discuss some example material pBSDFs.

Surfaces with varying-degrees of roughness

Let a wide-sense-stationary statistical surface be described by the Gaussian auto-correlation function $\mathcal{R}(\vec{\boldsymbol{d}}) = q^2 \mathrm{g}^{\sigma_s^2}(\vec{\boldsymbol{d}})$, with $\vec{\boldsymbol{d}} \in \mathbb{R}^2$ a distance on the surface, q then becomes the *root-mean-square (rms) roughness* and σ_s is the surface's *correlation length*. The PSD of the surface arises immediately:

$$\mathcal{P}\left(\vec{\boldsymbol{\zeta}}\right) = \mathscr{F}\left\{\mathcal{R}\right\}\left(\vec{\boldsymbol{\zeta}}\right) = q^2 \sigma_s^2 \mathrm{g}^{\sigma_s^{-2}}\left(\vec{\boldsymbol{\zeta}}\right) \ . \tag{4.25}$$

The scattering pBSDF that enters Theorem 4.2.1.1 for the surface is the Mueller form

of the Fresnel relations, viz.:

$$\boldsymbol{M} \triangleq \frac{1}{\lambda^2} \begin{pmatrix} m_{00} & m_{01} & & \\ m_{01} & m_{00} & & \\ & & m_{22} & m_{23} \\ & & -m_{23} & m_{22} \end{pmatrix} \boldsymbol{T} \ ,$$

$$\text{with} \qquad m_{00} = \frac{|\mathfrak{f}_{pp}|^2 + |\mathfrak{f}_{ss}|^2}{2} \ , \qquad m_{01} = \frac{|\mathfrak{f}_{pp}|^2 - |\mathfrak{f}_{ss}|^2}{2} \ ,$$

$$m_{22} = \mathrm{Re}\left\{\mathfrak{f}_{ss}\mathfrak{f}_{pp}^{\star}\right\} \quad \text{and} \quad m_{23} = \mathrm{Im}\left\{\mathfrak{f}_{ss}\mathfrak{f}_{pp}^{\star}\right\} \ . \tag{4.26}$$

The exit reference frame is defined as the standard s- and p-polarization frame: $\hat{\boldsymbol{x}}$ is perpendicular to both $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{y}}$ lies in the plane spanned by $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{s}}$; $\boldsymbol{T}$ is the Mueller rotation matrix that rotates from the (arbitrary) incident frame to the sp frame. The quantities $\mathfrak{f}_{ss}, \mathfrak{f}_{pp}$ are the well-known reflection or refraction Fresnel coefficients, which depend on the relative normal vector (the geometric half vector between $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{s}}$ in the case of reflective scattering) and on the relative ratio of the refractive indices of the media, denoted η (which might be complex and is typically wavelength-dependent).

This surface model is parametrized by the roughness q, the correlation length σ_s^2 (which, in term of optical response, inversely modulates the length of the "tail" of the highlight) as well as the wavelength-dependent refractive index η. This model is fully coherence-aware, and is easy to importance sample in a bi-directional path tracer. Nevertheless, real surfaces are rarely Gaussian, and typically their PSD admits a more "fractal" structure with a longer tail (more pronounced higher spatial frequencies). Deriving expressions for more realistic surface models is possible (as effectively the problem reduces to deriving closed-form expressions for the well-behaved diffraction operator Definition 3.6.1 and importance sampling it), and is left for future work.

Perfectly-diffuse surfaces

A perfectly-diffuse surface is a special case of the rough surface discussed above, and can be understood at the infinite roughness and zero correlation length limit, viz. $q \to \infty$ and $\sigma_s \to 0$ simultaneously. Then, Eq. (4.25) reduce to $\mathcal{P} = 1$. The scattering pBSDF of a perfectly-diffuse surface is

$$\boldsymbol{M} \triangleq \frac{m_{00}}{\pi\lambda^2}\boldsymbol{T} \ , \tag{4.27}$$

where $0 \leq m_{00} \leq 1$ is the mean scattering intensity, $\boldsymbol{T}$ is the Mueller rotation matrix, as before, and we normalize by $1/\pi$.

Diffraction grating

We use a similar formulation to Steinberg and Yan [44]. Consider a (perfect) sinusoidal phase grating, with pitch (period) Λ and grating height of b, residing on an infinite plane. Assume that the reflectivity is constant over the entire surface, hence the pBSDF is the same Fresnel Mueller matrix as in Eq. (4.26). The complex transmission function for a plane wave incident at angle ϑ_i is well-known [8]:

$$\phi(x) \triangleq \exp\left[\mathrm{i}k\frac{b}{2\cos\vartheta_i\cos\vartheta_o}\sin\left(2\pi\frac{x}{\Lambda}\right)\right] \ , \tag{4.28}$$

which quantifies the phase difference of the scattered wave due to the height differences across the grating, with x being the position on the grating. Applying the well-known identity of the Bessel function of the first kind, viz. $\mathrm{e}^{\mathrm{i}\upsilon\sin\varphi} = \sum_{n=-\infty}^{\infty} J_n(\upsilon)\mathrm{e}^{\mathrm{i}n\varphi}$, serves to rewrite the transmission function ϕ as a Fourier series with coefficients $J_n(\upsilon)$. The PSD

of a Fourier series is a series of Dirac deltas:

$$\mathcal{P}(\zeta) = \sum_{n=-\infty}^{\infty} J_n\left(k\tfrac{b}{2\cos\vartheta_i\cos\vartheta_o}\right)^2 \delta\left(\zeta - 2\pi\frac{n}{\Lambda}\right) . \tag{4.29}$$

Evaluated at $\zeta = k(\sin\vartheta_o + \sin\vartheta_i)$, i.e. the projection of $\vec{\boldsymbol{h}}$ onto the grating axis, the PSD of the grating above reproduces the known relation [8]:

$$\sin\vartheta_o + \sin\vartheta_i = \frac{n}{\Lambda}\lambda , \tag{4.30}$$

which governs the scattering angle of lobe n at wavenumber k. The intensity of each lobe is given by the Bessel function in Eq. (4.29).

Clearly, the DC lobe $n = 0$ (the direct, non-diffracted lobe) scatters at $\vartheta_i = \vartheta_o$, regardless of wavelength. The diffraction lobes, $n \neq 0$, scatter at an angle that is wavelength dependent. The convolution of the PSD with the coherence function, viz. Theorem 4.2.1.1, is then a series of these coherence functions, shifted to each respective lobe. It is important to note that the lobes generally overlap.

With moderately-coherent light, like sunlight, these lobe will be very sharp. Indeed, a diffraction grating describes a very difficult material to render and importance sample: A multitude of Dirac lobes, with each restricted to a single wavelength and carrying only a small fraction of energy.

4.3 Conclusion

The presented formulation of light-matter interactions generalizes classical, ray-optical BSDFs to wave-optical pBSDFs that quantify partially-coherent interactions. Theorem 4.2.1.1 describes the interaction via an energy-conserving, non-negative linear in-

tegral transform (convolution), retaining some of the analytic properties of the classical BSDF. It is easy to see, however, that Helmholtz reciprocity is lost due to the convolution, as illustrated in Section 3.6.

As a locally-stationary matter model may describe both statistical and deterministic scattering characteristics, essentially any material whose optical response can be explained electrodynamically can be modelled. The primary difficulty is then deriving an expression for the PSD $\mathcal{P}$, as well as the convolution of that PSD, as it appears in Theorem 4.2.1.1.

Chapter 5

Sampling via Generalized Rays

The previous chapters have outlined how to quantify the partial coherence of light, how to source and measure (sense) such light, how it propagates and diffracts. We also presented our locally-stationary matter—a description of matter that is amenable to analytic and numeric tools when describing light using its coherence functions, specifically intensity- or radiance-based cross-spectral density functions. The *sampling problem* (discussed in Chapter 1) inevitably arises: as the coherence properties of light are unknown when tracing partially-coherent light *backwards* (from the sensor to the light sources), practical application of our preceding theory is difficult. We will now resume our discussion of phase-space optics (Section 2.7), and formulate the sampling of the wave-optical phase space via *generalized rays*. We will show that partially-coherent light is a *ray bundle*: a collection of such rays; thereby making the connection to optical coherent theory, as well as all the tools that were developed over the previous chapters.

5.1 Wave-Optics Light Transport

The wave-optical phase space that arises via the *Wigner distribution function* (Eq. (2.193)) admits attractive properties: it facilitates performing phase-space queries in a manner similar to classical ray optics, and these "rays" transform inline with Liouville's equations for ABCD optical systems. However, the WDF is not non-negative, frustrating its interpretation as an energy density. Furthermore, the WDF tends to be highly oscillatory: a consequence of the Fourier-like relation in the definition of the WDF (Eq. (2.193)). As an example, consider a sample signal Φ, and its WDF $\mathscr{W}_\Phi$, and let a wave function be composed of two spatially- and frequency-shifted copies of this signal:

$$\psi(\vec{\boldsymbol{q}}) = \Phi(\vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_1)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{k}}_1\cdot\vec{\boldsymbol{q}}} + \Phi(\vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_2)\mathrm{e}^{\mathrm{i}\vec{\boldsymbol{k}}_2\cdot\vec{\boldsymbol{q}}}\,, \tag{5.1}$$

where $\vec{\boldsymbol{q}}_{1,2}$ and $\vec{\boldsymbol{k}}_{1,2}$ are the spatial and frequency shifts, respectively. Using the shift properties of the WDF (Property (IV)), the WDF of the wave function above is trivially:

$$\begin{aligned}\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right) =& \mathscr{W}_\Phi\left(\vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_1, \vec{\boldsymbol{k}} - \vec{\boldsymbol{k}}_1\right) + \mathscr{W}_\Phi\left(\vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_2, \vec{\boldsymbol{k}} - \vec{\boldsymbol{k}}_2\right) \\ &+ 2\,\mathrm{Re}\left[\mathrm{e}^{\mathrm{i}\left(\vec{\boldsymbol{k}}_1 - \vec{\boldsymbol{k}}_2\right)\cdot\vec{\boldsymbol{q}} - \mathrm{i}(\vec{\boldsymbol{q}}_1 - \vec{\boldsymbol{q}}_2)\cdot\vec{\boldsymbol{k}}'}\right]\mathscr{W}_\Phi\left(\vec{\boldsymbol{q}}', \vec{\boldsymbol{k}}'\right)\,,\end{aligned} \tag{5.2}$$

with the shorthands $\vec{\boldsymbol{q}}' = \vec{\boldsymbol{q}} - \frac{1}{2}(\vec{\boldsymbol{q}}_1 + \vec{\boldsymbol{q}}_2)$ and $\vec{\boldsymbol{k}}' = \vec{\boldsymbol{k}} - \frac{1}{2}(\vec{\boldsymbol{k}}_1 + \vec{\boldsymbol{k}}_2)$. Note the complex exponent in the cross-term above: it is a heavily oscillatory term at optical frequencies ($k \gg 0$), with frequencies that grow greater as the separation in phase space between the two Φ signals increases. Hence, if light is composed of multiple partially-coherent components, as these propagate and their separation increases, the WDF becomes increasingly oscillatory.

The WDF as a "generalized radiance" As a brief aside, we note that the WDF was used to derive wave-optical radiometric quantities, in particular the radiance, first by Walther [53]. Other definitions of such *generalized radiances* have been proposed, usually using other Cohen class joint space-frequency representations. A generalized radiance—in the form of the WDF—was also used in computer graphics to propagate partially-coherent fields.

However, it was shown that no such generalized radiance fulfils all the expected postulates: for example, it is not non-negative, or isn't conserved on non-paraxial propagation, or it is not a faithful representation of the signal (e.g., Property (II) does not hold). In general, such representations serve only as quasi-probability distributions. Furthermore, being "wasteful" representations (are of double the dimensionality of the represented signal), in practical applications only a restricted parametrized class of functions are used. But under this constraint, there is no value to using the WDF as opposed to the cross-spectral density of light directly (which then is parameterized by the Fourier-conjugated class of functions). Whichever representation of light we chose to use, once we decide to quantify partially-coherent fields explicitly, we always suffer from the "sampling problem" (see Steinberg et al. [46] and Chapter 1), where backward path tracing is difficult, as importance sampling light-matter interactions require information about the coherence of light.

Instead of using the WDF as a descriptor of light, we are interested in the phase space that arises via the WDF. We would like to find a wave-optical analogue of the classical ray, formally discuss when the wave-optical phase space can adequately sampled via such "rays", and analyze the dynamics of these "rays". These rays facilitate a coherent-mode decomposition of light, allowing us to reason about the partially-coherent light that is of primary interest for us in rendering in a "classical" manner.

5.1.1 Gaussian Beams as Rays

The Husimi Q representation To combat the unattractive cross-terms that arise in the WDF on superposition of waves, the WDF can be convolved in phase space with a kernel function (a *Cohen kernel*), masking out the interference terms and producing a different representation. It can be shown that a convolution with a multivariate Gaussian, with position and frequency variances satisfying the uncertainty relation (Eq. (2.188)), produces a representation that is strictly non-negative. The resulting distribution is known as the *Husimi Q distribution*:

$$\mathscr{Q}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\right) \triangleq \frac{1}{(\sqrt{2}\pi)^3}\int \mathrm{d}^3\vec{\boldsymbol{q}}'\,\mathrm{d}^3\vec{\boldsymbol{k}}'\,\mathscr{W}\left(\vec{\boldsymbol{q}}',\vec{\boldsymbol{k}}'\right)\mathrm{e}^{-\frac{1}{2}\vec{\boldsymbol{u}}'^{\intercal}\boldsymbol{\Sigma}^{-1}\vec{\boldsymbol{u}}'}\ , \tag{5.3}$$
$$\text{with}\quad \vec{\boldsymbol{u}}' \triangleq \begin{pmatrix}\vec{\boldsymbol{q}}-\vec{\boldsymbol{q}}'\\ \vec{\boldsymbol{k}}-\vec{\boldsymbol{k}}'\end{pmatrix} \qquad \text{and} \qquad |\boldsymbol{\Sigma}| = \frac{1}{2^3}\ ,$$

where $\boldsymbol{\Sigma}$ is any positive-definite covariance matrix of the Gaussian low-pass filter that fulfils the above.

A wave-optical "ray" Consider the WDF that takes the form of a Dirac delta in phase space, viz. $\mathscr{W}_{\mathfrak{r}} = \delta^3(\vec{\boldsymbol{q}}-\vec{\boldsymbol{q}}_0)\,\delta^3(\vec{\boldsymbol{k}}-\vec{\boldsymbol{k}}_0)$, which is an aphysical construct that represents an idealised "ray" at position $\vec{\boldsymbol{q}}_0$ with momentum $\vec{\boldsymbol{p}}_0 = \hbar\vec{\boldsymbol{k}}_0$. We stress that such a WDF is fictitious: it cannot arise from any physically-realizable wave function. However, its corresponding Husimi Q representation $\mathscr{Q}_{\mathfrak{r}}$, that arises from $\mathscr{W}_{\mathfrak{r}}$ via Eq. (5.3), is physical. Let the covariance take the block-diagonal form $\boldsymbol{\Sigma} = \mathrm{diag}\{\boldsymbol{\Sigma}_q, \boldsymbol{\Sigma}_k\}$, then:

$$\mathscr{Q}_{\mathfrak{r}}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}\ ;\ t_0\right) = \frac{1}{(\sqrt{2}\pi)^3}\mathrm{e}^{-\frac{1}{2}\vec{\boldsymbol{q}}'^{\intercal}\boldsymbol{\Sigma}_q^{-1}\vec{\boldsymbol{q}}'-\frac{1}{2}\vec{\boldsymbol{k}}'^{\intercal}\boldsymbol{\Sigma}_k^{-1}\vec{\boldsymbol{k}}'}\ , \tag{5.4}$$

which represent the phase space "picture" of the ray at an initial time t_0 of the system evolution. The shorthands $\vec{\boldsymbol{q}}' = \vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_0$ and $\vec{\boldsymbol{k}}' = \vec{\boldsymbol{k}} - \vec{\boldsymbol{k}}_0$ are the shifted coordinates. Clearly, this system is fully defined by its first 2 moments: the mean $\bar{\boldsymbol{u}}(t_0) = \left(\vec{\boldsymbol{q}}_0, \vec{\boldsymbol{k}}_0\right)$, and the moment matrix (Eq. (2.211)) $\boldsymbol{M}(t_0) = \boldsymbol{\Sigma}$. The time evolution follows Eq. (2.213):

$$\bar{\boldsymbol{u}}(t) = \boldsymbol{\mathcal{T}}(t, t')\bar{\boldsymbol{u}}(t') \ , \qquad \text{and} \tag{5.5}$$

$$\boldsymbol{M}(t) = \boldsymbol{\mathcal{T}}(t, t')\boldsymbol{M}(t')\boldsymbol{\mathcal{T}}(t, t')^{\mathsf{T}} \ , \tag{5.6}$$

for $t \geq t'$. For example, on propagation in a medium with a constant refractive-index η, viz.

$$\boldsymbol{\mathcal{T}}_{\text{propagation}}(t, t') = \begin{pmatrix} 1 & \frac{t-t'}{\eta}\hbar c \\ 0 & 1 \end{pmatrix} \ , \tag{5.7}$$

(elements represent 3×3 matrices) which represents a phase space horizontal shear, with c being the speed of light. Hence, the evolution effectively constitutes propagating the centre-of-mass in phase space (mean) in direction $\vec{\boldsymbol{k}}_0$ and spreading the spatial Gaussian footprint w.r.t. the variance in frequency.

Substitute the "smoothed WDF" $\mathscr{Q}_{\mathfrak{r}}$ of an idealised ray (Eq. (5.4)) into the definition of the WDF (Eq. (2.193)) and invert the transform:

$$\begin{aligned} \psi_{\mathfrak{r}}(\vec{\boldsymbol{q}} \ ; \ t_0) =& \frac{1}{\psi_{\mathfrak{r}}^{\star}(\vec{\boldsymbol{q}}_0 \ ; \ t_0)} \int \mathrm{d}^3\vec{\boldsymbol{k}}' \, \mathscr{Q}_{\mathfrak{r}}\left(\frac{\vec{\boldsymbol{q}} + \vec{\boldsymbol{q}}_0}{2}, \vec{\boldsymbol{k}}' \ ; \ t_0\right) \mathrm{e}^{\mathrm{i}\vec{\boldsymbol{k}}' \cdot (\vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_0)} \\ =& \sqrt{\frac{|\boldsymbol{\Sigma}_k|}{\pi^3}} \frac{\mathrm{e}^{-\mathrm{i}\vec{\boldsymbol{k}}_0 \cdot \vec{\boldsymbol{q}}'}}{\psi_{\mathfrak{r}}^{\star}(\vec{\boldsymbol{q}}_0 \ ; \ t_0)} \mathrm{e}^{-\frac{1}{8}\vec{\boldsymbol{q}}'^{\mathsf{T}}\boldsymbol{\Sigma}_q^{-1}\vec{\boldsymbol{q}}' - \frac{1}{2}\vec{\boldsymbol{k}}'^{\mathsf{T}}\boldsymbol{\Sigma}_k\vec{\boldsymbol{k}}'} \ . \end{aligned} \tag{5.8}$$

The value of the wave function at $\vec{\boldsymbol{q}}_0$ is (up to a constant phase factor) computed via the

respective marginal (Property (II) in Subsection 2.7.3):

$$|\psi_{\mathfrak{r}}(\vec{\boldsymbol{q}}_0\ ;\ t_0)|^2 = \int \mathrm{d}^3\vec{\boldsymbol{k}}'\, \mathscr{Q}_{\mathfrak{r}}\left(\vec{\boldsymbol{q}}_0, \vec{\boldsymbol{k}}'\ ;\ t_0\right) = \sqrt{\frac{|\boldsymbol{\Sigma}_k|}{\pi^3}}\ . \tag{5.9}$$

Plugging the above into Eq. (5.8) yields the wave function that is the closest analogue to the Dirac delta in phase space, i.e. a *generalized ray*:

$$\psi_{\mathfrak{r}}(\vec{\boldsymbol{q}}\ ;\ \bar{\boldsymbol{u}}) = \left(\frac{|\boldsymbol{\Sigma}_k|}{\pi^3}\right)^{\frac{1}{4}} \mathrm{e}^{\mathrm{i}\varphi} \mathrm{e}^{-\mathrm{i}\vec{\boldsymbol{k}}_0\cdot\vec{\boldsymbol{q}}'} \mathrm{e}^{-\frac{1}{8}\vec{\boldsymbol{q}}'^{\intercal}\left(\boldsymbol{\Sigma}_q^{-1}+4\boldsymbol{\Sigma}_k\right)\vec{\boldsymbol{q}}'}\ , \tag{5.10}$$

where we slightly abuse notation and make the mean $\bar{\boldsymbol{u}} = \left(\vec{\boldsymbol{q}}_0, \vec{\boldsymbol{k}}_0\right)$ at current time t explicit, with shifted position shorthand $\vec{\boldsymbol{q}}' = \vec{\boldsymbol{q}} - \vec{\boldsymbol{q}}_0$, as before, and $\varphi \in \mathbb{R}$ being an arbitrary initial phase. The evolution of that wave function to $t \geq t_0$ is dictated by Eqs. (5.5) and (5.6). The above should be understood as the *wave function that corresponds to the ray* $\vec{\boldsymbol{u}} = \bar{\boldsymbol{u}}$. Being a Gaussian beam, $\psi_{\mathfrak{r}}$ is indeed a subclass of AGSM beams, and it has the most compact support possible in phase space, as discussed.

Coherent-modes phase-space decomposition It is well-known that an arbitrary function in L^1 can be approximated arbitrary well by a finite sum of shifted Gaussians with identical variance (an immediate consequence of the Wiener's Tauberian theorem). In other words, multivariate Gaussians serve as an overcomplete functional basis. Therefore, the Husimi Q representation $\mathscr{Q}$ of an arbitrary WDF can be written as

$$\mathscr{Q} = \sum_{j=1}^{\infty} E_j\, \mathscr{Q}_{\mathfrak{r}}\Big|_{\bar{\boldsymbol{u}}_j, \boldsymbol{M}}\ , \tag{5.11}$$

i.e. a superposition of the Husimi Q representations of generalized rays, all with the same moment matrix $\boldsymbol{M}$ but shifted via different means $\bar{\boldsymbol{u}}_j$. The moment matrix must fulfil the Husimi Q condition $|\boldsymbol{M}| = \frac{1}{2^3}$, but otherwise is chosen at will, we may set $\boldsymbol{M}(t_0) = \frac{1}{\sqrt{2}}\boldsymbol{I}$

initially, for simplicity. As $\mathscr{Q}$ is non-negative, $E_j > 0$ is the energy contained in each generalized ray.

If we were to recover the corresponding cross-spectral density of the underlying field (via the inverse of Eq. (2.194) applied to $\mathscr{Q}$), then by the linearity of the transform it becomes: $\mathcal{C} = \sum_{j=1}^{\infty} E_j \, \mathcal{C}_j$, where $\mathcal{C}_j$ is simply the CSD of each of the shifted generalized rays. When a CSD can be written as a superposition of CSDs, the implication is that the light described by these CSDs is mutually incoherent. Therefore, we conclude that an arbitrary wave function ψ, with its phase space representation given by the husimi Q representation $\mathscr{Q}$, can be written as

$$\psi(\vec{\boldsymbol{q}}) = \mathrm{e}^{\mathrm{i}\varphi} \sum_{j=1}^{\infty} E_j \psi_{\mathfrak{r}}(\vec{\boldsymbol{q}} \; ; \; \bar{\boldsymbol{u}}_j) \; , \tag{5.12}$$

$$\text{such that} \qquad \langle \psi_{\mathfrak{r}}(\vec{\boldsymbol{q}}_1 \; ; \; \bar{\boldsymbol{u}}_l) \psi_{\mathfrak{r}}(\vec{\boldsymbol{q}}_2 \; ; \; \bar{\boldsymbol{u}}_j)^{\star} \rangle = \delta_{lj} \, \mathcal{C}_j \, (\vec{\boldsymbol{q}}_1, \vec{\boldsymbol{q}}_2) \; ,$$

for arbitrary spatial variables $\vec{\boldsymbol{q}}_{1,2} \in \mathbb{R}^3$, and where $\langle \cdot \rangle$ denotes the ensemble-averaging operator over same frequency realizations (as in Eq. (2.97)), and δ_{lj} is the Kronecker delta. That is, the above is a superposition of mutually-incoherent generalized rays, up to an irrelevant global phase term φ.

As a single generalized ray is perfectly coherent w.r.t. itself (the field admits only a single possible realization, viz. Eq. (5.10)), the above serves as a coherent-modes decomposition of an arbitrary wave function into a collection of Gaussian beams—generalized rays—with shifted origins (the mean at initial time t_0). We note again that this is only possible due to $\mathscr{Q}$ being non-negative, resulting in positive energies E_j.

We would like to point out that the decomposition in Eq. (5.12) is not the only possible way to understand Eq. (5.11). We could have used the inverse of the transform that defines the WDF, viz. Eq. (2.193), to recover a deterministic wave function, whose

Husimi Q representation is also as in Eq. (5.11). As a matter of fact, there are infinitely many such wave functions. There are also infinitely many stochastic processes, that would yield the desired Husimi Q representation. However, recall that in rendering we deal with time-averaged observables (we measure light over periods very long w.r.t. temporal frequency), therefore as all of the possible interpretations of Eq. (5.11) admit the same cross-spectral density function, they are all indistinguishable in terms of their time-averaged optical phenomena. Therefore, we may interpret Eq. (5.11) at will, and we chose the interpretation that is the most suitable for rendering: a coherent-mode decomposition into mutually-incoherent generalized rays.

5.1.2 Summary

The purpose of the discussion above is to find the closest, physically-realizable analogue of the ray, i.e. an impulse response in phase space, in the wave-optical context. We formally derive an expression for such a generalized ray, and furthermore show that an arbitrary Husimi Q distribution can be written as a coherent-modes decomposition of such generalized rays.

The Husimi Q representation (Eq. (5.3)) spreads every point of the WDF in phase space into the minimum footprint that can be resolved. We stress that while this procedure yields a strictly non-negative distribution (unlike the corresponding WDF $\mathscr{W}$), $\mathscr{Q}$ still remains only a quasi-probability density function, because it no longer reproduces the correct marginals (i.e., Property (II) no longer holds). That is, the low-pass filter induces errors into the distribution, and using $\mathscr{Q}$ in-place of $\mathscr{W}$ in Eq. (2.198) does not yield exact values for the expectation of wave-optical observables. The error can be understood as ignoring the uncertainty relation (Eq. (2.188)). When our observables of interest do not vary rapidly, as functions of $\vec{\boldsymbol{q}}$ and $\vec{\boldsymbol{k}}$, over a phase-space cell, then using

$\mathscr{Q}$ over the WDF serves as a valid approximation.

Thus, we regain the ability to perform phase space point-queries in order to do light transport via the mapping

$$\vec{\boldsymbol{u}} \mapsto \psi_{\tau}(\vec{\boldsymbol{q}} \; ; \; \vec{\boldsymbol{u}}) \; , \tag{5.13}$$

which maps between a classical ray and its corresponding wave function. The system evolution of these pair of constructs is similar, with respect to Liouville's equation (Property (VI)). Observe that Liouville's theorem (Eq. (2.176)) implies that the classical probability $\rho(\vec{\boldsymbol{u}})\,\mathrm{d}^6\vec{\boldsymbol{u}}$ that a ray is found in the phase-space volume $\mathrm{d}^6\vec{\boldsymbol{u}}$ is in-fact a property of the ray $\hat{\boldsymbol{u}}$, and remains invariant on system evolution. That is, the phase-space density around a ray transforms as the ray does. The generalized ray makes this fact explicit: the Gaussian spread of the generalized ray's energy around its mean is a property of the ray, and this spread transforms in an identical fashion to how ρ transforms.

Light transport operations can be classified into two categories:

1. **Simple** — propagation in homogeneous media with slowly-varying refractive-index, and perfect reflections and refractions of a smooth (potentially curved) interface. These interactions are accurately handled by the ABCD formalism (Eqs. (5.5) and (5.6)), and we continue to understand light as a ray. The implication being that classical rendering and sampling techniques apply, under these interactions, unaltered.

2. **Diffractive** — any interaction with an optical system that may not be classified as an ABCD system. For example, scattering by a rough surface, interaction with particles in media, and free-space diffraction. When dealing with diffractive interactions, we model the perfectly-coherent interaction of the wave function of the

generalized ray—the Gaussian beam—with the matter.

Note that generalized rays are mutually incoherent, hence they superpose linearly in terms of their intensities E, which can be equivalently rephrased in terms of radiometric quantities, just as classical rays do. Therefore, we do not need to propagate a wave's phase, nor do we deal with partially-coherent interactions. All interactions are either incoherent, on superposition of rays, or perfectly-coherent, on diffractive interactions between a singular generalized ray and matter.

5.2 Relation to Optical coherence

It is insightful to study how partially-coherent field effects arise under our formulation. Let $\mathcal{C}$ be the cross-spectral density of light, and $\mathscr{W}$ its corresponding WDF. The 3×3 spatial-coherence covariance matrix—termed the *shape matrix*—around a spatial point $\vec{\boldsymbol{q}}$ can be written as:

$$\begin{aligned}\boldsymbol{\Theta}(\vec{\boldsymbol{q}}) =& \frac{1}{\mathcal{C}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{q}}\right)} \int \mathrm{d}^3\vec{\boldsymbol{q}}'\, \vec{\boldsymbol{q}}'\vec{\boldsymbol{q}}'^{\mathsf{T}}\, \mathcal{C}\left(\vec{\boldsymbol{q}} - \frac{1}{2}\vec{\boldsymbol{q}}', \vec{\boldsymbol{q}} + \frac{1}{2}\vec{\boldsymbol{q}}'\right) \\ =& \frac{1}{\left|\psi(\vec{\boldsymbol{q}})\right|^2} \int \mathrm{d}^3\vec{\boldsymbol{q}}'\, \vec{\boldsymbol{q}}'\vec{\boldsymbol{q}}'^{\mathsf{T}} \int \mathrm{d}^3\vec{\boldsymbol{k}}'\, \mathscr{W}\left(\vec{\boldsymbol{q}}, \vec{\boldsymbol{k}}'\right) \mathrm{e}^{\mathrm{i}\vec{\boldsymbol{q}}'\cdot\vec{\boldsymbol{k}}'} \ . \end{aligned} \tag{5.14}$$

Formally-interchange the orders of integration, and note that

$$\int \mathrm{d}^3\vec{\boldsymbol{q}}'\, \vec{\boldsymbol{q}}'\vec{\boldsymbol{q}}'^{\mathsf{T}} \mathrm{e}^{\mathrm{i}\vec{\boldsymbol{q}}'\cdot\vec{\boldsymbol{k}}'} = -(2\pi)^3 \frac{\partial^2}{\partial \vec{\boldsymbol{k}}'^2}\, \delta^3\left(\vec{\boldsymbol{k}}'\right) \ , \tag{5.15}$$

i.e., the Hessian matrix of the Dirac delta. Then, for "well-behaved" $\mathscr{W}$:

$$\begin{aligned}\boldsymbol{\Theta}(\vec{\boldsymbol{q}}) &= -\frac{(2\pi)^3}{|\psi(\vec{\boldsymbol{q}})|^2}\int \mathrm{d}^3\vec{\boldsymbol{k}}'\,\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}'\right)\frac{\partial^2}{\partial\vec{\boldsymbol{k}}'^2}\,\delta^3\left(\vec{\boldsymbol{k}}'\right)\\ &= -\frac{(2\pi)^3}{|\psi(\vec{\boldsymbol{q}})|^2}\left[\frac{\partial^2}{\partial\vec{\boldsymbol{k}}'^2}\,\mathscr{W}\left(\vec{\boldsymbol{q}},\vec{\boldsymbol{k}}'\right)\right]_{\vec{\boldsymbol{k}}'=0}\,,\end{aligned} \tag{5.16}$$

that is, the Hessian (w.r.t. the frequency variable) of the WDF, evaluated at $\vec{\boldsymbol{q}}$ and $\vec{\boldsymbol{k}}=0$.

As mentioned, the cross-spectral density function and the WDF contain the same information (being Fourier-transform pairs), however we have shown that *spatial-coherence is dictated by the behaviour of the WDF in frequency-space only*, furthermore, for Gaussian signals the covariance of spatial coherence around a point $\vec{\boldsymbol{q}}$ is proportional to the inverse of covariance of angular spread of light at $\vec{\boldsymbol{q}}$.

Given a beam of light that consists of generalized rays (recall that we may always decompose any wave function into a superposition of such rays), the angular spread of these rays—quantified by the k-moments of the moment matrix (lower right 3×3 minor of $\boldsymbol{M}$ in Eq. (2.211))—is known as the *diffusivity* of light. Consider monochromatic light of wavelength λ, that propagates roughly in a preferred (mean) direction, say $\vec{\boldsymbol{z}}$. We define a *ray bundle* to be a collection of many generalized rays that compose this light. The diffusivity of this bundle is the variance of the geometric angular deviations of the generalized rays from the mean axis of propagation $\vec{\boldsymbol{z}}$. Let the paraxial k-moments that quantify diffusivity be the 2×2 positive-definite matrix $\boldsymbol{\Omega}$. The variance in the solid angle into which the bundle propagates is then $\Omega=|\boldsymbol{\Omega}|$. Then Eq. (5.16) immediately yields the following useful relation between coherence and diffusivity:

$$\boldsymbol{\Theta}=\lambda^2\boldsymbol{\Omega}^{-1}\,. \tag{5.17}$$

A highly convenient consequence of the presented light transport formalism is that it is devoid of the cumbersome two-point formulation of optical coherence theory, viz. the cross-spectral density $\mathcal{C}(\vec{\boldsymbol{q}}_1, \vec{\boldsymbol{q}}_2)$. In other words, we do not explicitly consider the spatial correlation between wavefronts. Instead, as we have shown, the variance in spatial coherence is fully described by the spatial-frequency curvature of the signal (2^{nd}-order derivative w.r.t. the k coordinate) in phase space.

When the cross-spectral density is used to describe light in-order to do light transport, partially-coherent beams are propagated, and the cross-section of these beams should be at least similar to the extent over which spatial coherence is retained. Backward path tracing (sensor to light source), as well as importance sampling light-matter interactions, is then complicated by the fact that the coherence properties of light are unknown a priori, before connecting a path to a light source. On the other hand, if we use generalized rays to sample the wave-optical phase space (the natural setting in light transport), then, as discussed, all our interactions are either classical (incoherent) or perfectly-coherent, and no a priori knowledge of the coherence of light is required. While we technically still propagate beams (viz. Eq. (5.10)), the extent of these beams in phase space is only dictated by the uncertainty relation, and is an order-of-magnitude less than the extent of partially-coherent beams.

Spatial-coherence effects are reproduced as we sample the diffusivity of light, i.e. the k space: for example, a perfectly collimated beam of zero diffusivity—where all the generalized rays have identical mean $\vec{\boldsymbol{k}}_0$ and propagate in the same direction—is perfectly coherent, while a beam with a positive variance in its mean wavevectors $\vec{\boldsymbol{k}}_0$, and hence positive diffusivity, is partially coherent. As such a beam interacts with matter, the fully-coherent interaction is blurred as generalized rays with different mean wavevectors are sampled, inducing partial coherence.

As an aside we note that a similar relation between the angular spread (diffusivity) of

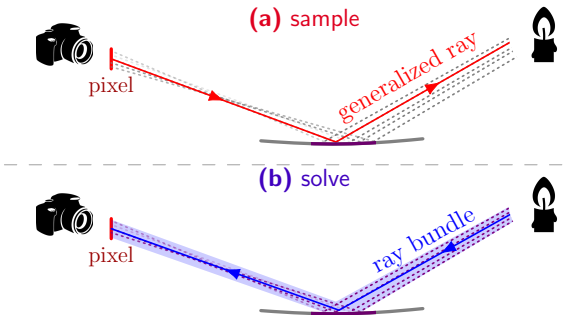
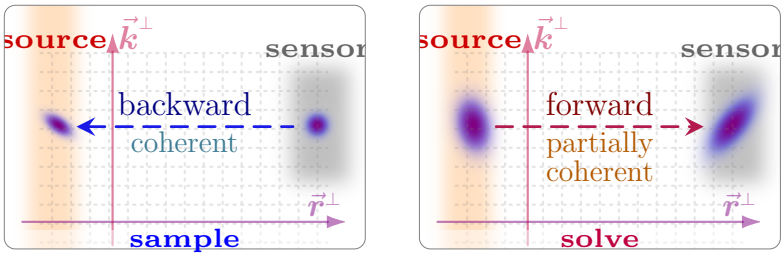


Figure 5.1: **Sample-solve.** Our path tracing algorithm (a) uses generalized rays (dotted lines) to *sample* paths through the scene. Because generalized rays are mutually incoherent, they superpose linearly, therefore classical sampling techniques apply essentially unchanged. Once a path is sampled (solid red path), we (b) *solve* for the partially-coherent transport from the light source to the sensor. As optical coherence is (inversely) related to the angular spread of rays in a ray bundle, the *solve* step should be understood as tracing a ray bundle (illustrated in blue) with a statistical distribution of generalized rays over the small solid angle subtended by the sensor from the source (across all intermediate interactions).

light and optical coherence is described by the van Cittert–Zernike (VCZ) theorem. However, a typical setting of the VCZ theorem is relating an incoherent source distribution to the mutual coherence at the destination. Eq. (5.16) is more general: the spatial variance in coherence of arbitrary light is related to the k moments in phase space. Furthermore, VCZ typically assumes a far-field setting where the implied decomposition is into plane waves. While the wavefront of a Gaussian beam can, in the far field (Fraunhofer region), be approximated by a plane wave, we do not make a far field assumption. Indeed, a mapping between a classical ray to a plane wave is highly unnatural: a ray is sharp impulse in phase space, while a plane wave has an equal probability of being measured throughout the entire space. Instead, by analysing the wave-optical phase space induced by the WDF, we map the classical ray to a Gaussian beam: a more general and a more pleasant analogue.



(a) path sampling via generalized rays **(b)** partially-coherent transport

Figure 5.2: **Light transport and dynamics in phase space.** A sensor's sensitivity function (gray) and a light source's emission function (orange) are plotted in the wave-optical phase space. (a) Generalized rays serve a similar purpose to ray-optical rays: they enable performing point queries of the wave-optical system. During the sample stage, a generalized ray (blue), occupying a single phase-space cell, is used to sample the sensor's sensitivity function. Then, we backward propagate this generalized ray (reverse its temporal dynamics) throughout the scene, until we hit a light source. (b) Solve stage: Once a path connecting a sensor to a light source has been sampled, we source partially-coherent light. This can be understood as widening the generalized ray's (that was used to sample the path) phase-space extent, thereby creating a *ray bundle* that occupies multiple phase-space cells. We then propagate that partially-coherent ray bundle back over the sampled path to the sensor. Observe that free-space propagation constitutes a horizontal shear of the phase space, formalised by Eq. (2.212), just as in the ray-optical case.

5.3 Sample-Solve

Sample We may now present the first stage of our *sample-solve* algorithm, see Fig. 5.1 for an overview. A measurable distribution of light (the convolved WDF discussed in Section 5.2) may always be sampled via generalized rays. The observable irradiance measured by a sensor is then the sum of all generalized rays that fall upon the sensor's sensitivity region in phase space (e.g., the Cartesian product of a pixel's spatial extent times its acceptance solid angle). From these generalized rays, we may sample a ray using classical techniques, for example uniformly, or importance sample with respect to the sensor's sensitivity function. We may do so because generalized rays are mutually incoherent, and hence superpose linearly in terms of their radiometric units.

Generalized rays traverse free space and homogeneous media in an identical fashion to classical, ray-optical rays (as these constitute simple interactions). Therefore, tracing

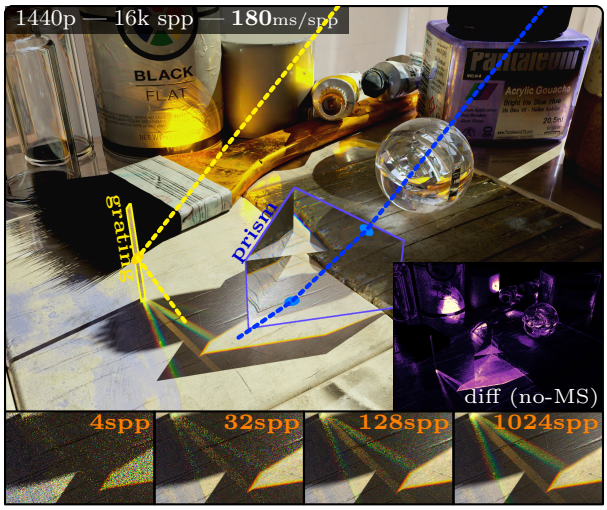


Figure 5.3: **Manifold sampling.** Example application of an advanced sampling technique. When performing next event estimation (NEE), manifold sampling (MS) [22, 62] enables finding a light path between a surface and a light source across one or more dielectric interfaces. In the rendered scene, such a sampled path, that refracts through the dispersive prism (outlined in blue), is visualized via the dotted blue line. We also employ MS for NEE on specular reflections: note the reflections off the paint brush's metal handle and off the paint tubes, as well as the thin diffraction grating (outlined in yellow) dispersing light into multiple diffraction lobes. (inset) Colour-coded difference compared to a without-MS rendering. (bottom) The diffraction grating lobes at increasing samples-per-pixel (spp).

generalized rays through space, tracing Eikonals through media with a slowly-varying refractive index, and reflecting or refracting at a smooth interface between media remains unchanged. We make the assumption that the spatial extent of a generalized ray is smaller than the scene's geometric details, therefore we may use ray tracing to propagate generalized rays. A similar assumption is made by PLT, though note that a generalized ray will typically have a smaller spatial extent compared to PLT's partially-coherent beams.

Finally, sampling diffractive interactions is done by importance sampling the coherent diffractive BSDFs that quantify these interactions. While classical BSDFs and diffractive BSDFs describe different physics, both fulfil similar postulates: they are non-negative density functions, energy is conserved over all scattering directions, and Helmholtz reci-

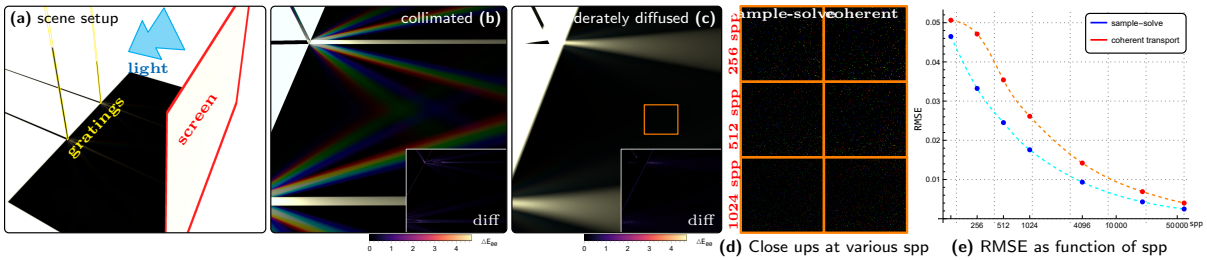


Figure 5.4: **PLT as variance reduction.** During the solve step, we apply PLT to solve for the partially-coherent light transport over a sampled light path. (a) Light arrives from the right, illuminating a simple scene. A screen (on the right, red outline) shadows a rectangular area. On the left a pair of thin diffraction gratings (yellow outline) reflect and disperse light. (b) When the light is highly collimated (subtends a very small solid angle from the scene), hence moderately coherent, the diffraction lobes are clearly visible. (c) As we increase the light's diffusivity (increase its solid angle), the reflection from the direct contribution lobe spreads out, while diffraction lobes mostly disappear, as expected. As the only light that arrives to the shadowed region is from the diffraction gratings, this is a good scene to study the benefits of the solve stage, and to do so we render the scene with only fully-coherent transport (no PLT applied). (b-c, insets) Difference images between the partially-coherent and fully-coherent sample-solve show that indeed both converge to an identical result. (d) However, close ups on the region outlined in orange in (c) show that, while the diffraction lobes are no longer visible, they still induce error, which the partially-coherent solve stage serves to reduce. (e) Plot of error in that area as function of sample count suggests that fully-coherent transport requires about 2-5 times the sample count to achieve similar-quality renderings.

procity is obeyed. Once an expression for a particular diffractive interaction is derived, seeking an importance sampling strategy for the resulting diffractive BSDF is no different from the classical case.

We stress that we are able to seamlessly and accurately transition between propagating a generalized ray through simple interactions to importance sampling diffractive interactions, because the generalized ray admits a well-defined wave function (unlike the classical ray, for which there is no clear way to define diffractive interactions). The above highlights again how generalized rays serve as the wave-optical analogue of the classical rays: sampling remains mostly unchanged, compared with classical formulations, with the primary difference being that we replace classical BSDFs with coherent diffractive BSDFs. To demonstrate that we are able to apply advanced sampling techniques for our

wave-optical path tracing, we have implemented manifold sampling, see Fig. 5.3.

Solve The sample stage above, as described, enables by itself accurate rendering. Though, at a price: we sample perfectly-coherent interactions. Coherent interactions can be imagined as illumination by a laser, giving rise to high-contrast, coherent optical speckle with rapidly-fluctuating intensities. As we sample many generalized rays propagating into different directions, we sample the angular spectrum of light—thus its optical coherence—correctly reproducing the *observable*, partially-coherent light transport through the scene. Hence, we may say that rendering via generalized rays is unbiased, but admits higher variance, see Fig. 5.4.

To combat this high variance, we *solve* for the partially-coherent transport over a path that was sampled in the sample stage. To do so we apply PLT, essentially unmodified. That is, while in the sample stage we trace a single generalized ray, the solve stage should be understood as tracing a ray bundle consisting of many generalized rays, all originating from the sampled light source, and diverging only a little in their direction. Our exact relation between the bundle's diffusivity and optical coherence then serves as the formal link between sampling and solving for the partially-coherent transport, and ensures that PLT can be accurately and rigorously applied. See Fig. 5.2 for an illustration.

In our domain of interest—rendering with partially-coherent optical light—applying PLT makes sense: The optical coherence of light serves as the primary factor that limits our ability to resolve diffraction effects, because partial coherence induces blurring (i.e. a convolution) of light-matter interactions. PLT then acts as a *variance-reduction technique*, by computing the observable optical phenomena over a sampled path, thereby eliminating the variance that arises due to the coherent sampling.

Other applications might find a different *solve* stage to be more appropriate: As an example, an SBR method (see Section 3.2) might apply the *sample* stage, as described,

to sample multiple paths that connect a source to a particular object, or an exit aperture. For the *solve*, the SBR method would remain virtually unchanged, and use these sampled paths in order to determine the induced electric currents on the object's surface, or compute the exit field's wavefront. Such applications are beyond the scope of this work, however they serve to highlight the generality and strength of our simple sample-solve approach: it bridges the gap between classical path tracing tools and wave optics, via the generalized ray concept, and enables the application of powerful sampling techniques in a much wider context.

5.3.1 Validity domain and limitations

We now briefly discuss the validity domain and enumerate the assumptions we make in this work.

1. **Ensemble averaging**. This is the core assumption that enables us to achieve mutual incoherence (formalised in Eq. (5.12)), and thus linearity, between generalized rays. For wave-optical rendering, we are only interested in ensemble-averaged—i.e. *observable*—effects. In other contexts, where sensors may no longer be understood to ensemble average, the instantaneous values of the fields come into play, and no linearity, at least in terms of radiometric units, can be regained.

2. **Spatial extent of a generalized ray**. Unlike classical rays, generalized rays have a positive spatial extent. We approximate the propagation of a generalized ray via ray tracing, i.e., ignore its spatial extent, and point sample the wave-optical phase space.

3. **Optical far field**. In practice, we make the far field assumption in order to simplify the expression for a generalized ray's wave function (resulting in a plane wave). This

is also a good approximation at optical frequencies, and it serves to significantly simplify the analysis of deriving analytic expressions for diffractive BSDFs, as well as importance sampling these BSDFs. Nevertheless, sample-solve, per sē, does not require any far-field assumptions, and can be formulated using the exact expression for the generalized ray's wave function (Eq. (5.10)).

The second assumption above is worth further discussion. We stress that as a generalized ray occupies a single wave-optical phase-space cell, resolving features smaller than a generalized ray is not possible. That is to say, geometrical reasoning does not apply beyond the limit of uncertainty, and a generalized ray cannot be broken up into smaller "parts". When a generalized ray is partially occluded by matter, the following occurs:

1. Part of the generalized ray's energy, proportional to the cross-sectional area masked by the matter, interacts with the matter. Therefore, on partial occlusion, a portion of the entire generalized ray interacts with matter (and not just the geometrical extent that is occluded, as geometrical reasoning does not hold). Hence, even though we ray trace only the mean of a generalized ray, as we sample many such generalized rays, we correctly sample all the occluders, just like the ray-optical setting.

2. The generalized ray's energy that does not interact with the occluder directly undergoes free-space diffraction. For practical reasons, such effects are left for future work. In our implementation we ignore free-space diffractions, and simply allow the unoccluded energy to continue to propagate unimpeded.

To conclude: making the assumption that we ignore the spatial extent of generalized rays is equivalent to ignoring the diffracted lobes of free-space diffracted generalized rays. This is a marked improvement over the state-of-the-art: in contrast to partially-

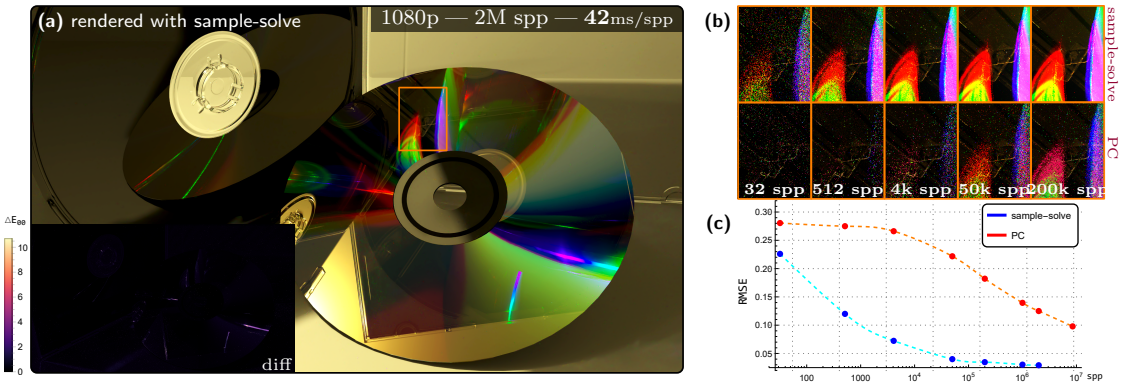


Figure 5.5: **Partially-coherent sampling.** (a) A scene contains a compact disk (CD) that rests next an open CD case, upon which another closed CD case is placed. A ceiling-mounted light source illuminates the scene. While simple, the scene admits interesting light transport. (b) We emulate partially-coherent (PC) sampling of BSDFs, in an identical manner to the state-of-the-art, which manifests the sampling problem: the diffraction lobes are very sharp lobes that are difficult to sample when doing backward path tracing (and the coherence properties of light are unknown) using existing tools. The close ups show the area marked in orange rendered at various sample counts. Note the sharp difference in noise between our proposed sample-solve approach and PC sampling. (c) Plot of noise as function of sample count, quantifying the drastic improvement in sampling performance: sampling using generalized rays reduces sample count by a factor of about 4000 for similar-quality rendering. (a, inset) A difference image between the two sampling strategies. With the exception of the very high-frequency diffraction effects, the differences are minor, and due to the PC rendering never converging.

coherent beams, whose geometric and coherence properties change on partial occlusion, generalized rays always remain perfectly coherent.

We stress that sample-solve is derived under a highly general validity domain. Up to the assumptions listed above, correctness is formally ensured: The measurable effects of any physically-realizable wave-optical phase space distribution can always be written, to arbitrary precision, as a finite sum of generalized rays (because Gaussians form an overcomplete functional basis). These generalized rays then act as "point queries" of light's behaviour, in a manner very similar to the classical case. These assumptions above are also less restrictive than the state-of-the-art. As a matter of fact, we extend PLT's applicability to light of any state-of-coherence: PLT assumes linearity due to light's partial coherence, while our generalized rays always remain linear, and hence sampling

is unconstrained by light's coherence properties. We understand partially-coherent light simply as a ray bundle, and ray bundles composed of distinct generalized rays must be mutually incoherent as well.

As a simple numeric validation of our implementation, we note that both perfectly-coherent transport, viz. Fig. 5.4, and partially-coherent sampling, viz. Fig. 5.5, produce identical results to the presented sample-solve, as expected.

5.4 Wave-Optical Rendering

We note a consequence of our contributions: Given a classical path tracer, the minimal change required to make that path tracer compatible with wave optics is replacing classical rays with generalized rays as the "point queries" of light's behaviour. In practice, doing so requires essentially only two changes:

1. transitioning to a spectral, vectorized (polarization-aware) path tracer; and

2. reformulating all the BSDFs in terms of generalized rays.

The former is simply a consequence of the fact that generalized rays and wave-optical BSDFs depend on light's wavelength and polarization.

Sample-solve takes a step beyond the bare minimum, and adds an additional pass that serves the role of a variance reduction method. This forward pass applies PLT in order to solve for the partially-coherent transport over the sampled path, and, as long as closed-form expressions for the partially-coherent BSDFs have been formulated, is very cheap. All images and videos in this work were rendered using our unidirectional path tracer that implements our sample-solve approach. In this section we will give a brief overview of our implementation.

As generalized rays are defined in (paraxial) phase space—the space spanned by the position and wavevector variables—their energy density is described in terms of flux per area per solid angle, i.e. units of the radiometric radiance. We therefore continue to use radiance as the primary quantity that describes energy transport throughout the scene. To formalize a generalized ray's polarimetric properties, we use Stokes parameters vectors (carrying units of radiance).

Initiating paths Consider a sensor. The sensor's sensitivity function $\mathcal{S}(\vec{\boldsymbol{r}}, \vec{\boldsymbol{k}})$ is a phase-space function that quantifies the sensor's observable response to radiation. As discussed, for this function to be measurable, it must satisfy the uncertainty relation (Eq. (2.188)), in the sense that it is possible to write it formally as an arbitrary function convolved with some minimum-uncertainty Gaussian. Then, it is always possible to write this sensitivity function as some superposition of generalized rays, viz.

$$\mathcal{S}\left(\vec{\boldsymbol{r}}, \vec{\boldsymbol{k}}\right) = \sum_n s^{(n)} \mathscr{W}_{\mathfrak{ray}}^{(n)} \ , \tag{5.18}$$

where s is sensitivity, and $\mathscr{W}_{\mathfrak{ray}}$ is the WDF of a generalized ray. That is, the sensor's extent in phase space is covered with generalized rays, and this is always possible, because a sensor must adhere to the uncertainty relation.

Each such generalized ray will serve as a single *sample* of the transport that light undergoes prior to arriving at the sensor. Denote $\vec{\boldsymbol{r}}_0^{(n)}, \vec{\boldsymbol{k}}_0^{(n)}$ as the mean positions and incident wavevectors (incident direction scaled by the wavenumber), at the sensor, of these generalized rays, where we use superscripts to indicate sample indices and subscripts for indices along a sampled path. In practice, positions and incidence directions are sampled with respect to the sensitivity function $\mathcal{S}$, in an identical manner to the classical setting. A wavelength for each sample is also needed (our implementation draws a random wave-

length uniformly from the visible spectrum). This wavelength drives the path sampling process, hence we refer to it as the "Hero wavelength" [56].

Sample stage This stage applies standard unidirectional path tracing from the sensor (i.e., backward path tracing). Once a generalized ray's terminal phase-space position has been sampled, we proceed with retracing the steps light took before arriving to the sensor, i.e. backward path tracing. The temporal dynamics of a generalized ray are entirely deterministic, hence time reversal and backward propagation are well defined. For backward propagation through free space (and other simple interactions, as discussed in Subsection 5.1.2), a generalized ray retains the classical dynamics, and we simply ray trace the generalized ray's mean position through space.

Once a generalized ray encounters matter, we attempt to connect the path to a light source via next-event-estimation (NEE). We also perform "Russian roulette" early path termination. To continue the path, we importance sample the diffractive BSDF that governs the diffractive interaction at the interaction region, yielding the next pair $\vec{\boldsymbol{r}}_j^{(n)}, \vec{\boldsymbol{k}}_j^{(n)}$, with $j \geq 1$ being the interaction index from the sensor. Sampling then proceeds recursively. Diffractive BSDFs will be discussed later.

As a design choice, our implementation ignores polarization during the sample stage. This is done for simplicity and performance: decomposing and inverting Mueller matrices, and other polarimetric interactions, can be cumbersome and expensive. No error is introduced, as the solve stage is fully vectorized.

Solve stage Every time a path has been connected to a light source, either organically or via NEE, the solve stage kicks in to compute the partially-coherent transport over the sampled path. If the sampled path does not contain dispersive delta segments (e.g., refraction through a smooth dielectric interface), in addition to the Hero wavelength, we

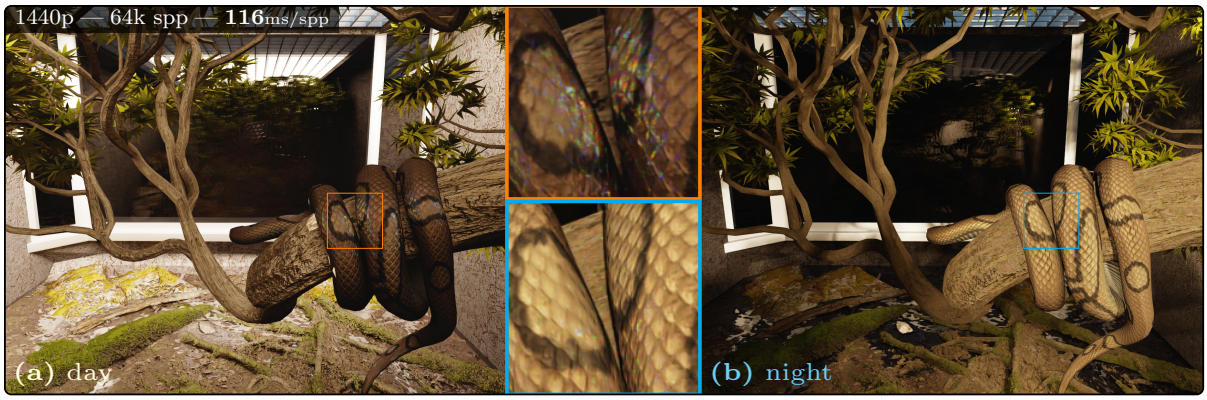


Figure 5.6: **Diffractive materials under different illumination conditions.** The appearance of diffractive materials depends on light's spectral, polarimetric and coherence properties. (a) During daytime, the illumination reaching the snake is dominated by indirect sunlight. At some angles, where incident light has a narrower angular spread (and thus is more coherent), the diffractive scales show clear interference patterns. (b) At night, direct light from the fluorescent lamps is the primary source of illumination, however that light is too incoherent to produce visible diffraction patterns.

also importance sample 3 additional wavelengths from the light's emission spectrum. We therefore transport one to four spectral samples per sampled path. Then, the solve stage applies PLT, as formalised in Chapter 3.

To source partially-coherent light, i.e. a ray bundle, from a light source, we need to quantify its coherence properties (for each spectral sample). This depends on the light source:

1. Distant light sources — These include analytic distant sources, with a predefined solid angle that they subtend from the scene, as well as environment maps, in which case light is sourced from a small cluster of a few pixels, which define the solid angle. The coherence shape matrix $\boldsymbol{\Theta}$ is then immediately defined via Eq. (5.17). Because these are distant emitters, the coherence shape matrix does not depend on the distance of propagation from the source, and only undergoes transformations due to interactions.

2. Emissive geometry light sources — We source light from a small area on the emitting

triangle. In our implementation we set this area to a constant $a = 10.00\,\mathrm{mm}^2$. Then, the solid angle subtended by the sourcing area at the first interaction region is $\Omega = a/r_1^2$, where r_1 is the distance from the source to that first interaction point. The coherence shape matrix, for the entire path, can then be written as

$$\boldsymbol{\Theta} = \frac{r^2}{\lambda^2}\boldsymbol{\Omega}^{-1} \,, \tag{5.19}$$

where r is the total distance of propagation from the source to each interaction region, and λ is the wavelength of light, as before. $\boldsymbol{\Omega}$ is the 2×2 matrix that quantifies the partially-coherent ray bundle's angular extent. Initially, we set $\boldsymbol{\Omega} = a\boldsymbol{I}$ and afterwards $\boldsymbol{\Omega}$ transforms on interactions.

Let $\vec{\boldsymbol{r}}_l^{(n)}, \vec{\boldsymbol{k}}_l^{(n)}$ be the mean position and wavevector, at the light source, of the sourced ray bundle, with $l > 0$ being the total path length and $l - 1$ the number of interactions along the path. We propagate this ray bundle over the sampled path, evaluate the partially-coherent BSDFs at each interaction in forward order, i.e. over interaction indices $l - 1, l - 2, \ldots, 1$. At each interaction we transform the coherence shape matrix, as described in Steinberg et al. [46, Section 3.5]. Finally, the ray bundle arrives at the sensor, and contributes to the observed intensities with respect to the (wavelength-dependant) sensor's sensitivities $s^{(n)}$. We also perform multiple importance sampling.

Diffractive BSDFs Note the light-matter interaction theorem, viz. Theorem 4.2.1.1: when light is perfectly coherent (that is, when the ray bundle is composed of collimated generalized rays), the convolution reduces to multiplication, because the Fourier transform of the coherence function is a Dirac delta. In other words: fully-coherent interactions reduce to multiplication by the matter's *diffractive BSDF*. It is easy to see that the diffractive BSDF is a function of incident and scattered directions, as well as

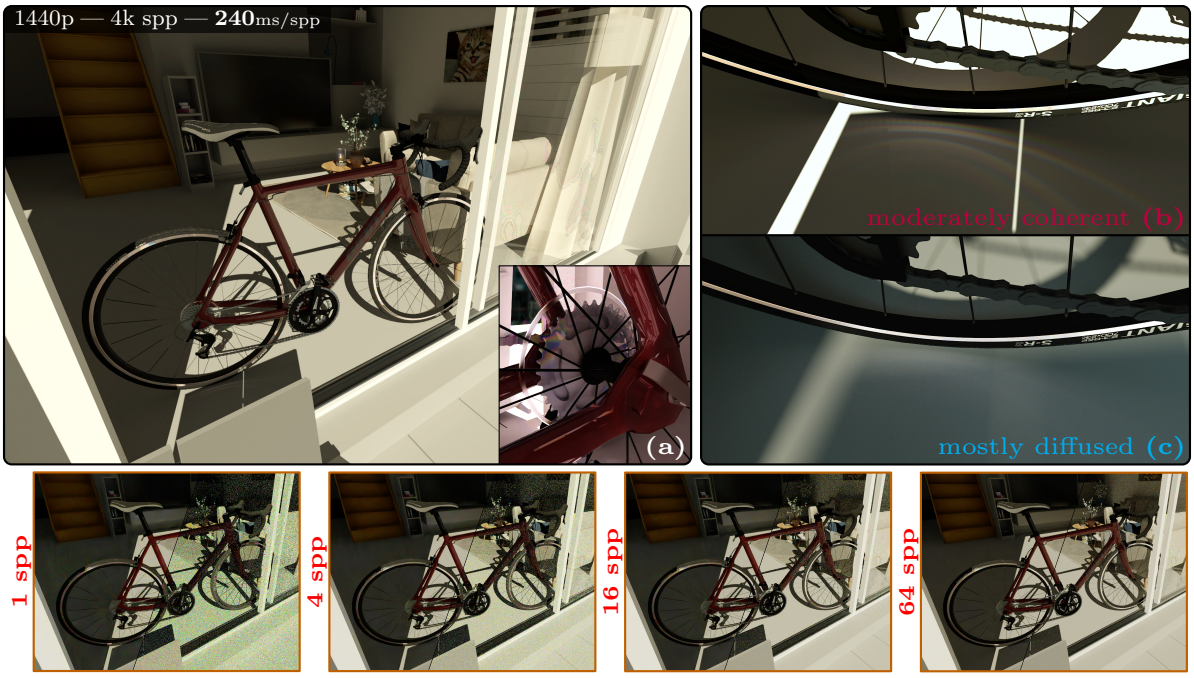


Figure 5.7: **Bike scene.** The bike scene from Steinberg et al. [46] rendered using sample-solve. The scene contains a few materials with visible wave-interference effects, most notably the birefringent dielectrics, (a) like the plastic wheel spoke guard, (b-c) as well as the diffraction grated wheel brake surface. (b) when illuminated by direct sunlight, this grating disperses light into visible diffraction lobes; however, (c) when sunlight passes through a diffuser (like clouds), only the direct lobe is visible. (bottom) Low spp renderings, with (left side) and without (right side) a denoiser, showcasing the interactive rendering performance of our wave-optical renderer.

incident and scattered wavelengths. Cross-wavelength scattering, i.e. when $|\vec{\boldsymbol{k}}_{\mathrm{o}}| \neq |\vec{\boldsymbol{k}}_{\mathrm{i}}|$, e.g., due to fluorescence or phosphorescence, can be formulated then, however we ignore such effects in our implementation (for simplicity). When restricted to fully-coherent, same-wavelength scattering, our rendering equation Definition 3.4.1 above reduces to a classical-like integral over a (scaled) unit sphere, because the diffraction operator reduces to multiplication.

The discussion above serves to highlight how the diffractive BSDF is the wave-optical analogue of the classical BSDF, and aims to solve the electrodynamics problem that answers the question: given a generalized ray, with wavevector $\vec{\boldsymbol{k}}_{\mathrm{i}}$, impinging upon the matter, how much energy is scattered into a generalized ray with wavevector $\vec{\boldsymbol{k}}_{\mathrm{o}}$? The

problem is simplified when we make the discussed far-field assumption, where the wavefront of the generalized ray becomes a plane wave.

Because the diffractive BSDF is reduced to a well understood and studied problem, we may apply well-known electromagnetism to devise a diffractive BSDF. Example materials were discussed in Subsection 4.2.2.

Finally, the partially-coherent form of the BSDF (used for the solve stage) is the convolution of the diffractive BSDF with light's diffusivity (i.e., its coherence properties). This is nothing more than the light-matter interaction theorem, viz. Theorem 4.2.1.1, made manifest by relation between coherence and diffusivity (viz. Eq. (5.17)). Also note that we do not require exact expressions for a partially-coherent BSDF: because partial coherence is naturally reproduced as we sample light's diffusivity.

5.5 Results

Our results are comprised of three main scenes:

1. **Snake enclosure**, Figs. 5.6 and 5.8. This scene is illuminated by multiple light sources: the sun, the sky (diffused sunlight), as well as a pair of industrial 4100.00 K fluorescent lamps with a decent colour rendering index of 82 located at the back of the enclosure. Sunlight and skylight arrive from the opening at the top. As most of the light that arrives at the different diffractive materials is indirect, this is a difficult scene to render.

2. **Manifold sampling**, Fig. 5.3. A highly-detailed scene that we use as our manifold-sampling playground.

3. **Bike**, Fig. 5.7. Adapted from Steinberg et al. [46], however appearance is not expected to match, as our materials are different.

In addition to the above, the CD scene is used for analysis of partially-coherent sampling, Fig. 5.5.

Performance metrics, i.e. rendering resolution and samples-per-pixel (spp) count are given in each figure. The figures in this thesis are high-quality, converged (i.e., very high samples-per-pixel) renders. Nevertheless, our method and implementation enable interactive wave-optical rendering at 1 spp. Similar to other modern GPU-accelerated path tracer, we use a denoiser for interactive rendering. This enables generating acceptable images at 1 spp, and allows the user to interact with and edit the scene in real-time.

Low-spp images of the bike scene, with and without the denoiser, are shown in Fig. 5.7 (bottom). Low sample count is sufficient for the vast majority of the materials, including the diffractive birefringent dielectrics. An exception is the dispersive diffraction lobes, which are visible on the floor, and arise due to diffraction grated wheel brake surface. This is due to a couple of reasons:

1. A diffraction grating scatters into many wavelength-dependant lobes, hence requires a moderate amount of spectral samples.

2. These are rendered via manifold sampling (MS), as discussed in Fig. 5.3. However, MS is only initiated during the sample stage when a path is scattered from the floor into the grating. The probability of finding such a connection organically is rather low (roughly about 1 in a few dozen), as the grated surface is quite small.

The difficulty of rendering the diffracted lobes of a diffraction grating is also analysed in Fig. 5.3 (bottom). This problem is similar to the classical problem of rendering dispersive caustics, e.g., on the bottom of a pool with a disturbed water surface; a problem where classical unidirectional path tracers struggle as well. We highlight again that our formalism reduces wave-optical rendering problem to classical sampling problem, to which classical tools apply.

Figure 5.8: **From ray optics to wave optics.** In this chapter we have presented the *generalized ray*: an extension of the classical ray to wave optics. The generalized ray retains the defining characteristics of the ray-optical ray: *locality* and *linearity*. These properties allow the generalized ray to serve as a "point query" of light's behaviour—the same purpose that the classical ray fulfils in rendering. By using such generalized rays, we enable the rendering of complex scenes, like the one shown, under rigorous wave-optical light transport. Materials admitting diffractive optical phenomena are visible: (a) a Bornite ore with a layer of copper oxide causing interference; (b) a Brazilian Rainbow Boa, whose scales are biological diffraction grated surfaces; and (c) a Chrysomelidae beetle, whose colour arises due to naturally-occurring multilayered interference reflectors in its elytron. Our formalism serves as a link between path tracing techniques and wave optics, and admits a highly general validity domain. Therefore, we are able to apply sophisticated sampling techniques, and achieve performance that surpasses the state-of-the-art by orders-of-magnitude. We indicate resolution and samples-per-pixel (spp) count in all figures rendered using our method. While these figures showcase converged (high spp) results, our implementation also allows interactive rendering of all these scenes at 1 spp. Frame times (at 1 spp) for interactive rendering are indicated.

Chapter 6

Conclusion

We presented a comprehensive, computationally-tractable theory of partially-coherent light transport. Complete formulae, for sourcing, propagation, diffraction were derived, as well as light-matter interaction framework with locally-stationary matter. We have shown how a wave-optical system can be sampled using generalized rays, and how optical coherence relates to a collection of such rays—a ray bundle.

The generalized ray—being a local and linear descriptor of light—serves as the tool that allows us to apply classical sampling and path tracing techniques, in order to sample the wave-optical phase space. As that generalized ray admits a well-defined wave function, its electrodynamics and diffractive interactions with matter are also well-defined. Thereby, the generalized ray spans a link between path tracing and computational optics, formalised via our sample-solve method. This serves as our primary contribution. We have also shown that generalized rays sample the partial coherence of light naturally: as we sample the angular spectrum of light. Therefore, the partially-coherent analysis that happens in our solve step does not need to be exact, and useful simplifications can be made. This is important, because designing partially-coherent BSDFs, as mandated by PLT, can be analytically difficult, and being able to make analytic approximations is of

real practical value.

We showcase the power of our formalism, by applying sample-solve for wave-optical rendering. For our use case, the solve stage is used as a variance-reduction technique that applies PLT in order to solve for the partially-coherent transport (i.e., a ray bundle) over a sampled path. As a consequence of our generalized ray formalism, the challenges faced by designing such a wave-optical path tracer, e.g., sampling, importance sampling strategies for diffractive BSDFs, challenges imposed by spectral rendering, and effects like fluorescence, are essentially all classical in nature. That is, electromagnetism needs to be explicitly considered only when devising diffractive BSDFs. Furthermore, we show that we are able to achieve interactive wave-optical rendering of complex scenes, at a performance that is orders-of-magnitude faster than the state-of-the-art. Our formalism is highly general, with a validity domain that subsumes the state-of-the-art.

6.1 Future Work

This work has focused on wave-optical light transport, at optical frequencies. The majority of this work carries over to other frequencies, where unique challenges arise. At other frequencies, some of the assumptions (see Subsection 5.3.1) might need to be revisited, most notably the omission of edge diffractions in this work. Also, for other applications a different, more appropriate solve stage might be suitable.

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, ninth dover printing, tenth gpo printing edition, 1964.

[2] T. Auzinger, W. Heidrich, and B. Bickel. Computational design of nanostructural color for additive manufacturing. *ACM Transactions on Graphics*, 37(4):1–16, Aug 2018. doi: 10.1145/3197517.3201376.

[3] C. Bar, M. Alterman, I. Gkioulekas, and A. Levin. A monte carlo framework for rendering speckle statistics in scattering media. *ACM Transactions on Graphics*, 38 (4):1–22, Jul 2019. doi: 10.1145/3306346.3322950.

[4] C. Bar, I. Gkioulekas, and A. Levin. Rendering near-field speckle statistics in scattering media. *ACM Transactions on Graphics*, 39(6):1–18, Nov 2020. doi: 10.1145/3414685.3417813.

[5] M. Bass, C. DeCusatis, J. M. Enoch, V. Lakshminarayanan, G. Li, C. A. MacDonald, V. N. Mahajan, and E. W. Van Stryland. *Handbook of optics: Geometrical and physical optics, polarized light, components and instruments(set) v. 1*. McGraw-Hill Professional, New York, NY, 3 edition, Oct. 2009.

[6] M. J. Bastiaans. The wigner distribution function applied to optical signals and systems. *Opt. Commun.*, 25(1):26–30, Apr. 1978.

[7] L. Belcour and P. Barla. A practical extension to microfacet theory for the modeling of varying iridescence. *ACM Trans. Graph.*, 36(4):65:1–65:14, July 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073620.

[8] M. Born and E. Wolf. *Principles of optics : electromagnetic theory of propagation, interference and diffraction of light*. Cambridge University Press, Cambridge New York, 1999. ISBN 978-0521642224.

[9] H. A. Buchdahl. *An introduction to Hamiltonian optics*. Courier Corporation, 1993.

[10] J. Burgess. RTX on—the NVIDIA turing GPU. *IEEE Micro*, 40(2):36–44, Mar. 2020.

[11] R. Carminati and J.-J. Greffet. Near-field effects in spatial coherence of thermal sources. *Phys. Rev. Lett.*, 82:1660–1663, Feb 1999. doi: 10.1103/PhysRevLett.82.1660.

[12] M. Charnotskii. Coherence of radiation from incoherent sources: I sources on a sphere and far-field conditions. *Journal of the Optical Society of America A*, 36(8): 1433, Jul 2019. doi: 10.1364/josaa.36.001433.

[13] L. Cohen. Generalized phase-space distribution functions. *J. Math. Phys.*, 7(5): 781–786, May 1966.

[14] L. Cohen. *Time frequency analysis*. Prentice-Hall signal processing series. Prentice Hall, Philadelphia, PA, Dec. 1994.

[15] T. Cuypers, T. Haber, P. Bekaert, S. B. Oh, and R. Raskar. Reflectance model for diffraction. *ACM Transactions on Graphics*, 31(5):1–11, Aug 2012. doi: 10.1145/2231816.2231820.

[16] D. Dhillon, J. Teyssier, M. Single, I. Gaponenko, M. Milinkovitch, and M. Zwicker. Interactive diffraction from biological nanostructures. *Computer Graphics Forum*, 33 (8):177–188, 2014. doi: 10.1111/cgf.12425. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12425`.

[17] V. Falster, A. Jarabo, and J. R. Frisvad. Computing the bidirectional scattering of a microstructure using scalar diffraction theory and path tracing. *Computer Graphics Forum*, 39(7):231–242, Oct 2020. doi: 10.1111/cgf.14140.

[18] A. T. Friberg. On the existence of a radiance function for finite planar sources of arbitrary states of coherence. *Journal of the Optical Society of America*, 69(1):192, Jan 1979. doi: 10.1364/JOSA.69.000192.

[19] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics*, 34(4):1–14, Jul 2015. doi: 10.1145/2766928.

[20] J. Goodman. *Statistical optics*. John Wiley & Sons Inc, Hoboken, New Jersey, 2015. ISBN 978-1119009450.

[21] I. Guillén, J. Marco, D. Gutierrez, W. Jakob, and A. Jarabo. A general framework for pearlescent materials. *ACM Transactions on Graphics*, 39(6), 2020. doi: 10.1145/3414685.3417782.

[22] J. Hanika, M. Droske, and L. Fascione. Manifold next event estimation. *Comput. Graph. Forum*, 34(4):87–97, July 2015.

[23] J. T. Kajiya. The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques - SIGGRAPH '86*. ACM Press, 1986. doi: 10.1145/15922.15902.

[24] T. Kneiphof, T. Golla, and R. Klein. Real-time image-based lighting of microfacet brdfs with varying iridescence. *Computer Graphics Forum*, 38(4):77–85, 2019. doi: 10.1111/cgf.13772. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13772`.

[25] M. Koivurova, H. Partanen, J. Turunen, and A. T. Friberg. Grating interferometer for light-efficient spatial coherence measurement of arbitrary sources. *Applied Optics*, 56(18):5216, Jun 2017. doi: 10.1364/ao.56.005216.

[26] O. Korotkova. *Random Light Beams*. CRC Press, Boca Raton, 2017. ISBN 9781351833875.

[27] O. Korotkova and E. Wolf. Changes in the state of polarization of a random electromagnetic beam on propagation. *Optics Communications*, 246(1):35 – 43, 2005. ISSN 0030-4018. doi: https://doi.org/10.1016/j.optcom.2004.10.078.

[28] A. Kotwal, A. Levin, and I. Gkioulekas. Interferometric transmission probing with coded mutual intensity. *ACM Transactions on Graphics*, 39(4), Jul 2020. doi: 10.1145/3386569.3392384.

[29] A. Krywonos. *Predicting surface scatter using a linear systems formulation of non-paraxial scalar diffraction*. PhD thesis, University of Central Florida, Jan 2006.

[30] A. Levin, D. Glasner, Y. Xiong, F. Durand, W. Freeman, W. Matusik, and T. Zickler. Fabricating brdfs at high spatial resolution using wave optics. *ACM Transactions on Graphics*, 32(4):1–14, Jul 2013. doi: 10.1145/2461912.2461981.

[31] C. F. Loan. The ubiquitous kronecker product. *Journal of Computational and Applied Mathematics*, 123(1-2):85–100, Nov 2000. doi: 10.1016/s0377-0427(00)00393-9. URL `https://linkinghub.elsevier.com/retrieve/pii/S0377042700003939`.

[32] L. Mandel and E. Wolf. *Optical coherence and quantum optics*. Cambridge University Press, Cambridge, 1995. ISBN 978-0521417112.

[33] H. Mashaal, A. Goldstein, D. Feuermann, and J. M. Gordon. First direct measurement of the spatial coherence of sunlight. *Optics Letters*, 37(17):3516, Aug 2012. doi: 10.1364/ol.37.003516.

[34] S. Miller. *Probability and Random Processes : With Applications to Signal Processing and Communications*. Elsevier Science, Burlington, 2012. ISBN 978-0123869814.

[35] A. B. Murphy and E. Tam. Thermodynamic properties and transport coefficients of arc lamp plasmas: argon, krypton and xenon. *Journal of Physics D: Applied Physics*, 47(29):295202, Jun 2014. doi: 10.1088/0022-3727/47/29/295202.

[36] A. Musbach, G. W. Meyer, F. Reitich, and S. H. Oh. Full wave modelling of light propagation and reflection. *Computer Graphics Forum*, 32(6):24–37, Feb 2013. doi: 10.1111/cgf.12012. URL `http://dx.doi.org/10.1111/cgf.12012`.

[37] S. B. Oh, S. Kashyap, R. Garg, S. Chandran, and R. Raskar. Rendering wave effects with augmented light field. *Computer Graphics Forum*, 29(2):507–516, May 2010. doi: 10.1111/j.1467-8659.2009.01620.x.

[38] I. Sadeghi, A. Munoz, P. Laven, W. Jarosz, F. Seron, D. Gutierrez, and H. W. Jensen. Physically-based simulation of rainbows. *ACM Transactions on Graphics*, 31(1):1–12, Jan 2012. doi: 10.1145/2077341.2077344.

[39] S. N. Savenkov. *Jones and Mueller matrices: structure, symmetry relations and information content*, page 71–119. Springer Berlin Heidelberg, 2009. doi: 10.1007/978-3-540-74276-0_3. URL `http://dx.doi.org/10.1007/978-3-540-74276-0_3`.

[40] R. Silverman. Locally stationary random processes. *IEEE Transactions on Information Theory*, 3(3):182–187, Sep 1957. doi: 10.1109/tit.1957.1057413. URL `http://dx.doi.org/10.1109/TIT.1957.1057413`.

[41] J. Stam. Diffraction shaders. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*. ACM Press, 1999. doi: 10.1145/311535.311546.

[42] S. Steinberg. Analytic spectral integration of birefringence-induced iridescence. *Computer Graphics Forum*, 38(4):97–110, Jul 2019. doi: 10.1111/cgf.13774.

[43] S. Steinberg. Accurate rendering of liquid-crystals and inhomogeneous optically anisotropic media. *ACM Transactions on Graphics*, 39(3):1–23, Jun 2020. doi: 10.1145/3381748.

[44] S. Steinberg and L.-Q. Yan. Physical light-matter interaction in hermite-gauss space. *ACM Trans. Graph.*, 40(6), dec 2021. ISSN 0730-0301. doi: 10.1145/3478513.3480530. URL `https://doi.org/10.1145/3478513.3480530`.

[45] S. Steinberg and L.-Q. Yan. Rendering of subjective speckle formed by rough statistical surfaces. *ACM Trans. Graph.*, 41(1), feb 2022. ISSN 0730-0301. doi: 10.1145/3472293. URL `https://doi.org/10.1145/3472293`.

[46] S. Steinberg, P. Sen, and L.-Q. Yan. Towards practical physical-optics rendering. *ACM Transactions on Graphics*, 41(4):1–13, Jul 2022. doi: 10.1145/3528223.3530119.

[47] P. Sysel and P. Rajmic. Goertzel algorithm generalized to non-integer multiples of fundamental frequency. *EURASIP Journal on Advances in Signal Processing*, 2012 (1), Mar 2012. doi: 10.1186/1687-6180-2012-56. URL `http://dx.doi.org/10.1186/1687-6180-2012-56`.

[48] M. Testorf, B. Hennelly, and J. Ojeda-Castañeda. *Phase-space optics: fundamentals and applications*. McGraw-Hill Education, 2010.

[49] A. Toisoul and A. Ghosh. Practical acquisition and rendering of diffraction effects in surface reflectance. *ACM Transactions on Graphics*, 36(5):1–16, Jul 2017. doi: 10.1145/3012001.

[50] A. Toisoul, D. S. Dhillon, and A. Ghosh. Acquiring spatially varying appearance of printed holographic surfaces. *ACM Trans. Graph.*, 37(6), Dec. 2018. ISSN 0730-0301. doi: 10.1145/3272127.3275077.

[51] A. Torre. *Linear ray and wave optics in phase space: bridging ray and wave optics via the Wigner phase-space picture*. Elsevier, 2005.

[52] Z. Velinov, S. Werner, and M. B. Hullin. Real-time rendering of wave-optical effects on scratched surfaces. *Computer Graphics Forum*, 37(2):123–134, 2018. doi: 10.1111/cgf.13347. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13347`.

[53] A. Walther. Radiometry and coherence. *JOSA*, 58(9):1256–1259, 1968. doi: 10.1364/JOSA.58.001256.

[54] S. Werner, Z. Velinov, W. Jakob, and M. Hullin. Scratch iridescence: Wave-optical rendering of diffractive surface structure. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 36(6), Nov. 2017. doi: 10.1145/3130800.3130840.

[55] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40:749–759, Jun 1932. doi: 10.1103/PhysRev.40.749. URL `https://link.aps.org/doi/10.1103/PhysRev.40.749`.

[56] A. Wilkie, S. Nawaz, M. Droske, A. Weidlich, and J. Hanika. Hero wavelength spectral sampling. *Comput. Graph. Forum*, 33(4):123–131, July 2014.

[57] E. Wolf. Coherence and radiometry. *JOSA*, 68(1):6–17, 1978. doi: 10.1364/JOSA.68.000006.

[58] E. Wolf. New theory of partial coherence in the space–frequency domain. part i: spectra and cross spectra of steady-state sources. *J. Opt. Soc. Am.*, 72(3):343–351, Mar 1982. doi: 10.1364/JOSA.72.000343.

[59] E. Wolf. *Introduction to the theory of coherence and polarization of light*. Cambridge University Press, Cambridge, 2007. ISBN 978-0521822114.

[60] L.-Q. Yan, M. Hašan, B. Walter, S. Marschner, and R. Ramamoorthi. Rendering specular microgeometry with wave optics. *ACM Trans. Graph.*, 37(4):75:1–75:10, July 2018. ISSN 0730-0301. doi: 10.1145/3197517.3201351.

[61] A. Zangwill. *Modern electrodynamics*. Cambridge University Press, Cambridge, 2013. ISBN 978-0521896979.

[62] T. Zeltner, I. Georgiev, and W. Jakob. Specular manifold sampling for rendering high-frequency caustics and glints. *ACM Trans. Graph.*, 39(4), Aug. 2020.