A Generic Framework for Physical Light Transport

SHLOMI STEINBERG, University of California, Santa Barbara, USA
LING-QI YAN, University of California, Santa Barbara, USA

Fig. 1. The ability of a light beam to produce observable wave interference phenomena evolves globally: A small but powerful white LED source (marked by a yellow circle) illuminates a scene. The light falls upon the head of a desk lamp made of brushed aluminum, however (a) no diffractive effects are visible because the lamp is close to the source. The light beam is then incident upon a Venus de Milo statue made of scratched bronze. The illumination of the upper part of the statue is dominated by direct incident light, and (b) visible interference patterns arise. On the other hand, light reaching the lower parts of the statue is diffused by a large decorative vase filled with water, altering the coherence properties of the light and (c) diminishing the observable diffraction effects. Rendering is done using a bi-directional path tracer that propagates coherence information, under our formalism, from the light sources. The lamp head and the scratches on the statue are rendered using the tools developed in Subsection 8.1, propagation of partially-coherent light through participating media is discussed in Subsection 8.3 and the other surfaces are rendered classically. See Subsection 8.4 for additional implementation details. The insets were individually rendered and have adjusted contrast and exposure for visualization proposes.

Physically accurate rendering often calls for taking the wave nature of light into consideration. In computer graphics, this is done almost exclusively locally, i.e. on a micrometre scale where the diffractive phenomena arise. However, the statistical properties of light, that dictate its coherence characteristics and its capacity to give rise to wave interference effects, evolve globally: these properties change on, e.g., interaction with a surface, diffusion by participating media and simply by propagation. In this paper, we derive the first global light transport framework that is able to account for these properties of light and, therefore, is fully consistent with Maxwell's electromagnetic theory. We show that our framework is a generalization of the classical, radiometry-based light transport—prominent in computer graphics—and retains some of its attractive properties. Finally, as a proof of concept, we apply the presented framework to a few practical problems in rendering and validate against well-studied methods in optics.

CCS Concepts: • **Computing methodologies** → **Rendering**; **Computer graphics**; • **Applied computing** → *Physics*.

Additional Key Words and Phrases: wave-optics, light transport, path tracing, optical coherence, radiometry, rendering, electromagnetism, interference, diffraction, scattering, rendering equation, volumetric rendering, radiative transfer

ACM Reference Format:
Shlomi Steinberg and Ling-Qi Yan. 2021. A Generic Framework for Physical Light Transport. *ACM Trans. Graph.* 40, 4, Article 139 (August 2021), 20 pages. https://doi.org/10.1145/3450626.3459791

Authors' addresses: Shlomi Steinberg, p@shlomisteinberg.com, University of California, Santa Barbara, 2119 Harold Frank Hall, Santa Barbara, California, 93106, USA; Ling-Qi Yan, lingqi@cs.ucsb.edu, University of California, Santa Barbara, 2119 Harold Frank Hall, Santa Barbara, California, 93106, USA.

1 INTRODUCTION

Light transport is a theoretical underpinning of 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, light transport have been formulated under the context of classical radiometry, with the core quantities being values of intensity and radiance. However, it has been conclusively shown that radiometry is not consistent with more physical foundations of optics, such as Maxwell's electromagnetic theory. Indeed, a radiometric foundation lacks the rigour to perceive the wave-nature of light, and derived frameworks are not able to accurately account for the grating patterns that appear on compact disks (CD), holographs and LCD screens; goniochromatic surfaces and iridescence effects in coated or painted materials and in layered materials such as the human skin; certain atmospheric and liquid scattering effects; pleochroism and birefringent effects in molded plastics and framed glass; the unique visual response of cloth and fabric; lens- and aperture-induced artefacts; as well as other effects and phenomena that arise due to diffraction.

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 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 (see illustrative example in Fig. 1). Furthermore, temporal coherence effects are only resolved during the observation of light, that is at the eye or camera and not locally at the source of diffraction.

This restricted coherence of light arises due to the fact that the natural light that we observe daily can be physically regarded as a collection of very many individual electromagnetic waves, each acting independently but together orchestrating the different properties of each beam of natural light. This greatly hinders the success of any deterministic attempt to model natural light, and instead calls for a stochastic formalism. Over the past few decades such optical formulations that deal with observable, measurable quantities took shape in the optical literature. An important theoretical conclusion from that work is that single-point stochastic descriptions, such as the classical radiometric radiance (which, being a time-averaged quantity, can be considered as a first-order moment), are insufficient to physically describe the propagation and diffraction of natural light [Wolf 2007]. Rather, a two-point formalism, which provides richer information and can quantify the statistical correlation of an *ensemble* of many waves at two spacetime points, is required.

Building upon those foundations, we present the first general-purpose global light transport theory that deals exclusively with partially-coherent natural light. Our guiding physical principle is the (non-relativistic) electromagnetic theory, with which our framework is fully consistent. To be able to describe the two-point coherence information of light, our basic quantities are functions, and sourcing, propagation, diffraction and light-matter interactions are modelled in terms of operators that act upon those functions. Our contributions can be summarised as follows:

- Our theory retains some of convenience of classical radiance and light transport is formalised via the *spectral-density transport equation*, a "wave optics"-analogue of the rendering equation (Section 4).
- We formally show that our framework reduces in the short-wavelength limit to the classical radiometric formulation of light ("geometric optics"), thus in essence generalizing traditional light transport (Appendix B).
- To analytically quantify the radiation we call "natural light", we define a model of what constitutes a natural light source and formally derive the analytic expressions for electromagnetic radiation sourced from that model (Section 5). Despite the variety of such light sources, natural and artificial, they are all driven by the same quantum mechanical principle–spontaneous emission—and hence our model is able to capture the defining characteristic of virtually all such sources. We show good agreement with well-known theory and experimental results.
- We discuss the propagation of partially-coherent light under our formalism and derive the relevant diffraction formulae (Section 6).
- Finally, we present a few practical applications of our framework, and validate against a deterministic model formulated using established methods in electromagnetism (Section 8).

This paper is intended to serve as a theoretical foundation. The focus of our discussion is set on a general-purpose framework and ignores practical technical considerations. While we provide a complete formulation of free-space propagation and sourcing of natural light, due to the breadth and complexity of a comprehensive light-matter discussion, such a discussion is mostly left for future work. Hence, we derive light-matter interaction operators only for simple and idealised cases (Section 7), which limits the immediate practical applicability of our work. Nevertheless, we do demonstrate a couple of applications (Section 8) that show that our theory is practically suitable to describing local and global transport problems.

1.1 Motivation for Our Work

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 traditional 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" (see Appendix B). 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 (see examples in Section 8).

Generalized radiance. Similar notions have been studied before: In optical literature, the seminal paper by Walther [1968] 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 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. (12)) 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 [Friberg 1979].

This non-negativity prohibits the interpretation of the WDF 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. Being a space and spatial frequency joint distribution, the WDF is also more difficult to reason about. This motivates us to drop a search for a "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 4), giving rise to a general purpose framework of partially-coherent light transport.

2 RELATED WORK

Of particular relevance is the work by Oh et al. [2010], 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. Furthermore, a WDF is a more analytically difficult quantity to work with (compared to the cross-spectral density) and even so does not adhere to all the characteristics of radiometric radiance. The shortcomings of this approach were discussed in Subsection 1.1.

In an attempt to quantify the limited spatial coherence of radiation, a Gaussian footprint is used to modulate coherent contributions for the purpose of rendering diffractive scratches [Werner et al. 2017], measuring diffraction grating patterns using Jones calculus [Toisoul et al. 2018] and the rendering of biological diffractive surfaces [Dhillon et al. 2014]. 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 [Falster et al. 2020; Yan et al. 2018], and simple square kernels for the construction of high-resolution BRDFs [Levin et al. 2013]. Two-point information has been used by Steinberg and Yan [2021], 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. [2015]; Kotwal et al. [2020] 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 [Bar et al. 2019, 2020]. A large body of work has applied the electromagnetic theory locally, for the study of a particular optical effect. Stam [1999] presented the first work that considered the electromagnetic nature of light. Additional works studied diffraction-aware BSDFs [Cuypers et al. 2012; Toisoul and Ghosh 2017; Velinov et al. 2018], iridescence in pearlescent materials [Guillén et al. 2020], thin-film interference [Belcour and Barla 2017; Kneiphof et al. 2019] and real-time rendering of birefringent materials [Steinberg 2019]. Sadeghi et al. [2012] 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 [2020] renders liquid-crystal micrographs via an approximative analytic solution to a light transport differential system. Musbach et al. [2013] 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. [2018].

The "coherency matrix" was employed in computer graphics for studying some local cases of coherent light transport [Guy and Soler 2004; Wilkie et al. 2001], though only to quantify the state of polarization of light and not in-order to propagate partially-coherent light.

3 THEORETICAL FOUNDATIONS

The purpose of this section is threefold: (i) Outline the required theoretical foundations in brief. The curious reader is referred to the more comprehensive and rigorous background that is presented in our supplemental material. (ii) Summarise contemporary light transport formalisms, as well as related work that have been employed in computer graphics and computational optics. (iii) Finally, discuss the shortcomings of the state-of-the-art. We also use this opportunity to familiarize the reader with our notation.

Notation. See Table 1 for a list of symbols and notation. Most of our notation is self-explanatory, however we would like to point out two potential sources of confusion. The common conventions for the definition of the Fourier transform vary in terms of the kernel and normalization factors. Our definition for the (spatial) n-dimensions transform is the non-unitary, angular frequency version commonly employed in optical literature:

$$\mathscr{F}\left\{f\right\}\left(\vec{\mathbf{k}}\right) \triangleq \int_{\mathbb{R}^n} \mathrm{d}^n\vec{\mathbf{r}}'\, f\left(\vec{\mathbf{r}}'\right) e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}'} \tag{1}$$

$$\mathscr{F}^{-1}\left\{g\right\}\left(\vec{\mathbf{r}}\right) \triangleq \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathrm{d}^n\vec{\mathbf{k}}'\, g\left(\vec{\mathbf{k}}'\right) e^{i\vec{\mathbf{k}}'\cdot\vec{\mathbf{r}}} \tag{2}$$

This choice is justified by the fact that in this form the inverse Fourier transform above represents the superposition of plane-wave modes that compose an analytic signal. In addition, note also the following vector notation (borrowed from [Zangwill 2013]) which we use throughout the paper: Given some vector $\vec{\mathbf{r}}$, we denote $\hat{\mathbf{r}}$ as the unit vector in direction of $\vec{\mathbf{r}}$ and its magnitude as $r = |\vec{\mathbf{r}}|$.

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* [Kajiya 1986], to which we refer as the *light transport equation* (LTE):

$$L_o\left(\vec{\mathbf{p}}, \Omega_o, \omega, t\right) = L_e\left(\vec{\mathbf{p}}, \Omega_o, \omega, t\right) + \int_{\mathcal{S}^2} \mathrm{d}^2\Omega'\, f_r\left(\vec{\mathbf{p}}, \Omega', \Omega_o, \omega, t\right) L_i\left(\vec{\mathbf{p}}, \Omega', \omega, t\right) \left|\Omega'\cdot\hat{\mathbf{n}}\right| \tag{3}$$

where L_o is the exitant spectral radiance, L_e is the emitted radiance, L_i is the incident spectral radiance, $\mathcal{S}^2$ is the unit sphere centred at the point of interest $\vec{\mathbf{p}}$, 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 [Wolf 1978] and, as we shall formally demonstrate in Section 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,

Table 1. List of symbols and notation (location of definition on the right)

NOTATION AND SYMBOLS

Notation

$\vec{\mathbf{r}}$	Vectors: arrow accented boldface, typically lower-case latin letters	
$r = \|\vec{\mathbf{r}}\|$	Vector's magnitude: scalars sharing the same letter as a vector	
$\hat{\mathbf{r}} = \vec{\mathbf{r}}/r$	Unit vectors: hat accented boldface	
$\boldsymbol{A}$	Matrices and linear operators: boldface, capital letters	
$\mathscr{L}$	Operators: script, typically capital latin letters	
$\vec{\mathbf{r}}'$	Primed variables usually denote local or integration variables	

Symbols

$\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$	The cartesian unit vectors spanning $\mathbb{R}^3$	
$\vec{\mathbf{E}}$	Electric field	(Page 5)
$\vec{\mathbf{B}}$	Magnetic field	(Page 5)
c	The speed of light	
λ	Wavelength	(Page 5)
$\vec{\mathbf{k}}$	Wavevector	(Page 5)
$k = \frac{2\pi}{\lambda}$	Wavenumber	(Page 5)
$\omega = ck$	Wave's angular frequency	(Page 5)
J_ν	Bessel function of the first kind of order ν	
$\mathcal{S}^2$	Unit sphere centred at the origin	
Γ	Mutual coherence function	(Eq. (11))
$\mathcal{W}, \boldsymbol{\mathcal{W}}$	Cross-spectral density (scalar and matrix forms)	(Eqs. (12) and (14))
$\mathcal{L}, \boldsymbol{\mathcal{L}}$	RCSD: cross-spectral density with units of radiance	(Eq. (18))

Operators

$\delta(x)$	Dirac delta	
$\mathscr{F}\{\cdot\}$	Fourier transform operator	(Eq. (1))
$z^\star$	Complex conjugate of $z \in \mathbb{C}$	
$\boldsymbol{A}^\dagger$	Conjugate transpose of $\boldsymbol{A}$	
$*$	Convolution operator	
$\langle\cdot\rangle$	Ensemble average	(Subsection 3.3)
$\langle\cdot\rangle_\omega$	Ensemble average over same frequency realizations	(Subsection 3.3)

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.

3.2 Electromagnetic Waves

We begin with a very brief overview of relevant Maxwellian electromagnetism. It is a well-known fact that observed light is electromagnetic energy, and optical phenomena are well described by Maxwell's phenomenological theory when the atomic structure of matter plays no role [Born and Wolf 1999]. Hence, rigorous (classical) optical theories and results can always be traced to Maxwell's celebrated equations, which in free-space (away from charge and current) and under Gaussian-cgs units take the following form:

$$\nabla \cdot \vec{\mathbf{E}} = 0 \qquad \nabla \cdot \vec{\mathbf{B}} = 0 \tag{4}$$

$$\nabla \times \vec{\mathbf{E}} = -\frac{1}{c}\frac{\partial \vec{\mathbf{B}}}{\partial t} \qquad \nabla \times \vec{\mathbf{B}} = \frac{1}{c}\frac{\partial \vec{\mathbf{E}}}{\partial t} \tag{5}$$

where t is time and c is the speed of light. The intrinsic electromagnetic quantities are the vectors $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$, which are the electric and magnetic fields, respectively. It is these fields that describe the induced excitations in space, and the interaction between these fields, quantified by Maxwell's equations, results in self-supporting propagating fields that oscillate in unison and give rise to time-harmonic electromagnetic waves, i.e. light. Microscopic charge and current are ignored as we focus on free-space propagation of light, and because sourcing of natural light is better explained quantum mechanically rather than classically (we will discuss sourcing in Section 5).

By taking the curl of each of the dynamic Maxwell equations above (Eq. (5)) we arrive at the (homogeneous) wave equations governing the propagation of electromagnetic waves:

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\mathbf{E}} = 0 \qquad \left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\mathbf{B}} = 0 \tag{6}$$

where $\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$ is known as the *d'Alembert wave operator*. Any physically-realizable electromagnetic waves, formalised in terms of their $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields, must satisfy Eqs. (4) and (6). For this reason, we treat the set of equations above as the first principles of optics and the foundations of physical light transport.

An important class of solutions to these equations are plane waves, which are a form of transverse electromagnetic waves. Those kind of waves admit simple and intuitive geometric and analytic properties, making them ubiquitous in practical applications, like computer rendering. A (monochromatic) plane wave takes the following form:

$$\vec{\mathbf{E}}(\vec{\mathbf{r}}, t) = \vec{\mathbf{a}}_\perp e^{i(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}} - \omega t)} \tag{7}$$

$$\vec{\mathbf{B}}(\vec{\mathbf{r}}, t) = -(\hat{\mathbf{k}} \times \vec{\mathbf{a}}_\perp) e^{i(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}} - \omega t)} \tag{8}$$

where $\vec{\mathbf{a}}_\perp$ is a constant vector describing both the field's peak amplitude and the wave's (linear) polarization direction, $\vec{\mathbf{k}}$ is the *wavevector*—the plane wave's direction of propagation—and its magnitude, $k = |\vec{\mathbf{k}}|$, is known as the *wavenumber* and is related to the wavelength λ via the relation $k = 2\pi/\lambda$. For a plane wave, in order to satisfy Eq. (4), the wavevector must be perpendicular to the polarization direction, viz. $\vec{\mathbf{k}} \cdot \vec{\mathbf{a}}_\perp = 0$. We also define the unit vector $\hat{\mathbf{k}} = \vec{\mathbf{k}}/k$ and $\omega = ck$ as the angular frequency. Observe that $\vec{\mathbf{a}}_\perp$ and $\vec{\mathbf{k}}$ fully define a plane wave, with has important consequences for the study of far-field propagation: the electric and magnetic fields do not need to be quantified individually. Plane waves extend to infinity in the transverse direction (perpendicular to the wavevector), hence are an aphysical construct. Nonetheless, they are a good local approximation to "real" electromagnetic waves in a confined region of space, far from the radiation's source.

Our sensors that observe light (e.g., the eye, camera) measure not the field strength but its intensity, and as these sensors observe electromagnetic radiation over periods long with respect to the angular frequency ω, it is the time-averaged quantities that are of significance to us and not instantaneous values. For a plane wave, the time-averaged intensity (irradiance impinging on a differential surface area perpendicular to $\vec{\mathbf{k}}$, not to be confused with radiometric intensity) becomes [Zangwill 2013]:

$$I(\vec{\mathbf{r}}) = \frac{c}{8\pi}|\vec{\mathbf{a}}_\perp|^2 \tag{9}$$

i.e. proportional to the field strength squared.

Limitations. The formalism briefly outlined above admits a few major shortcomings frustrating its practical use in computer graphics applications: First, apart from the most simple settings, solutions to Eqs. (4) and (6) are difficult to compute; and, crucially, natural light is not easily described by its electric and magnetic fields. The last point arises from the fact that natural light is highly chaotic, being composed of very many different electromagnetic waves of distinct wavelengths, polarizations, field strengths and wavevectors. Hence, even when tractable solutions are found, they fail to account for that disorganized, i.e. *incoherent*, behaviour of natural light, meaning that those solutions are typically only able to predict observable optical phenomena over very short spatio-temporal lengths.

3.3 Wave Ensembles

The intricate inter-play between the electromagnetic wave constituents that form natural light dictates how light energy propagates and interacts with light and matter, and thus its observable and measurable properties. Those constituents are typically numerous, and as fluctuations at optical frequencies are not directly measured, the observable quantities are time-averaged values. This prompts us to shed a deterministic analysis in favour of a stochastic one. Typically, a detailed study of each constituent is neither feasible nor even desirable, instead we model light as an ensemble of very many individual waves and study its statistics. This is the domain of optical coherence theory, which can be thought of as the study of the measurable quantities of optical phenomena, and this subsection serves as background. Deep understanding of stochastic processes or optical coherence theory is not required for this paper, therefore we try to keep the discussion qualitative and as concise as possible. The reader is encouraged to review Goodman [2015]; Wolf [2007] or our supplemental material for a more complete and rigorous discussion.

A stochastic process is a functional analogue of a random variable. While a random variable maps the sample space to a value, a stochastic process is a mapping from the sample space to a function space. Consider such a stochastic process, denoted U, describing the scalar electromagnetic disturbances over spacetime. Being a scalar value, polarization is ignored and only the field strength is considered. This is done in-order to simplify the following discussion, however a unified theory of polarization and coherence will be presented later. The functions that U maps to, i.e. the possible outcomes of experiments that measure U, are functions from spacetime to scalar real values describing field strength, and each is an appropriate solution to Maxwell equations (up to a chosen field direction). We call each such function a *realization* of the stochastic process, and the collection—the *ensemble*—of all such realization together with their joint probability density functions fully describes U.

A restricted class of stochastic processes, known as *ergodic* processes, possess the property where a single realization fully describes the statistics of the entire process. That is, the expected outcome of an experiment measuring U, i.e. the *ensemble average* over the entire ensemble of realizations, is equal to the infinite time-average of a single realization. Formally, if we denote the ensemble average of the signal U as $\langle U(\vec{\mathbf{r}},t)\rangle$, then for ergodic processes it holds that

$$\langle U(\vec{\mathbf{r}},t)\rangle \triangleq \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\mathrm{d}t\,U(\vec{\mathbf{r}},t) \tag{10}$$

Ergodicity applies similarly to higher-order moments: For an ergodic process ensemble moments are equivalent to time moments. Of particular importance is the second-order moment, the *mutual coherence function*:

$$\Gamma(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) \triangleq \left\langle U(\vec{\mathbf{r}}_1,t)U(\vec{\mathbf{r}}_2,t+\tau)^\star\right\rangle \tag{11}$$

which describes the statistical similarity between the wave ensemble at two different space points and with a time shift of τ. Note that $\langle U(\vec{\mathbf{r}},t)\rangle$ and in turn $\left\langle U(\vec{\mathbf{r}}_1,t)U(\vec{\mathbf{r}}_2,t+\tau)^\star\right\rangle$ are independent of t.

Consider a lamp producing polychromatic natural light: Typically, the expected observed instantaneous intensity at some point is simply the average intensity over a single, sufficiently long measurement. Furthermore, that measurement also fully captures the source's emitted spectra and the temporal coherence (the correlation between the wave ensemble compared to a time-shifted version of itself) at that point. Likewise, the (instantaneous) correlation between waves reaching a pair of spacetime points would equal the correlation between a pair of measured realizations at those points (with the appropriate time-shift). That is, most physical processes that are of relevance to us are indeed ergodic. Nevertheless, the results in this paper apply, either directly or with some additional work, to wide-sense stationary processes, a less restrictive class. However, for simplicity, we make no formal distinction between those classes of stochastic processes.

Space-frequency formulation. An important result in optical coherence theory is the *coherent-modes representation* [Wolf 1982]: A construction of an ensemble that consists of monochromatic waves that is faithful to the original wave ensemble (that is, it describes the same stochastic process). Such an ensemble of monochromatic waves can be formally expressed as $\{u(\vec{\mathbf{r}},\omega)e^{-i\omega t}\}$ with different values of ω—the angular frequency—and where u are the time-independent realizations. Ensemble averages over such realizations give rise to the primary quantity of interest, the *cross-spectral density*:

$$\mathcal{W}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) \triangleq \left\langle u(\vec{\mathbf{r}}_1,\omega)u(\vec{\mathbf{r}}_2,\omega)^\star\right\rangle_\omega \tag{12}$$

where the $\langle\cdot\rangle_\omega$ operator should be read as "ensemble average over same frequency realizations", which implies the decomposition into monochromatic waves. Physically, the cross-spectral density should be understood as the correlation between same-frequency components of the field (components that oscillate exactly as $e^{-i\omega t}$ with respect to time) at two different space points. Mathematically, the cross-spectral density is a positive semi-definite function that is the spectral decomposition of the mutual coherence function, formally $\mathcal{W}=\mathscr{F}\{\Gamma\}$ when the Fourier transform exists and with the transform being with respect to time. It is noteworthy that such a spectral decomposition into same-frequency waves always exists in practice, and is established by the Wiener–Khinchin theorem (even when the mutual coherence function is not Fourier-transformable) [Goodman 2015].

In the far-field, the wave ensemble can be regarded as a collection of plane-waves with an essentially identical wavevector $\vec{\mathbf{k}}$, in which case a realization $U(\vec{\mathbf{r}},t)$ describes the instantaneous transverse field strength and $|\vec{\mathbf{a}}_\perp|^2 = \left\langle UU^\star\right\rangle$ (see Eq. (9)). Hence, as the cross-spectral density is the spectral decomposition of the mutual coherence function, it is easy to show that the time-averaged observed intensity of the wave ensemble is

$$I(\vec{\mathbf{r}}) = \frac{c}{8\pi}\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}},0) = \frac{c}{8\pi^2}\int_0^\infty \mathrm{d}\omega\,\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}},\omega) \tag{13}$$

which implies that distinct frequencies must be uncorrelated in an ergodic wave ensemble, and we sum up those contributions incoherently. This makes the space-frequency formulation of optical coherence of great convenience for our purpose: As typically is the case with spectral rendering, we treat and render each spectral component separately. To accurately reproduce diffractive phenomena, statistical knowledge of the same-frequency constituents only is sufficient, and different frequencies are summed up on an intensity basis.

Polarization and coherence. So far in this section we have discussed a scalar wave ensemble. In order to achieve full agreement with electromagnetism, we need to consider a stochastic process of the vector field $\vec{\mathbf{E}}(\vec{\mathbf{r}},t)$, which quantifies the electric field distribution over spacetime. This is done by decomposing a wave ensemble into orthogonal transverse components and considering the second-order statistics between these components. This gives rise to a 2×2 matrix representation of the cross-spectral density, that describes both the coherence and the polarization properties of the wave ensemble, as well as the interactions between these properties. Assume that the direction of propagation of the wave described by $\vec{\mathbf{E}}$ is concentrated along the direction $\hat{\mathbf{k}}$. Denote $\{\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2\}$ as the orthonormal basis of the orthogonal complement of $\hat{\mathbf{k}}$, and let $E_{1,2}=\hat{\mathbf{e}}_{1,2}\cdot\vec{\mathbf{E}}$ be the mutually-orthogonal transverse components of the electric field $\vec{\mathbf{E}}$. Then, the cross-spectral density matrix is defined as [Wolf 2007]

$$\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) \triangleq \begin{bmatrix} \left\langle E_1(\vec{\mathbf{r}}_1)E_1(\vec{\mathbf{r}}_2)^\star\right\rangle_\omega & \left\langle E_1(\vec{\mathbf{r}}_1)E_2(\vec{\mathbf{r}}_2)^\star\right\rangle_\omega \\ \left\langle E_2(\vec{\mathbf{r}}_1)E_1(\vec{\mathbf{r}}_2)^\star\right\rangle_\omega & \left\langle E_2(\vec{\mathbf{r}}_1)E_2(\vec{\mathbf{r}}_2)^\star\right\rangle_\omega \end{bmatrix} \tag{14}$$

When evaluated at $\vec{\mathbf{r}}_1 = \vec{\mathbf{r}}_2$, the matrix above is Hermitian and positive semi-definite and thus fulfils the conditions of a *polarization matrix*, sometimes referred to as the "coherency matrix" [Born and Wolf 1999] (which is a misnomer as the polarization matrix provides only single-point information). See Appendix A for a summary of the properties of those matrices. The diagonal elements in the matrix above express the coherence properties of the light, while the off-diagonal elements describe the polarization characteristics.

It is noteworthy that the notation of $\boldsymbol{\mathcal{W}}$ abstracts the underlying orthonormal basis that dictates the decomposition into transverse components. Cross-spectral density matrices of the same wave ensemble but expressed using different decomposition bases are all linearly related to each other (see Section 7) and they encode the same information. Nonetheless, some operators that act upon those matrices expect a specific decomposition, consequently when dealing with cross-spectral density matrices it is of necessity to bear in mind that a decomposition basis is implied.

Discussion. Treating light as a stochastic process and committing to a stochastic analysis gives rise to a formalism that can account for the chaotic nature of natural light. The presented formalism models the electromagnetic field disturbances as an ensemble of very many realizations, each a solution to Maxwell's homogeneous free-space equations (Eqs. (4) and (6)) and by the superposition principle any linear combination of those realizations is also a solution. A question that is left to be answered is: are second-order statistics sufficient in describing natural light? To see why that is truly the case, recall that as already discussed, our sensors that observe light do so over long periods. We are now in a position to elaborate on that statement: If t is the exposure time of a camera or the rendering time of a frame, then in practice it would typically hold that $t \gg \tau_0$, where τ_0 is the characteristic time-length of the observed light—more formally, it is the period over which $\Gamma(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \tau_0)$ remains non-negligible. Indeed, with $\mathcal{W}$ and Γ being Fourier-transform pairs, for natural, wide-spectrum polychromatic light τ_0 would be small (direct calculations show that τ_0 would be on the order of nanoseconds for relevant blackbody radiation). Therefore, over the exposure time many uncorrelated waveforms would contribute to the observed intensity, and thus by the central limit theorem the observed intensity is asymptotically (point-wise) normally distributed, and the second-order moments are sufficient [Goodman 2015].

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 $\delta\vec{\mathbf{p}} \subset \mathbb{R}^3$ (the notation $\delta\vec{\mathbf{p}}$ is used to describe a small two-dimensional patch or some small volume around the point $\vec{\mathbf{p}}$) and then propagated to a point $\vec{\mathbf{r}} \in \mathbb{R}^3$. 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

$$\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) \approx I(\vec{\mathbf{r}}, \omega)\gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) \tag{15}$$

for points $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$ close to $\vec{\mathbf{r}}$, i.e. $\vec{\mathbf{r}}_{1,2} \in \delta\vec{\mathbf{r}}$. 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: $|\gamma| \leq 1$, $\gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) = \gamma(\vec{\mathbf{r}}_2, \vec{\mathbf{r}}_1)^\star$ and $\gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) = 1$ when $\vec{\mathbf{r}}_1 = \vec{\mathbf{r}}_2$. 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. (12)). Hence, akin to time-averaged intensity of an electromagnetic wave (Eq. (9)), 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 $\delta\vec{\mathbf{r}}$ at the source $\vec{\mathbf{p}}$, via $\mathrm{d}A_r = r^2\,\mathrm{d}\Omega$ (with $r = |\vec{\mathbf{r}}|$, as usual), thus

$$I = \frac{\partial\Phi}{\partial A_r} = \frac{1}{r^2}\frac{\partial\Phi}{\partial\Omega} \tag{16}$$

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

$$r^2\frac{\partial I}{\partial A_s\cos\theta} = \frac{\partial^2\Phi}{\partial\Omega\partial A_s\cos\theta} \tag{17}$$

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 *radiance cross-spectral density* (RCSD) function as

$$\mathcal{L}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) \triangleq r^2\frac{\partial\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega)}{\partial A_s\cos\theta} \tag{18}$$

and the matrix RCSD $\boldsymbol{\mathcal{L}}$ is defined similarly. Substitute the decomposition outlined in Eq. (15) 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: $\mathcal{L}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ 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 $\delta\vec{\mathbf{p}}$ with normal $\hat{\mathbf{n}}$. The irradiance, in terms of a cross-spectral density matrix, impinging upon that surface patch can be written as the superposition of the incident RCSD matrices, viz.

$$\boldsymbol{\mathcal{W}}\Big|_{\vec{\mathbf{p}}} = \int_{\mathcal{S}^2}\mathrm{d}^2\Omega'\,\big|\Omega'\cdot\hat{\mathbf{n}}\big|\,\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{p}}} \tag{19}$$

where $\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{p}}}$ is the RCSD of the incident radiation that arrived from a point $\vec{\mathbf{r}}'$ (that lies in direction Ω' from $\vec{\mathbf{p}}$). 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 $\mathscr{D}_{\delta\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}}$: an abstract diffraction operator—the "wave optics" generalization of the classical BSDF—that diffracts and propagates the cross-spectral density on the surface patch $\delta\vec{\mathbf{p}}$ to a region in space $\delta\vec{\mathbf{r}}$. Then,

$$\begin{aligned}\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}} &= \mathscr{D}_{\delta\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}}\left\{\boldsymbol{\mathcal{W}}\Big|_{\vec{\mathbf{p}}}\right\}\\ &= \mathscr{D}_{\delta\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}}\left\{\int_{\mathcal{S}^2}\mathrm{d}^2\Omega'\left|\Omega'\cdot\hat{\mathbf{n}}\right|\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{p}}}\right\}\end{aligned}\tag{20}$$

with $\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}}$ being the quantity of interest: the RCSD of the radiation scattered from $\vec{\mathbf{p}}$ and propagated to region $\delta\vec{\mathbf{r}}$. Making the assumption that $\mathscr{D}$ is a linear operator and formally interchangeable with the integral in Eq. (20), we are now ready to formulate the *spectral-density transport equation* (SDTE):

Definition 4.1 (Spectral-Density Transport Equation). The SDTE governs the transport of light at the boundary between two media:

$$\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}} = \boldsymbol{\mathcal{L}}^{(e)}_{\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}} + \int_{\mathcal{S}^2}\mathrm{d}^2\Omega'\left|\Omega'\cdot\hat{\mathbf{n}}\right|\mathscr{D}_{\delta\vec{\mathbf{p}}\to\delta\vec{\mathbf{r}}}\left\{\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{p}}}\right\}\tag{21}$$

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

Clear similarities can be drawn between the classical LTE (Eq. (3)) 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. (19) 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 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 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. Once we have developed our theory, we return to this discussion in Appendix B to formally show what that at the short-wavelength limit the RCSD functions can be understood as classical radiance and thus verify our claim.

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. However, our framework can be adapted to essentially any geometry in practice (see the derivations our supplemental material).

Those assumptions do not always hold for real light source: Voltage and temperature shift across the source, and especially

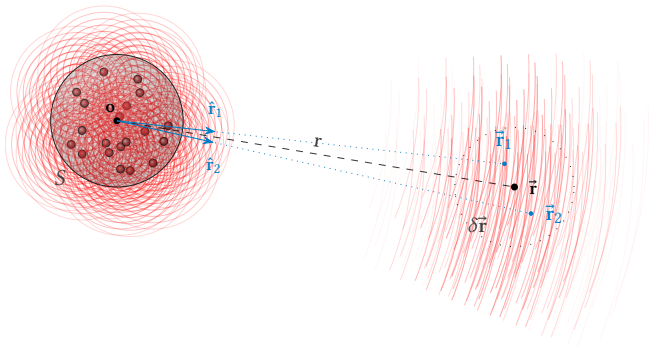

Fig. 2. A spherical natural light source of volume S is centred at the origin $\mathbf{o}$. The source consists of very many *elementary radiators* (illustrated as red balls) that radiate isotropically (wavefronts illustrated as red curves). This gives rise to the highly chaotic radiation that we call *natural light*. To quantify the statistical properties of such light that has propagated a distance r to a small region $\delta\vec{\mathbf{r}}$ far from the source, we examine the statistical similarity of the wave ensemble at some points $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$ (in directions $\hat{\mathbf{r}}_1$, $\hat{\mathbf{r}}_2$ from the center of the source) in that region.

towards the edges, producing spatially-varying emission characteristics. Nonetheless, it is reasonable to assume that the spatially-varying effects of certain sources usually have only a small impact on the produced radiation: limb darkening and a drop in emission power at the edge of the source can be accounted for by appropriately reducing the size of the model source. Those effects are also source-dependent, and as our aim is to derive a general theory that is applicable across a wide range of light sources, we choose to ignore those spatially-varying peculiarities and leave discussion of particular light sources for future work. Anyhow, we capture the defining characteristic of all spontaneous emission natural sources: a large collection of elementary radiators, collectively emitting light but radiating independently [Guryev 2012], and we will show good agreement with experimental data.

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. 2 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 integral of the first kind [Born and Wolf 1999]:

$$u(\vec{\mathbf{r}}) = \frac{1}{2\pi}\int_{\Sigma} \mathrm{d}^2\vec{\mathbf{r}}'\, u(\vec{\mathbf{r}}')\nabla_{\hat{\mathbf{r}}}\left[\frac{e^{ik|\vec{\mathbf{r}}-\vec{\mathbf{r}}'|}}{|\vec{\mathbf{r}}-\vec{\mathbf{r}}'|}\right] \tag{22}$$

where Σ is a diffracting aperture (centred at the origin), as before k is the wavenumber and $\nabla_{\hat{\mathbf{r}}}$ is the directional derivative in direction $\hat{\mathbf{r}}$, i.e. the direction of the point-of-interest $\vec{\mathbf{r}}$ from the source. A scalar diffraction formalism is sufficient because elementary radiators radiate with random polarization. Applying Eq. (22) to the definitions of the cross-spectral density (Eq. (12)) and the RCSD (Eq. (18)), 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:

$$\boldsymbol{\mathcal{L}}\left(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega\right) = \frac{1}{\lambda^2}\boldsymbol{I}e^{ik(r_1-r_2)} \times \frac{\partial}{\partial A_s}\int_S \mathrm{d}^3\vec{\mathbf{r}}_1'\, e^{-ik\hat{\mathbf{r}}_1\cdot\vec{\mathbf{r}}_1'}\int_S \mathrm{d}^3\vec{\mathbf{r}}_2'\,\mathcal{W}'\left(\vec{\mathbf{r}}_1',\vec{\mathbf{r}}_2',\omega\right)e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{r}}_2'} \tag{23}$$

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. See our supplemental material for a detailed derivation of Eq. (23). The RCSD above is defined for $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2 \in \delta\vec{\mathbf{r}}$ for some point $\vec{\mathbf{r}}$ away from the source.

Consider the quantity $\mathcal{W}'\left(\vec{\mathbf{r}}_1',\vec{\mathbf{r}}_2',\omega\right)$ that appears in Eq. (23). 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 [Murphy and Tam 2014]), 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 $\vec{\mathbf{r}}_1'$, $\vec{\mathbf{r}}_2'$ increases. However, observe that if we were to set $\mathcal{W}'\left(\vec{\mathbf{r}}_1',\vec{\mathbf{r}}_2',\omega\right) = 0$ for $\vec{\mathbf{r}}_2' \neq \vec{\mathbf{r}}_2'$ inside the source, then Eq. (23) 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 [1999] 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:

$$\mathcal{W}'\left(\vec{\mathbf{r}}_1',\vec{\mathbf{r}}_2',\omega\right) = \Lambda(\omega)e^{-\frac{1}{2}\left[\left(\frac{1}{2}\lambda\right)^{-1}\left|\vec{\mathbf{r}}_2'-\vec{\mathbf{r}}_1'\right|\right]^2} = \Lambda(\omega)e^{-\frac{k^2\left|\vec{\mathbf{r}}_2'-\vec{\mathbf{r}}_1'\right|^2}{2\pi^2}} \tag{24}$$

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. (24) effectively constraints the integration volume of the inner integral in Eq. (23) 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 5.1 (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:

$$\boldsymbol{\mathcal{L}}\left(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega\right)=\mathcal{S}\{\Lambda,\rho\}\triangleq \boldsymbol{I}\frac{\Lambda(\omega)}{\sqrt{\rho}}e^{ik(r_1-r_2)}\frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|\right]}{|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|^{\frac{3}{2}}} \tag{25}$$

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 $\hat{\mathbf{r}}_{1,2}$ 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. See our supplemental material for more details as well as rigorous derivations of Eqs. (22), (23) and (25).

Looking at the limit $\vec{\mathbf{r}}_2\to\vec{\mathbf{r}}_1$, we observe that the radiance carried by the RCSD formalised by Definition 5.1 is independent of the distance r, viz.

$$\lim_{\vec{\mathbf{r}}_2\to\vec{\mathbf{r}}}\boldsymbol{\mathcal{L}}\left(\vec{\mathbf{r}},\vec{\mathbf{r}}_2,\omega\right)=\lim_{\vec{\mathbf{r}}_2\to\vec{\mathbf{r}}}\mathcal{S}\{\Lambda,\rho\}\propto \boldsymbol{I}\rho k^{\frac{3}{2}}\Lambda(\omega) \tag{26}$$

as expected. Also, by rewriting $|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|\approx\sin\alpha$, where α is the separation angle between $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$, 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 [Born and Wolf 1999]. If we were to plug a wavelength of $\lambda=0.50\,\mu\mathrm{m}$, the solar radius $\rho\approx 696\,000\,\mathrm{km}$ and the distance to earth's surface from the sun, $r\approx 1\,\mathrm{au}$, into Definition 5.1, and then solve for the first zero of the Bessel function (≈ 4.49341), this would yield a coherence radius of $|\vec{\mathbf{r}}_1-\vec{\mathbf{r}}_2|\approx 77\,\mu\mathrm{m}$, showing excellent agreement between our model and experimental measurements performed by Mashaal et al. [2012] of sunlight's coherence on earth. Likewise, we also found very good agreement with measurements of the coherence of a white LED [Koivurova et al. 2017].

Our derivations employ far-field diffraction theory (Eq. (22)) that makes the paraxial approximation. To study the validity domain of this approximation, Charnotskii [2019] 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 inside the source (Eq. (24)), 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.

See Appendix B for a discussion of the short-wavelength limit of the sourcing operator (Definition 5.1).

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 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 [Born and Wolf 1999]:

$$\begin{aligned}\vec{\mathbf{E}}(\vec{\mathbf{r}})&=ik\frac{e^{ikr}}{2\pi r}\hat{\mathbf{r}}\times\int_{\Sigma}\mathrm{d}^2\vec{\mathbf{r}}'_{\perp}\left[\hat{\mathbf{n}}\times\vec{\mathbf{E}}(\vec{\mathbf{r}}'_{\perp})\right]e^{-ik\hat{\mathbf{r}}\cdot\vec{\mathbf{r}}'_{\perp}}\\&=ik\frac{e^{ikr}}{2\pi r}\hat{\mathbf{r}}\times\mathscr{F}_{\Sigma}\left\{\hat{\mathbf{n}}\times\vec{\mathbf{E}}\right\}(\hat{\mathbf{r}})\end{aligned} \tag{27}$$

where Σ is an "aperture"—a small planar region in space upon which the electric field $\vec{\mathbf{E}}$ is known, $\hat{\mathbf{n}}$ 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 $\{\hat{\mathbf{e}}'_1,\hat{\mathbf{e}}'_2\}$, with the implied assumption being that the incident radiation is concentrated in some direction $\hat{\mathbf{k}}'$ and thus $\hat{\mathbf{e}}'_1\cdot\hat{\mathbf{k}}'=\hat{\mathbf{e}}'_2\cdot\hat{\mathbf{k}}'=0$ and naturally $\hat{\mathbf{e}}'_1\cdot\hat{\mathbf{e}}'_2=0$. Likewise, the arbitrary orthonormal basis under which the propagated light is decomposed is denoted as $\{\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2\}$, such that $\hat{\mathbf{e}}_1\cdot\hat{\mathbf{r}}=\hat{\mathbf{e}}_2\cdot\hat{\mathbf{r}}=0$ as well as $\hat{\mathbf{e}}_1\cdot\hat{\mathbf{e}}_2=0$. See Fig. 3 for an illustration. By making the Born approximation [Zangwill 2013], 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. (27). 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 $\mathcal{W}'$. We are interested in deriving an expression for the cross-spectral density matrix $\mathcal{W}$ of the radiation that diffracted and propagated to $\vec{\mathbf{r}}$. An element $\mathcal{W}_{\xi\zeta}$ (indexed by $\xi, \zeta \in \{1, 2\}$) of that cross-spectral density matrix (Eq. (14)) can be written as

$$\mathcal{W}_{\xi\zeta}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) = \left\langle \left[\hat{\mathbf{e}}_\xi \cdot \vec{\mathbf{E}}(\vec{\mathbf{r}}_1)\right] \left[\hat{\mathbf{e}}_\zeta \cdot \vec{\mathbf{E}}(\vec{\mathbf{r}}_2)\right]^\star \right\rangle_\omega \tag{28}$$

We now apply the far-field Smythe's diffraction formula, Eq. (27), to each of the field components in the equation above. Exercising some vector calculus, formally interchanging the orders of ensemble averaging and integration and applying a series of elementary algebraic manipulations finally yields the general formula for the diffracted and propagated RCSD:

Definition 6.1 (Diffraction Operator). Let $\mathcal{W}'$ 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:

$$\mathscr{D}\{\mathcal{W}', \Sigma\} \triangleq \frac{e^{ik(r_1 - r_2)}}{\lambda^2 A_\Sigma} \boldsymbol{H}_1 \boldsymbol{\Psi}(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \omega) \boldsymbol{H}_2^T \tag{29}$$

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

$$\boldsymbol{H}_j \triangleq \begin{bmatrix} (\hat{\mathbf{e}}_1 \times \hat{\mathbf{r}}_j) \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_1') & (\hat{\mathbf{e}}_1 \times \hat{\mathbf{r}}_j) \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_2') \\ (\hat{\mathbf{e}}_2 \times \hat{\mathbf{r}}_j) \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_1') & (\hat{\mathbf{e}}_2 \times \hat{\mathbf{r}}_j) \cdot (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_2') \end{bmatrix} \tag{30}$$

with the index $j \in \{1, 2\}$. And, the matrix $\boldsymbol{\Psi}$ is the central quantity of interest: the (double) two-dimensional Fourier transform of the cross-spectral density matrix on the diffracting region, viz.

$$\begin{aligned} \boldsymbol{\Psi}(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \omega) &\triangleq \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}' \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}'' \, \mathcal{W}'(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega) e^{-ik(\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}' - \hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}'')} \\ &= \mathscr{F}_\Sigma\{\mathscr{F}_\Sigma\{\mathcal{W}'(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega)\}(-k\hat{\mathbf{r}}_2)\}(k\hat{\mathbf{r}}_1) \end{aligned} \tag{31}$$

It is easy to verify that both $\boldsymbol{\Psi}(\hat{\mathbf{r}}, \hat{\mathbf{r}}, \omega)$ and, in turn, the RCSD produced by the diffraction operator are polarization matrices (see Appendix A), as expected. See our supplemental material for detailed derivations of Eq. (27) and Definition 6.1. Definition 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 $\hat{\mathbf{r}}_{1,2}$, in contrast to the points $\vec{\mathbf{r}}_{1,2}$, indicating again that spatial coherence stems from propagation.

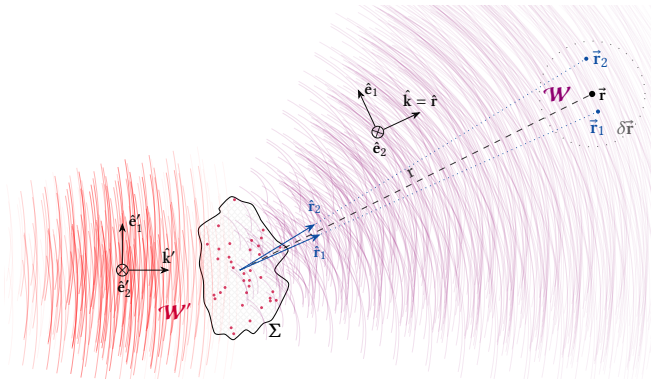

Fig. 3. Diffraction of natural light by an obstacle Σ. The obstacle may be any matter or medium or, equivalently, an opening in a conductive screen (i.e. an aperture). An incident field (wavefronts illustrated as red lines), with its coherence properties quantified by $\mathcal{W}'$, induces electromagnetic excitations in Σ (small purple balls), giving rise to *diffracted* radiation (wavefronts illustrated as purple lines). The solution to the diffraction problem is then $\mathcal{W}$, which quantifies the diffracted wave ensemble that has propagated to some point $\vec{\mathbf{r}}$. The wavevectors $\hat{\mathbf{k}}$ and the orthonormal bases $\hat{\mathbf{e}}$ are illustrated as well. Note, as the diffracted wave ensemble propagates the wavefronts become more similar, illustrating that light gains coherence on 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 6.1) to the diffracting region $\Sigma = \Sigma_1 \cup \Sigma_2$. The matrix $\boldsymbol{\Psi}$ (Eq. (31)) then can be written as:

$$\begin{aligned} \boldsymbol{\Psi} &= \int_\Sigma \int_\Sigma \mathcal{W}'(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega) e^{-ik(\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}' - \hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}_\perp'')} \\ &= \sum_{\xi,\zeta} \int_{\Sigma_\xi} \int_{\Sigma_\zeta} \mathcal{W}'(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega) e^{-ik(\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}' - \hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}_\perp'')} \end{aligned} \tag{32}$$

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 $\mathcal{W}'(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega) \approx 0$ whenever $\xi \neq \zeta$. In that case, we say that the the sources are statistically independent, and the sum in Eq. (32) 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 6.1 is equivalent to its scalar counterpart, Eq. (23) (where in general we need to also account for inclination), only at normal incidence and when $\hat{\mathbf{r}} \approx \hat{\mathbf{n}}$. 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 $f_r(\Omega_i, \Omega_o, \omega) = f_r(\Omega_o, \Omega_i, \omega)$. In our supplemental material we show that the "geometrical" terms in the diffraction operator, i.e. all the terms except the matrix $\boldsymbol{\Psi}$, do adhere to that notation of reciprocity. Clearly, however, $\boldsymbol{\Psi}$ which describes the diffraction process is not reciprocal, and indeed physical optics does not admit that property. Nonetheless, we state without proving that reciprocity is retained across sufficiently large solid angles.

Furthermore, in Appendix B we also discuss the short-wavelength limit of the diffraction formulae derived above, and we show that under that limit energy diffracts and propagates along rays, in line with "geometrical optics". We also conclude that in that limit, the diffraction operator does indeed obey the classical reciprocity condition.

7 LIGHT-MATTER INTERACTION

In this section we discuss and formalise a few simple cases of light-matter interaction. We start with arguably the most fundamental operation in imaging and computer graphics: the measurement of the intensity of light incident upon a sensor. In practice, light is measured by a photodetector (or, typically, an array of such detectors which together constitute a camera's sensor) or biological photoreceptors. Those detectors respond to an incident electromagnetic flux, and have varying sensitivity as a function of wavelength.

Definition 7.1 (Measurement Operator). Let $\boldsymbol{\mathcal{L}}$ be an incident RCSD, and $\chi(\omega)$ the responsivity curve of the measurement device. Then, the observed radiance at a point $\vec{\mathbf{r}}$ is

$$\mathscr{L}\{\boldsymbol{\mathcal{L}}, \chi, \vec{\mathbf{r}}\} \triangleq \frac{c}{8\pi^2} \int_0^\infty \mathrm{d}\omega\, \chi(\omega)\, \mathrm{tr}\, \boldsymbol{\mathcal{L}}\,(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega) \tag{33}$$

As discussed in Subsection 3.3, the integrals $\int \mathcal{L}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ and $\int \mathrm{tr}\, \boldsymbol{\mathcal{L}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ compute time-averaged quantities, which are exact when averaging over a $T \to \infty$ period of time and in the far-field (plane wave) limit. For polychromatic natural light, which exhibits very short temporal coherence lengths, Definition 7.1 is an excellent approximation. However, when dealing with quasi-monochromatic radiation that gives rise to significant coherence times, or when rendering images with ultra-low exposure time (as is done by, e.g., Jarabo et al. [2014]), the assumption that we average over infinite time no longer holds. In such a scenario we would be forced to consider the higher-order statistics of the wave ensemble, which is beyond the scope of this paper.

Jones calculus operators. When dealing with perfectly coherent light, a well-studied optical formalism, known as *Jones calculus*, is sufficient to model many simple light-matter interactions, such as basic reflection and refraction of smooth surfaces, phase retarders like quarter- and half-wave plates and polarizers. Under that formalism, the transverse components of a coherent electromagnetic wave are represented using a *Jones vector*, a 2-dimensional complex valued vector $[\, E'_1 \;\; E'_2 \,]^T$, and light-matter interactions are modelled via *Jones matrices*—2×2 complex matrices that model an optical element and act upon those Jones vectors. We can introduce those Jones matrices directly into our formalism:

Definition 7.2 (Jones Calculus Operator). Let $\boldsymbol{A}$ be a Jones matrix representing an optical element. Then, the action of that optical element upon a wave ensemble is described via the following operator

$$\mathscr{J}_{\boldsymbol{A}}\{\boldsymbol{\mathcal{L}}\} \triangleq \boldsymbol{A}\, \boldsymbol{\mathcal{L}}\, \boldsymbol{A}^\dagger \tag{34}$$

The correctness of the above stems from the following analysis: Let $\boldsymbol{A} = \left[a_{\xi\zeta}\right]$ ($\xi, \zeta \in \{1, 2\}$ are indices) be a Jones matrix, let $E'_{1,2}$ be the transverse components of a realization of a wave ensemble and denote $E_{1,2}$ as the components of the realization after interaction with the optical element (that is, after being acted upon by $\boldsymbol{A}$). Then,

$$\begin{aligned}\mathcal{L}_{\xi\zeta} &= \left\langle E_\xi E_\zeta^\star \right\rangle_\omega = \left\langle \left(a_{\xi 1} E'_1 + a_{\xi 2} E'_2\right)\left(a_{\zeta 1} E'_1 + a_{\zeta 2} E'_2\right)^\star \right\rangle_\omega \\ &= \sum_{p,q \in \{1,2\}} a_{\xi p}\, \mathcal{L}'_{pq}\, a^\star_{\zeta q}\end{aligned} \tag{35}$$

where $\mathcal{L}'_{pq} = \left\langle E'_p E'^\star_q \right\rangle_\omega$ and $\mathcal{L}_{\xi\zeta}$ are elements of the RCSD matrices of the radiation before and after the interaction, respectively. Finally, it is easy to verify that the RCSD produced by the operator above is Hermitian and positive semi-definite, when evaluated at $\vec{\mathbf{r}}_1 = \vec{\mathbf{r}}_2$, and thus is indeed a polarization matrix.

Note that Jones calculus deals exclusively with perfectly coherent radiation. However, by acting upon each realization of a wave ensemble (as formalised by Eq. (35)), we have shown that that domain of validity can be extended (via Definition 7.2) to light of any state of coherence and polarization. A few other useful matrices that model specific cases of reflections and refractions will be introduced in the rest of this section.

Basis rotation. Before we proceed with additional light-matter interaction operators, note that, as discussed, RCSDs of the same wave ensemble expressed under different decomposition bases yield different matrices. Evidently, operators formulated via Jones matrices are sensitive to that choice of basis. As a consequence, a change of basis of an RCSD matrix is a common practical need and it can be easily accomplished by the use of a simple rotation matrix

$$\boldsymbol{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{36}$$

where θ is any (counter-clockwise) rotation angle. $\boldsymbol{R}_\theta$ is a Jones matrix, therefore basis rotation of an RCSD is done via the corresponding operator $\mathscr{J}_{\boldsymbol{R}_\theta}$ (Definition 7.2).

7.1 Reflection and Refraction Operators

The well-known Fresnel equations relate the amplitudes of an incident field to the reflected and refracted fields at an interface between a pair of media. These relations have seen wide application in the realm of computer graphics, therefore we provide only a brief summary of the Fresnel equations in Appendix C and in this subsection we proceed to present their application under our formalism. We assume that the RCSD matrix of the radiation incident upon the

surface is decomposed using an s-polarization (where the field is strictly perpendicular to the plane of incidence) and p-polarization (where the field is strictly parallel to the plane of incidence) basis. Then, the physical properties of the interface give rise to the *Fresnel coefficients*, which also depend on the incidence angle: $r_{s,p}$, which are the amplitude-ratios of the reflected s- and p-polarized fields, respectively, and $t_{s,p}$, the amplitude-ratios for the refracted s- and p-polarized fields, respectively. Clearly, the diagonal matrices $\boldsymbol{R} = \text{diag}\left\{r_s, r_p\right\}$ and $\boldsymbol{T} = \text{diag}\left\{t_s, t_p\right\}$ are Jones matrices describing the action of the interface. Due to their importance we choose to define dedicated Fresnel reflection and refraction operators:

Definition 7.3 (Fresnel Reflection and Refractions Operators). Let $\boldsymbol{\mathcal{L}}$ be the RCSD impinging upon an interface between two media. Then, the reflected RCSD is

$$\mathscr{R}_{\text{Fresnel}}\{\boldsymbol{\mathcal{L}}\} \triangleq \begin{bmatrix} r_s & r_p \end{bmatrix} \cdot \boldsymbol{\mathcal{L}} \cdot \begin{bmatrix} r_s & r_p \end{bmatrix}^{\dagger} \tag{37}$$

and the refracted RCSD is

$$\mathscr{T}_{\text{Fresnel}}\{\boldsymbol{\mathcal{L}}\} \triangleq \begin{bmatrix} t_s & t_p \end{bmatrix} \cdot \boldsymbol{\mathcal{L}} \cdot \begin{bmatrix} t_s & t_p \end{bmatrix}^{\dagger} \tag{38}$$

It is assumed that the cross-spectral density matrix $\boldsymbol{\mathcal{L}}$ is formulated with respect to a decomposition into s- and p-polarized incident waves.

The Church polarization factor. While the Fresnel equations are formally exact, their use implies the assumption that the surface is perfectly smooth. Of practical interest are a set of similar equations that take residual surface roughness into account. These relations appear to have been first derived in the modern radar scattering literature by Ruck [1970], and introduced into applied optics and surface scattering by Church et al. [1977]. They were later borrowed by the Rayleigh-Rice small perturbation surface theory as well as the Harvey-Shack generalized scattering theory, and have also seen some use in computer graphics [Holzschuch and Pacanowski 2017].

The Church polarization factor consists of four complex-valued coefficients, which are the amplitude-ratios between the incident and the scattered s- and p-polarized fields and together give rise to the Jones matrix $\boldsymbol{Q}$.

Definition 7.4 (Church Reflection Operator). With $\boldsymbol{\mathcal{L}}$ being the incident RCSD, the reflected RCSD under the Church reflection formulation is

$$\mathscr{R}_{\text{Church}}\{\boldsymbol{\mathcal{L}}\} \triangleq \boldsymbol{Q}\,\boldsymbol{\mathcal{L}}\,\boldsymbol{Q}^{\dagger} \tag{39}$$

Decomposition into s- and p-polarized waves is assumed. The formulae for the matrix $\boldsymbol{Q}$ as well as the Church polarization coefficients are detailed in Appendix D.

8 APPLICATIONS AND VALIDATION

In this section we present a couple of sample applications of our framework as well as validate our derivations.

8.1 Rendering Diffractions on Surface Scattering

We consider the case of scatter of partially-coherent light by an explicitly defined microscopic patch of surface. Consider a surface described by some height function, $h : \Sigma \to \mathbb{R}$, where the height deviations are from the mean surface plane, $\Sigma \subset \{z = 0\}$, assumed to be located on the xy-plane and centred at the origin. The surface normal is then $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, and we denote a point on the surface as $\vec{\mathbf{r}}' = \vec{\mathbf{r}}'_{\perp} + h(\vec{\mathbf{r}}'_{\perp})\hat{\mathbf{n}}$, with $\vec{\mathbf{r}}'_{\perp} \in \Sigma$ being a point on the mean surface plane. Assume that radiation arrives from a natural light source of radius ρ, emitted power spectral density Λ and position $\vec{\mathbf{s}}$. We would like to measure the observed intensity of light scattering of the surface patch and arriving at an imaging device (pinhole camera for simplicity) at position $\vec{\mathbf{e}}$.

Applying the SDTE (Definition 4.1), our quantity of interest can be written as

$$L = \mathscr{L}\left\{\int_{\mathcal{S}^2} \mathrm{d}^2\Omega' \left|\Omega' \cdot \hat{\mathbf{n}}\right| \mathscr{D}\left\{e^{ik(\Omega'+\hat{\mathbf{e}})\cdot(\vec{\mathbf{r}}'-\vec{\mathbf{r}}'')} \mathscr{R}\left\{\boldsymbol{\mathcal{W}}'\left(\vec{\mathbf{r}}', \vec{\mathbf{r}}'', \omega\right)\right\}\right\}\right\} \tag{40}$$

where $\mathscr{L}$ is the radiance measurement operator (Definition 7.1), $\vec{\mathbf{r}}', \vec{\mathbf{r}}''$ are points on the surface, $\mathscr{D}$ is the diffraction operator (Definition 6.1) that operates with respect to the points $\vec{\mathbf{r}}', \vec{\mathbf{r}}''$ and $\boldsymbol{\mathcal{L}}'$ is the RCSD sourced from incident direction Ω'. For simplicity and in a like manner to the Harvey-Shack surface scatter theory [Krywonos 2006], we assume that the reflected field strength ratio is constant across the surface patch and apply the Church reflection operator $\mathscr{R}$ (Definition 7.4) to the incident RCSD. 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 a diffracting aperture. The familiar wave propagation term, $\exp[ik(\Omega' + \hat{\mathbf{e}}) \cdot (\vec{\mathbf{r}}' - \vec{\mathbf{r}}'')]$, that appears in Eq. (40), can be regraded as the propagator that propagates incident waves to the mean plane Σ, taking the height fluctuations into account. We refer to it as the *frequency transmission function* [Born and Wolf 1999] that arises due to residual surface roughness. This expression can also be found in other related works that deal with rendering surface-induced diffractions [Holzschuch and Pacanowski 2017; Krywonos 2006; Yan et al. 2018].

Making the small-angle paraxial approximation and a few algebraic simplifications allows us to write the (far-field) coherence properties of the incident radiation (that is, the non-constant and non-phase terms in Definition 5.1) as

$$\mathcal{P}\left(\vec{\mathbf{r}}'_{\perp} - \vec{\mathbf{r}}''_{\perp}, \omega\right) \triangleq \frac{s^{\frac{3}{2}} \mathrm{J}_{\frac{3}{2}}\left[\frac{\rho k}{s}\left|\Omega' \times \left(\vec{\mathbf{r}}'_{\perp} - \vec{\mathbf{r}}''_{\perp}\right)\right|\right]}{\left|\Omega' \times \left(\vec{\mathbf{r}}'_{\perp} - \vec{\mathbf{r}}''_{\perp}\right)\right|^{\frac{3}{2}}} \tag{41}$$

with s being the distance to the light source from the surface patch. The above expression is termed the *impulse response function* induced by incident the wave ensemble. The frequency transmission function fully quantifies the height fluctuations of the surface geometry, while the impulse response function serves as a complete description of the incident light's coherence properties on the surface. Together, they supply the information necessary to compute the diffracted wave ensemble's RCSD function.

In the far-field, i.e. $s \gg \rho$, a long, but straightforward, derivation (see our supplemental material) yields a simplified expression for the observed intensity of partially-coherent light reflected from the

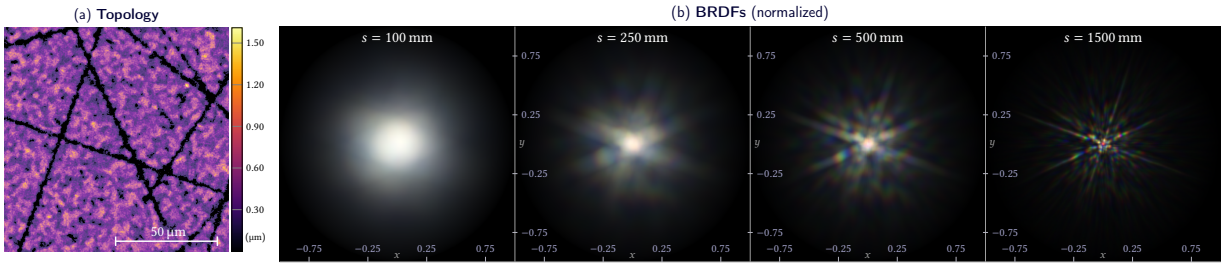

Fig. 4. (a) Topology of a uniformly scratched, moderately rough aluminum surface that is used to render (b) BRDF visualizations (normalized intensity) of the diffraction effects that arise when natural, partially-coherent light scatters of that surface. Illumination is done by a D65 light source of radius 8 mm that is positioned directly above the surface. By varying s, the distance of the source from the surface, we change the coherence properties of the incident wave ensemble, resulting in starkly different scattering behaviour.

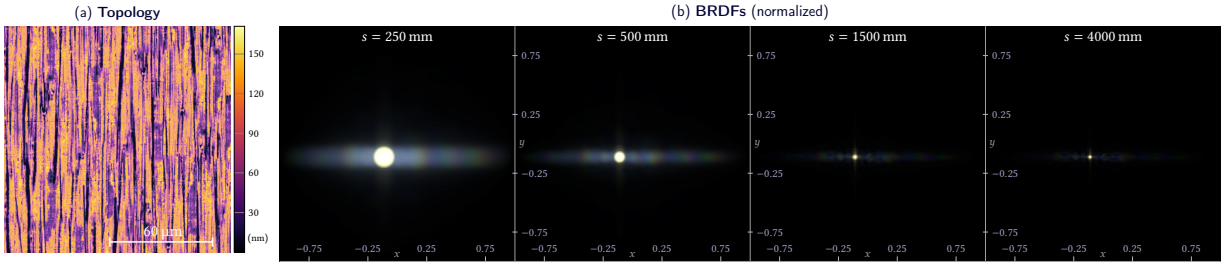

Fig. 5. (a) An aluminum surface that has undergone electric polish resulting in low intrinsic roughness and highly anisotropic details. (b) BRDF visualizations (normalized intensity) under partially-coherent illumination produced by a fluorescent F1 model light source at various distances to the surface.

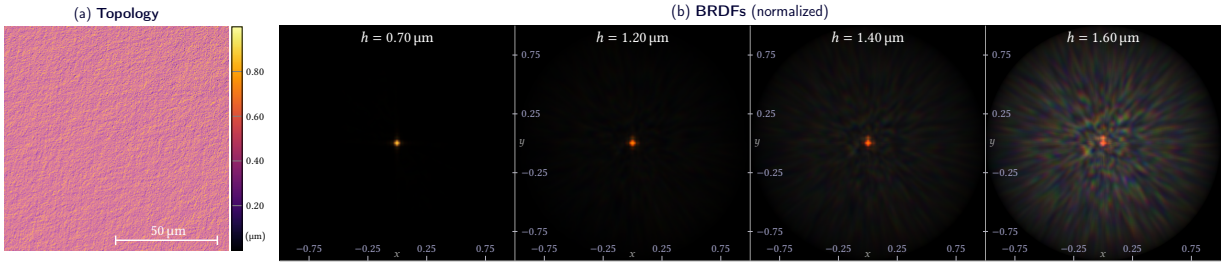

Fig. 6. (a) A sandpaper polished gold surface with isotropic statistics. (b) We vary the height multiplier of the surface, h, to compare the generated BRDFs (D65 illuminant and with intensity normalized) between different levels of roughness: For a very smooth surface (on the left), the only visible details are the reflection of the light source, as expected, and as we increase the roughness (left to right) isotropic surface diffraction effects appear.

surface patch:

$$L = \frac{\Theta}{\sqrt{\rho}} \frac{1}{A_\Sigma |\hat{\mathbf{e}} \cdot \hat{\mathbf{n}}|} \int_0^\infty \frac{\mathrm{d}\omega}{\lambda^2} \Lambda(\omega) \times \mathrm{tr}\left[\boldsymbol{H}_1 \boldsymbol{Q} (\boldsymbol{H}_2 \boldsymbol{Q})^\dagger \right] \left(\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} \right)(k\hat{\mathbf{e}}) \qquad (42)$$

where $\Theta = 2\pi/s\ (s - \sqrt{s^2 - \rho^2})$ is the solid angle subtended by the source and the Fourier transform of the impulse response function, $\mathscr{F}\{\mathcal{P}\}$, is with respect to the variable $\vec{\mathbf{d}}' = \vec{\mathbf{r}}'_\perp - \vec{\mathbf{r}}''_\perp$. The quantity $\mathcal{T}^2(\vec{\xi})$ is denoted as the *frequency response function* and is the complex magnitude squared of the Fourier transform of the frequency transmission function, viz.

$$\mathcal{T}^2\left(\vec{\xi}\right) \triangleq \left| \mathscr{F}_\Sigma \left\{ e^{ik(\Omega' + \hat{\mathbf{e}}) \cdot [\vec{\mathbf{r}}'_\perp + \hat{\mathbf{n}} h(\vec{\mathbf{r}}'_\perp)]} \right\} (\vec{\xi}) \right|^2 \qquad (43)$$

where the Fourier transforms is confined to Σ and is with respect to the variable $\vec{\mathbf{r}}'_\perp \in \Sigma$. Eq. (42) expresses a 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.

We have shown in this section that under general conditions the scattered RCSD can be accurately formulated as the convolution between a function describing the coherence properties of the incident wave ensemble and a function describing the surface-induced perturbations. Some parallels can be drawn between this conclusion and known results in optics regarding transmission of mutual intensities through an optical system [Born and Wolf 1999].

Numeric evaluation. The evaluation of Eq. (42) will typically be done numerically, and it is more suitable to rewrite the convolution as multiplication in Fourier space, viz.

$$\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} = \mathscr{F}\left\{\mathscr{F}^{-1}\{\mathcal{T}^2\} \cdot \mathcal{P}\right\} \tag{44}$$

Then, $\mathscr{F}^{-1}\{\mathcal{T}^2\}$ is evaluated via a pair of fast Fourier transforms (FFT). Only a single frequency of the last transform is needed, thus the Goertzel algorithm (generalized to arbitrary frequencies [Sysel and Rajmic 2012]) is used as it is faster, compared to an FFT, and provides a sinc interpolated result.

Using the theory derived above BRDF lobes visualizing the scatter of a few sample surfaces were rendered, see Figs. 4 to 6. Additional results, as well as a comparison with Yan et al. [2018], are available in our supplemental material. Furthermore, we have rendered more complete scenes showing a few diffractive surfaces, Figs. 1 and 8. See Subsection 8.4 for more details regarding the rendering of these scenes as well as discussion about integration of our framework into a path tracer.

8.2 Validation

The results derived in Subsection 8.1 serve as a convenient and simple test bed to validate our formalism. We produce a ground truth by solving the surface scatter problem using classical coherent diffraction theory for each elementary radiator in our natural light source. Each elementary radiator can be readily considered as a point source, therefore it produces perfectly spatially-coherent radiation and we use the Smythe's diffraction formula (Eq. (27)) to evaluate the surface scattered radiation for a single elementary radiator. This makes for a good validation, as: (1) the Smythe diffraction formula is a virtually exact solution to Maxwell's equations in the far-field [Zangwill 2013]; and (2) the ground truth is devoid of any optical coherence formalism. Thus, we are validating the derivations performed in Sections 4 to 6.

Because a high density of elementary radiators is required, numerical evaluation of the ground truth using classical diffraction theory is exceedingly expensive and only feasible in flatland. We employ a natural light source with a radius of $\rho = 1\,\text{mm}$ and with elementary radiators spaced at 8 µm intervals (the density in a real natural source would be greater, however this is sufficient for convergence), implying a light source composed of 49 089 elementary radiators. Note that in $\mathbb{R}^2$ the derivations in Subsection 8.1 change very slightly due to the source's geometry now being a disk instead of a ball. The rest of the analysis remains the same. Validation results are plotted in Fig. 7, and very good agreement is demonstrated. The small errors (absolute deviations from the ground truth of a few percent) are mostly due to numeric reasons. The computation time required to render a 1024 pixel flatland BRDF using a modern desktop computer is a fraction of a second using our method, but over a couple of hours for the ground truth. Derivations of the ground truth formulae as well as additional validation results are available in our supplemental material.

8.3 Partially-Coherent Radiative Transfer

The radiative transfer equation (RTE) can be traced back to Chandrasekhar [1960]; Lommel [1889] and provides a physical basis for light transport in participating media. The steady-state, isotropic RTE that commonly appears in computer graphics literature is a integro-differential equation that governs the distribution of radiance L in a participating medium:

$$\begin{aligned}\nabla_\Omega L(\vec{\mathbf{p}}, \Omega) =& \sigma_s \int_{\mathcal{S}^2} \mathrm{d}^2\Omega'\, f_p(\Omega, \Omega') L(\vec{\mathbf{p}}, \Omega') \\ &+ \sigma_a L_e(\vec{\mathbf{p}}, \Omega) - (\sigma_s + \sigma_a) L(\vec{\mathbf{p}}, \Omega)\end{aligned} \tag{45}$$

with ∇_Ω being the directional derivative, $\vec{\mathbf{p}}$ a point in the medium, Ω a direction, σ_a and σ_s the (possibly spatially varying) absorption and scattering coefficients (units of reciprocal length, i.e. the total absorption or scattering cross section over volume), respectively, and f_p the scattering phase function (assumed to be normalized). The dependence on wavelength of all quantities is implied. In an analogous fashion to our derivation of the SDTE (Definition 4.1), we replace the phase function-driven scattering of radiance above with an abstract diffraction operator:

Definition 8.1 (Spectral-Density Radiative Transfer Equation). For a homogeneous, isotropic medium the equation governing the distribution of the RCSD is

$$\begin{aligned}\nabla_\Omega \boldsymbol{\mathcal{L}}_\Omega =& \sigma_s \int_{\mathcal{S}^2} \mathrm{d}^2\Omega'\, \mathscr{D}_{\Omega' \to \Omega}\{\boldsymbol{\mathcal{L}}_{\Omega'}\} \\ &+ \sigma_a \boldsymbol{\mathcal{L}}^{(e)}_\Omega - (\sigma_s + \sigma_a)\, \boldsymbol{\mathcal{L}}_\Omega\end{aligned} \tag{46}$$

where $\sigma_a = 2k\kappa$, with κ being the imaginary part of the medium's refractive index, is the absorption coefficient that arises due to the medium's conductivity and $\mathscr{D}$ quantifies the interaction of light with the participating medium. The position and wavelength dependences are omitted for brevity.

As with the SDTE, the equation above should be understood in the sense of propagation of independent beams of light, in contrast to rays. The spectral-density radiative transfer equation is an integro-differential equation that formalises the volumetric light transport under our formalism, and the SDTE (Definition 4.1) serves as its boundary condition.

Rayleigh scattering. Finding analytic solutions to the radiative transfer diffraction problem that arises above can be difficult. We restrict ourselves to the simplest case of scattering by very small (with respect to wavelength) spherical particles in a homogenous medium, i.e. Rayleigh scattering. Because the particles are small, an incident electric field, $\vec{\mathbf{E}}'$, is almost constant across a scattering particle, leading to a quasistatic approximation. Thus, the scattering is dominated by the time-harmonic dipole moments that arise [Zangwill 2013]. These dipoles produce a far-field electric field $\vec{\mathbf{E}}$, whose field strength is related to the incident field's strength as follows:

$$r^2 \left|\vec{\mathbf{E}}\right|^2 = \sigma_s k^4 (1 + \left|\Omega \cdot \Omega'\right|^2) \left|\vec{\mathbf{E}}'\right|^2 \tag{47}$$

where the constant σ_s depends on the medium's electric polarizability and particle size and r is the propagated distance from the particle. The total scattering cross section is then $\frac{16\pi}{3}\sigma_s k^4$. The scattering region in the medium is typically large with respect to the spatial coherence of the incident light and contains very many

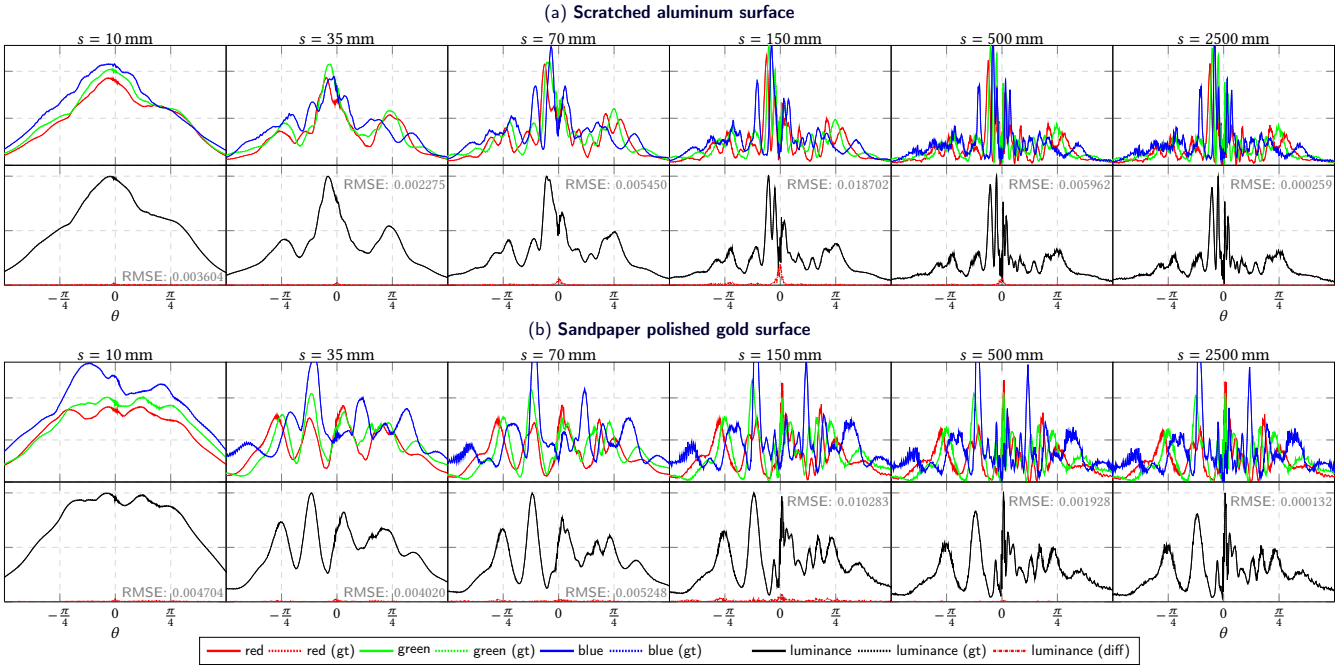

Fig. 7. Numeric validation of our method in flat-land for the (a) scratched surface (see Fig. 4) and (b) polished gold surface (see Fig. 6). The radius of the light source is 1 mm and the ground truth is generated using 49 089 elementary radiators. Rendering was done using 32 spectral samples. For each surface: (first row) we compare each primary colour component by plotting results computed using our method (solid plots) against the ground truth (dashed plots); similarly, (second row) we compare the total luminance and also plot the absolute luminance difference (in dash-dotted red) as well as the RMSE.

scattering particles. This is a similar situation to sourcing from a natural light source (Section 5): The scattering particles act as elementary radiators (with a radiation differential cross section that is no longer isotropic, but described by Eq. (47)) and the correlation between the produced fields depends on the coherence properties of the incident light. We make the approximation that the scattering region is much greater than the distance over which the incident light remains coherent, this implies that the scattering region acts as a perfect diffuser and the transfer equation becomes (neglecting emission)

$$\nabla_\Omega \mathcal{L}_\Omega = \sigma_s k^4 \int_{\mathcal{S}^2} \mathrm{d}^2\Omega' \left(1 + \left|\Omega \cdot \Omega'\right|^2\right) \mathcal{S}\{\Lambda', \rho\} - \left(\frac{16\pi}{3}\sigma_s k^4 + \sigma_a\right) \mathcal{L}_\Omega \qquad (48)$$

where $\mathcal{S}$ is the sourcing operator (Definition 5.1) and Λ' is the power spectral density of the radiation incident from direction Ω'. That is, the scattering medium acts as a secondary light source.

8.4 Rendering and Implementation

The scene rendered in Fig. 8 demonstrates diffractive effects arising on scattering of brushed aluminum and scratched copper bars. Furthermore, Fig. 1 showcases an example of the global effects of optical coherence: As light passes through a vase filled with water, it is diffused and gives rise to weaker interference patterns. As described in Subsection 8.3, the scattering water can be thought of as a

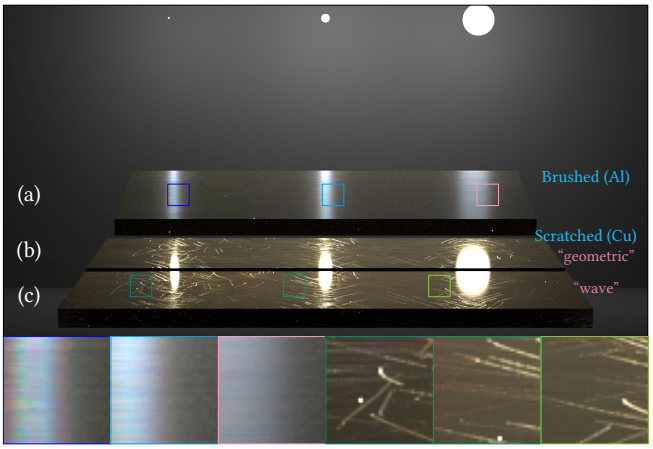

Fig. 8. A scene illuminated by three white natural sources of diameters (right to left) 0.10 mm, 1 mm and 1 cm, positioned about 6 cm above bars made of (a) brushed aluminum and (b,c) isotropically scratched copper. The scratches on the bottom copper bar (c) are rendered using our method, while the top copper bar (b) is rendered using "geometric optics" only for comparison. Interference effects are most visible when reflecting the more coherent light from the smallest source. Magnification insets are drawn at the bottom (colour coded).

large, and hence less optically coherent, secondary light source. The bars in Fig. 8 as well as the statue and the lamp head in Fig. 1 are

rendered using precomputed data of Eq. (42) for various scattering directions and values of r—the distance of travel of the beam of light from the source. The other surfaces are rendered classically. It is assumed that reflections and refractions of smooth surfaces (the glass, plastic surfaces and other conductors) do not affect the coherence properties of light, i.e. the interactions are fully described by the Fresnel relations (Definition 7.3). The rough surfaces are assumed to fully diffuse the light, meaning they act as secondary natural sources. In practice, when tracing rays we track the travel distance from the light source, and water scatter events or reflections of a diffuse surface reset that travel distance to zero. Thus, in our implementation, the coherence properties of a light beam only depend on the total effective distance travelled from the source.

Clearly, some of those assumptions do not hold in general: Interaction with media and matter will typically alter the shape of the coherence function and scattering through water is not confined to scatter by small particles. However, the illumination of the diffractive surfaces in Figs. 1 and 8 is dominated by either direct illumination or by light propagating through the vase. The refraction through the smooth glass surfaces is readily described by the Fresnel equations, and the water will scatter and diffuse light to some degree, therefore those simplifications do not greatly affect the accuracy of the rendered images. Anyhow, our implementation serves merely as a proof of concept, demonstrating the global transport of partially-coherent light.

A major difficulty in path tracing partially-coherent light is the breakdown of the symmetry of the path integral. The classical Helmholtz reciprocity property clearly no longer applies when dealing with diffraction (see Section 6 and our supplemental material), but the problem is more acute: The light's coherence information dictates the light-matter interaction behaviour. Therefore, the angular distribution of the scattered energy morphs (potentially greatly so) as the coherence properties of light mutate (e.g., see Figs. 4 to 6). This implies that tracing paths from the camera is not directly possible, as the coherence properties depend on the source and can only be traced from the source. Unfortunately, this symmetry of the path integral is important for many path tracing algorithms and their performance. This remains an open problem and the development of more sophisticated approaches and methods for path tracing partially-coherent beams are left for future work.

Integration into path tracers. The first barrier to integration of our framework into a modern path tracers is the requirement to propagate coherence information from the sources. Rendering of Figs. 1 and 8 was done using a patched version of Mitsuba, where we extended the bi-directional integrator `bdpt` to quantify the optical coherence of a traced beam when tracing from a light source. Naïvely tracing from the sources only results in slow convergence and some residual noise remains in these figures despite high sample counts and rendering times (the rendering time of Fig. 1 was about 2 days using a modern personal computer).

Second, the manner in which the interaction of light with matter and media alters those coherence properties of light needs to be quantified. As mentioned, our implementation assumes that all traced beams retain the Bessel beam shape (as formalised by Eq. (25)), and the total beam travel distance r is sufficient to describe the coherence properties of the light. In our implementation, light-matter interaction is limited to idealised diffusion, by participating media or diffuse surfaces, and interaction with perfectly smooth interfaces.

9 CONCLUSION AND FUTURE WORK

We have presented a global light transport framework that deals exclusively with partially coherent light. By deriving a theory that is firmly grounded in concrete physical foundations, the presented framework is able to fully account for the electromagnetic nature of light and reproduce wave-related optical phenomena that arise with natural light and can be observed using typical (time-averaged) sensors. By dealing with units of radiance, we have shown that our framework can be considered as a generalization of the classical radiometric light transport, and is mostly consistent with the way modern path tracers operate. The driving assumptions for the rendering equations introduced in this paper (Definitions 4.1 and 8.1) are that

(i) beams of weakly-coherent natural light produced by distinct (primary or secondary) sources are mutually uncorrelated, and thus their superposition is linear (in terms of intensity and RCSD functions); and
(ii) observation is over very long periods of time (with respect to the temporal coherent of light).

Clearly, these assumptions break down if we render ultra-low exposure time images or deal with strongly spatially- or temporally-coherent light: such as when lasing or using narrow-band filtered (i.e. quasi-monochromatic) light.

In this paper our focus has been on deriving a complete and generally-applicable theory of the free-space sourcing, propagation and diffraction of natural (partially-coherent) light. These derived formulae are essentially exact in the optical far-field. Subsequently, we demonstrated the application of these formulae to the study of scatter of explicitly defined surfaces. Due to the generality of these formulations, computing the scattered light of an arbitrary surface is both expensive and requires highly detailed, sub-micron surface topologies. However, for some special cases, e.g., a scratch in a flat surface, a closed form expression can be found for the frequency response function $\mathcal{T}^2$, greatly simplifying the numerical calculations. Such special cases are left for future work. Likewise, as surfaces are typically described statistically and not explicitly in computer graphics, it is also of interest to discuss the scattering by statistical surface models under our formalism.

In addition, we have also derived some basic, idealized light-matter interaction operators: interaction with perfectly smooth interfaces and scattering by very small particles in a homogeneous participating medium. It is of practical importance to formulate more interesting interaction operators: for example, more general scattering in participating media (e.g., non-homogeneous media, large or non-spherical particles); or propagation of light through a lens system, where partial coherence plays an important role in quantifying the various diffraction effects that arise. We leave this more comprehensive analysis of light-matter interactions for future work.

REFERENCES

Thomas Auzinger, Wolfgang Heidrich, and Bernd Bickel. 2018. Computational design of nanostructural color for additive manufacturing. *ACM Transactions on Graphics* 37, 4 (Aug 2018), 1–16. https://doi.org/10.1145/3197517.3201376

Chen Bar, Marina Alterman, Ioannis Gkioulekas, and Anat Levin. 2019. A Monte Carlo framework for rendering speckle statistics in scattering media. *ACM Transactions on Graphics* 38, 4 (Jul 2019), 1–22. https://doi.org/10.1145/3306346.3322950

Chen Bar, Ioannis Gkioulekas, and Anat Levin. 2020. Rendering near-field speckle statistics in scattering media. *ACM Transactions on Graphics* 39, 6 (Nov 2020), 1–18. https://doi.org/10.1145/3414685.3417813

Laurent Belcour and Pascal Barla. 2017. A Practical Extension to Microfacet Theory for the Modeling of Varying Iridescence. *ACM Trans. Graph.* 36, 4, Article 65 (July 2017), 14 pages. https://doi.org/10.1145/3072959.3073620

Max Born and Emil Wolf. 1999. *Principles of optics : electromagnetic theory of propagation, interference and diffraction of light.* Cambridge University Press, Cambridge New York.

Rémi Carminati and Jean-Jacques Greffet. 1999. Near-Field Effects in Spatial Coherence of Thermal Sources. *Phys. Rev. Lett.* 82 (Feb 1999), 1660–1663. Issue 8. https://doi.org/10.1103/PhysRevLett.82.1660

S Chandrasekhar. 1960. *Radiative transfer.* Dover Publications, New York.

Mikhail Charnotskii. 2019. 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 (Jul 2019), 1433. https://doi.org/10.1364/josaa.36.001433

E. L. Church, H. A. Jenkinson, and J. M. Zavada. 1977. Measurement of the Finish of Diamond-Turned Metal Surfaces By Differential Light Scattering. *Optical Engineering* 16, 4 (Aug 1977). https://doi.org/10.1117/12.7972054

Tom Cuypers, Tom Haber, Philippe Bekaert, Se Baek Oh, and Ramesh Raskar. 2012. Reflectance model for diffraction. *ACM Transactions on Graphics* 31, 5 (Aug 2012), 1–11. https://doi.org/10.1145/2231816.2231820

D. S. Dhillon, J. Teyssier, M. Single, I. Gaponenko, M. C. Milinkovitch, and M. Zwicker. 2014. Interactive Diffraction from Biological Nanostructures. *Computer Graphics Forum* 33, 8 (2014), 177–188. https://doi.org/10.1111/cgf.12425

V. Falster, A. Jarabo, and J. R. Frisvad. 2020. Computing the Bidirectional Scattering of a Microstructure Using Scalar Diffraction Theory and Path Tracing. *Computer Graphics Forum* 39, 7 (Oct 2020), 231–242. https://doi.org/10.1111/cgf.14140

Ari T. Friberg. 1979. 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 (Jan 1979), 192. https://doi.org/10.1364/JOSA.69.000192

Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. 2015. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics* 34, 4 (Jul 2015), 1–14. https://doi.org/10.1145/2766928

Joseph Goodman. 2015. *Statistical optics.* John Wiley & Sons Inc, Hoboken, New Jersey.

Ibón Guillén, Julio Marco, Diego Gutierrez, Wenzel Jakob, and Adrian Jarabo. 2020. A General Framework for Pearlescent Materials. *ACM Transactions on Graphics* 39, 6 (2020). https://doi.org/10.1145/3414685.3417782

M.V. Guryev. 2012. Detailed description of spontaneous emission. *Journal of Modern Optics* 59, 14 (Aug 2012), 1278–1282. https://doi.org/10.1080/09500340.2012.712728

Stephane Guy and Cyril Soler. 2004. Graphics gems revisited. In *ACM SIGGRAPH 2004 Papers on - SIGGRAPH '04*. ACM Press. https://doi.org/10.1145/1186562.1015708

Nicolas Holzschuch and Romain Pacanowski. 2017. A Two-scale Microfacet Reflectance Model Combining Reflection and Diffraction. *ACM Trans. Graph.* 36, 4, Article 66 (July 2017), 12 pages. https://doi.org/10.1145/3072959.3073621

Adrian Jarabo, Julio Marco, Adolfo Muñoz, Raul Buisan, Wojciech Jarosz, and Diego Gutierrez. 2014. A framework for transient rendering. *ACM Transactions on Graphics* 33, 6 (Nov 2014), 1–10. https://doi.org/10.1145/2661229.2661251

James T. Kajiya. 1986. The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques - SIGGRAPH '86*. ACM Press. https://doi.org/10.1145/15922.15902

Tom Kneiphof, Tim Golla, and Reinhard Klein. 2019. Real-time Image-based Lighting of Microfacet BRDFs with Varying Iridescence. *Computer Graphics Forum* 38, 4 (2019), 77–85. https://doi.org/10.1111/cgf.13772

Matias Koivurova, Henri Partanen, Jari Turunen, and Ari T. Friberg. 2017. Grating interferometer for light-efficient spatial coherence measurement of arbitrary sources. *Applied Optics* 56, 18 (Jun 2017), 5216. https://doi.org/10.1364/ao.56.005216

Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. 2020. Interferometric transmission probing with coded mutual intensity. *ACM Transactions on Graphics* 39, 4 (Jul 2020). https://doi.org/10.1145/3386569.3392384

Andrey Krywonos. 2006. *Predicting surface scatter using a linear systems formulation of non-paraxial scalar diffraction.* Ph.D. Dissertation. University of Central Florida.

Anat Levin, Daniel Glasner, Ying Xiong, Frédo Durand, William Freeman, Wojciech Matusik, and Todd Zickler. 2013. Fabricating BRDFs at high spatial resolution using wave optics. *ACM Transactions on Graphics* 32, 4 (Jul 2013), 1–14. https://doi.org/10.1145/2461912.2461981

Eugene Lommel. 1889. Die Photometrie der diffusen Zurückwerfung. *Annalen der Physik* 272, 2 (1889), 473–502.

Heylal Mashaal, Alex Goldstein, Daniel Feuermann, and Jeffrey M. Gordon. 2012. First direct measurement of the spatial coherence of sunlight. *Optics Letters* 37, 17 (Aug 2012), 3516. https://doi.org/10.1364/ol.37.003516

Anthony B Murphy and Eugene Tam. 2014. Thermodynamic properties and transport coefficients of arc lamp plasmas: argon, krypton and xenon. *Journal of Physics D: Applied Physics* 47, 29 (Jun 2014), 295202. https://doi.org/10.1088/0022-3727/47/29/295202

A. Musbach, G. W. Meyer, F. Reitich, and S. H. Oh. 2013. Full Wave Modelling of Light Propagation and Reflection. *Computer Graphics Forum* 32, 6 (Feb 2013), 24–37. https://doi.org/10.1111/cgf.12012

Se Baek Oh, Sriram Kashyap, Rohit Garg, Sharat Chandran, and Ramesh Raskar. 2010. Rendering Wave Effects with Augmented Light Field. *Computer Graphics Forum* 29, 2 (May 2010), 507–516. https://doi.org/10.1111/j.1467-8659.2009.01620.x

George Ruck. 1970. *Radar cross section handbook.* Plenum Press, New York.

Iman Sadeghi, Adolfo Munoz, Philip Laven, Wojciech Jarosz, Francisco Seron, Diego Gutierrez, and Henrik Wann Jensen. 2012. Physically-based simulation of rainbows. *ACM Transactions on Graphics* 31, 1 (Jan 2012), 1–12. https://doi.org/10.1145/2077341.2077344

Jos Stam. 1999. Diffraction shaders. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*. ACM Press. https://doi.org/10.1145/311535.311546

Shlomi Steinberg. 2019. Analytic Spectral Integration of Birefringence-Induced Iridescence. *Computer Graphics Forum* 38, 4 (Jul 2019), 97–110. https://doi.org/10.1111/cgf.13774

Shlomi Steinberg. 2020. Accurate Rendering of Liquid-Crystals and Inhomogeneous Optically Anisotropic Media. *ACM Transactions on Graphics* 39, 3 (Jun 2020), 1–23. https://doi.org/10.1145/3381748

Shlomi Steinberg and Lingqi Yan. 2021. Rendering of Subjective Speckle Formed by Rough Statistical Surfaces. *ACM Transactions on Graphics* (2021), To appear.

John Stover. 2012. *Optical scattering : measurement and analysis.* SPIE Press, Bellingham, Wash. (1000 20th St. Bellingham WA 98225-6705 USA.

Petr Sysel and Pavel Rajmic. 2012. Goertzel algorithm generalized to non-integer multiples of fundamental frequency. *EURASIP Journal on Advances in Signal Processing* 2012, 1 (Mar 2012). https://doi.org/10.1186/1687-6180-2012-56

Antoine Toisoul, Daljit Singh Dhillon, and Abhijeet Ghosh. 2018. Acquiring Spatially Varying Appearance of Printed Holographic Surfaces. *ACM Trans. Graph.* 37, 6, Article 272 (Dec. 2018), 16 pages. https://doi.org/10.1145/3272127.3275077

Antoine Toisoul and Abhijeet Ghosh. 2017. Practical Acquisition and Rendering of Diffraction Effects in Surface Reflectance. *ACM Transactions on Graphics* 36, 5 (Jul 2017), 1–16. https://doi.org/10.1145/3012001

Z. Velinov, S. Werner, and M. B. Hullin. 2018. Real-Time Rendering of Wave-Optical Effects on Scratched Surfaces. *Computer Graphics Forum* 37, 2 (2018), 123–134. https://doi.org/10.1111/cgf.13347

A Walther. 1968. Radiometry and coherence. *JOSA* 58, 9 (1968), 1256–1259. https://doi.org/10.1364/JOSA.58.001256

Sebastian Werner, Zdravko Velinov, Wenzel Jakob, and Matthias Hullin. 2017. Scratch Iridescence: Wave-Optical Rendering of Diffractive Surface Structure. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 36, 6 (Nov. 2017). https://doi.org/10.1145/3130800.3130840

Alexander Wilkie, Robert F. Tobler, and Werner Purgathofer. 2001. *Combined Rendering of Polarization and Fluorescence Effects.* Springer Vienna, 197–204. https://doi.org/10.1007/978-3-7091-6242-2_18

Emil Wolf. 1978. Coherence and radiometry. *JOSA* 68, 1 (1978), 6–17. https://doi.org/10.1364/JOSA.68.000006

Emil Wolf. 1982. 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 (Mar 1982), 343–351. https://doi.org/10.1364/JOSA.72.000343

Emil Wolf. 2007. *Introduction to the theory of coherence and polarization of light.* Cambridge University Press, Cambridge.

Ling-Qi Yan, Miloš Hašan, Bruce Walter, Steve Marschner, and Ravi Ramamoorthi. 2018. Rendering Specular Microgeometry with Wave Optics. *ACM Trans. Graph.* 37, 4, Article 75 (July 2018), 10 pages. https://doi.org/10.1145/3197517.3201351

Andrew Zangwill. 2013. *Modern electrodynamics.* Cambridge University Press, Cambridge.

A PROPERTIES OF POLARIZATION MATRICES

Let

$$\boldsymbol{J} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \left\langle \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}^\dagger \right\rangle = \begin{bmatrix} \left\langle E_1 E_1^\star \right\rangle & \left\langle E_1 E_2^\star \right\rangle \\ \left\langle E_2 E_1^\star \right\rangle & \left\langle E_2 E_2^\star \right\rangle \end{bmatrix} \tag{49}$$

be a polarization matrix, with $E_{1,2}$ being some field transverse components. Clearly $J_{21} = J_{12}^\star$, and $\boldsymbol{J}$ is Hermitian, i.e. $\boldsymbol{J}^\dagger = \boldsymbol{J}$. By

Schwarz's inequality the following must hold

$$|J_{12}| = |J_{21}| \leq \sqrt{|J_{11}|}\sqrt{|J_{22}|} \tag{50}$$

therefore $\boldsymbol{J}$ is positive semi-definite, i.e. $\boldsymbol{J} \succeq 0$, and thus $\det \boldsymbol{J} \geq 0$ and all its eigenvalues are real and non-negative. Any matrix that fulfils the constraints above is a polarization matrix, notably $\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ and $\boldsymbol{\mathcal{L}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$. Note that the determinant and the trace of $\boldsymbol{J}$ are invariant under the choice of the basis that decomposes the field into the orthogonal transverse components $E_{1,2}$.

The polarization matrix lends insight into the state of polarization of the electromagnetic wave. The *degree of correlation* is defined as

$$j_{12} \triangleq \frac{1}{\sqrt{J_{11}}\sqrt{J_{22}}} J_{12} \tag{51}$$

Clearly, $|j_{12}| \leq 1$ (at least in the limit sense). When $|j_{12}| = 0$ and $J_{11} = J_{22}$, we say the light is *randomly polarized* or *unpolarized*. This can be summarised as (1) under any choice of basis, the transverse field components are uncorrelated; and (2) energy is evenly distributed in all transverse directions. On the other hand, $|j_{12}| = 1$ implies that the transverse oscillations are fully correlated and such light is completely polarized. In that case, it trivially follows that

$$\det \boldsymbol{J} = J_{11}J_{22} - J_{12}(J_{12})^{\star} = J_{11}J_{22} - \left|\sqrt{J_{11}}\sqrt{J_{22}} j_{12}\right|^2 = 0 \tag{52}$$

The degree of correlation j_{12} is also invariant under the choice of transverse basis.

It is always possible to decompose light into an exact superposition of completely polarized and unpolarized radiation. That is, a polarization matrix $\boldsymbol{J}$ can always be decomposed into

$$\boldsymbol{J} = \boldsymbol{J}_{\mathrm{u}} + \boldsymbol{J}_{\mathrm{p}} \tag{53}$$

where each matrix is a polarization matrix and in addition the following holds: $\det \boldsymbol{J}_{\mathrm{p}} = 0$ and $\boldsymbol{J}_{\mathrm{u}} \propto \boldsymbol{I}$. Furthermore, that decomposition is unique. As a corollary it is easy to show that a completely polarized matrix can not be non-trivially decomposed (a trivial decomposition is $\boldsymbol{J} = \alpha\boldsymbol{J} + (1-\alpha)\boldsymbol{J}$).

B AT THE SHORT-WAVELENGTH LIMIT

The purpose of this appendix is to formally show that at the short-wavelength limit, $k \to \infty$, the RCSD functions can be interpreted as classical radiance.

Sourcing. Looking at $k \to \infty$ limit of the absolute value of the sourced RCSD function (Definition 5.1), we immediately deduce that it takes a form of radiance at the short-wavelength limit:

$$\boldsymbol{\mathcal{L}}_{(k\to\infty)} = \lim_{k\to\infty} |\mathcal{S}\{\Lambda, \rho\}| = \frac{\rho}{3}\sqrt{\frac{2}{\pi}}\boldsymbol{I}\Lambda(\omega)\delta^2(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2) \tag{54}$$

where δ is the Dirac delta. The absolute value is taken to discard phase, as the term e^{ik} does not converge. The two-dimensional Dirac term suggests that for any $\hat{\mathbf{r}}_1$, energy arises in the far-field only in direction $\hat{\mathbf{r}}_2 = \hat{\mathbf{r}}_1$, and vanishes otherwise. The physical implausibility of the singular nature of an RCSD in the form presented above should not deter the reader: It describes an aphysical process—propagation of electromagnetic radiation in the short-wavelength limit—and should only be understood in the sense of a mathematical representation that encodes radiance information.

Diffraction. We assume an extended aperture. To keep the analysis simple, we give a planar aperture a uniform constant thickness of $L > 0$, assumed to be small, and denote the extended aperture as $\Sigma \subset \mathbb{R}^3$. The following analysis can be generalized to any realistic aperture or diffracting region. We substitute the short-wavelength limit expression that we derived for the RCSD matrix of sourced natural radiation (Eq. (54)) into the diffraction expression of the matrix $\boldsymbol{\Psi}$ (Eq. (31)), and reintroduce the necessary phase information $e^{ik(r_1 - r_2)}$. Then, as before, we take the $k \to \infty$ limit of the absolute value, yielding (up to a constant):

$$\begin{aligned}
&\boldsymbol{\Psi}_{(k\to\infty)}(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \omega) \\
&= \boldsymbol{I}\Lambda(\omega) \lim_{k\to\infty} \left| \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}' \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}'' \, \delta^2(\hat{\mathbf{r}}' - \hat{\mathbf{r}}'') e^{ik(r' - r'')} e^{-ik(\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}' - \hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}'')} \right| \\
&= \boldsymbol{I}\Lambda(\omega) \lim_{k\to\infty} \left| \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}' \, e^{-ik(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)\cdot\vec{\mathbf{r}}'} \int_{l_0(\vec{\mathbf{r}}')}^{l_0(\vec{\mathbf{r}}') + L(\Omega_i)} \mathrm{d}l \, e^{ikl} e^{ikl\hat{\mathbf{r}}_2 \cdot \Omega_i} \right| \\
&= \boldsymbol{I}\Lambda(\omega) \lim_{k\to\infty} \left| \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}' \, e^{-ik(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)\cdot\vec{\mathbf{r}}'} \right| \frac{\left|1 - e^{ikL(\Omega_i)(\hat{\mathbf{r}}_2 \cdot \Omega_i + 1)}\right|}{k|\hat{\mathbf{r}}_2 \cdot \Omega_i + 1|}
\end{aligned} \tag{55}$$

where Ω_i is the direction to the source, $L(\Omega_i) = L/|\Omega_i \cdot \hat{\mathbf{n}}|$ (with $\hat{\mathbf{n}}$ being the aperture normal) is the effective aperture thickness as viewed from direction Ω_i. We used the fact the Dirac term $\delta^2(\hat{\mathbf{r}}' - \hat{\mathbf{r}}'')$ constraints the integration to a path integral through the aperture, viz. $\vec{\mathbf{r}}'' = \vec{\mathbf{r}}' - l\Omega_i$, and l_0 is the term that ensures we integrate over the depth of the aperture. The characteristic function of Σ is L^1 integrable, and by a direct application of the Riemann–Lebesgue lemma we deduce:

$$\lim_{k\to\infty} \left| \int_\Sigma \mathrm{d}^3\vec{\mathbf{r}}' \, e^{-ik(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)\cdot\vec{\mathbf{r}}'} \right| = \begin{cases} 0 & \text{if} \quad \hat{\mathbf{r}}_1 \neq \hat{\mathbf{r}}_2 \\ LA_\Sigma & \text{otherwise} \end{cases} \tag{56}$$

where A_Σ is the cross-sectional area of the aperture, as before. The second factor vanishes trivially at the limit, unless $\hat{\mathbf{r}}_2 \cdot \Omega_i = -1$:

$$\lim_{k\to\infty} \frac{\left|1 - e^{ikL(\Omega_i)(\hat{\mathbf{r}}_2 \cdot \Omega_i + 1)}\right|}{k|\hat{\mathbf{r}}_2 \cdot \Omega_i + 1|} = \begin{cases} 0 & \text{if} \quad \hat{\mathbf{r}}_2 \cdot \Omega_i \neq -1 \\ \frac{L}{|\Omega_i \cdot \hat{\mathbf{n}}|} & \text{otherwise} \end{cases} \tag{57}$$

Using the above, we finally arrive at the short-wavelength limit the diffraction operator:

$$\boldsymbol{\mathcal{L}}_{(k\to\infty)} = \lim_{k\to\infty} |\mathscr{D}\{\boldsymbol{\mathcal{W}}, \Sigma\}| \propto \Lambda(\omega)\delta^2(2\Omega_i + \hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2) \tag{58}$$

(up to a polarization matrix constant). The Dirac term vanishes unless $\hat{\mathbf{r}}_1 = \hat{\mathbf{r}}_2 = -\Omega_i$, implying both that spatial coherence is exhibited only at $\hat{\mathbf{r}}_1 = \hat{\mathbf{r}}_2$, i.e. the perfectly incoherent "geometrical optics" approximation, and that far-field radiation only arises in the direction $\hat{\mathbf{r}} = -\Omega_i$. This means that no diffraction happens at the aperture, and the incident radiance continues to propagate along the incidence ray unaffected, as expected. Also note that the above is essentially the same representation of radiance as the sourced RCSD (Eq. (54)) in the short-wavelength limit. Furthermore, we can also conclude that in the short-wavelength limit the diffraction operator does obey the classical Helmholtz reciprocity condition.

C FRESNEL REFLECTION AND REFRACTION

The Fresnel kinematic equations arise due to the matching conditions implied by Maxwell's equations at an interface [Zangwill

2013]. Let $z = 0$ be the flat interface between a pair of media. Let μ_1, μ_2 be the magnetic permeabilities and ϵ_1, ϵ_2 be the permittivities of the $z > 0$ medium and the $z < 0$ medium, respectively. Those values may be complex. Then, $\eta_{1,2} = c\sqrt{\mu_{1,2}\epsilon_{1,2}}$ denote the (possibly complex) *indices of refraction* of the media. Assume that a wave with wavevector $\vec{\mathbf{k}}$ is incident upon the interface from the $z > 0$ half-space. Let $\vec{\mathbf{k}}_R = \vec{\mathbf{k}} - 2(\vec{\mathbf{k}} \cdot \hat{\mathbf{z}})\hat{\mathbf{z}}$ be the reflected wavevector. *Snell's law of refraction* relates the incident angle θ_1 that the wavevector $\vec{\mathbf{k}}$ makes with the interface, to the angle θ_2 between the refracted wave's wavevector, denoted $\vec{\mathbf{k}}_T$, and the interface:

$$\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2 \tag{59}$$

The three wavevectors $\vec{\mathbf{k}}, \vec{\mathbf{k}}_R, \vec{\mathbf{k}}_T$ are co-planar.

As mentioned, the polarization of the incident wave plays a role in its reflection and refraction. It is the common convention to consider the decomposition of the incident wave into an *s-polarized* and a *p-polarized* component. That is, if $\{\hat{\mathbf{e}}_s, \hat{\mathbf{e}}_p\}$ is the orthogonal basis, then $\hat{\mathbf{e}}_s \cdot \hat{\mathbf{z}} = 0$ and $\hat{\mathbf{e}}_p \cdot (\vec{\mathbf{k}} \times \hat{\mathbf{z}}) = 0$. Denoting $r_{s,p}$ and $t_{s,p}$ as the (possibly complex) amplitude ratios of the reflected and transmitted s- and p-polarized waves, respectively, the Fresnel equations become [Zangwill 2013]:

$$r_s = \frac{Z_2\cos\theta_1 - Z_1\cos\theta_2}{Z_2\cos\theta_1 + Z_1\cos\theta_2} \qquad r_p = \frac{Z_1\cos\theta_1 - Z_2\cos\theta_2}{Z_1\cos\theta_1 + Z_2\cos\theta_2} \tag{60}$$

$$t_s = \frac{2Z_2\cos\theta_1}{Z_2\cos\theta_1 + Z_1\cos\theta_2} \qquad t_p = \frac{2Z_2\cos\theta_1}{Z_1\cos\theta_1 + Z_2\cos\theta_2} \tag{61}$$

where $Z_{1,2} = \frac{\eta_{1,2}}{c\epsilon_{1,2}}$ are the *intrinsic impedances*. For non-magnetic media, i.e. $\mu_1 = \mu_2 = 1$ (or μ_0 under SI-units, where μ_0 is the free-space permeability), $Z_{1,2} = \frac{c}{\eta_{1,2}}$ and thus by writing $\frac{Z_2}{Z_1} = \frac{\eta_1}{\eta_2}$, Eqs. (60) and (61) can be rewritten in terms of the indices of refraction $\eta_{1,2}$.

D THE CHURCH POLARIZATION FACTOR

Let θ_i and θ_s be the angles of incidence and reflection, and let ϕ_s be the azimuthal angle of scattering (the angle between the projections of the incident and reflected wave-vectors onto the surface plane). Unlike the Fresnel equations which mandate $\theta_i = \theta_s$ and ϕ_s, we now choose the reflected direction at will. Denote the relative permittivity between the surfaces as $\epsilon = \epsilon_2/\epsilon_1$. Then, the four Church polarization coefficients are [Stover 2012]

$$q_{ss} \triangleq \frac{(\epsilon-1)\cos\phi_s}{\left(\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{62}$$

$$q_{sp} \triangleq \frac{(\epsilon-1)\sqrt{\epsilon - \sin^2\theta_s}\sin\phi_s}{\left(\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\epsilon\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{63}$$

$$q_{ps} \triangleq \frac{(\epsilon-1)\sqrt{\epsilon - \sin^2\theta_i}\sin\phi_s}{\left(\epsilon\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{64}$$

$$q_{pp} \triangleq \frac{(\epsilon-1)\left(\sqrt{\epsilon - \sin^2\theta_s}\sqrt{\epsilon - \sin^2\theta_i}\cos\phi_s - \epsilon\sin\theta_i\sin\theta_s\right)}{\left(\epsilon\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\epsilon\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{65}$$

where the first subscript denotes the polarization of the incident wave and the second subscript the polarization of the scattered wave. Note that when the reflection is in the plane of incidence, i.e. $\phi_s = 0$, the cross-polarization terms vanish, viz. $q_{sp} = q_{ps} = 0$. Likewise, when the reflection is specular, i.e. $\phi_s = 0$ and $\theta_i = \theta_s$ as mandated by the law of reflection, the expressions reduce to the Fresnel equations, that is $q_{ss} = r_s$ and $q_{pp} = r_p$. It is convenient to gather those coefficients into a matrix, viz.

$$\boldsymbol{Q} = \begin{bmatrix} q_{ss} & q_{ps} \\ q_{sp} & q_{pp} \end{bmatrix} \tag{66}$$

The total scattered energy is then know as the Church polarization factor and can be written as $Q \triangleq \sum |q_{\xi\zeta}|^2 = \mathrm{tr}\left(\boldsymbol{Q}\boldsymbol{Q}^\dagger\right)$.