Generic Framework for Physical Light Transport - Derivations

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

The purpose of this manuscript is to provide a complete set of derivations of our formulae as well as an extensive background. We start with an overview of relevant topics in mathematical analysis (Section 2) and stochastic processes (Section 3), with the discussion serving as a brief review of the required background. We proceed with presenting the necessary theoretical foundations of modern electrodynamics and optical coherence theory in a comprehensive fashion, with full derivations of most relevant results and theorems. Finally, in Section 7, the preceding discussions converge and we start presenting our primary results.

Background

1 PRELIMINARIES

1.1 Notation

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

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

$$\operatorname{Re} z = a \qquad \operatorname{Im} z = b \tag{1.1}$$

Then, absolute value of a complex number is

$$|z| = \sqrt{\operatorname{Re}\{z\}^2 + \operatorname{Im}\{z\}^2} = \sqrt{a^2 + b^2} \tag{1.2}$$

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

$$z = re^{i\theta} = r(\cos\theta + i\sin\theta) \tag{1.3}$$

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

$$z^\star = a - ib \tag{1.4}$$

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

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

$$\vec{\mathbf{v}} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}} \tag{1.5}$$

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.

where the shorthand $v_e = \hat{\mathbf{e}} \cdot \vec{\mathbf{v}}$ is used to express a component of $\vec{\mathbf{v}}$ is some arbitrary basis. The gradient, divergence, curl and Laplacian operators, in cartesian coordinates, are then, respectively:

$$\boldsymbol{\nabla} = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z} \tag{1.6}$$

$$\boldsymbol{\nabla} \cdot \vec{\mathbf{v}} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \tag{1.7}$$

$$\boldsymbol{\nabla} \times \vec{\mathbf{v}} = \hat{\mathbf{x}}\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) + \hat{\mathbf{y}}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) + \hat{\mathbf{z}}\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) \tag{1.8}$$

$$\nabla^2\vec{\mathbf{v}} = \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_z}{\partial z^2} \tag{1.9}$$

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

$$\vec{\mathbf{a}} \times \left(\vec{\mathbf{b}} \times \vec{\mathbf{c}}\right) = \vec{\mathbf{b}}(\vec{\mathbf{a}} \cdot \vec{\mathbf{c}}) - \vec{\mathbf{c}}\left(\vec{\mathbf{a}} \cdot \vec{\mathbf{b}}\right) \tag{1.10}$$

$$\vec{\mathbf{a}} \cdot \left(\vec{\mathbf{b}} \times \vec{\mathbf{c}}\right) = \vec{\mathbf{b}} \cdot (\vec{\mathbf{c}} \times \vec{\mathbf{a}}) = \vec{\mathbf{c}} \cdot \left(\vec{\mathbf{a}} \times \vec{\mathbf{b}}\right) \tag{1.11}$$

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \vec{\mathbf{v}}) = 0 \tag{1.12}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = 0 \tag{1.13}$$

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

$$\int_{\mathbb{F}} \mathrm{d}x\, |f(x)| < \infty \tag{1.14}$$

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

$$\int_{\mathbb{F}} \mathrm{d}x\, |f(x)|^2 < \infty \tag{1.15}$$

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

$$f(w) = \phi_{\mathrm{r}}(w) + i\phi_{\mathrm{i}}(w) \tag{1.16}$$

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

$$f(w)^\star = f(-w) \tag{1.17}$$

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

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

$$\mathscr{L}\psi = \lambda\psi \tag{1.18}$$

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

$\cos x$ is the eigenfunction of the 2nd-order derivative operator $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, with -1 being the associated eigenvalue.

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

$$\left|\int_{\mathbb{R}} fg\right| \leq \left(\int_{\mathbb{R}} |f|^2\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |g|^2\right)^{\frac{1}{2}} \tag{1.19}$$

which is the functional analogue of the triangle inequality.

1.1.4 Kronecker and Dirac delta. We denote the *Kronecker delta* as

$$\delta_{ij} \triangleq \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.20}$$

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

$$\forall x \neq 0 \rightarrow \delta(x) = 0 \qquad \text{but} \qquad \int_{-L}^{L} \mathrm{d}x\, \delta(x) = 1 \tag{1.21}$$

where L is any real positive value and can be extended to infinity. Of course, strictly speaking those definitions are contradictory: Any function that vanishes almost everywhere, up to a set of zero measure, will integrate (if it is integrable) to 0, both in the Lebesgue and Riemann sense. From a strictly mathematical perspective, the Dirac delta is usually defined as a limit of a convergent sequence of functions. However, such subtleties belong to the realm of distribution theory, and for our purposes is sufficient to say that the Dirac delta is a generalized function such that it diverges at 0, and admits the "filtering" property:

$$f(x) = \int_{-\infty}^{\infty} \mathrm{d}x'\, f(x')\delta(x' - x) \tag{1.22}$$

We also trivially generalize the Dirac delta to higher dimensions. For example, the three dimensional Dirac delta in cartesian coordinates is denoted as

$$\delta^3(\vec{\mathbf{r}}) = \delta(r_x)\delta(r_y)\delta(r_z) \tag{1.23}$$

The Dirac delta is typically considered to have units of inverse length.

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

1.1.6 Phasors and wavelets. A *phasor*, $ae^{i\phi}$, is a complex number representing a disturbance with a given amplitude and phase. a is the peak amplitude of the phasor, and ϕ is the phase. A one-dimensional monochromatic plane-wave can be represented as a simple real *wavelet* that describes the space and time dependent wave disturbance:

$$w(t) = a\cos\left(\frac{2\pi}{\lambda}x - \omega t + \varphi_0\right) \tag{1.24}$$

with ω being the angular frequency, λ the spatial wavelength and φ_0 a static phase-shift. It is often more convenient to deal with the wavelet in phasor form:

$$u(t) = ae^{i\left(\frac{2\pi}{\lambda}x - \omega t + \varphi_0\right)} \tag{1.25}$$

and clearly $\mathrm{Re}\{u\} = w$. u is called the *analytic representation* of the real wave disturbance w.

1.1.7 Additional definitions. The convolution between two functions $\mathbf{f}$ and $\mathbf{g}$ is defined as follows:

$$(f * g)(t) \triangleq \int_{-\infty}^{\infty} \mathrm{d}t'\, f(t - t')g(t') \tag{1.26}$$

Commonly used in optics are rectangular and circular aperture functions:

$$\mathrm{rect}\,(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \qquad \mathrm{circ}\,(r) = \begin{cases} 1 & r \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.27}$$

2 FUNCTIONAL ANALYSIS

2.1 Fourier Analysis

Fourier analysis is a crucial tool across a broad range of scientific and engineering applications, in particular in the analysis of time-harmonic fields and their statistics, which is at the centre of our discussion. The *Fourier transform* is a form of an integral transform, where a function is mapped from its original domain to another domain, or put equivalently: From one representation to another. The Fourier transform decomposes a function, or signal, into the frequencies that build up the original function. There are multiple conflicting definitions of the Fourier transform, with different disciplines favouring different normalization constants and units. To be consistent with the notation of a wavelet, viz. $e^{i(\vec{\mathbf{r}}\cdot\vec{\mathbf{k}} - \omega t)}$, we assume units of time t and angular frequency ω for the 1-dimensional transform pairs and use Fourier kernels with a positive sign, $e^{i\omega t}$. For higher dimensions (spatial transforms) we use kernels with a negative sign, $e^{-i\vec{\mathbf{r}}\cdot\vec{\mathbf{k}}}$, where we typically denote $\vec{\mathbf{r}}$, $\vec{\mathbf{k}}$ as spatial units and spatial frequency, respectively. The definition of (temporal) Fourier transform pairs is then:

$$\check{f}(\omega) = \mathscr{F}\left\{f\right\} \triangleq \int_{-\infty}^{\infty} \mathrm{d}t\, f(t)e^{i\omega t} \tag{2.1}$$

$$f(t) = \mathscr{F}^{-1}\left\{\check{f}\right\} \triangleq \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \check{f}(\omega)e^{-i\omega t} \tag{2.2}$$

This definition is only valid when the integrals converge and a sufficient conditions is that f and its Fourier transform are L^1-functions (absolutely integrable) on $\mathbb{R}$. The Fourier transform pair will generally be denoted with a wedge diacritic. The Fourier transform results (in general) in a complex-valued function that describes the magnitude and phase-shift of each frequency that is present in the original function. Regarding terminology: We say that the original function $f(t)$ is in the *time-domain*, while its transform pair $\check{f}(\omega)$ is in the *Fourier-domain* or *frequency-domain*, used interchangeably.

At times, when the integral in Eq. (2.1) does not converge but f is an L^2-function (square integrable, i.e. has finite average power), it is nonetheless possible to define a generalized Fourier transform as

$$\check{f}(\omega) = \mathscr{F}\left\{f\right\} \triangleq \lim_{R\to\infty}\int_{-R}^{R} \mathrm{d}t\, f(t)e^{i\omega t} \tag{2.3}$$

and similarly for $\mathscr{F}^{-1}$. A generalized Fourier transform for functions neither in L^1 not L^2 can also exist when the integral exists in terms of generalized functions. In those cases we say f is *Fourier transformable in the generalized sense.*

For (spatial) n-dimensional functions the definitions are extended in a natural manner, expect we conjugate the kernels:

$$\check{f}\left(\vec{\mathbf{k}}\right) = \mathscr{F}\left\{f\right\} \triangleq \int_{\mathbb{R}^n} \mathrm{d}^n\vec{\mathbf{r}}\ f(\vec{\mathbf{r}})e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{2.4}$$

$$f(\vec{\mathbf{r}}) = \mathscr{F}^{-1}\left\{\check{f}\right\} \triangleq \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \mathrm{d}^n\vec{\mathbf{k}}\ \check{f}\left(\vec{\mathbf{k}}\right)e^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{2.5}$$

and for L^2-functions:

$$\check{f}\left(\vec{\mathbf{k}}\right) = \mathscr{F}\left\{f\right\} \triangleq \lim_{R\to\infty}\int_{B_R} \mathrm{d}^n\vec{\mathbf{r}}\ f(\vec{\mathbf{r}})e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{2.6}$$

where $B_R = \{\vec{\mathbf{v}} \in \mathbb{R}^n : |\vec{\mathbf{v}}| < R\}$ is a ball centred at the origin and the limit is taken in the L^2-norm.

2.1.1 Real functions. Let f be a real-valued signal and $\check{f}$ its Fourier transform. Clearly $\mathscr{F}^{-1}\{\check{f}\}^\star = \mathscr{F}^{-1}\{\check{f}\}$, hence

$$\begin{aligned}\mathscr{F}^{-1}\left\{\check{f}\right\}^\star &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\check{f}(\omega)^\star e^{i\omega t}\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\check{f}(-\omega)e^{i\omega t} = \mathscr{F}^{-1}\left\{\check{f}\right\}\end{aligned} \tag{2.7}$$

and we deduce that $\check{f}(\omega)^\star = \check{f}(-\omega)$, i.e. $\check{f}$ is Hermitian. As the Fourier kernels $e^{-i\omega t}$ are also Hermitian, it follows that $\check{f}(\omega)e^{-i\omega t}$ is Hermitian as well. Thus, the real and imaginary parts of $\check{f}(\omega)e^{-i\omega t}$ are even and odd, respectively, and when integrating the odd part vanishes. Therefore,

$$f(t) = \mathscr{F}^{-1}\left\{\check{f}\right\} = \frac{1}{\pi}\,\mathrm{Re}\int_0^\infty \mathrm{d}\omega\,\check{f}(\omega)e^{-i\omega t} \tag{2.8}$$

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

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

Theorem 2.1 (Parseval's Theorem).

$$\int_{-\infty}^{\infty}\mathrm{d}t\,f(t)g(t)^\star = \int_{-\infty}^{\infty}\mathrm{d}\omega\,\check{f}(\omega)\check{g}(\omega)^\star \tag{2.9}$$

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

Theorem 2.2 (Convolution Theorem).

$$\mathscr{F}\{f * g\} = \mathscr{F}\{f\}\,\mathscr{F}\{g\} \tag{2.10}$$

$$\mathscr{F}\{fg\} = \mathscr{F}\{f\} * \mathscr{F}\{g\} \tag{2.11}$$

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

$$\mathrm{Corr}\{f,g\}(t) \triangleq \int_{-\infty}^{\infty}\mathrm{d}\tau\,f(t+\tau)g(\tau)^\star \tag{2.12}$$

Theorem 2.3 (Cross-Correlation Theorem).

$$\mathrm{Corr}\{f,g\}(t) = \mathscr{F}^{-1}\left\{\mathscr{F}\{f\}\,\mathscr{F}\{g\}^\star\right\} \tag{2.13}$$

Proof. Using the convolution theorem:

$$\begin{aligned}\mathrm{Corr}\{f,g\}(t) &= \int_{-\infty}^{\infty}\mathrm{d}\tau\,f(t+\tau)g(\tau)^\star = f(t) * g(-t)^\star\\ &= \mathscr{F}^{-1}\left\{\mathscr{F}\{f\}\,\mathscr{F}\{g\}^\star\right\}\end{aligned} \tag{2.14}$$

as desired. □

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

$$\mathscr{F}\{\delta(t-a)\} = e^{ia} \tag{2.15}$$

$$\mathscr{F}\left\{e^{iat}\right\} = 2\pi\delta(a+\omega) \tag{2.16}$$

$$\mathscr{F}\{\sin(at)\} = i\pi[\delta(\omega-a) - \delta(\omega+a)] \tag{2.17}$$

$$\mathscr{F}\{\cos(at)\} = \pi[\delta(\omega-a) + \delta(\omega+a)] \tag{2.18}$$

$$\mathscr{F}\left\{e^{-(at)^2}\right\} = \frac{\sqrt{\pi}}{|a|}e^{\frac{-\omega^2}{4a^2}} \tag{2.19}$$

Transform pairs of simple aperture functions:

$$\mathscr{F}\{\mathrm{rect}(t)\} = \mathrm{sinc}\left(\frac{\omega}{2}\right) \tag{2.20}$$

$$\mathscr{F}\{\mathrm{circ}\,(|\vec{\mathbf{r}}|)\} = \frac{2\pi}{|\vec{\mathbf{k}}|}\mathrm{J}_1\left(|\vec{\mathbf{k}}|\right) \tag{2.21}$$

where J_1 is the Bessel function of first kind. The transform of the circular aperture is known as the *Airy disk* and is generalized to higher dimensions in Appendix A.1. See Appendix A for additional identities involving the Fourier transform of apertures of different geometries.

2.2 Mercer's Theorem

The *Mercer's theorem* is a functional analysis tool, that decomposes a positive semi-definite Hilbert-Schmidt kernel into a sum of orthogonal components. It can be considered as the functional analogue of the principal component analysis in machine learning. We present it without proof, instead the reader is referred to Jorgens [1982]. We start with a few definitions.

Definition 2.4. Let $\Omega \subset \mathbb{R}^n$ be a closed set ($n \geq 1$). A function $\mathscr{K} : \Omega \times \Omega \to \mathbb{C}$ is called a *Hilbert–Schmidt kernel* if the following conditions are satisfied:

(1) $\mathscr{K}$ is L^2 in the domain $\Omega \times \Omega$

$$\int_\Omega\int_\Omega \mathrm{d}x\,\mathrm{d}y\,|\mathscr{K}(x,y)|^2 < \infty$$

(2) $\mathscr{K}$ is self-adjoint, i.e. $\mathscr{K}(x,y) = \mathscr{K}(y,x)^\star$

Remark: Ω can be relaxed to be any compact Hausdorff space.

Definition 2.5 (Hilbert–Schmidt Integral Operator). Given a kernel $\mathscr{K}$ and its domain Ω, then the associated *Hilbert–Schmidt integral operator* $T_{\mathscr{K}}$ acts upon a function f as follows:

$$T_{\mathscr{K}} f \triangleq \int_{\Omega} \mathrm{d}y\, \mathscr{K}(x,y) f(y)$$

Definition 2.6 (Mercer's Condition). A kernel $\mathscr{K} : \Omega \times \Omega \to \mathbb{C}$ is *positive semi-definite* if for any L^2-function f

$$\int_{\Omega}\int_{\Omega} \mathrm{d}x\, \mathrm{d}y\, \mathscr{K}(x,y) f(x) f(y)^{\star} \geq 0$$

Theorem 2.7 (Mercer's Theorem). *Assume a continuous positive semi-definite Hilbert–Schmidt kernel $\mathscr{K}$ (that fulfils Mercer's condition) on domain Ω. Then, there exists a countable set of orthonormal eigenfunctions, $\{\phi_j\}_{j=1}^{\infty}$, and non-negative eigenvalues, $\{\lambda_j\}_{j=1}^{\infty}$, of the associated Hilbert–Schmidt integral operator $T_{\mathscr{K}}$, viz.*

(1) $T_{\mathscr{K}}\phi_j = \lambda_j \phi_j$
(2) $\int_{\Omega} \mathrm{d}x\, \phi_n(x)\phi_m(x) = \delta_{nm}$

and $\mathscr{K}$ can be decomposed as follows:

$$\mathscr{K}(x,y) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \phi_j(y)^{\star}$$

where the convergence is absolute and uniform on $\Omega \times \Omega$.

2.3 Method of Green Functions

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

$$\mathscr{L}\ \Psi(\vec{\mathbf{r}}, t) = -f(\vec{\mathbf{r}}, t) \tag{2.22}$$

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

$$\mathscr{L}\ G\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}'; t, t'\right) = -\delta^3(\vec{\mathbf{r}} - \vec{\mathbf{r}}')\delta(t - t') \tag{2.23}$$

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

$$\Psi(\vec{\mathbf{r}}, t) = -\int_{-\infty}^{\infty} \mathrm{d}t' \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\mathbf{r}}'\, G\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}'; t, t'\right) f\left(\vec{\mathbf{r}}', t'\right) \tag{2.24}$$

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

$$\Psi(\vec{\mathbf{r}}, t) = -\iiiint \mathrm{d}t'\, \mathrm{d}^3\vec{\mathbf{r}}'\, G\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'; t - t'\right) f\left(\vec{\mathbf{r}}', t'\right) = -G * f \tag{2.25}$$

2.3.1 Retarded Green function. The wave operator, $\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$, is an important linear differential operator that will be employed extensively later on when we discuss electromagnetic waves. The Green function for the wave operator admits the following analytic form (see Zangwill [2013] for derivations)

$$G_{\pm}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}'; t, t'\right) = \frac{\delta\left(t - t' \pm \frac{1}{c}|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|\right)}{4\pi|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|} \tag{2.26}$$

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

$$\begin{aligned}\Psi(\vec{\mathbf{r}}, t) &= -\int_{-\infty}^{\infty} \mathrm{d}t' \int_{\mathbb{R}^3} \mathrm{d}^3\vec{\mathbf{r}}'\, G\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'; t - t'\right) f\left(\vec{\mathbf{r}}', t'\right) \\ &= \frac{1}{4\pi}\int_{\mathbb{R}^3} \mathrm{d}^3\vec{\mathbf{r}}'\, \frac{f\left(\vec{\mathbf{r}}', t - \frac{1}{c}|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|\right)}{|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|}\end{aligned} \tag{2.27}$$

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

$$\lim_{r\to\infty} r\left(\frac{\partial}{\partial r} - ik\right) G_0 = 0 \tag{2.28}$$

where the derivative is in spherical coordinates. The Sommerfeld radiation condition implies that a point impulse solution must behave like an outgoing propagating spherical wave. Noting that the Helmholtz operator is also translation invariant, the free-space Green function can be shown to be [Zangwill 2013]

$$G_0\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'\right) = \frac{1}{4\pi|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|} e^{ik|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|} \tag{2.29}$$

3 STOCHASTIC PROCESSES

Random variables naturally generalize to random (or stochastic) processes, where the uncertain values—the possible outcomes from an experiment—are functions instead of numbers. Consider a physical process that undergoes random fluctuations. For example, a gas-discharge lamp contains a large collection of excited atoms which randomly drop to a lower energy state and in the process emit light energy in the form of a photon. Such spontaneous emission is a stochastic process where photons with distinct energies have different probabilities—possibly varying over time—of being emitted (and those probabilities define the emission spectrum). Curious readers are encouraged to refer to more comprehensive sources: Miller [2012] for an introductory textbook on random processes, and Goodman [2015] for a statistical optics perspective.

We now formally define a stochastic process: Let Ω be a (finite or infinite) sample space with an accompanying probability measure. Then, a stochastic process $U(t)$ is defined as an ensemble of sample functions, called *realizations*, and can be written as $U = \{^{\zeta}u(t)\}$

with $\zeta \in \Omega$. Note that a stochastic process implies the existence of the underlying state space and measure, however for simplicity we omit this dependence from the notation. Each realization ${}^{\zeta}u(t)$ can be thought of as a possible outcome of a singular experiment that measures the stochastic process (e.g., the observed photons and their energy emitted over time), and is typically a function of time. Generalization to spatial functions (or other ordered index sets) is straightforward though of little use for our discussion. Being time-dependent processes, it is natural to define the *time-averaging operator*:

$$\left\langle {}^{\zeta}u \right\rangle_{\mathrm{t}} \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t\, {}^{\zeta}u(t) \tag{3.1}$$

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

$$\begin{aligned} \tilde{\Gamma}_{\zeta\xi}(\tau) &\triangleq \left\langle {}^{\zeta}u(t+\tau)\, {}^{\xi}u(t)^{\star} \right\rangle_{\mathrm{t}} \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t\, {}^{\zeta}u(t+\tau)\, {}^{\xi}u(t)^{\star} \end{aligned} \tag{3.2}$$

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

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

$$\langle U(t) \rangle \triangleq \int_{\Omega} \mathrm{d}u\, u p_U(u; t) \tag{3.3}$$

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

$$\begin{aligned} \Gamma_{UV}(t_1, t_2) &\triangleq \left\langle U(t_1) V(t_2)^{\star} \right\rangle \\ &= \iint \mathrm{d}u\, \mathrm{d}v\, u v^{\star} p_{UV}(u, v; t_1, t_2) \end{aligned} \tag{3.4}$$

In similar fashion to the time-based case, we denote the *statistical autocorrelation function* as Γ_{UU}. We choose not to consider higher-order moments.

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

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

$$\langle U \rangle = \left\langle {}^{\zeta}u \right\rangle_{\mathrm{t}} \tag{3.5}$$

$$\Gamma_{UU}(t_1, t_2) = \Gamma_{UU}(\tau) = \tilde{\Gamma}_{\zeta\zeta}(\tau) \tag{3.6}$$

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

As an example consider again our gas discharge lamp. The lifetime of such a lamp is typically far greater than the characteristic time of emission, hence we can consider the process to be stationary: If we were to perform a couple of experiments, each starting at a different time point and each measuring an emission realizations, we'd expect the produced sample functions to be statistically similar. Put differently, the probability of some photon energy being emitted remains constant over time. This process is also ergodic: Each realization will measure the lamp's entire spectrum of emission.

3.0.2 Spectral analysis. Each realization $u(t)$ of a stochastic process U is a function of time, thus we are encouraged to study the spectral decomposition $\mathscr{F}\{u\}$. Unfortunately, for many physical processes the realizations are not Fourier transformable, not even in the generalized sense (Eq. (2.3)). This holds in general for stationary processes as well, where by definition energy is distributed over all time. However, the ensemble averaged decomposition does typically exist. We assume that U is stationary, at least in the wide sense, and define the truncated, normalized Fourier transform-like of a realization to be

$$\bar{u}_T(\omega) = \frac{1}{\sqrt{2T}} \int_{-T}^{T} \mathrm{d}t\, u(t) e^{i\omega t} \tag{3.7}$$

By Parseval's theorem (Theorem 2.1), the total energy is preserved in transform pairs, therefore $|\bar{u}_T|^2$ being normalized describes the power in each spectral component of the truncated signal. At the limit, we define the *power spectral density* (PSD) of the process U:

$$S_{UU}(\omega) \triangleq \lim_{T\to\infty} \left\langle |\bar{u}_T(\omega)|^2 \right\rangle \tag{3.8}$$

which has units of power per unit frequency. We now ready to present the celebrated *Wiener–Khinchin theorem*:

Theorem 3.1 (Wiener–Khinchin Theorem). *The spectral decomposition of the autocorrelation function of a stationary process is its power spectral density.*

Proof. With the definition of the PSD (Eq. (3.8)) as the starting point, note that integration and ensemble averaging commute and by using the properties of a stationary process we deduce:

$$\begin{aligned} S_{UU}(\omega) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \int_{-T}^{T} \mathrm{d}t' \left\langle u(t) u(t')^{\star} \right\rangle e^{i\omega(t-t')} \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \int_{-T-t}^{T-t} \mathrm{d}\tau\, \Gamma_{UU}(\tau) e^{i\omega\tau} \\ &= \lim_{T\to\infty} \int_{-T}^{T} \mathrm{d}\tau\, \Gamma_{UU}(\tau) e^{i\omega\tau} = \mathscr{F}\{\Gamma_{UU}\} \end{aligned} \tag{3.9}$$

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

This is a surprising result: The PSD and the autocorrelation function are intrinsically linked as Fourier transform pairs for any stationary stochastic process! The Wiener–Khinchin theorem can be regarded as a special case of the cross-correlation theorem (Theorem 2.3), however it applies under a much more general setting. The definition above is trivially extended to become *cross power spectral density* of a couple of stochastic processes U and V:

$$S_{UV}(\omega) \triangleq \lim_{T\to\infty} \left\langle \bar{u}_T(\omega)\bar{v}_T(\omega)^\star \right\rangle = \mathscr{F}\{\Gamma_{UV}\} \tag{3.10}$$

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

It can be shown that even when the limit in Eq. (3.8) does not converge, a spectral decomposition still exists in form of some power spectrum distribution. Furthermore, the theorem can also be generalized to a wider class of stochastic processes. We won't prove either of those claims, however, as the above formulation is sufficient for our use. The Wiener–Khinchin theorem will be implicitly used throughout this manuscript. When we consider a spectral decomposition of the autocorrelation we assume its existence is given by the Wiener–Khinchin theorem.

4 ELECTROMAGNETIC WAVES IN VACUUM

Light energy is propagated by electromagnetic waves. The intrinsic quantities that give rise to those waves are the electric, $\vec{\mathbf{E}}$, and magnetic, $\vec{\mathbf{B}}$, fields whose behaviour is governed by Maxwell's famous set of equations. Those equations are the fundamental principles that describe the interaction of those fields with charge and current, as well as the time-dependent interaction between the fields themselves. It is that interaction between the electric and magnetic fields that results in self-supporting fields that oscillate in unison and give rise to time-harmonic electromagnetic waves. We provide a very brief discussion about the relevant classical electrodynamics and we elect to use the Gaussian-cgs units in this write-up. The justification to that choice of a units-system is that while the SI units are more popular and familiar, Gaussians units are simpler and make fundamental physical properties and insights clearer—e.g., the pesky SI constants ϵ_0 and μ_0 are artefacts of SI units and not physical properties of free-space.

We start with the general Maxwell equations:

$$\boldsymbol{\nabla}\cdot\vec{\mathbf{E}} = 4\pi\rho \qquad \boldsymbol{\nabla}\cdot\vec{\mathbf{B}} = 0 \tag{4.1}$$

$$\boldsymbol{\nabla}\times\vec{\mathbf{E}} = -\frac{1}{c}\frac{\partial\vec{\mathbf{B}}}{\partial t} \qquad \boldsymbol{\nabla}\times\vec{\mathbf{B}} = \frac{4\pi}{c}\vec{\mathbf{j}} + \frac{1}{c}\frac{\partial\vec{\mathbf{E}}}{\partial t} \tag{4.2}$$

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

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\vec{\mathbf{E}} = -\frac{1}{c}\frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\times\vec{\mathbf{B}}\right) = -\frac{1}{c^2}\frac{\partial^2\vec{\mathbf{E}}}{\partial t^2} - \frac{4\pi}{c^2}\frac{\partial\vec{\mathbf{j}}}{\partial t} \tag{4.3}$$

and by using the vector triple product identity (1.10), we deduce that the electric and magnetic fields satisfy the inhomogeneous wave equations:

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\mathbf{E}} = 4\pi\left(\frac{1}{c}\frac{\partial\vec{\mathbf{j}}}{\partial t} + \boldsymbol{\nabla}\rho\right) \tag{4.4}$$

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{\mathbf{B}} = -\frac{4\pi}{c}\boldsymbol{\nabla}\times\vec{\mathbf{j}} \tag{4.5}$$

The operator on the left-hand side in Eqs. (4.4) and (4.5) is known as the *d'Alembert operator*, or the wave operator, and we denote it via

$$\triangle_{\mathrm{m}} = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \tag{4.6}$$

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

$$\triangle_{\mathrm{m}}\vec{\mathbf{E}} = 0 \qquad \triangle_{\mathrm{m}}\vec{\mathbf{B}} = 0 \tag{4.7}$$

While Eqs. (4.4) and (4.5) govern the sourcing of electromagnetic waves, the homogeneous partial differential equations above govern the propagation of those waves. Sourcing of high-frequency electromagnetic waves is better explained quantum mechanically and not classically, making the general sourcing Maxwell equations of little use for us in computer rendering. The propagation of electromagnetic waves in space and matter, on the other hand, is the foundation of light transport and the homogeneous wave equations above are the driving principle.

4.0.1 Transverse electromagnetic waves. An important class of solutions to the free-space wave Equations (4.7) are *transverse electromagnetic* waves (TEM). This simple kind of waves are commonly used in practical applications (like computer rendering), provide some physical insight and are useful to study propagation in matter and polarization, which will be discussed later.

TEM waves are waves where the electric and magnetic fields are perpendicular to the direction of propagation. Let $\vec{\mathbf{k}}$ be the wave's direction of propagation, i.e. the *wavevector*, and without loss of generality we set $\vec{\mathbf{k}} = k\hat{\mathbf{z}}$, with $k = |\vec{\mathbf{k}}|$ being the *wavenumber*. The spatial frequency is then $2\pi k$ and the wavelength is related to the wavenumber via $\lambda = \frac{2\pi}{k}$. Then, we are looking for fields $\vec{\mathbf{E}}(z,t), \vec{\mathbf{B}}(z,t)$ that satisfy the homogeneous wave equations as well as Maxwell's source-free equations. From Eqs. (4.1) and (4.2) we immediately deduce that

$$\frac{\partial E_z}{\partial z} = 0 \qquad \frac{\partial E_z}{\partial t} = c\hat{\mathbf{z}}\cdot\boldsymbol{\nabla}\times\vec{\mathbf{B}} = c\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) = 0 \tag{4.8}$$

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

$$\vec{\mathbf{E}} = \mathbf{f}_\perp(z-ct) + \mathbf{g}_\perp(z+ct) \tag{4.9}$$

Further, using Eq. (4.2) we write

$$\begin{aligned}\frac{\partial\vec{\mathbf{B}}}{\partial t} &= -c\boldsymbol{\nabla}\times\vec{\mathbf{E}} = -c\hat{\mathbf{z}}\times\frac{\partial}{\partial z}(f_\perp(z-ct) + \mathbf{g}_\perp(z+ct)) \\ &= \hat{\mathbf{z}}\times\frac{\partial}{\partial t}(f_\perp(z-ct) - \mathbf{g}_\perp(z+ct))\end{aligned} \tag{4.10}$$

And thus

$$\vec{\mathbf{B}} = \hat{\mathbf{z}} \times \mathbf{f}_\perp(z - ct) - \hat{\mathbf{z}} \times \mathbf{g}_\perp(z + ct) \tag{4.11}$$

up to a constant which we can neglect. Noting again that $\vec{\mathbf{k}} \parallel \hat{\mathbf{z}}$ we deduce that indeed $\vec{\mathbf{E}} \cdot \vec{\mathbf{k}} = \vec{\mathbf{B}} \cdot \vec{\mathbf{k}} = 0$, and hence Eqs. (4.9) and (4.11) describe a TEM wave. However, observe that it does not hold in general that $\vec{\mathbf{E}} \cdot \vec{\mathbf{B}} = 0$ (such waves always admit some standing wave characteristic).

The most simple form of a TEM wave is the monochromatic plane-wave. Those kind of waves have very simple geometric properties, which is why they are ubiquitously used in computer rendering. Let $\mathbf{f}_\perp \equiv 0$, then the solution to the homogeneous wave equations becomes:

$$\vec{\mathbf{E}}(\vec{\mathbf{r}}, t) = \mathcal{E}_\perp e^{i(\vec{\mathbf{k}} \cdot \vec{\mathbf{r}} - \omega t)} \tag{4.12}$$

$$\vec{\mathbf{B}}(\vec{\mathbf{r}}, t) = -(\hat{\mathbf{k}} \times \mathcal{E}_\perp) e^{i(\vec{\mathbf{k}} \cdot \vec{\mathbf{r}} - \omega t)} \tag{4.13}$$

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

$$u_{\text{EM}} = \frac{1}{8\pi}\left[\left(\operatorname{Re}\vec{\mathbf{E}}\right)^2 + \left(\operatorname{Re}\vec{\mathbf{B}}\right)^2\right] \tag{4.14}$$

that is, the complex parts of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ contribute no energy. The complex representation is used solely for convenient and ease of computation. As plane-waves extend to infinity in the transverse direction (perpendicular to $\vec{\mathbf{k}}$), they are clearly aphysical! Nonetheless, plane-waves are a good local approximation to real waves in a confined region of time and space, and they admit useful properties that make them easy to use and reason about: Note that for plane-waves it always holds that $\vec{\mathbf{E}} \cdot \vec{\mathbf{B}} = 0$ hence the vectors $\{\vec{\mathbf{k}}, \vec{\mathbf{E}}, \vec{\mathbf{B}}\}$ give rise to an orthogonal triad, with $\vec{\mathbf{B}}$ being uniquely defined by the former two vectors. This means we do not need to concern ourselves with the $\vec{\mathbf{B}}$ field. The *Poynting vector*—the energy current density and direction of the electromagnetic flux—is defined as

$$\vec{\mathbf{S}} = \frac{c}{4\pi}\left(\vec{\mathbf{E}} \times \vec{\mathbf{B}}\right) \tag{4.15}$$

for any electromagnetic wave. In the case of plane-waves, $\vec{\mathbf{S}}$ is clearly aligned with $\vec{\mathbf{k}}$ and the time-averaged energy density and Poynting vector are easily calculated:

$$\langle u_{\text{EM}} \rangle_t = \frac{1}{8\pi}|\mathcal{E}_\perp|^2 \tag{4.16}$$

$$\left\langle \vec{\mathbf{S}} \right\rangle_t = \langle u_{\text{EM}} \rangle_t \, c\hat{\mathbf{k}} \tag{4.17}$$

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

$$I(\vec{\mathbf{r}}) = \left|\operatorname{Re}\left\langle \vec{\mathbf{S}}(\vec{\mathbf{r}}) \right\rangle_t\right| \tag{4.18}$$

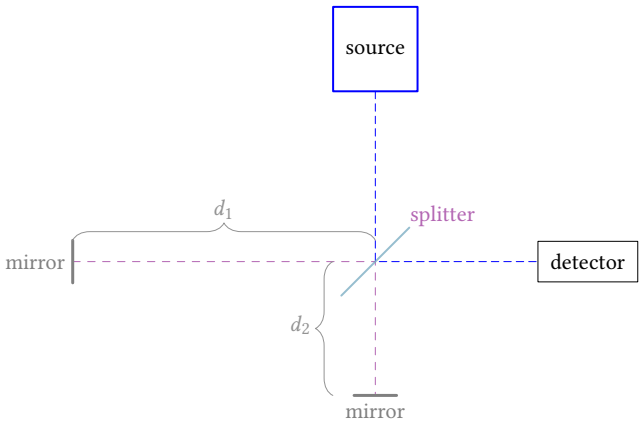

Fig. 1. Michelson interferometer: Light is emitted by a source and is then split into two beams. Each beam travels until it is reflected by a mirror, then the beams are recombined and are observed by the detector. The distance difference between the travelled beams, $d = d_2 - d_1$, is controllable by moving the mirrors. This distance difference induces a time-delay of $\tau = \frac{d}{c}$ between the beams which enables us to measure the light's ability to interfere with a time-delayed version of itself, i.e. its temporal coherence.

5 OPTICAL COHERENCE THEORY

So far we have been concerned with deterministic fields, waves and wave packets. Real-life radiation from physical sources, on the other hand, can not be adequately explained classically and is fundamentally a quantum mechanics process. Thermal light sources, gas discharge lamps, light-emitting diodes (LEDs) all emit light by the spontaneous emission of a photon when an excited particle transitions to a lower energy state. Even stimulated emission, such as laser radiation, contains some randomness due to contributions from spontaneous emission as well as uncontrollable environment factors, e.g., temperature and mechanical vibrations. Furthermore, fields at visible spectra oscillate too rapidly for the oscillations to be directly measured, therefore it is only the time-averaged quantities that are observed. This calls for a statistical treatment of those wave packets. The optical coherence theory, which can be considered as the study of observable optical quantities, has been pioneered by Emil Wolf, who also remained a dominant figure and contributor to the field. This section loosely follows Wolf [2007]. Additional resources are Born and Wolf [1999]; Goodman [2015]; Mandel and Wolf [1995].

5.0.1 Temporal coherence. To simplify the discussion we start by ignoring polarization. Polarization does play an important role in optical coherence and will be revisited later in Section 6. Let $u(\vec{\mathbf{r}}, t)$ be a scalar wave packet. Consider an experiment involving the Michelson interferometer (Fig. 1), a device where light is split into two paths, then recombined inducing a time-delay between the recombined beams. The recombined beam is then observed by a

detector. The distance difference between the paths is denoted as d resulting in a $\tau = \frac{d}{c}$ time-delay between the recombined beams. The amplitude of the field at the observation point at some time t, denoted $w(\tau)$, is then the superposition of the recombined beams:

$$w(\tau) = \alpha_1 u(t) + \alpha_2 u(t+\tau) \tag{5.1}$$

where $\alpha_{1,2}$ are real constants that describe the field strength that remains in each beam after interacting with the Michelson interferometer. If the splitting and recombining machinery introduces no phase-shifts then we can assume that $\alpha_{1,2}$ are real values. Further, if we assume that the energy lost in the process is negligible, then $\alpha_1^1 + \alpha_2^2 = 1$. As mentioned, the fluctuations of the field u are not directly observable, we therefore treat u as a *wave ensemble*—a (stationary, in the wide sense, and ergodic) stochastic process. Denote $u_1 = u(t)$, $u_2 = u(t+\tau)$, then the observed intensity of w by the detector is the time-averaged quantity (up to a constant):

$$\begin{aligned} I(w(\tau)) &= \left\langle w(\tau) w(\tau)^\star \right\rangle_\mathrm{t} \\ &= \alpha_1^2 \left\langle u_1 u_1^\star \right\rangle_\mathrm{t} + \alpha_2^2 \left\langle u_2 u_2^\star \right\rangle_\mathrm{t} + 2\,\mathrm{Re}\left\{\alpha_1\alpha_2 \left\langle u_1 u_2^\star \right\rangle_\mathrm{t}\right\} \\ &= I(u) + 2\alpha_1\alpha_2\,\mathrm{Re}\,\tilde{\Gamma}_{uu}(\tau) \end{aligned} \tag{5.2}$$

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

$$\gamma_{uu}(\tau) \triangleq \frac{\Gamma_{uu}(\tau)}{\Gamma_{uu}(0)} \tag{5.3}$$

Notice that $|\gamma_{uu}| \leq 1$, and we can rewrite Eq. (5.2) in the form of the *temporal interference law for stationary fields*:

$$I(u_1 + u_2) = I(u_1) + I(u_2) + 2\sqrt{I(u_1)}\sqrt{I(u_2)}\,\mathrm{Re}\,\gamma_{uu}(\tau) \tag{5.4}$$

for any stationary u and any τ, the time-delay between u_1 and u_2. When $I(w) = 0$ total *destructive interference* has occurred, while when $I(w) = 2I(u)$ total *constructive interference* has occurred. The significance of the Michelson interferometer is its ability to measure the *temporal coherence* of the observed light: By slowly changing the distance difference between the paths, and thus the time-delay τ, we measure the ability of the light beam to interfere with a time-delayed version of itself. This physical property is quantified by the *degree of temporal coherence* γ_{uu}.

5.0.2 Spatial coherence. Our discussion of temporal coherence implicitly assumed that the beams enter the Michelson interferometer at a singular point, giving rise to a clearly aphysical perfect point light source. We now dispense with that assumption and consider the spatial extent of the light source. Consider Young's famous double slit experiment (Fig. 2) where light passes through two thin slits situated in close proximity to each other and then is observed on a screen (we ignore and do not formalise some practical requirements that are needed to perform the experiment, such as the size of the slits and distance from the source). Assume the following geometry: An opaque screen, with two slits notched in it at points $\vec{\mathbf{r}}_1$, $\vec{\mathbf{r}}_2$, and

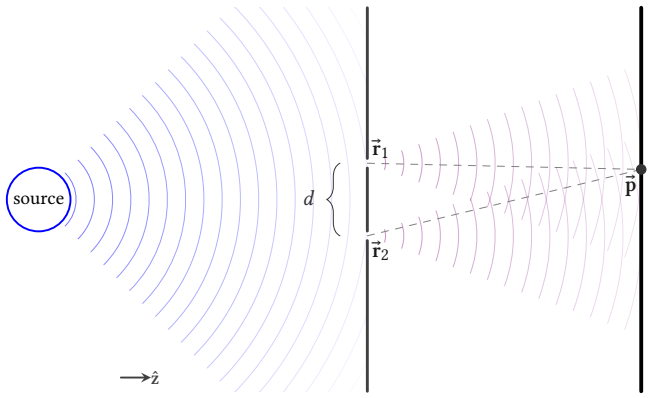

Fig. 2. Young's double slit experiment: Light is incident upon an opaque screen (center) with two thin slits notched in it. The slits are positioned at points $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$. The light diffracts through the slits and produces an interference pattern that is observed on a screen (right). The intensity of the interference pattern at an observation point $\vec{\mathbf{p}}$ depends both on the temporal coherence properties of the light, due to time difference between the light arriving from $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$; as well as the spatial coherence properties that come into effect due to distance between the slits $d = |\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2|$.

a screen with an observation point $\vec{\mathbf{p}}$. Both screens are parallel to the xy-axis. A light source is placed such that the screen containing the slits is between the source and the observation screen, and the centre of the light source, the middle point between the slits and $\vec{\mathbf{p}}$ all lie on the z-axis. Denote the distance between the slits $d = |\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2|$. As before, let u be the amplitude of the emitted light and the amplitude at the point of interest $\vec{\mathbf{p}}$ can be written as the superposition

$$u(\vec{\mathbf{p}}, t) = K_1 u\left(\vec{\mathbf{r}}_1, t - \frac{|\vec{\mathbf{p}} - \vec{\mathbf{r}}_1|}{c}\right) + K_2 u\left(\vec{\mathbf{r}}_2, t - \frac{|\vec{\mathbf{p}} - \vec{\mathbf{r}}_2|}{c}\right) \tag{5.5}$$

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

$$\begin{aligned} I(u(\vec{\mathbf{p}}, t)) =& |K_1|^2 I\big(u(\vec{\mathbf{r}}_1, t')\big) + |K_2|^2 I\big(u(\vec{\mathbf{r}}_2, t'')\big) \\ &+ 2\,\mathrm{Re}\left\{K_1 K_2^\star \left\langle u(\vec{\mathbf{r}}_1, t') u(\vec{\mathbf{r}}_2, t'')^\star \right\rangle_\mathrm{t}\right\} \end{aligned} \tag{5.6}$$

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

$$\Gamma_{uu}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \tau) \triangleq \left\langle u(\vec{\mathbf{r}}_1, t) u(\vec{\mathbf{r}}_2, t+\tau)^\star \right\rangle_\mathrm{t} \tag{5.7}$$

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

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

$$\gamma_{uu}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \tau) \triangleq \frac{\Gamma_{uu}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \tau)}{\sqrt{\Gamma_{uu}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_1, 0)}\sqrt{\Gamma_{uu}(\vec{\mathbf{r}}_2, \vec{\mathbf{r}}_2, 0)}} \tag{5.8}$$

and it is trivially verifiable that $|\gamma| \leq 1$ (by Schwarz's inequality, Eq. (1.19)). Using the degree of coherence we can rewrite Eq. (5.6) as the *interference law for stationary fields*

$$I(u(\vec{\mathbf{r}}_1, t) + u(\vec{\mathbf{r}}_2, t+\tau)) = I(u(\vec{\mathbf{r}}_1, t)) + I(u(\vec{\mathbf{r}}_2, t)) + 2\sqrt{I(u(\vec{\mathbf{r}}_1, t))}\sqrt{I(u(\vec{\mathbf{r}}_2, t))}\,\mathrm{Re}\,\gamma_{uu}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \tau) \quad (5.9)$$

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

To understand how the spatial extent of the light source plays an effect on the observed intensity, we model the light source as a collection of a great many independent elementary radiators, an accurate model for spontaneous emission sources. Let ${}^{j}u(\vec{\mathbf{r}}, t)$ be a singular wavelet emitted by a radiator. The wave ensemble is then a collection of all such emitted wavelets. Then, we can formulate the amplitude that diffracts through the slits and is measured at the observation point as the ensemble average $\langle u(\vec{\mathbf{p}}, t)\rangle$. The instantaneous intensity is then:

$$I(u(\vec{\mathbf{p}}, t)) = \left\langle u(\vec{\mathbf{p}}, t) u(\vec{\mathbf{p}}, t)^{\star} \right\rangle \quad (5.10)$$

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

From a physical perspective, it makes little difference how we formulate the sample functions of the stochastic process. In the example above we formulated the wavelet emitted from each elementary radiator as a sample function. Alternative we can consider a sample function as a realization of an experiment measuring the amplitude. Some formulations are easier to reason about or serve a pedagogical purpose, however from a purely physical standpoint they are equivalent. This justifies the decision to drop the subscripts from the coherence functions, but it is important to remember that an underlying stochastic process is implied nonetheless.

5.1 Coherence in the Spectral Domain

The space-time formulation of the mutual coherence presented in the previous subsection is useful when the light is monochromatic, or satisfies some monochromaticity conditions. When studying the interaction of light and (dispersive) matter, it is more natural to study each frequency component independently in-place of the time-dependant field. To that end, given a polychromatic wave ensemble, the (by now) firmly recognizable game plan is to consider the Fourier transform of the mutual coherence function. In addition to rendering some of the mathematics more tractable, an apparent advantage is that by treating each spectral component individually, the coherence time is effectively infinite and we can ignore temporal effects. We will show formally that it is indeed the case and discuss caveats.

First, we formally present the space-frequency formulation of coherence: In Section 3 we discussed the power spectral density of a stationary process and its relation to the autocorrelation. This relation is described by the Wiener–Khinchin theorem (Theorem 3.1) which states that the transform pair of the autocorrelation function is the cross power spectral density, known as the *cross-spectral density* under the context of optical coherence theory. Viz.

$$\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) \triangleq \mathscr{F}\{\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, t)\} \quad (5.11)$$

Just as the mutual coherence Γ is the central quantity in the space-time formulation of coherence, the cross-spectral density $\mathcal{W}$ is the central quantity in the space-frequency formulation of coherence. To show that different spectral components are uncorrelated, at least for ergodic processes, consider a truncated Fourier transform of the signal (an unnormalized version of Eq. (3.7)):

$$\bar{u}_T(\vec{\mathbf{r}}, \omega) = \int_{-T}^{T} \mathrm{d}t\, u(\vec{\mathbf{r}}, t) e^{i\omega t} \quad (5.12)$$

Then, consider the following ensemble average:

$$\lim_{T\to\infty} \left\langle \bar{u}_T(\vec{\mathbf{r}}, \omega) \bar{u}_T(\vec{\mathbf{r}}', \omega')^{\star} \right\rangle = \lim_{T\to\infty} \int_{-T}^{T} \mathrm{d}t\, e^{i(\omega-\omega')t} \times \int_{-T+t}^{T+t} \mathrm{d}\tau \left\langle u(\vec{\mathbf{r}}, t) u(\vec{\mathbf{r}}', t+\tau)^{\star} \right\rangle e^{-i\omega'\tau} \quad (5.13)$$

Due to stationarity the integrand of the second integral is independent of t. Then, at the limit the second integral is the (inverse) Fourier transform of $\Gamma(\vec{\mathbf{r}}, \vec{\mathbf{r}}', \tau)$, i.e. the cross-spectral density, while the first integral is calculated trivially using the Fourier identity (2.16), resulting in the expression

$$\lim_{T\to\infty} \left\langle \bar{u}_T(\vec{\mathbf{r}}, \omega) \bar{u}_T(\vec{\mathbf{r}}', \omega')^{\star} \right\rangle = 2\pi\delta(\omega - \omega')\, \mathcal{W}(\vec{\mathbf{r}}, \vec{\mathbf{r}}', -\omega') \quad (5.14)$$

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

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

$$w(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) \triangleq \frac{\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega)}{\sqrt{\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_1, \omega)}\sqrt{\mathcal{W}(\vec{\mathbf{r}}_2, \vec{\mathbf{r}}_2, \omega)}} \quad (5.15)$$

The observed intensity of a wave ensemble at a given point is simply its self coherence, i.e. $I = \Gamma(\vec{\mathbf{r}}, \vec{\mathbf{r}})$. Note that the same information is contained in the cross-spectral density if we were to integrate over the relevant frequencies:

$$\Gamma(\vec{\mathbf{r}}, \vec{\mathbf{r}}, 0) = \mathscr{F}^{-1}\{\mathcal{W}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)\}\Big|_{t=0} = \frac{1}{\pi}\int_0^{\infty} \mathrm{d}\omega\, \mathcal{W}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega) \quad (5.16)$$

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

5.1.1 Coherent-modes representation of the spectrum. An important result presented by Wolf [1982] is the decomposition of the cross-spectral density. Let $D \subset \mathbb{R}^3$ be a closed set, large enough to be the spatial domain of interest. Note that for any ω the following holds

$$\mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) = \mathcal{W}\left(\vec{\mathbf{r}}_2, \vec{\mathbf{r}}_1, \omega\right)^\star \tag{5.17}$$

$$\int_D \int_D \mathrm{d}\vec{\mathbf{r}}\, \mathrm{d}\vec{\mathbf{r}}' \left|\mathcal{W}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}', \omega\right)\right|^2 < \infty \tag{5.18}$$

where the integral inequality holds because the cross-spectral density is point-wise finite (for a physical process) and D is a finite domain. The above are the conditions of the Hilbert–Schmidt kernel (Definition 2.4), therefore $\mathcal{W}$ with domain $D \times D$ is a kernel function. It is also a positive semi-definite kernel (See Wolf [1982] for a proof), therefore Mercer's condition is satisfied as well. Then, by Mercer's theorem (Theorem 2.7), there exists a decomposition:

$$\mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) = \sum_{n=1}^{\infty} \alpha_n(\omega) \phi_n(\vec{\mathbf{r}}_1) \phi_n(\vec{\mathbf{r}}_2)^\star \tag{5.19}$$

where $\alpha_n \geq 0$ are eigenvalues and ϕ_n are eigenfunctions of the integral operator $T_{\mathcal{W}}$ (Definition 2.5). The set of eigenfunctions form an orthonormal basis, viz. $\int_D \phi_n \phi_m = \delta_{nm}$, and is a countable set at most. Such decomposition is called the *coherent-modes* decomposition, where $\mathcal{W}^{(n)}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) = \alpha_n(\omega) \phi_n(\vec{\mathbf{r}}_1) \phi_n(\vec{\mathbf{r}}_2)^\star$ are the coherent-modes. Notice now that by taking the inverse Fourier transform of each side in Eq. (5.11), we get:

$$\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, t) = \mathscr{F}^{-1}\{\mathcal{W}\} = \sum_n \mathscr{F}^{-1}\left\{\alpha_n(\omega) \phi_n(\vec{\mathbf{r}}_1) \phi_n(\vec{\mathbf{r}}_2)^\star\right\} \tag{5.20}$$

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

A theoretical result immediately follows:

Corollary 5.1. *A wave ensemble is spatially coherent everywhere in a domain D if and only if the coherent-mode decomposition of its cross-spectral density in D consists of a single mode.*

Proof. Consider a wave ensemble where the coherent-modes decomposition can be written as a single mode, i.e. $\mathcal{W} \equiv \mathcal{W}^{(1)}$. Then, it is trivially verifiable that in the entire domain D it holds $|w| \equiv 1$. The other direction follows from $|w| \equiv 1$ and some algebra. □

That is, we have a derived a sufficient and necessary condition for the formation of a fully spatially coherent field.

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

$$v(\vec{\mathbf{r}}, \omega) = \sum_{n=1}^{\infty} a_n(\omega) \phi_n(\vec{\mathbf{r}}) \tag{5.21}$$

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

$$\mathrm{E}\left[a_n(\omega) a_m(\omega)^\star\right] = \delta_{nm} \alpha_n(\omega) \quad \text{and} \quad \sum_n |a_n(\omega)| < \infty \tag{5.22}$$

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

$$\begin{aligned}\left\langle v(\vec{\mathbf{r}}_1, \omega) v(\vec{\mathbf{r}}_2, \omega)^\star \right\rangle_\omega &\triangleq \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mathrm{E}\left[a_n(\omega) a_m(\omega)^\star\right] \phi_n(\vec{\mathbf{r}}_1) \phi_m(\vec{\mathbf{r}}_2)^\star \\ &= \mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right)\end{aligned} \tag{5.23}$$

indicating that the space-frequency ensemble average over the ensemble of monochromatic realizations, as defined, indeed reproduces the cross-spectral density. In addition, using Eq. (5.16) we immediately observe that the observed intensity at a point becomes:

$$\pi I(\vec{\mathbf{r}}) = \int_0^\infty \mathrm{d}\omega \left\langle v(\vec{\mathbf{r}}, \omega) v(\vec{\mathbf{r}}, \omega)^\star \right\rangle_\omega = \int_0^\infty \mathrm{d}\omega\, \mathcal{W}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega\right) \tag{5.24}$$

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

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

$$v(\vec{\mathbf{p}}, \omega) e^{-i\omega t} = v(\vec{\mathbf{q}}_1, \omega) e^{-i\omega \frac{r_1}{c}} e^{-i\omega t} + v(\vec{\mathbf{q}}_2, \omega) e^{-i\omega \frac{r_2}{c}} e^{-i\omega t} \tag{5.25}$$

Denoting $\mathscr{I} = \mathcal{W}(\vec{\mathbf{p}}, \vec{\mathbf{p}}, \omega)$ and $\mathscr{I}'_{1,2} = \mathcal{W}(\vec{\mathbf{r}}_{1,2}, \vec{\mathbf{r}}_{1,2}, \omega)$, the spectral density at the observation point is then

$$\mathscr{I} = \mathscr{I}'_1 + \mathscr{I}'_2 + 2\sqrt{\mathscr{I}'_1}\sqrt{\mathscr{I}'_2}\, \mathrm{Re}\left[w(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) e^{-i\omega \frac{r_1 - r_2}{c}}\right] \tag{5.26}$$

which is the spectral interference law. While the interference law (Eq. (5.9)) describes the intensity of a superposition of waves of any state of coherence, the spectral interference law describes the spectral density of the superposition of monochromatic wavelets of any state of coherence.

5.1.2 Separability of temporal and spatial coherence. So far we have shown that under the space-frequency formulation, distinct frequencies can be treated independently while ignoring temporal effects. Clearly temporal effects do not just disappear, and we will show now that temporal coherence gets reintroduced when taking the intensity in the space-frequency domain. Consider the spectral interference law (Eq. (5.26)), applying Eq. (5.24) we integrate both sides to calculate the intensity:

$$I = I'_1 + I'_2 + \frac{2}{\pi}\, \mathrm{Re} \int_0^\infty \mathrm{d}\omega\, \mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) e^{-i\omega \frac{r_1 - r_2}{c}} \tag{5.27}$$

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

$$\frac{2}{\pi}\, \mathrm{Re} \int_0^\infty \mathrm{d}\omega\, \mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) e^{-i\omega \frac{r_1 - r_2}{c}} = 2\, \mathrm{Re}\, \Gamma\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \frac{r_1 - r_2}{c}\right) \tag{5.28}$$

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

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

$$\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \tau) = \Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2)\tilde{\Gamma}(\tau) \tag{5.29}$$

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

$$\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) = \Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2)\,\mathscr{F}\{\tilde{\Gamma}(\tau)\} \tag{5.30}$$

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

5.2 The Van Cittert–Zernike Theorem

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

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

$$u(\vec{\mathbf{r}}, t) = \int_{V(S)} \mathrm{d}^3\vec{\mathbf{s}}\; {}^{\mathbf{s}}u(\vec{\mathbf{r}}, t) \tag{5.31}$$

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

Theorem 5.2 (Van Cittert–Zernike Theorem). *Assume a light source as described above. Given a pair of points $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$ the (equal-time) mutual coherence is as follows*

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

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

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

$${}^{\mathbf{s}}u(\vec{\mathbf{r}}, t) = {}^{\mathbf{s}}u\left(\vec{\mathbf{s}}, t - \frac{|\vec{\mathbf{r}} - \vec{\mathbf{s}}|}{c}\right)\frac{1}{|\vec{\mathbf{r}} - \vec{\mathbf{s}}|}e^{-i\omega\frac{|\vec{\mathbf{r}}-\vec{\mathbf{s}}|}{c}} \tag{5.32}$$

We assume simple spherical waves and hence the amplitude decays as the inverse of the distance from the source. We also make the natural assumption that the radiators are zero-meaned, i.e. $\left\langle {}^{\mathbf{s}}u(\vec{\mathbf{r}}, t)\right\rangle_\mathrm{t} = 0$. Then, an almost incoherent ensemble that corresponds to a propagating wave can be written as [Goodman 2015]

$$\begin{aligned}\left\langle {}^{\mathbf{s}}u(\vec{\mathbf{s}}, t)\,{}^{\mathbf{q}}u(\vec{\mathbf{q}}, t+\tau)^\star\right\rangle_\mathrm{t} &= \begin{cases}\left\langle {}^{\mathbf{s}}u(\vec{\mathbf{s}}, t)\,{}^{\mathbf{q}}u(\vec{\mathbf{q}}, t+\tau)^\star\right\rangle_\mathrm{t} & , \mathbf{s} = \mathbf{q}\\ \left\langle {}^{\mathbf{s}}u(\vec{\mathbf{s}}, t)\right\rangle_\mathrm{t}\left\langle {}^{\mathbf{q}}u(\vec{\mathbf{q}}, t+\tau)\right\rangle_\mathrm{t}^\star & , \mathbf{s} \neq \mathbf{q}\end{cases}\\ &= \delta^3(\vec{\mathbf{s}} - \vec{\mathbf{q}})\left\langle {}^{\mathbf{s}}u(\vec{\mathbf{s}}, t)\,{}^{\mathbf{s}}u(\vec{\mathbf{s}}, t+\tau)^\star\right\rangle_\mathrm{t}\end{aligned} \tag{5.33}$$

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

$$\begin{aligned}\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) &= \left\langle u(\vec{\mathbf{r}}_1, t)u(\vec{\mathbf{r}}_2, t)^\star\right\rangle_\mathrm{t}\\ &= \int_{V(S)} \mathrm{d}^3\mathbf{s}\int_{V(S)} \mathrm{d}^3\mathbf{q}\left\langle {}^{\mathbf{s}}u(\vec{\mathbf{r}}_1, t)\,{}^{\mathbf{q}}u(\vec{\mathbf{r}}_2, t)^\star\right\rangle_\mathrm{t}\\ &= \int_{V(S)} \mathrm{d}^3\mathbf{s}\;\frac{e^{-i\omega\frac{r_\mathbf{s}-r'_\mathbf{s}}{c}}}{r_\mathbf{s} r'_\mathbf{s}}\left\langle {}^{\mathbf{s}}u\left(\vec{\mathbf{s}}, t - \frac{r_\mathbf{s}}{c}\right)\,{}^{\mathbf{s}}u\left(\vec{\mathbf{s}}, t - \frac{r'_\mathbf{s}}{c}\right)^\star\right\rangle_\mathrm{t}\end{aligned} \tag{5.34}$$

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

$$\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) = \int_{V(S)} \mathrm{d}^3\mathbf{s}\;\frac{e^{-i\omega\frac{r_\mathbf{s}-r'_\mathbf{s}}{c}}}{r_\mathbf{s} r'_\mathbf{s}} I(\vec{\mathbf{s}}) \tag{5.35}$$

as desired. □

Let R_1, R_2 be the distances from the source to $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$, respectively. If we are far from the source, where $R_1 \approx r_\mathbf{s}$ and $R_2 \approx r'_\mathbf{s}$ are a reasonable approximation for any $\vec{\mathbf{s}} \in V(S)$, then the Van Cittert–Zernike theorem simplifies to:

$$\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) \approx \frac{1}{R_1 R_2}\int_{V(S)} \mathrm{d}^3\mathbf{s}\, I(\vec{\mathbf{s}})e^{-ik(r_\mathbf{s} - r'_\mathbf{s})} \tag{5.36}$$

where $\omega = kc$ as usual. Assume that we image the light arriving from the source, and would like to study its coherence properties. The imaging effectively transforms the source into a two-dimensional plane, which without loss of generality we set to coincide with the xy-plane. Our points of interest are now placed on the "image" plane, i.e. the imaging sensor, which we assume to be co-planar to the source and located at $z = R \approx R_{1,2}$. See Fig. 3 for a depiction of the considered geometry.

Theorem 5.3 (Far-Field Van Cittert–Zernike Theorem). *Given the planar geometry described above and $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$ on the "image" plane, we denote $I(\xi, \zeta)$ as the radiance that leaves the imaged source at planar coordinates (ξ, ζ) and is observed by the sensor. The implied convention is that $I(\xi, \zeta) \equiv 0$ outside the light source. Then,*

$$\Gamma(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2) = e^{-ik\vartheta}\,\mathscr{F}\{I\}\left(k\Delta_x, k\Delta_y\right)$$

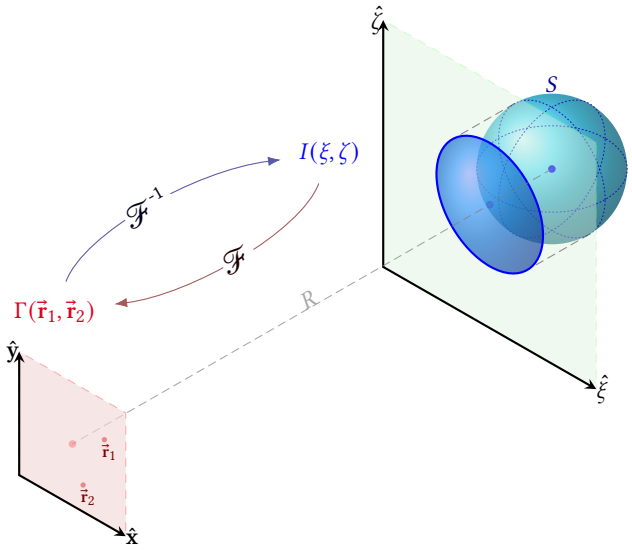

Fig. 3. A light source S is imaged by a far-away observer. The light source gives rise to an intensity distribution on the source plane (right), denoted $I(\xi,\zeta)$, and we are interested in the mutual coherence, Γ, of the light that is observed on the image plane (left). The Far-Field Van Cittert–Zernike Theorem relates the two quantities as Fourier transform pairs.

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

Proof. Start with Eq. (5.36). Given (ξ,ζ), coordinates on the source plane, we adopt the following approximation:

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

where we power expanded the square root and neglected the $O\left(\frac{1}{R^3}\right)$ terms. A similar expression for $r'_{\mathbf{s}}$ follows. Define $\Delta_x = \frac{x_2-x_1}{R}$ and $\Delta_y = \frac{y_2-y_1}{R}$ then the distance difference becomes:

$$\begin{aligned} r_{\mathbf{s}} - r'_{\mathbf{s}} &= \frac{(\xi-x_1)^2 + (\zeta-y_1)^2}{2R} - \frac{(\xi-x_2)^2 + (\zeta-y_2)^2}{2R} \\ &= \vartheta + \xi\Delta_x + \zeta\Delta_y \end{aligned} \tag{5.37}$$

with

$$\vartheta = \frac{x_1^2 + y_1^2 - x_2^2 - y_2^2}{2R} = \frac{1}{2R}\left(|\vec{\mathbf{r}}_1|^2 - |\vec{\mathbf{r}}_2|^2\right) \tag{5.38}$$

And we conclude that the mutual coherence is:

$$\begin{aligned} \Gamma(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2) &\approx e^{-ik\vartheta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}\xi\,\mathrm{d}\zeta\, I(\xi,\zeta) e^{-ik(\xi\Delta_x+\zeta\Delta_y)} \\ &= e^{-ik\vartheta}\,\mathscr{F}\{I\}(k\Delta_x, k\Delta_y) \end{aligned} \tag{5.39}$$

The R^{-2} factor from Eq. (5.36) was eliminated due to $I(\xi,\zeta)$ having units of radiance (flux per solid angle per area). □

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

Before we conclude, it is worthwhile to clarify some terminology: We sometimes refer to a light source as "incoherent" or "coherent", however strictly speaking only wave ensembles at particular regions in time and space can be described as coherent or incoherent. An "incoherent" light source, which are essentially all natural sources, are incoherent in the sense that they are composed of uncorrelated radiators. Nonetheless, radiation from such sources can still be coherent. The spatial region over which a wave ensemble remains somewhat coherent can be quantified, and we will do so next.

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

We define the *spatial coherence area* as

$$A_{\mathrm{C}}(\vec{\mathbf{r}}) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}x\,\mathrm{d}y\,|\gamma(x,y)|^2 \tag{5.40}$$

The degree of coherence $\gamma(x,y)$ is the intensity normalized version of the planar form of the mutual coherence, similar to the far-field Van Cittert–Zernike theorem (Theorem 5.3).

Lemma 5.4. *Given a monochromatic disk-shaped planar source with radius ζ and with a uniform radiant flux per unit area (i.e. constant radiosity). The spatial coherence area A_{C} satisfies the following proportionality relation:*

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

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

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

$$\begin{aligned} A_{\mathrm{C}}(\vec{\mathbf{r}}) &= \iint \mathrm{d}x\,\mathrm{d}y \left|\frac{\Gamma(x,y)}{I(\vec{\mathbf{r}})}\right|^2 \\ &\approx \iint \mathrm{d}x\,\mathrm{d}y \left|\frac{I_0 R^{-2}\iint_{\mathrm{disk}} \mathrm{d}\xi\,\mathrm{d}\zeta\, e^{-ik\frac{\xi x+\zeta y}{R}}}{\pi\zeta^2 I_0 R^{-2}}\right|^2 \\ &= \frac{1}{\pi^2\zeta^4}\iint \mathrm{d}x\,\mathrm{d}y \left|\frac{2\pi\zeta \mathrm{J}_1\left(k\zeta R^{-1}\sqrt{x^2+y^2}\right)}{kR^{-1}\sqrt{x^2+y^2}}\right|^2 \\ &= \frac{8\pi R^2}{k^2\zeta^2}\int_0^{\infty} \mathrm{d}\rho\,\rho \left|\frac{\mathrm{J}_1\left(k\zeta R^{-1}\rho\right)}{\rho}\right|^2 \end{aligned} \tag{5.41}$$

where we used the circular aperture Fourier transform identity (2.21), and transformed from cartesian coordinates to polar coordinates via $\mathrm{d}x\,\mathrm{d}y = \mathrm{d}\theta\,\rho\,\mathrm{d}\rho$. The Bessel function of the first kind, $\mathrm{J}_1(a)$ admits its first zero at $a_0 \approx 3.8317$ [Abramowitz and Stegun 1965]. It makes

practical sense to limit the integration till a_0, and thereby we get the expression

$$A_C(\vec{\mathbf{r}}) \approx \frac{8\pi R^2}{k^2\zeta^2}\int_0^{a_0}\frac{\mathrm{d}\rho'}{\rho'}\left|\mathrm{J}_1(\rho')\right|^2$$
$$= \frac{4\pi R^2}{k^2\zeta^2}\left(1-\mathrm{J}_0(a_0)^2-\mathrm{J}_1(a_0)^2\right) \approx \frac{4R^2\lambda^2}{5\pi\zeta^2} \propto \frac{\lambda^2}{\Omega_S} \tag{5.42}$$

where we used the following Bessel integral identity [Born and Wolf 1999]: $\int_0^a \rho^{-1}\mathrm{J}_1(\rho)^2 = \frac{1}{2}\left(1-\mathrm{J}_0(a)^2-\mathrm{J}_1(a)^2\right)$, and the approximation $1-\mathrm{J}_0(a_0)^2 \approx \frac{4}{5}$. Also, $k = 2\pi\lambda^{-1}$, as usual. □

Lemma 5.4 presents an order-of-magnitude relation: The spatial coherence area induced by a perfectly circular light source is proportional to the inverse solid angle subtended by the source. It is also implied that longer wavelength light will generally remain coherent over greater spatial regions compared to light with shorter wavelengths. This approximative relation is conserved, in practice, for sources of other shapes.

It is of interest to briefly examine the behaviour of the degree of coherence, $|\gamma|$, when $\rho > a_0$ in the integral in Eq. (5.41). For $a \gg 1$ the following asymptotic expansion of the Bessel function of the first kind is known [Abramowitz and Stegun 1965]:

$$\mathrm{J}_1(a) \approx \sqrt{\frac{2}{\pi a}}\left[\cos\left(a-\frac{3\pi}{4}\right)+\mathcal{O}\left(|\rho|^{-1}\right)\right] \tag{5.43}$$

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

5.3 Propagation of Coherence Functions

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

$$\nabla^2 u(\vec{\mathbf{r}},t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(\vec{\mathbf{r}},t) \tag{5.44}$$

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

$$\nabla^2_{\vec{\mathbf{r}}}\left\langle u(\vec{\mathbf{r}},t)u(\vec{\mathbf{r}}',t')^\star\right\rangle_{\mathrm{t}} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left\langle u(\vec{\mathbf{r}},t)u(\vec{\mathbf{r}}',t')^\star\right\rangle_{\mathrm{t}} \tag{5.45}$$

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

$$\nabla^2_{\vec{\mathbf{r}}}\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}}',t'-t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}}',t'-t) \tag{5.46}$$

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

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

$$\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}}',\tau) = \mathscr{F}^{-1}\{\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}',\omega)\} \tag{5.47}$$

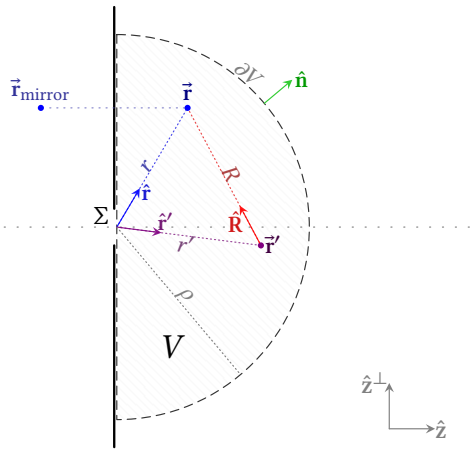

Fig. 4. Propagation of cross-spectral density.

by applying the wave operator (Eq. (4.6)) to each side of the equation above, and formally interchanging the order of integration and differentiation, we get

$$\triangle_{\mathrm{m}}\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}}',\tau) = \triangle_{\mathrm{m}}\mathscr{F}^{-1}\{\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}',\omega)\}$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\triangle_{\mathrm{m}}\left[\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}',\omega)e^{-i\omega\tau}\right]$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\left[\nabla^2+\frac{\omega^2}{c^2}\right]\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}',\omega)e^{-i\omega\tau} \tag{5.48}$$

which can only vanish for all time differences, τ, if the integrand is identically 0. The same conclusion can also be trivially reached by considering the spectral decomposition of u into elementary monochromatic wavelets, $\{v(\vec{\mathbf{r}},\omega)e^{-i\omega t}\}$, as discussed in Subsection 5.1.1. Applying the wave operator to each wavelet in the ensemble and taking the space-frequency ensemble average immediately shows that indeed the cross-spectral density obeys the Helmholtz equation.

Hence, we can conclude:

$$\triangle_{\mathrm{m}}\Gamma(\vec{\mathbf{r}},\vec{\mathbf{r}}',\tau) = 0 \tag{5.49}$$

$$\left[\nabla^2+\frac{\omega^2}{c^2}\right]\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}',\omega) = 0 \tag{5.50}$$

that is, the propagation of the mutual coherence is dictated by the wave operator. And, as expected, the cross-spectral density—being the Fourier transform pair of the mutual coherence—propagates in accordance to the Helmholtz equation, which is generally easier to solve.

5.3.1 Propagation as a diffraction problem. As the mutual coherence and the cross-spectral density functions propagate as waves, we can formulate the propagation of those coherence functions as a diffraction problem. We focus on the cross-spectral density. The cross-spectral density propagates according to the Helmholtz equation, which calls for a Green function treatment. Recall the free-space Green function G_0 (Eq. (2.29)):

$$\left[\nabla^2+k^2\right]G_0(\vec{\mathbf{r}}-\vec{\mathbf{r}}') = -\delta^3(\vec{\mathbf{r}}-\vec{\mathbf{r}}') \tag{5.51}$$

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

$$\left[\nabla_1'^2 + k^2\right]\left[\nabla_2'^2 + k^2\right] G_0\left(\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_1'\right)^\star G_0\left(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_2'\right) = \delta^3\left(\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_1'\right)\delta^3\left(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_2'\right) \tag{5.52}$$

We denote $\nabla^2_{1,2}$, $\nabla'^2_{1,2}$ as the laplacians with respect to $\vec{\mathbf{r}}_{1,2}$ and $\vec{\mathbf{r}}'_{1,2}$, respectively. Multiply both sides by $\mathcal{W}$ and integrate twice over the space $V = \{z > 0\} \cap B_\rho$, where B_ρ is a ball of some radius ρ centred at the origin (see Fig. 4 for a visualization of the geometry):

$$\mathcal{W}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega) = \int_V \mathrm{d}^3\vec{\mathbf{r}}_1' \left[\nabla_1'^2 + k^2\right] G_0\left(\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_1'\right)^\star \times \int_V \mathrm{d}^3\vec{\mathbf{r}}_2' \left[\nabla_2'^2 + k^2\right] G_0\left(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_2'\right) \mathcal{W}\left(\vec{\mathbf{r}}_1', \vec{\mathbf{r}}_2', \omega\right) \tag{5.53}$$

Consider the second integral on the right-hand side: By noting that $k^2\mathcal{W} = -\nabla_2'^2\mathcal{W}$ (directly from Eq. (5.50), the Helmholtz equation) and applying Green's second identity we get

$$\int_V \mathrm{d}^3\vec{\mathbf{r}}_2' \left[\nabla_2'^2 + k^2\right] G_0\,\mathcal{W} = \int_V \mathrm{d}^3\vec{\mathbf{r}}_2' \left[\mathcal{W}\,\nabla_2'^2 G_0 - G_0 \nabla_2'^2 \mathcal{W}\right] = \oint_{\partial V} \mathrm{d}^2\vec{\mathbf{r}}_2' \,\hat{\mathbf{n}} \cdot \left[\mathcal{W}\,\boldsymbol{\nabla}_2' G_0 - G_0 \boldsymbol{\nabla}_2' \mathcal{W}\right] \tag{5.54}$$

where $\boldsymbol{\nabla}_2'$ is the gradient with respect to $\vec{\mathbf{r}}_2'$ and ∂V is th boundary of V with $\hat{\mathbf{n}}$ being its outward normal at a point. Denote the shorthand $\vec{\mathbf{R}}_2 = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_2'$. The gradient of the free-space Green function follows immediately from rewriting Eq. (2.29) in spherical coordinates:

$$\boldsymbol{\nabla}' G_0\left(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_2'\right) = \hat{\mathbf{R}}_2 \frac{\partial}{\partial R_2} G_0(R_2) = \hat{\mathbf{R}}_2 \frac{ikR_2 - 1}{4\pi R_2^2} e^{ikR_2} \tag{5.55}$$

where $R_2 = |\vec{\mathbf{R}}_2|$ and $\hat{\mathbf{R}}_2 = \vec{\mathbf{R}}_2 / R_2$ is the radius basis vector in spherical coordinates (see Fig. 4). The Sommerfeld radiation condition (Eq. (2.28)) applies to outgoing spherical wavelets, and therefore it applies to the cross-spectral density as well. Therefore $\mathcal{W} \to 0$ as $\vec{\mathbf{r}}_2' \to \infty$ at least faster than $1/R_2$, and thus the integrand on the right-hand side in Eq. (5.54) decays to 0 faster than R_2^{-2}. We conclude that the integrand vanishes everywhere at the limit $\vec{\mathbf{r}}_2' \to \infty$. Hence, by taking the limit $\rho \to \infty$ (recall ρ is the radius of the half-sphere ∂V) and substituting the expressions for G_0 and its gradient into Eq. (5.54), we get

$$\lim_{\rho\to\infty} \int_V \mathrm{d}^3\vec{\mathbf{r}}_2' \left[\nabla_2'^2 + k^2\right] G_0\,\mathcal{W} = -\int_{\hat{\mathbf{z}}^\perp} \mathrm{d}^2\vec{\mathbf{r}}_2' \frac{e^{ikR_2}}{4\pi R_2} \hat{\mathbf{z}} \cdot \left[\hat{\mathbf{R}}_2 \frac{ikR_2 - 1}{R_2}\mathcal{W} - \boldsymbol{\nabla}_2'\mathcal{W}\right] \tag{5.56}$$

where $\hat{\mathbf{z}}^\perp$ is the orthogonal complement of the z-axis, i.e. the xy-plane. The integral above is a superposition of spherical wavelets that propagate from the plane $\hat{\mathbf{z}}^\perp$. Indeed, it is a mathematical formulation of Huygens' principle and the integral itself is nothing more than the Kirchhoff diffraction integral. Proceeding with the same approach we also apply Kirchhoff's approximation and assume that the electric field (and in turn, the cross-spectral density) vanishes everywhere on $\hat{\mathbf{z}}^\perp$, except an area $\Sigma \subset \hat{\mathbf{z}}^\perp$ which can be considered either as an aperture or a light source. Kirchhoff's boundary conditions produce an overdetermined problem by also assuming that the derivative of $\mathcal{W}$ vanishes on $\hat{\mathbf{z}}^\perp$. This is not necessary, and we follow Zangwill [2013] by replacing the free-space Green function with a mirror-image variant:

$$\tilde{G}_0\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}'\right) \triangleq G_0\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'\right) - G_0\left(\vec{\mathbf{r}}_{\text{mirror}} - \vec{\mathbf{r}}'\right) \tag{5.57}$$

where in the second Green function $\vec{\mathbf{r}}_{\text{mirror}} = \hat{\mathbf{x}} r_x + \hat{\mathbf{y}} r_y - \hat{\mathbf{z}} r_z$, i.e. $\vec{\mathbf{r}}$ mirrored with respect to the $\hat{\mathbf{z}}^\perp$ plane. Clearly, $\tilde{G}_0$ vanishes on $\vec{\mathbf{r}} \in \hat{\mathbf{z}}^\perp$, which dispenses with the $\boldsymbol{\nabla}_2'\mathcal{W}$ term—the Neumann boundary condition—in Eqs. (5.54) and (5.56). When acted upon by the Helmholtz operator $\tilde{G}_0$ produces two Dirac deltas, however one of them lies outside of V and does not contribute to Eq. (5.53). Hence, by replacing G_0 with $\tilde{G}_0$ in Eq. (5.54) we are left with a mathematically well-defined problem and Eq. (5.56) reduces to

$$\lim_{\rho\to\infty} \int_V \mathrm{d}^3\vec{\mathbf{r}}_2' \left[\nabla_2'^2 + k^2\right] \tilde{G}_0\,\mathcal{W} = -\int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}_2' \,\mathcal{W} \frac{\partial}{\partial z}\left[\frac{e^{ikR_2}}{2\pi R_2}\right] \tag{5.58}$$

which is the Rayleigh-Sommerfeld diffraction integral of the first kind [Born and Wolf 1999]. The integral above is a solution of the Helmholtz equation in the $z > 0$ half-space expressed in terms of the boundary values on Σ.

In an entirely analogous manner the same process is readily applied to the first integral (over $\vec{\mathbf{r}}_1' \in V$) in Eq. (5.53). Note that $\frac{\partial}{\partial z} = \frac{\partial}{\partial R_j}\frac{\partial}{\partial z} R_j = \frac{z}{R_j}\frac{\partial}{\partial R_j}$ and $\frac{\partial}{\partial R_j}\frac{e^{ikR_j}}{R_j} = \frac{e^{ikR_j}}{R_j}[ik - 1/R_j]$, for $j \in \{1, 2\}$. At which point we finally end up with the following expression

$$\mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) = \frac{1}{4\pi^2} \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}_1' \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}_2' \,\mathcal{W}\left(\vec{\mathbf{r}}_1', \vec{\mathbf{r}}_2', \omega\right) \times \frac{z_1 z_2}{R_1^2 R_2^2} e^{ik(R_1 - R_2)}\left(-ik - \frac{1}{R_1}\right)\left(ik - \frac{1}{R_2}\right) \tag{5.59}$$

where z_1 and z_2 are the z-components of $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$, respectively. Denote $\hat{\mathbf{r}}_j$ as the unit vector in direction of $\vec{\mathbf{r}}_j$ and r_j the distance to $\vec{\mathbf{r}}_j$, and similarly with primed variables (see Fig. 4). Then, in the Fresnel-Fraunhofer diffraction region—far from the source with respect to the characteristic size of the source—we can approximate $R_1 \approx r_1$ and $R_2 \approx r_2$ (for $\vec{\mathbf{r}}_j' \in \Sigma$). Furthermore, for visible light clearly $kr_{1,2} \gg 1$ in that region, and thus the fractions $1/R_j$ are negligible. In the high-frequency exponent a better approximation is required and we power expand $R_{1,2}$ (in similar manner to the proof of Theorem 5.3):

$$R_j = \left|\vec{\mathbf{r}}_j - \vec{\mathbf{r}}_j'\right| = r_j + \frac{r_j'^2 - 2\vec{\mathbf{r}}_j \cdot \vec{\mathbf{r}}_j'}{2r_j} + O\left(\frac{1}{r_j^2}\right) \approx r_j - \hat{\mathbf{r}}_j \cdot \vec{\mathbf{r}}_j' \tag{5.60}$$

Applying all of the above, Eq. (5.59) simplifies to

$$\mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) \approx \frac{\cos\theta_1 \cos\theta_2}{\lambda^2 r_1 r_2} e^{ik(r_1 - r_2)} \times \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}_1' \, e^{-ik\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}_1'} \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}_2' \,\mathcal{W}\left(\vec{\mathbf{r}}_1', \vec{\mathbf{r}}_2', \omega\right) e^{ik\hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}_2'} \tag{5.61}$$

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

6 POLARIZATION

The electric and magnetic fields supporting propagating electromagnetic waves generally admit transverse components. This is suggested by the Poynting theorem and the Poynting vector (Eq. (4.15)), where we can see that positive energy flow must be perpendicular to the (real) fields. The wave's *polarization* describes the geometry of the transverse oscillations of the supporting fields, and note that the polarization of the electric field is generally not the same as the polarization of the magnetic field (for example, the beam-like wave packet presented previously).

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

$$\boldsymbol{J} \triangleq \frac{c}{8\pi} \begin{bmatrix} \left\langle E_1 E_1^\star \right\rangle_\mathrm{t} & \left\langle E_1 E_2^\star \right\rangle_\mathrm{t} \\ \left\langle E_2 E_1^\star \right\rangle_\mathrm{t} & \left\langle E_2 E_2^\star \right\rangle_\mathrm{t} \end{bmatrix} \tag{6.1}$$

The polarization matrix was introduced by Born and Wolf in older literature as the "coherency matrix", before the advent of optical coherence theory, however the term is a misnomer: The polarization matrix elements are the (equal-time) cross-correlations of the transverse fields at a single spatial position, and the matrix describes the polarization characteristics of the wave packet at that position. Coherence studies the cross-correlation between points in multiple space and time points. Denote the matrix elements J_{xy}, with $x, y \in \{1, 2\}$, and we first briefly consider the mathematical properties of the polarization matrix. The matrix is clearly Hermitian, i.e. $\boldsymbol{J}^\dagger = \boldsymbol{J}$. By applying Schwarz's inequality (Eq. (1.19)) we deduce that

$$|J_{12}| = |J_{21}| \leq \sqrt{|J_{11}|}\sqrt{|J_{22}|} \tag{6.2}$$

and it follows that $\boldsymbol{J}$ is also positive semi-definite, i.e. $\boldsymbol{J} \succeq 0$, and thus $\det \boldsymbol{J} \geq 0$ and all the eigenvalues are real and non-negative. The diagonal elements of the equal-time polarization matrix are always real, and are the time-averaged energy density carried by the electric field. For the idealized case of a plane-wave, that is exactly the observed intensity (Eq. (4.18)) of each orthogonal component of the wave, and thus the intensity of the wave is the sum of the diagonal elements, i.e.

$$I = \operatorname{tr} \boldsymbol{J} \tag{6.3}$$

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

$$\begin{bmatrix} E_1' \\ E_2' \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} E_1 \cos\theta - E_2 \sin\theta \\ E_1 \sin\theta + E_2 \cos\theta \end{bmatrix} \tag{6.4}$$

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

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

$$j_{12} \triangleq \frac{1}{\sqrt{J_{11}}\sqrt{J_{22}}} J_{12} \tag{6.5}$$

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

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

$$\boldsymbol{J}_\mathrm{u} = \frac{1}{2} \begin{bmatrix} I_\mathrm{u} & \\ & I_\mathrm{u} \end{bmatrix} \tag{6.6}$$

where I_u is the intensity of the wave (for a plane wave).

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

$$\det \boldsymbol{J} = J_{11}J_{22} - J_{12}(J_{12})^\star = J_{11}J_{22} - \left|\sqrt{J_{11}}\sqrt{J_{22}} j_{12}\right|^2 = 0 \tag{6.7}$$

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

$$\boldsymbol{J}_\mathrm{p} = \frac{c}{8\pi} \begin{bmatrix} J_{11} & \sqrt{J_{11}}\sqrt{J_{22}} e^{i\varphi} \\ \sqrt{J_{11}}\sqrt{J_{22}} e^{-i\varphi} & J_{22} \end{bmatrix} \tag{6.8}$$

where φ is any real number.

Corollary 6.1. *A completely polarized polarization matrix has eigenvalues* 0 *and* $J_{11} + J_{22}$*, and its Cholesky decomposition is* $\boldsymbol{L}\boldsymbol{L}^\dagger$ *with*

$$\boldsymbol{L} = \begin{bmatrix} \sqrt{J_{11}} & 0 \\ \sqrt{J_{22}} e^{-i\varphi} & 0 \end{bmatrix} \tag{6.9}$$

Observe from Eq. (6.8) that the phase difference between E_1 and E_2, φ, must be independent of time, as otherwise E_1 and E_2 are not fully correlated. When $\varphi = m\pi$, for some integer m, then the transverse oscillations are fully in-phase or delayed by π (which is equivalent to fully in-phase with a negated E_2), and we deduce that such a wave is completely *linearly polarized*. A necessary and sufficient condition for completely linearly polarized light is then $\det \boldsymbol{J} = 0$ and $\operatorname{Im} j_{12} = 0$. When $\operatorname{Im} j_{12} \neq 0$ the wave will generally be a (correlated) superposition of linearly polarized light and non-linearly polarized light. Under the context of a monochromatic plane-wave, a typical treatment of the subject of optical polarization further classifies the states of polarization into circular polarization

($\varphi = \pm i$), and elliptical polarization (the general complete polarization state, with linear and circular being special cases). However, we deal predominantly with polychromatic light, and circular and elliptical polarizations do not generally arise in polychromatic wave packets, because the different wavelengths admit distinct spatial frequencies. Instead, we focus on analysing what fraction of the energy carried by a wave packet is polarised, and in what transverse direction most of the polarized energy is concentrated.

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

Lemma 6.2. *For any polarization matrix, the angle between the major axis and $\hat{\mathbf{e}}_1$ is ψ, where*

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

Proof. See Born and Wolf [1999]. □

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

As a side note, the geometric shapes traced by transverse field lines of polychromatic light are called "Lissajous" figures. Such a figure is well-defined and closes upon itself with a regular period when the frequencies composing the wave packet are commensurable, that is $\omega_{j+1}/\omega_j \in \mathbb{Q}$ for all (finite) frequencies ω_j. Clearly a highly artificial scenario. Those intricacies arise even before we consider the random fluctuations that accompany light from any natural source, in which case the traced Lissajous figures become highly irregular and generally are of little physical insight or observable consequence. For more information on geometric characteristic of "Lissajous" figures, see Freund [2003].

6.0.1 Decomposition of light of any state of polarization. Consider again a polychromatic wave packet propagating in direction $\hat{\mathbf{z}}$. Its polarization matrix $\boldsymbol{J}$, under the most general conditions, is as presented in Eq. (6.1). We will now show that it is possible to decompose the wave packet into a superposition of unpolarized light and completely polarized light.

Lemma 6.3 (Canonical Decomposition of the Polarization Matrix). *Any polarization matrix $\boldsymbol{J}$ can always be decomposed into*

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

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

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

Furthermore, this decomposition is unique.

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

$$\boldsymbol{J} = \frac{I_\mathrm{u}}{2}\boldsymbol{I} + \boldsymbol{J}_\mathrm{p} \tag{6.10}$$

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

$$I_\mathrm{u} = J_{11} + J_{22} \pm \sqrt{(J_{11} - J_{22})^2 + 4J_{11}J_{22}|j_{12}|^2} \tag{6.11}$$

The completely polarized fraction of the energy would then be

$$\boldsymbol{J}_\mathrm{p} = \boldsymbol{J} - \boldsymbol{J}_\mathrm{u} = \begin{bmatrix} J'_{11} & J_{12} \\ J^{\star}_{12} & J'_{22} \end{bmatrix} \tag{6.12}$$

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

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

Corollary 6.4. *Let the following be a decomposition of a completely polarized matrix $\boldsymbol{J}_\mathrm{p}$ into a couple of polarization matrices:*

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

We state, without proving, that the only possible such decomposition is a trivial decomposition, i.e. $\boldsymbol{J}'_\mathrm{p} = \alpha\boldsymbol{J}_\mathrm{p}$ and $\boldsymbol{J}''_\mathrm{p} = (1-\alpha)\boldsymbol{J}_\mathrm{p}$ for some $0 \le \alpha \le 1$. It is easy to check that a completely polarized matrix admits no decomposition with a partially polarized or unpolarized matrix. Of more consequence is the conclusion that the phase-shifts, i.e. φ in Eq. (6.8), of each of the decomposing matrices is identical to the phase-shift of $\boldsymbol{J}$.

That is, a completely polarized wave ensemble is fully correlated, and therefore can not be decomposed into different states of polarization. For example, a completely polarized matrix can not be decomposed into a completely linearly polarized part and the non-linearly polarized part, even under a high degree of linear polarization.

To conclude, we define the important *degree of polarization* as the ratio between the polarized fraction of the energy and the total energy carried by the wave:

$$\mathcal{P} \triangleq \frac{\mathrm{tr}\,\boldsymbol{J}_\mathrm{p}}{\mathrm{tr}\,\boldsymbol{J}} = \frac{I_\mathrm{p}}{I} = \sqrt{1 - \frac{4\det \boldsymbol{J}}{(\mathrm{tr}\,\boldsymbol{J})^2}} \tag{6.13}$$

where the identity $I_\mathrm{p} = (\mathrm{tr}\,\boldsymbol{J})^2 - 4\det \boldsymbol{J}$ follows directly from Eq. (6.11). When $\mathcal{P} = 1$ the light is completely polarized, while

$\mathcal{P} = 0$ indicates unpolarized light. When $0 < \mathcal{P} < 1$ we say the light is *partially polarized*.

6.1 Unified Theory of Polarization and Coherence

We now make contact with optical coherence theory presented in Section 5. The discussion in this section outlines a unified theory of optical coherence that takes polarization into account and presents a few interesting results. Our discussion is based on Wolf [2007].

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

$$\begin{aligned}\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) &\triangleq \frac{c}{8\pi}\begin{bmatrix}\Gamma_{11}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) & \Gamma_{12}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau)\\ \Gamma_{21}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) & \Gamma_{22}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau)\end{bmatrix}\\ &= \frac{c}{8\pi}\begin{bmatrix}\langle E_1(\vec{\mathbf{r}}_1,t)E_1(\vec{\mathbf{r}}_2,t+\tau)^\star\rangle_t & \langle E_1(\vec{\mathbf{r}}_1,t)E_2(\vec{\mathbf{r}}_2,t+\tau)^\star\rangle_t\\ \langle E_2(\vec{\mathbf{r}}_1,t)E_1(\vec{\mathbf{r}}_2,t+\tau)^\star\rangle_t & \langle E_2(\vec{\mathbf{r}}_1,t)E_2(\vec{\mathbf{r}}_2,t+\tau)^\star\rangle_t\end{bmatrix}\end{aligned} \tag{6.14}$$

where Γ_{xy} is the mutual coherence function relating the orthogonal field components. The above is the space-time formulation of optical coherence, and the space-frequency readily follows with the cross-spectral density matrix:

$$\begin{aligned}\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) &\triangleq \frac{c}{8\pi}\begin{bmatrix}\mathcal{W}_{11}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) & \mathcal{W}_{12}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)\\ \mathcal{W}_{21}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) & \mathcal{W}_{22}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)\end{bmatrix}\\ &= \frac{c}{8\pi}\begin{bmatrix}\langle E_1(\vec{\mathbf{r}}_1)E_1(\vec{\mathbf{r}}_2)^\star\rangle_\omega & \langle E_1(\vec{\mathbf{r}}_1)E_2(\vec{\mathbf{r}}_2)^\star\rangle_\omega\\ \langle E_2(\vec{\mathbf{r}}_1)E_1(\vec{\mathbf{r}}_2)^\star\rangle_\omega & \langle E_2(\vec{\mathbf{r}}_1)E_2(\vec{\mathbf{r}}_2)^\star\rangle_\omega\end{bmatrix}\end{aligned} \tag{6.15}$$

where the ensemble averages are over same-frequency wavelets (see Eq. (5.23)). The relation between the matrix-versions of the mutual coherence and cross-spectral density is preserved, with them being Fourier transform pairs, viz:

$$\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) = \mathscr{F}\{\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau)\} \tag{6.16}$$

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

The mutual coherence and cross-spectral density matrices are the akin to their scalar counterparts discussed in Section 5. Under matrix formulation, the energy carried by a wave becomes $I(\vec{\mathbf{r}}) = \operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}},\vec{\mathbf{r}})$, where we again abuse notation and drop the time-delay for equal-time mutual coherence, and similarly the spectral density becomes $\mathcal{I}(\vec{\mathbf{r}},\omega) \triangleq \operatorname{tr}\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}},\vec{\mathbf{r}},\omega)$. Then, in an analogous way, we define the matrix degree of coherence (matrix analogue of Eq. (5.8)) and the matrix spectral degree of coherence (matrix analogue of Eq. (5.15)) as

$$\boldsymbol{\gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) \triangleq \frac{\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau)}{\sqrt{\operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_1)}\sqrt{\operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_2,\vec{\mathbf{r}}_2)}} \tag{6.17}$$

$$\boldsymbol{\mathcal{w}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) \triangleq \frac{\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)}{\sqrt{\operatorname{tr}\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_1,\omega)}\sqrt{\operatorname{tr}\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_2,\vec{\mathbf{r}}_2,\omega)}} \tag{6.18}$$

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

The interference laws also mostly retain their scalar form. Let $\vec{\mathbf{E}}(\vec{\mathbf{r}},t)$ be the electric field of a wave ensemble. We are interested in the observed intensity of a superposition between the energy arriving from different points, $\vec{\mathbf{E}}'(\vec{\mathbf{r}},t) = \vec{\mathbf{E}}(\vec{\mathbf{r}}_1,t_1) + \vec{\mathbf{E}}(\vec{\mathbf{r}}_2,t_2)$. The expression for the intensity is the matrix form of the interference law for stationary fields (analogue of Eq. (5.9)):

$$\begin{aligned}I'(\vec{\mathbf{r}},t) =& \operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_1) + \operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_2,\vec{\mathbf{r}}_2)\\ &+ 2\sqrt{\operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_1)}\sqrt{\operatorname{tr}\mathbf{\Gamma}(\vec{\mathbf{r}}_2,\vec{\mathbf{r}}_2)}\operatorname{Re}\operatorname{tr}\boldsymbol{\gamma}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,t_2-t_1)\end{aligned} \tag{6.19}$$

In the spectral domain, the matrix form of the spectral interference law (Eq. (5.26)) takes on the familiar form:

$$\begin{aligned}\mathcal{I}(\vec{\mathbf{r}},\omega) =& \operatorname{tr}\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}},\vec{\mathbf{r}},\omega) = \mathcal{I}(\vec{\mathbf{r}}_1,\omega) + \mathcal{I}(\vec{\mathbf{r}}_2,\omega)\\ &+ 2\sqrt{\mathcal{I}(\vec{\mathbf{r}}_1,\omega)}\sqrt{\mathcal{I}(\vec{\mathbf{r}}_2,\omega)}\operatorname{Re}\left[\operatorname{tr}\boldsymbol{\mathcal{w}}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)e^{-i\omega\frac{r_1-r_2}{c}}\right]\end{aligned} \tag{6.20}$$

where r_1 and r_2 are the distances from $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$, respectively, to the observation point $\vec{\mathbf{r}}$. The derivation is identical to the derivations that preceded Eq. (5.26).

The matrix-based space-time and the space-frequency formalisms of optical coherence serve to merge polarization and coherence into a single entity: The diagonal elements express the coherence properties, while the off-diagonal elements describe the polarization characteristics. This is not done solely in the name of convenience—both polarization and coherence affect observable properties of electromagnetic waves and guide their interaction with matter. Indeed, polarization and coherence might appear as unrelated aspects at first, but nonetheless they are very closely intertwined in describing the behaviour of a wave ensemble. This surprisingly tight relationship is demonstrated by the observation that the coherence properties of the wave induce polarization changes on propagation [Korotkova and Wolf 2005].

6.2 Diffraction of Polarized Light

In Subsection 5.3.1 we formulated the propagation of the cross-spectral density as a scalar diffraction problem. A scalar diffraction formulation is useful when sourcing waves from a light source, as polarization is essentially always either strictly linear (lasing) or random (spontaneous emission sources). Further, scalar diffraction operates on a single scalar quantity, which might be assumed to be the electric or magnetic field of a plane wave. However, unless the plane wave is at perfect normal incidence, applying scalar diffraction (Eq. (5.50)) to the electric and magnetic fields separately does not produce fields that comply with Maxwell's equations. We now formulate the diffraction problem for electromagnetic waves under the context of the more rigorous vector diffraction theory.

Under vector diffraction theory we treat incoming light as an electromagnetic incident plane wave, with its electric and magnetic components known on some aperture Σ, and we seek a unique solution that satisfies Maxwell's equations (see Fig. 4). As before, the geometry is an aperture that is located on $z = 0$ and we solve for $\vec{\mathbf{r}} = (x, y, z \geq 0)$. We follow Zangwill [2013] by looking for solutions formulated as a superposition of monochromatic spherical waves, each varying as $e^{-i\omega t}$, with $\omega = ck$ being the angular frequency of interest. That approach gives rise to *Smythe's diffraction formula*:

$$\vec{\mathbf{E}}(\vec{\mathbf{r}}) = 2\boldsymbol{\nabla}\times\int_{z'=0}\mathrm{d}^2\vec{\mathbf{r}}'_\perp G_0(\vec{\mathbf{r}}-\vec{\mathbf{r}}'_\perp)\hat{\mathbf{z}}\times\vec{\mathbf{E}}(\vec{\mathbf{r}}'_\perp) \tag{6.21}$$

$$\vec{\mathbf{B}}(\vec{\mathbf{r}}) = -\frac{i}{k_0}\boldsymbol{\nabla}\times\vec{\mathbf{E}}(\vec{\mathbf{r}}) \tag{6.22}$$

with G_0 being the free-space Green function. The above expression for the magnetic field arises immediately from Maxwell equations, Eq. (4.2). It can be verified that if $\vec{\mathbf{E}}$ satisfies Maxwell's divergence equation, $\boldsymbol{\nabla}\cdot\vec{\mathbf{E}} = 0$ (Eq. (4.1)), on the aperture, then $\boldsymbol{\nabla}\cdot\vec{\mathbf{E}} = 0$ on $z > 0$ as well. Smythe's formula is a manifestation of Huygens' principle, with every point on the aperture being regarded a source of spherical wavelets. With a little algebra we can rewrite the expression for the electric field as

$$\vec{\mathbf{E}}(\vec{\mathbf{r}}) = -2\int_{z'=0} \mathrm{d}^2\vec{\mathbf{r}}'_\perp \left[\hat{\mathbf{z}} \times \vec{\mathbf{E}}(\vec{\mathbf{r}}'_\perp)\right] \times \boldsymbol{\nabla} G_0 \tag{6.23}$$

We proceed in a fashion similar to the scalar diffraction case. At the Fresnel-Fraunhofer diffraction region, far from the aperture with respect to the wavelength and characteristic size of the aperture, we can simplify considerably. The gradient of the free-space Green function was computed previously (Eq. (5.55)):

$$\boldsymbol{\nabla} G_0\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'_\perp\right) = \hat{\mathbf{R}}\frac{ikR - 1}{4\pi R^2} e^{ikR} \tag{6.24}$$

with $\vec{\mathbf{R}} = \vec{\mathbf{r}} - \vec{\mathbf{r}}'_\perp$ and $R = |\vec{\mathbf{R}}|$, $\hat{\mathbf{R}} = \frac{\vec{\mathbf{R}}}{R}$, as usual. Using the power expansion $R \approx r - \hat{\mathbf{r}}\cdot\vec{\mathbf{r}}'_\perp$ (Eq. (5.60)), as well as the fact that $kR \ggg 1$, we get

$$\boldsymbol{\nabla} G_0\left(\vec{\mathbf{r}} - \vec{\mathbf{r}}'_\perp\right) \approx ik\hat{\mathbf{r}}\frac{e^{ikr}e^{-ik\hat{\mathbf{r}}\cdot\vec{\mathbf{r}}'_\perp}}{4\pi r} \tag{6.25}$$

As with scalar diffraction, we apply the Kirchhoff approximation, that is $\hat{\mathbf{z}} \times \vec{\mathbf{E}} = 0$ everywhere on $z = 0$ except the aperture $\Sigma \subset \{z = 0\}$. This is a good approximation at the high-frequency limit with wavelengths small compared with the aperture size. Finally, by substituting Eq. (6.25) into Eq. (6.23) the far-field Smythe's diffraction formula under the Kirchhoff approximation becomes

$$\vec{\mathbf{E}}(\vec{\mathbf{r}}) = ik\frac{e^{ikr}}{2\pi r}\hat{\mathbf{r}} \times \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}'_\perp \left[\hat{\mathbf{z}} \times \vec{\mathbf{E}}(\vec{\mathbf{r}}'_\perp)\right] e^{-ik\hat{\mathbf{r}}\cdot\vec{\mathbf{r}}'_\perp} \tag{6.26}$$

Generic Framework for Physical Light Transport

7 RENDERING EQUATIONS

In this section we derive a set of generic rendering equations for free-space light transport.

7.1 Sourcing of Electromagnetic Waves

In this subsection, we derive formal expressions for the coherence functions of a wave ensemble radiated by a source, and we start by discussing a natural ("incoherent") light source. As previously discussed, such a natural source would be a spontaneous emission source—essentially a collection of independent, elementary radiators (molecules, atoms, electrons).

Let S be the light source and $V \subset \mathbb{R}^3$ its volume. The power spectral density of the light source—which describes the spectrum of the emitted light by the source—is denoted as $\Lambda(\omega)$. It was shown in Subsection 5.1.1 that a wave ensemble can always be written as a decomposition into monochromatic wavelets, viz. $\{v(\vec{\mathbf{r}},\omega)e^{-i\omega t}\}$. Further, let ${}^{\mathfrak{s}}v\,e^{-i\omega t}$ be a realization that describes the radiation from an elementary radiator at point $\vec{\mathbf{s}} \in V$. The source radiates incoherently, therefore it might seem justified to let $\left\langle {}^{\mathfrak{s}}v(\vec{\mathbf{s}})\,{}^{\mathfrak{q}}v(\vec{\mathbf{q}})^\star\right\rangle_\omega \equiv 0$ for every $\vec{\mathbf{s}} \neq \vec{\mathbf{q}}$. However, in that case we can immediately deduce that $\mathcal{W} \equiv 0$ everywhere away from the source (directly via the cross-spectral density propagation Eq. (5.61)). Indeed, a perfectly incoherent wave does not give rise to a radiation field and does not propagate! A propagating wave must admit spatial coherence over an extent of, roughly, a wavelength [Goodman 2015]. Experiments performed by Carminati and Greffet [1999] show that the near-field spatial coherence of thermal (natural) light is roughly $\frac{1}{2}\lambda$ when resonant surface waves, most notably surface-plasmon polaritons, are ignored (those surface resonances decay sharply away from the source and do not give rise to a Fresnel-region radiation field). With that in mind, we elect to model the spatial coherence (inside the source) as a Gaussian with a deviation of half a wavelength:

$$\mathcal{W}(\vec{\mathbf{s}},\vec{\mathbf{q}},\omega) = \left\langle {}^{\mathfrak{s}}v(\vec{\mathbf{s}})\,{}^{\mathfrak{q}}v(\vec{\mathbf{q}})^\star\right\rangle_\omega = \frac{1}{4\pi}\Lambda(\omega)e^{-\frac{1}{2}\left(\frac{|\vec{\mathbf{s}}-\vec{\mathbf{q}}|}{\lambda/2}\right)^2} \tag{7.1}$$

where $\vec{\mathbf{s}},\vec{\mathbf{q}} \in V$ are elementary radiators and $\frac{1}{4\pi}$ the isotropic phase factor, as each elementary radiator in a spontaneous emission source radiates isotropically. In addition to experimental evidence, we also justify the order-of-magnitude guess of the spatial coherence length (inside the source) via the observation that it only serves to scale the final expression for the cross-spectral density of the sourced light by a constant, and does not affect its shape.

The cross-spectral density above is defined only in the volume of the light source. In-order to extend the definition to the entire space we proceed by substituting $\mathcal{W}$ into the free-space propagation of a sourced cross-spectral density, Eq. (5.61):

$$\begin{aligned}\mathcal{W}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) = {}& \frac{\cos\theta_1\cos\theta_2}{\lambda^2 r_1 r_2}e^{ik(r_1-r_2)} \\ & \times \int_V \mathrm{d}^3\vec{\mathbf{s}}\, e^{-ik\hat{\mathbf{r}}_1\cdot\vec{\mathbf{s}}} \int_V \mathrm{d}^3\vec{\mathbf{q}} \left\langle {}^{\mathfrak{s}}v(\vec{\mathbf{s}})\,{}^{\mathfrak{q}}v(\vec{\mathbf{q}})^\star\right\rangle_\omega e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{q}}}\end{aligned} \tag{7.2}$$

where the points of interest are $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$, and as usual, $\omega = kc$ is the angular frequency. It is typically very safe to assume that the characteristic length of the light source is large compared to wavelength. Hence, due to the Gaussian term in Eq. (7.1) being a function of the distance between the points $\vec{\mathbf{s}}$ and $\vec{\mathbf{q}}$ only, the second integral only contributes in a spatial region very small compared to V. Then, by rewriting the second integral as an integral over the entire space centred at $\vec{\mathbf{s}}$, the following becomes a very good approximation:

$$\int_V \mathrm{d}^3\vec{\mathbf{q}}\,\left\langle {}^{\mathrm{s}}v(\vec{\mathbf{s}})\,{}^{\mathrm{q}}v(\vec{\mathbf{q}})^{\star}\right\rangle_{\omega} e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{q}}} \approx \frac{\Lambda(\omega)}{4\pi}\int_{\mathbb{R}^3}\mathrm{d}^3\vec{\mathbf{q}}'\,e^{-2\frac{|\vec{\mathbf{q}}'|^2}{\lambda^2}}e^{ik\hat{\mathbf{r}}_2\cdot(\vec{\mathbf{q}}'+\vec{\mathbf{s}})} \tag{7.3}$$

where we performed the variable change $\vec{\mathbf{q}}' = \vec{\mathbf{q}} - \vec{\mathbf{s}}$ to transform the origin to $\vec{\mathbf{s}}$. Expand the second integral in spherical coordinates, with ξ' being the radial coordinate, ϕ' the azimuthal angle and ϑ' the polar angle, such that the direction when $\vartheta' = \phi' = 0$ aligns with $\hat{\mathbf{r}}_2$. Then, $\hat{\mathbf{r}}_2\cdot\vec{\mathbf{q}}' = \xi'\cos\vartheta'$ and we integrate yielding an elementary closed-form solution

$$\begin{aligned}\int_{\mathbb{R}^3}\mathrm{d}^3\vec{\mathbf{q}}'\,e^{-2\frac{|\vec{\mathbf{q}}'|^2}{\lambda^2}}e^{ik\hat{\mathbf{r}}_2\cdot(\vec{\mathbf{q}}'+\vec{\mathbf{s}})}&\approx e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{s}}}\int_0^{\infty}\mathrm{d}\xi'\,\xi'^2e^{-2\frac{\xi'^2}{\lambda^2}}\int_0^{2\pi}\mathrm{d}\phi'\int_0^{\pi}\mathrm{d}\vartheta'\sin\vartheta' e^{ik\xi'\cos\vartheta'}\\&= 2\sqrt{2}e^{-\frac{\pi^2}{2}}\pi^{9/2}\frac{e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{s}}}}{k^3}\approx\pi\frac{e^{ik\hat{\mathbf{r}}_2\cdot\vec{\mathbf{s}}}}{k^3}\end{aligned} \tag{7.4}$$

where we simplify by remembering that $k = 2\pi/\lambda$. The exact value of the leading constant above matters little as it only scales the cross-spectral density and does not affect its shape.

Having integrated the inner integral, we are now ready to make the usual assumption that $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$ are close to each other but far form the source, hence setting $\theta_1 \approx \theta_2 \approx 0$. Substituting Eqs. (7.3) and (7.4) into Eq. (7.2) and simplifying allows us to rewrite Eq. (7.2) as a spatial Fourier transform:

$$\begin{aligned}\mathcal{W}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)&\approx\frac{\Lambda(\omega)}{4k^3}\frac{e^{ik(r_1-r_2)}}{r_1r_2}\int_V\mathrm{d}^3\vec{\mathbf{s}}\,e^{-ik(\hat{\mathbf{r}}_2-\hat{\mathbf{r}}_1)\cdot\vec{\mathbf{s}}}\\&=\frac{\Lambda(\omega)}{4k^3}\frac{e^{ik(r_1-r_2)}}{r_1r_2}\mathscr{F}\{\mathbb{1}_V\}[k(\hat{\mathbf{r}}_2-\hat{\mathbf{r}}_1)]\end{aligned} \tag{7.5}$$

where $\mathbb{1}_V$ is the characteristic function of the volume V, i.e. $\mathbb{1}_V(\vec{\mathbf{r}}) = 1$ when $\vec{\mathbf{r}}\in V$ and $\mathbb{1}_V(\vec{\mathbf{r}}) = 0$ when $\vec{\mathbf{r}}\notin V$. Thus, the cross-spectral density is proportional to the three-dimensional spatial Fourier transform of the source's geometry. If that geometry is, or can be approximated by, a ball of radius ρ, then the Fourier transform admits a closed-form expression:

$$\mathcal{W}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)\approx\frac{1}{\sqrt{2}}(\pi\rho)^{\frac{3}{2}}\Lambda(\omega)e^{ik(r_1-r_2)}\frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|\right]}{r_1r_2|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|^{\frac{3}{2}}} \tag{7.6}$$

and, for convenience, we have folded the wavelength-dependent $1/k^{3/2}$ term into the Λ.

An important theoretical observation is that the spatial frequencies that enter the Fourier transform in Eq. (7.5) are $\vec{\mathbf{f}}_{1,2} \triangleq k\hat{\mathbf{r}}_{1,2}$. As $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ are unit-vectors, the following holds: $|\vec{\mathbf{f}}_1 - \vec{\mathbf{f}}_2| \le 2k$. That is, spectral components with frequencies greater than $\frac{4\pi}{\lambda}$ do not give rise to propagating energy. This is known as the *evanescent waves* phenomenon [Mandel and Wolf 1995].

The above Eq. (7.6) is our sourcing equation for natural spherical light sources and we briefly examine its behaviour now. We can see that spatial coherence indeed behaves as an Airy disk and the spatial coherence area increases with distance from the source (due to the implicit $\frac{1}{r}$ factor admitted by $\hat{\mathbf{r}} = \vec{\mathbf{r}}/r$ in the Bessel function), confirming our findings in Subsection 5.2.1. The induced spectral density at a point is then simply the limit

$$\mathcal{J}(\vec{\mathbf{r}},\omega) = \lim_{\vec{\mathbf{r}}_2\to\vec{\mathbf{r}}}\mathcal{W}(\vec{\mathbf{r}},\vec{\mathbf{r}}_2,\omega)\propto\frac{\Lambda(\omega)}{r^2} \tag{7.7}$$

and indeed, we can confirm that the spectral density, and in-turn the intensity, of the sourced radiation decays as $\frac{1}{r^2}$, as desired and expected. The argument of the Bessel function can generally be large, even in the far-field where $|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1| \ll r_{1,2}$, due to the wavenumber k being large in the visible spectrum. For large arguments the Bessel function admits the asymptotic form [Abramowitz and Stegun 1965]:

$$\mathrm{J}_{\frac{3}{2}}(x)\sim-\sqrt{\frac{2}{x\pi}}\cos x + O\left(x^{-\frac{3}{2}}\right) \tag{7.8}$$

Substituting the above expression into the sourcing equation, Eq. (7.6), we get

$$\mathcal{W}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)\propto\Lambda(\omega)\frac{\cos\left[\rho k|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|\right]}{r_1r_2|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|^2} \tag{7.9}$$

which portrays the oscillatory behaviour of the cross-spectral density. In addition, the first zero is at about $|\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2| \approx \frac{b_0}{k\rho}$, where $b_0 \approx 4.49340$ is the first zero of the Bessel function of the first kind of order $\frac{3}{2}$ [Abramowitz and Stegun 1965]. This expression is also consistent with the expression for the first zero that was found in Subsection 5.2.1 (with the only difference being due to the order of the Bessel function).

The relation between the propagated field's cross-spectral density and its Fourier transform is a well-known result in Fourier optics [Mandel and Wolf 1995]. We build upon that by deriving analytic expressions for the cross-spectral density of light sourced from a natural source that is assumed to be composed of very many independent radiators. Eq. (7.6) is exact, in-practice, under very natural assumptions ($r_{1,2}$ are large with respect to the characteristic size of the source and the distance between the points of interest) and for spherical light sources that conform to the described model. Some other geometries, most notably any geometry that can be approximated as a simple polytope, also admit exact closed-form expressions (see Appendix A).

Gaussian Schell-model radiation. We now briefly discuss sourcing from a coherent source, like a laser. We assume a planar circular source, denoted Σ, of radius ρ, located on the xy-plane and centred around the origin. The source radiates at a tight beam around the $+\hat{\mathbf{z}}$ direction, and with a Gaussian distribution across the source of both its spatial coherence as well as the radiating power spectral

density, i.e.

$$\Lambda\left(\rho', \omega\right) = \Lambda_0(\omega) e^{-\frac{\rho'^2}{2\sigma_s^2}} \tag{7.10}$$

$$\mathcal{W}'\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) = \sqrt{\Lambda(|\vec{\mathbf{r}}_1|, \omega)}\sqrt{\Lambda(|\vec{\mathbf{r}}_2|, \omega)} e^{-\frac{|\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2|^2}{2\sigma_c^2}} \tag{7.11}$$

on the source, with Λ_0 being the power spectral density at the centre, σ_s the standard deviation of distribution of the radiating power across the source and σ_c the spatial coherence standard deviation. Such sources are known as Gaussian Schell-model sources [Wolf 2007] and have been widely used in theoretical study due to their very simple and convenient analytic form. In practice, it holds that the size of the source is large compared to the standard deviations of the power spectral density, i.e. $a \gg \sigma_s$. Plugging the expressions above into the scalar propagation equation (Eq. (5.61)), simplifying using the assumptions above and considering the limit $\sigma_c \to \infty$, i.e. assuming a fully coherent source (a perfect laser), results in [Wolf 2007]

$$\mathcal{W}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) \approx \left(2k\sigma_s^2\right)^2 \Lambda_0(\omega) e^{-2(k\sigma_s\theta)^2} e^{ik(r_1 - r_2)} \tag{7.12}$$

where the assumption is that $\theta \approx \theta_1 \approx \theta_2$. It is immediately evident that such a coherent source radiates in a very tight beam, roughly up to $|\theta| \sim \frac{1}{k\sigma_s}$.

7.2 Diffraction of Electromagnetic Waves

In the previous subsection we considered the cross-spectral density of light emitted from an incoherent light source. We now study the diffraction of the sourced cross-spectral density by an aperture, using the Smythe's diffraction formulae for monochromatic electromagnetic waves that were derived in Subsection 6.2.

As before, we assume that $\Sigma \subset \{z = 0\}$ is an aperture with normal $\hat{\mathbf{z}}$. A wave ensemble is incident to the aperture, diffracts and propagates. Let $\{\vec{\mathbf{E}}(\vec{\mathbf{r}}, \omega) e^{-i\omega t}\}$ be a decomposition of that wave ensemble into a collection of monochromatic wavelets, each a (vector) electric field treated as stationary, at least in the wide sense, stochastic process. To study propagation, we propagate the individual coherent-modes (the monochromatic wavelets), and recombine the beam. This has been done implicitly so far, and we now do so explicitly. The electric field of any such monochromatic wavelet can be decomposed into its mutually-orthogonal transverse components under some some orthonormal basis. Assume that the incident electric field decomposes under some given basis $\{\hat{\mathbf{e}}'_1, \hat{\mathbf{e}}'_2\}$, and the incident field's cross-spectral density matrix is denote as $\boldsymbol{\mathcal{W}}'$. It is given that $\boldsymbol{\mathcal{W}}'$ is known on the aperture. Likewise, the orthonormal basis under which the propagated light is decomposed is assumed to be $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ and we are interested in deriving an expression for the cross-spectral density between a couple of points $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$ (with $z > 0$), i.e. $\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega)$, of the propagated light that has diffracted through the aperture. The Fresnel-Fraunhofer region assumptions is retained as well: We assume that the points are far from the aperture, with respect to the distance between each other, the characteristic size of the aperture and the wavelength, and denote $\vec{\mathbf{r}} \approx \vec{\mathbf{r}}_1 \approx \vec{\mathbf{r}}_2$. This implies that $\vec{\mathbf{r}} \cdot \hat{\mathbf{e}}_j = 0$.

The cross-spectral density between the points $\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2$ can be written as the ensemble average over the same-frequency wavelets (see Eq. (5.23)), and each such constituent $\vec{\mathbf{E}}$ diffracts and propagates in accordance to Smythe's diffractions formula (Eq. (6.23)) or its far-field counterpart (Eq. (6.26)). Therefore, an element $\mathcal{W}_{xy}$ (indexed by $x, y \in \{1, 2\}$) of the cross-spectral density matrix (Eq. (6.15)) can be written as

$$\begin{aligned}
\mathcal{W}_{xy}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) &= \left\langle \left[\hat{\mathbf{e}}_x \cdot \vec{\mathbf{E}}(\vec{\mathbf{r}}_1)\right] \left[\hat{\mathbf{e}}_y \cdot \vec{\mathbf{E}}(\vec{\mathbf{r}}_2)\right]^\star \right\rangle_\omega \\
&= \frac{k^2 e^{ik(r_1 - r_2)}}{4\pi^2 r_1 r_2} \left\langle \left[\hat{\mathbf{e}}_x \cdot \left(\hat{\mathbf{r}}_1 \times \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}'_\perp \left(\hat{\mathbf{z}} \times \vec{\mathbf{E}}(\vec{\mathbf{r}}'_\perp)\right) e^{-ik\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}'_\perp}\right)\right] \right. \\
&\quad \left. \times \left[\hat{\mathbf{e}}_y \cdot \left(\hat{\mathbf{r}}_2 \times \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}'_\perp \left(\hat{\mathbf{z}} \times \vec{\mathbf{E}}(\vec{\mathbf{r}}'_\perp)\right) e^{-ik\hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}'_\perp}\right)\right]^\star \right\rangle_\omega
\end{aligned} \tag{7.13}$$

where we applied the far-field Smythe's diffraction formula to each of the field components. And, as usual, $r_j = |\vec{\mathbf{r}}_j|$ and $\hat{\mathbf{r}}_j = \vec{\mathbf{r}}_j / r_j$. To bring out the cross-spectral density on the aperture, $\boldsymbol{\mathcal{W}}'$, in the expression above, we decompose the electric field into its (incident) transverse components, viz. $\vec{\mathbf{E}} = \hat{\mathbf{e}}'_1 E_1 + \hat{\mathbf{e}}'_2 E_2$. Then, the vector expressions, inside the ensemble average, are rewritten as follows

$$\begin{aligned}
\hat{\mathbf{e}}_j \cdot \left[\hat{\mathbf{r}}_l \times \left(\hat{\mathbf{z}} \times \vec{\mathbf{E}}\right)\right] &= \hat{\mathbf{e}}_j \cdot \left[\hat{\mathbf{z}}\left(\hat{\mathbf{r}}_l \cdot \vec{\mathbf{E}}\right) - \left(\hat{\mathbf{r}}_l \cdot \hat{\mathbf{z}}\right)\vec{\mathbf{E}}\right] \\
&= \left(\hat{\mathbf{e}}_j \cdot \hat{\mathbf{z}}\right)\left(\hat{\mathbf{r}}_l \cdot \vec{\mathbf{E}}\right) - \left(\hat{\mathbf{r}}_l \cdot \hat{\mathbf{z}}\right)\left(\hat{\mathbf{e}}_j \cdot \vec{\mathbf{E}}\right) \\
&= \left(\hat{\mathbf{e}}_j \cdot \hat{\mathbf{z}}\right)\hat{\mathbf{r}}_l \cdot \left(\vec{\mathbf{e}}'_1 E_1 + \vec{\mathbf{e}}'_2 E_2\right) - \left(\hat{\mathbf{r}}_l \cdot \hat{\mathbf{z}}\right)\hat{\mathbf{e}}_j \cdot \left(\vec{\mathbf{e}}'_1 E_1 + \vec{\mathbf{e}}'_2 E_2\right)
\end{aligned} \tag{7.14}$$

where we applied the triple product vector identities (Eqs. (1.10) and (1.11)). Denote the common factors

$$\begin{aligned}
{}_j\mathscr{H}_{\xi,\zeta} &= \left(\hat{\mathbf{e}}_j \cdot \hat{\mathbf{z}}\right)\left(\hat{\mathbf{r}}_\xi \cdot \hat{\mathbf{e}}'_\zeta\right) - \left(\hat{\mathbf{r}}_\xi \cdot \hat{\mathbf{z}}\right)\left(\hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}'_\zeta\right) \\
&= \left(\hat{\mathbf{e}}_j \times \hat{\mathbf{r}}_\xi\right) \cdot \left(\hat{\mathbf{z}} \times \hat{\mathbf{e}}'_\zeta\right)
\end{aligned} \tag{7.15}$$

with $j, \xi, \zeta \in \{1, 2\}$ being indices. Then, by substituting Eqs. (7.14) and (7.15) into Eq. (7.13), formally interchanging the orders of ensemble averaging and integration and applying a series of elementary algebraic manipulations, the propagated cross-spectral density finally becomes

$$\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) = \frac{1}{\lambda^2} \frac{e^{ik(r_1 - r_2)}}{r_1 r_2} \boldsymbol{H}_1 \boldsymbol{\Psi} \boldsymbol{H}_2^T \tag{7.16}$$

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

$$\boldsymbol{H}_\xi \triangleq \begin{bmatrix} {}_1\mathscr{H}_{\xi,1} & {}_1\mathscr{H}_{\xi,2} \\ {}_2\mathscr{H}_{\xi,1} & {}_2\mathscr{H}_{\xi,2} \end{bmatrix} \tag{7.17}$$

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

$$\begin{aligned}
\boldsymbol{\Psi} &\triangleq \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}'_\perp \int_\Sigma \mathrm{d}^2\vec{\mathbf{r}}''_\perp \, \boldsymbol{\mathcal{W}}'\left(\vec{\mathbf{r}}'_\perp, \vec{\mathbf{r}}''_\perp, \omega\right) e^{-ik\left(\hat{\mathbf{r}}_1 \cdot \vec{\mathbf{r}}'_\perp - \hat{\mathbf{r}}_2 \cdot \vec{\mathbf{r}}''_\perp\right)} \\
&= \mathscr{F}\left\{\mathbb{1}_\Sigma\left(\vec{\mathbf{r}}'_\perp\right) \mathscr{F}\left\{\mathbb{1}_\Sigma\left(\vec{\mathbf{r}}''_\perp\right) \boldsymbol{\mathcal{W}}'\left(\vec{\mathbf{r}}'_\perp, \vec{\mathbf{r}}''_\perp, \omega\right)\right\}(-k\hat{\mathbf{r}}_2)\right\}(k\hat{\mathbf{r}}_1)
\end{aligned} \tag{7.18}$$

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

We briefly verify that the propagated cross-spectral density matrix, $\boldsymbol{\mathcal{W}}$, is a polarization matrix when evaluated at a single point. Indeed, when $\vec{\mathbf{r}}_1 = \vec{\mathbf{r}}_2$, it is easy to see that $\boldsymbol{H}_1 = \boldsymbol{H}_2$ and thus $(\boldsymbol{H}_1 \boldsymbol{\Psi} \boldsymbol{H}_2^T)^\dagger = \boldsymbol{H}_1 \Psi^\dagger \boldsymbol{H}_2^T$. As $\boldsymbol{\Psi} = \boldsymbol{\Psi}^\dagger$, due to $\boldsymbol{\Psi}$ being the Fourier transform of a Hermitian matrix, we deduce that $\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ is

Hermitian. Further, $\boldsymbol{\Psi}$ is positive semi-definite, thus we can write $\boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{A}^\dagger$ for some matrix $\boldsymbol{A}$ and then $\boldsymbol{H}_1\boldsymbol{\Psi}\boldsymbol{H}_2^T = (\boldsymbol{H}_1\boldsymbol{A})(\boldsymbol{H}_1\boldsymbol{A})^\dagger$, therefore $\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega) \succeq 0$. We conclude that if $\boldsymbol{\mathcal{W}}'(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ is a polarization matrix, then $\boldsymbol{\mathcal{W}}(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega)$ is a polarization matrix.

Eq. (7.16) (together with Eqs. (7.15), (7.17) and (7.18)) formalises the diffraction and propagation of the cross-spectral density matrix of light of any state of coherence, polarization and spectrum, by an arbitrary aperture, under the context of full electromagnetism. To our knowledge, this expression has not been previously derived, either in optics or computer graphics literature. A quick comparison with the scalar formalism (Subsection 5.3.1) reveals that Eq. (7.16) is equivalent to its scalar counterpart, Eq. (5.61) (applied per transverse component), only at normal incidence and when $\hat{\mathbf{r}} \approx \hat{\mathbf{z}}$, in which case indeed $\boldsymbol{H}_\xi \equiv \boldsymbol{I}$. However, away from normal incidence and exitance, the vectorized theory provides a more accurate formalism. Another advantage is that the polarization state of the wave ensemble becomes a first-class citizen under a vectorized formalism, which is important as polarization plays a major role in light-matter interactions.

7.3 Superposition of Light of any State of Coherence

Consider the superposition of light that arrives at an area from a couple of apertures or sources, Σ_1 and Σ_2, such that $\Sigma_1 \cap \Sigma_2 = \emptyset$. We assume that $\boldsymbol{\mathcal{W}}'$—the cross-spectral density on $\Sigma = \Sigma_1 \cup \Sigma_2$—is known. The cross-spectral density $\boldsymbol{\mathcal{W}}$ of light that propagated away from Σ and can be computed using the tools we have developed so far, namely one of the propagation equations, Eqs. (5.61) and (7.16). Let $\mathfrak{I}$ be a linear integration operator that serves as a shorthand for the propagation of the cross-spectral density, e.g.,

$$\mathfrak{I}_{\xi,\zeta}\{\boldsymbol{\mathcal{W}}'\} = \int_{\Sigma_\xi} \mathrm{d}^2\vec{\mathbf{s}}_1 \int_{\Sigma_\zeta} \mathrm{d}^2\vec{\mathbf{s}}_2\, \boldsymbol{\mathcal{W}}'\left(\vec{\mathbf{s}}_1', \vec{\mathbf{s}}_2', \omega\right) e^{-ik\left(\hat{\mathbf{r}}_1\cdot\vec{\mathbf{s}}_1' - \hat{\mathbf{r}}_2\cdot\vec{\mathbf{s}}_2'\right)} \tag{7.19}$$

or any other diffraction integral, e.g., Eq. (6.26). Then, the propagated cross-spectral density function can be succinctly written as the superposition

$$\begin{aligned}\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) &\propto \left(\mathfrak{I}_{1,1} + \mathfrak{I}_{2,2} + \mathfrak{I}_{1,2} + \mathfrak{I}_{2,1}\right)\{\boldsymbol{\mathcal{W}}'\} \\ &= \left(\mathfrak{I}_{1,1} + \mathfrak{I}_{2,2} + 2\,\mathrm{Re}\,\mathfrak{I}_{1,2}\right)\{\boldsymbol{\mathcal{W}}'\}\end{aligned} \tag{7.20}$$

with constants emitted both for brevity and as they depend on the exact formalism of choice. The equality on the second line holds because $\boldsymbol{\mathcal{W}}'$ is a Hilbert–Schmidt kernel, thus $\boldsymbol{\mathcal{W}}'(\vec{\mathbf{s}}_1, \vec{\mathbf{s}}_2, \omega) = \boldsymbol{\mathcal{W}}'(\vec{\mathbf{s}}_2, \vec{\mathbf{s}}_1, \omega)^\star$ and $\mathfrak{I}_{1,2} = \mathfrak{I}_{2,1}^\star$.

Incoherent super-position. When the contributions from Σ_1 and Σ_2 are mutually-independent, e.g. one is a natural source, then the double-integral over Σ_1 and Σ_2, $\mathfrak{I}_{1,2}\{\boldsymbol{\mathcal{W}}'\}$, is essentially zero. This holds, in general, when the apertures (or sources) are far from each other, with respect to the distance over which the light remains spatially coherent, i.e. $|\mathrm{tr}\,\boldsymbol{\mathcal{W}}'(\vec{\mathbf{s}}_1, \vec{\mathbf{s}}_2, \omega)| \approx 0$ when $\vec{\mathbf{s}}_1 \in \Sigma_1, \vec{\mathbf{s}}_2 \in \Sigma_2$ (or vice versa). Then,

$$\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) \propto \left(\mathfrak{I}_{1,1} + \mathfrak{I}_{2,2}\right)\{\boldsymbol{\mathcal{W}}'\} \tag{7.21}$$

We conclude that the cross-spectral density of the superposition of wave ensembles, sourced (or propagated) from different, independent sources (or apertures) is simply the superposition of each ensemble's cross-spectral density (computed using the formulae in Subsections 7.1 and 7.2).

The formulations above are trivially extended to any finite collection of disjoint apertures or sources, $\{\Sigma_1, \Sigma_2, \ldots, \Sigma_n\}$. Our superposition of cross-spectral densities formulae for coherent and incoherent super-positions, respectively, are then:

$$\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) \propto \left(\sum_{j=1}^{n} \mathfrak{I}_{j,j} + 2\,\mathrm{Re}\sum_{m=j+1}^{n} \mathfrak{I}_{j,m}\right)\{\boldsymbol{\mathcal{W}}'\} \tag{7.22}$$

$$\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1, \vec{\mathbf{r}}_2, \omega\right) \propto \sum_{j=1}^{n} \mathfrak{I}_{j,j}\{\boldsymbol{\mathcal{W}}'\} \tag{7.23}$$

7.4 Measurement

Measurement is an operation that analyses the wave ensemble to deduce its observed intensity. It is assumed that observation is done over a period of time long compared to the temporal coherence of the measured radiation, see Subsection 7.7 for additional discussion. In space-frequency formulation of optical coherence, we use Eq. (5.24) to compute the intensity of the wave ensemble characterized by the cross-spectral density above, that is

$$I(\vec{\mathbf{r}}) = \frac{1}{\pi}\int_0^\infty \mathrm{d}\omega\, \mathcal{W}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega\right) \tag{7.24}$$

or, for cross-spectral density expressed in matrix form:

$$I(\vec{\mathbf{r}}) = \frac{1}{\pi}\int_0^\infty \mathrm{d}\omega\, \mathrm{tr}\,\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega\right) \tag{7.25}$$

In general, however, it is the spectral response of an imaging system, with respect to the spectral stimulus, that interests us. Computation of that quantity is accomplished by integrating the above over the support of some response function (e.g. the XYZ colour matching functions), and we do not expect to achieve a closed-form expression in that case. However, as the spectrum will generally be a well-behaved function, we resign ourselves to numerical integration over the visible spectrum. Note that typically during rendering we only measure intensity once per image element (or pixel).

Special types of spectra. We briefly discuss a few spectral densities, Λ, that arise in practical applications.

Emission from low-pressure gas discharge lamps (e.g., fluorescent lamps and neon lighting) admits a spectral density that is readily approximated by a Gaussian [Hollas 2004]:

$$\Lambda_{\mathrm{lp}}(\omega) \triangleq \sqrt{\frac{\ln 2}{\pi}}\frac{2}{\Delta\omega}e^{-4\ln 2\left(\frac{\omega-\bar{\omega}}{\Delta\omega}\right)^2} \tag{7.26}$$

where $\bar{\omega}$ is the mean and $\Delta\omega$ is the standard deviation of the Gaussian spectrum, also known as *half-power bandwidth*. When $\Delta\omega$ is very small the Gaussian spectrum is also a very good approximation for (single-mode) laser radiation.

Another useful spectrum is the *Lorentzian spectrum*, which approximates the spectra of high-pressure gas discharge lamps (e.g., some automotive lamps and powerful outdoor lighting lamps) [Hollas 2004]:

$$\Lambda_{\mathrm{hp}}(\omega) \triangleq \frac{2}{\pi\Delta\omega}\frac{1}{1 + \left(2\frac{\omega-\bar{\omega}}{\Delta\omega}\right)^2} \tag{7.27}$$

with $\bar{\omega}$ being the mean and $\Delta\omega$ the half-power bandwidth, as before.

Finally, a body that absorbs all incident radiant energy and remains at thermal equilibrium is known as a *blackbody* (e.g., incandescent bulbs and the sun). Such a radiator admits a spectral density described by *Planck's Law*:

$$\Lambda_{\mathrm{bb}}(\omega) \triangleq \frac{2h}{c^2}\frac{\omega^3}{e^{\frac{\hbar\omega}{k_{\mathrm{B}}T}}-1} \tag{7.28}$$

where T is the temperature (in Kelvin), k_{B} is the Boltzmann constant and h is the Planck constant.

7.5 Fresnel Equations

While we avoid discussion and formalisation of light-matter interaction, we briefly mention the well-known Fresnel equations, which relate the amplitudes of an incident field to the reflected and refracted fields at an interface between two materials. 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 ensemble is incident upon the interface from the $z > 0$ half-space. As before, we consider a constituent $\vec{\mathbf{E}}(\vec{\mathbf{r}},\omega) = \vec{\mathbf{E}}_0 e^{i(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}-\omega t)}$ of the decomposition of the incident ensemble into a collection of monochromatic wavelets. $\vec{\mathbf{k}}$ is the wavevector of the incident field, and two solutions to Maxwell's equation arise at the interface: a reflected field and a refracted field. The matching conditions implied by the Maxwell equations readily allow us to derive the famous *Snell's law of refraction* [Zangwill 2013]. Snell's law 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:

$$\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2 \tag{7.29}$$

The *law of reflection* states that the angle of reflection is equal to the angle of incidence, thus $\vec{\mathbf{k}}_R = \vec{\mathbf{k}} - 2(\vec{\mathbf{k}}\cdot\hat{\mathbf{z}})\hat{\mathbf{z}}$ is the reflected wavevector. The three wavevectors $\vec{\mathbf{k}}, \vec{\mathbf{k}}_R, \vec{\mathbf{k}}_T$ are co-planar.

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* component (where the field is strictly perpendicular to the plane of incidence), and a *p-polarized* component (where the field is strictly parallel to the plane of incidence). 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, respectively, s- and p-polarized waves, Fresnel equations become [Zangwill 2013]:

$$r_s = \frac{Z_2\cos\theta_1 - Z_1\cos\theta_2}{Z_2\cos\theta_1 + Z_1\cos\theta_2} \qquad r_p = \frac{Z_1\cos\theta_1 - Z_2\cos\theta_2}{Z_1\cos\theta_1 + Z_2\cos\theta_2} \tag{7.30}$$

$$t_s = \frac{2Z_2\cos\theta_1}{Z_2\cos\theta_1 + Z_1\cos\theta_2} \qquad t_p = \frac{2Z_2\cos\theta_1}{Z_1\cos\theta_1 + Z_2\cos\theta_2} \tag{7.31}$$

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. (7.30) and (7.31) can be rewritten in terms of the indices of refraction $\eta_{1,2}$.

We are now ready to derive an expression for the cross-spectral density matrix, $\boldsymbol{\mathcal{W}}$, of light that was reflected or refracted at the interface between two media. Assume that $\boldsymbol{\mathcal{W}}'$ is the (known) cross-spectral density matrix of the incident light. Both matrices are formulated with respect to a decomposition into s- and p-polarized fields. Let $E_{s,p} = \vec{\mathbf{E}}_0 \cdot e_{s,p}$ be the amplitudes of the s- and p-polarized components of the incident field, and likewise let

$$E_s^{(R)} = r_s E_s \quad E_p^{(R)} = r_p E_p \quad E_s^{(T)} = t_s E_s \quad E_p^{(T)} = t_p E_p \tag{7.32}$$

be the amplitudes of the respective components of the reflected and refracted fields. Then, an element of the cross-spectral density of the reflected field is

$$\begin{aligned}\boldsymbol{\mathcal{W}}^{(R)}_{\xi,\zeta}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega) &= \left\langle \left(E_\xi^{(R)} e^{i(\vec{\mathbf{k}}_R\cdot\vec{\mathbf{r}}_1-\omega t)}\right)\left(E_\zeta^{(R)} e^{i(\vec{\mathbf{k}}_R\cdot\vec{\mathbf{r}}_2-\omega t)}\right)^\star \right\rangle_\omega \\ &= \left\langle r_\xi E_\xi r_\zeta^\star E_\zeta^\star \right\rangle_\omega = r_\xi r_\zeta^\star \boldsymbol{\mathcal{W}}'_{\xi,\zeta}(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)\end{aligned} \tag{7.33}$$

where $\xi, \zeta \in \{s, p\}$ and are also used to index the cross-spectral density matrices, and $\vec{\mathbf{r}}_{1,2}$ are points on the surface. The cross-spectral density of the refracted field is deduced analogously, and we conclude:

$$\boldsymbol{\mathcal{W}}^{(R)} = \begin{bmatrix} r_s & \\ & r_p \end{bmatrix} \cdot \boldsymbol{\mathcal{W}}' \cdot \begin{bmatrix} r_s & \\ & r_p \end{bmatrix}^\dagger \tag{7.34}$$

$$\boldsymbol{\mathcal{W}}^{(T)} = \begin{bmatrix} t_s & \\ & t_p \end{bmatrix} \cdot \boldsymbol{\mathcal{W}}' \cdot \begin{bmatrix} t_s & \\ & t_p \end{bmatrix}^\dagger \tag{7.35}$$

It is easy to verify that the resulting cross-spectral density matrices are indeed polarization matrices.

The Church polarization factor. The Fresnel equations above are derived elementary from Maxwell divergence equations, Eq. (4.1) (as a side note, the relations described by the Fresnel equations are purely kinematic, in the sense that the dynamic Maxwell curl equations, Eq. (4.2), are neither used nor needed), however they assume a perfectly smooth surface. Of practical interest are similar equations that take residual surface roughness and surface wave excitation (e.g, surface-plasmon polaritons) into account. Such relations appear to have been first derived in the modern radar scattering literature by Ruck [1970] and then introduced into applied optics and surface scattering by Church et al. [1977] (also see Stover [2012] for additional discussion). We use Church's notation, and denote the following relations as the *Church polarization factor Q*. This is the same factor that was borrowed both by the Rayleigh-Rice small perturbation surface theory and the Harvey-Shack generalized scattering theory [Krywonos 2006].

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 expressions that give rise to the Church polarization factor are

$$q_{ss} \triangleq \frac{(\epsilon-1)\cos\phi_s}{\left(\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{7.36}$$

$$q_{sp} \triangleq \frac{(\epsilon-1)\sqrt{\epsilon - \sin^2\theta_s}\sin\phi_s}{\left(\cos\theta_i + \sqrt{\epsilon - \sin^2\theta_i}\right)\left(\epsilon\cos\theta_s + \sqrt{\epsilon - \sin^2\theta_s}\right)} \tag{7.37}$$

$$q_{ps} \triangleq \frac{(\epsilon-1)\sqrt{\epsilon-\sin^2\theta_i}\sin\phi_s}{\left(\epsilon\cos\theta_i+\sqrt{\epsilon-\sin^2\theta_i}\right)\left(\cos\theta_s+\sqrt{\epsilon-\sin^2\theta_s}\right)} \tag{7.38}$$

$$q_{pp} \triangleq \frac{(\epsilon-1)\left(\sqrt{\epsilon-\sin^2\theta_s}\sqrt{\epsilon-\sin^2\theta_i}\cos\phi_s-\epsilon\sin\theta_i\sin\theta_s\right)}{\left(\epsilon\cos\theta_i+\sqrt{\epsilon-\sin^2\theta_i}\right)\left(\epsilon\cos\theta_s+\sqrt{\epsilon-\sin^2\theta_s}\right)} \tag{7.39}$$

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 $q_{pp} = r_p$. The total reflected energy is then the Church polarization factor Q and can be written as $Q = \sum_{xy} |q_{xy}|^2$. We define the following matrix, which we call the *Church reflection matrix*:

$$\boldsymbol{Q} \triangleq \begin{bmatrix} q_{ss} & q_{ps} \\ q_{sp} & q_{pp} \end{bmatrix} \tag{7.40}$$

which can be used in-place of the reflection matrix in Eq. (7.34), viz.

$$\boldsymbol{\mathcal{W}}^{(R)} = \boldsymbol{Q}\,\boldsymbol{\mathcal{W}}'\,\boldsymbol{Q}^{\dagger} \tag{7.41}$$

to describe the reflected cross-spectral density matrix in an arbitrary direction.

7.6 Space-Time Formulation

In Subsection 5.1.2 we discussed conditions for the separability of the spatial and temporal coherence. We deduced that the mutual coherence can be written as a product of spatial and temporal coherence functions if and only if the light's spectrum is identical in the (spatial and temporal) region of interest. This condition only applies in the immediate vicinity of a light source that radiates with a constant spectrum. For an incoherent light source the cross-spectral density is given by Eq. (7.1), and indeed the spatial coherence term is separate from the spectral term. By the Wiener–Khinchin Theorem—Theorem 3.1—the power spectral density is the Fourier transform pair of the temporal coherence function, i.e.

$$\tilde{\Gamma}(\tau) = \mathscr{F}^{-1}\{\Lambda(\omega)\} \tag{7.42}$$

Then, the cross-spectral density relates to the mutual coherence via Eq. (5.30). Applying the above yields the expression for the *sourcing of the mutual coherence* from a natural (i.e. incoherent) light source:

$$\Gamma(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\tau) = e^{-\frac{|\vec{\mathbf{r}}_1-\vec{\mathbf{r}}_2|^2}{\bar{\lambda}^2}}\,\mathscr{F}^{-1}\{\Lambda(\omega)\}(\tau) \tag{7.43}$$

with $\bar{\lambda}$ being the mean wavelength, as before. Once propagation-induced spectral changes arise the mutual coherence can be recovered from the cross-spectral density via an inverse Fourier transform. We do not employ the mutual coherence in our formalisms, and therefore do not investigate further.

7.7 Higher-order Statistics

Our discussion and formulations have been centred around the second-order statistics of a wave ensemble—i.e. the mutual coherence or cross-spectral density functions. While the wave fields are typically not Gaussian and higher-order statistics would be required to reconstruct the ensemble constituents, we are ultimately only interested in accurate representation of the intensity of the measured (observed) field. For ergodic process (discussed in Subsection 3.0.1) the infinite-time average (Eq. (3.2)) equals to the ensemble average. Integrating over infinite time can be physically interpreted as observing light over very long periods. Specifically, very long with respect to the coherence time of the measured wave ensemble, and this is typically the case when observing light sourced from a natural source and imaging using a camera or the eye. When integrating over periods very long compared to the coherence time, many independent fluctuations contribute to the final intensity, and thus, by the central limit theorem, the statistics of the intensity are (point-wise) asymptotically Gaussian.

However, when dealing with quasi-monochromatic radiation that may exhibit very long coherence times, or when rendering images with ultra-low exposure time, the assumption that we average over infinite time no longer holds, and higher-order statistics might be required. Those higher-order statistics are more difficult to work with, and for this body-of-work, we simply assume that the second-order functions provide sufficient information for our applications. A more comprehensive discussion of higher-order coherence is presented by Goodman [2015].

8 GENERIC LIGHT PROPAGATORS

In this section we use the rendering equations developed in Section 7 to formalise physically-accurate light transport operators. Those are general-purpose operators, we make no spectral assumption or geometric requirements, except assuming Fresnel-Fraunhofer region, and, for simplicity, we only formalise for spherical sources. However, as mentioned in Section 7 and shown in Appendix A, other shapes will produce useful closed-form expressions as well.

8.1 The Spectral-Density Transport Equation

Consider the light transport equation (LTE) [Pharr et al. 2016] that governs the propagation and (equilibrium) distribution of radiance in a rendered scene:

$$L_o(\vec{\mathbf{x}},\Omega_o,\omega,t) = L_e(\vec{\mathbf{x}},\Omega_o,\omega,t) + \int_{\mathcal{S}^2} \mathrm{d}^2\Omega'\, f_r(\vec{\mathbf{x}},\Omega',\Omega_o,\omega,t) L_i(\vec{\mathbf{x}},\Omega',\omega,t)\left|\Omega'\cdot\hat{\mathbf{n}}\right| \tag{8.1}$$

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 $\vec{\mathbf{x}}$, ω is the radiation angular frequency, t is time, f_r is the bidirectional reflectance distribution function (BRDF), $\Omega \equiv (\theta,\phi)$ 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. The LTE written above ignores volumetric scattering and participating media, however it can be generalized to account for those phenomena as well. Expressions similar to the LTE that govern the transport of the coherence functions, specifically the cross-spectral density, can also be formulated, and that is the goal of this section. We begin with a short discussion of radiometry.

Radiometry. The basic quantity of traditional radiometry 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 LTE above deals exclusively with quantities of radiance. Under "geometric optics", radiance is conserved along rays in an optical system (in absence of transmission losses), however such optics are fundamentally paradoxical: as discussed previously, incoherent radiation does not propagate (into the far-field) and the concept of radiance is inconsistent with Maxwellian Electromagnetism [Wolf 1978]. A *generalized radiance* was introduced by Walther [1968] (and then discussed by a large body of work), which is a propagated coherence function. As we have rigorously shown, propagation of coherence functions is a diffraction problem, and we will now show that we can continue to use units of radiance to describe partially-coherent light propagation.

The cross-spectral density is the ensemble average of the equal-frequency constituents of the coherent-modes decomposition of a wave ensemble (as discussed in Subsection 5.1.1):

$$\mathcal{W}\left(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega\right)=\left\langle u(\vec{\mathbf{r}}_1,\omega)u(\vec{\mathbf{r}}_2,\omega)^\star\right\rangle_\omega \tag{8.2}$$

where u is a scalar monochromatic wavelet. The expression that enters into the ensemble average above, $u(\vec{\mathbf{r}}_1,\omega)u(\vec{\mathbf{r}}_2,\omega)^\star$, is the mutual intensity of u between points $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$, i.e. the equal-time mutual coherence (see Subsection 5.0.2). u has units of field strength, therefore the mutual intensity describes directional power flow (analogously to the Poynting vector, Eq. (4.15)). When dividing by the intrinsic impedance $Z=\sqrt{\mu/\epsilon}$ (encountered in Subsection 7.5), the result is the (mutual) instantaneous power density P that flows through the medium [Goodman 2015], viz.

$$P=\frac{u(\vec{\mathbf{r}}_1,\omega)u(\vec{\mathbf{r}}_2,\omega)^\star}{2Z} \tag{8.3}$$

P has units of spectral power per area, $\frac{\text{erg}}{\text{m}^2\,\text{s}\,\mu\text{m}}$ or $\frac{\text{W}}{\text{m}^2\,\mu\text{m}}$ (in SI-units), i.e. spectral irradiance. Therefore, the cross-spectral density normalized by the intrinsic impedance, $\mathcal{W}/(2Z)$, also has units of spectral irradiance, and $\frac{1}{2Z}\frac{\partial^2\mathcal{W}}{\partial\Omega\partial A\cos\theta}$ has units of spectral radiance (power per solid angle per projected source area). The normalization by $2Z$ is typically ignored.

Our transport equation. While the LTE (Eq. (8.1)) deals with values, we propagate functions. Hence, our formalism is described in terms of operators that formalise physical interactions by acting upon and generating cross-spectral density functions. As the cross-spectral density is a function of two spatial points, we propagate a cross-spectral density function not point-to-point, but between primary or secondary (aperture, surface patch, etc.) sources or an imaging device, each of which occupies a small spatial area, $\delta\vec{\mathbf{x}}\subset\mathbb{R}^3$. Typically the only constraint is the Fresnel-Fraunhofer region assumption, i.e. that the characteristic size of those patches $\delta\vec{\mathbf{x}}$ is large compared to wavelength and small compared to the propagation distance between the patches. Otherwise, our transport equation remains very similar to the classical LTE. Following the dimensional analysis above and using the relation between the differential surface area and differential solid angle, viz. $\mathrm{d}A_r=r^2\,\mathrm{d}\Omega$, in order to work with units of radiance we define the *radiance cross-spectral density* (RCSD):

$$\boldsymbol{\mathcal{L}}\left(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega\right)\triangleq r^2\frac{\partial\boldsymbol{\mathcal{W}}\left(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega\right)}{\partial A_s\cos\theta} \tag{8.4}$$

i.e. the cross-spectral density per solid angle per projected source area. Changes in medium impedance are accounted for by the Fresnel equations (discussed in Subsection 7.5), which modulate the power carried by an electromagnetic wave when light propagates from one medium to another.

We are now ready to formulate the *spectral-density transport equation* (SDTE):

$$\boldsymbol{\mathcal{L}}^{(o)}_{\vec{\mathbf{x}}\to\delta\vec{\mathbf{r}}}=\boldsymbol{\mathcal{L}}^{(e)}_{\vec{\mathbf{x}}\to\delta\vec{\mathbf{r}}}+\int_{\mathcal{S}^2}\mathrm{d}^2\Omega'\left|\Omega'\cdot\hat{\mathbf{n}}\right|\mathscr{D}_{\delta\vec{\mathbf{x}}\to\delta\vec{\mathbf{r}}}\left\{\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{x}}}\right\} \tag{8.5}$$

where the exitant, emitted and incident radiances, L, in the LTE (Eq. (8.1)) are replaced with their RCSD counterparts. $\boldsymbol{\mathcal{L}}^{(i)}_{\vec{\mathbf{r}}'\to\delta\vec{\mathbf{x}}}$ is the RCSD incident to the point $\vec{\mathbf{x}}$ from direction Ω_i and $\mathscr{D}_{\delta\vec{\mathbf{x}}\to\delta\vec{\mathbf{r}}}$ is a context-dependant diffraction operator that, given an incident RCSD impinging upon the patch $\delta\vec{\mathbf{x}}$ from direction Ω_i, generates the diffracted RCSD in direction Ω_o. The time dependence is implicit.

The rest of the section is dedicated to formalising the cross-spectral density transport operators that are at the core of our formalism. Note that some of the formulae derived in Section 7 have been implicitly formulated as radiant intensity (power per solid angle), which is the natural way to formulate diffraction. In this section we will rewrite those formulae as RCSD.

8.2 Propagation Operators

Definition 8.1 (Incoherent Sourcing Operator). Given Λ, the emitted power spectral density, and ρ, the source radius, the *incoherent sourcing operator* produces the sourced RCSD, propagating in some direction Ω (with constants folded into Λ):

$$\mathscr{S}_{\text{inc}}\{\Lambda,\rho\}\triangleq\boldsymbol{I}\sqrt{\frac{\pi}{\rho}}\Lambda(\omega)\frac{e^{ik(r_1-r_2)}}{r_1r_2}\frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|\right]}{|\hat{\mathbf{r}}_1-\hat{\mathbf{r}}_2|^{\frac{3}{2}}} \tag{8.6}$$

with $\bar{\lambda}$ being the mean wavelength, $\boldsymbol{I}$ the identity matrix.

Note that there is no directional dependence in the function produced by the incoherent sourcing operator, that is due to the isotropic nature of the elementary radiators that assemble a natural light source. We divided by the cross-sectional area of the source $\pi\rho^2$ to produce an RCSD. The off-diagonal elements of the matrix generated by $\mathscr{S}_{\text{inc}}$ are 0: We assume randomly polarized emitted energy and we can trivially verify that the degree-of-polarization of this matrix (using Eq. (6.13)) is zero.

Similarly, we also define a sourcing operator for coherent Gaussian Schell-model sources (single-mode lasers).

Definition 8.2 (Coherent Gaussian Schell-Model Beam Sourcing Operator). Given Λ, the emitted power spectral density, and σ_s, the standard deviation of distribution of the radiating power across the source, the *Gaussian Schell-model sourcing operator* is defined as follows:

$$\mathscr{S}_{\text{gsm}}\{\Lambda,\sigma_s\}\triangleq\begin{bmatrix}1&0\\0&0\end{bmatrix}\frac{2}{\pi\rho^2}\left(k\sigma_s^2\right)^2\Lambda(\omega)e^{-2(k\sigma_s\theta)^2}e^{ik(r_1-r_2)} \tag{8.7}$$

with θ being the inclination angle relative to the source, which is assumed to be a flat disk. The expression above describes fully linearly polarized radiation.

An important free-space propagation operator is the diffraction operator:

Definition 8.3 (Diffraction Operator). Let $\boldsymbol{\mathcal{W}}'$ be a RCSD function of the radiation impinging upon a planar aperture $\Sigma \subset \mathbb{R}^3$ with normal vector $\hat{\mathbf{n}}$. The diffracted and propagated RCSD in direction Ω_o from the aperture is generated by the *diffraction operator*:

$$\mathscr{D}\{\boldsymbol{\mathcal{W}}', \Omega_o, \Sigma\} \triangleq \frac{e^{ik(r_1-r_2)}}{\lambda^2 A_\Sigma |\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}|} \boldsymbol{H}_1 \, \mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{W}}'\} \boldsymbol{H}_2^T \tag{8.8}$$

with $(|\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}| A_\Sigma)$ being the projected aperture area and

$$\boldsymbol{H}_\xi \triangleq \begin{bmatrix} \left(\hat{\mathbf{e}}_1 \times \hat{\mathbf{r}}_\xi\right) \cdot \left(\hat{\mathbf{n}} \times \hat{\mathbf{e}}'_1\right) & \left(\hat{\mathbf{e}}_1 \times \hat{\mathbf{r}}_\xi\right) \cdot \left(\hat{\mathbf{n}} \times \hat{\mathbf{e}}'_2\right) \\ \left(\hat{\mathbf{e}}_2 \times \hat{\mathbf{r}}_\xi\right) \cdot \left(\hat{\mathbf{n}} \times \hat{\mathbf{e}}'_1\right) & \left(\hat{\mathbf{e}}_2 \times \hat{\mathbf{r}}_\xi\right) \cdot \left(\hat{\mathbf{n}} \times \hat{\mathbf{e}}'_2\right) \end{bmatrix} \tag{8.9}$$

The transverse basis vectors $\hat{\mathbf{e}}_{1,2}$ and $\hat{\mathbf{e}}'_{1,2}$ as well as the double Fourier transform over the aperture are as described in Subsection 7.2.

The Helmholtz reciprocity of the diffraction operator is discussed in Appendix B.

Finally, the following operator measures the radiance of the wave ensemble at a point.

Definition 8.4 (Observation Operator). Let $\boldsymbol{\mathcal{L}}$ be the RCSD. The observed radiance at a point $\vec{\mathbf{r}}$ is then

$$\mathscr{S}\{\boldsymbol{\mathcal{L}}, \vec{\mathbf{r}}\} \triangleq \frac{1}{\pi} \int_0^\infty \mathrm{d}\omega \, \mathrm{tr} \, \boldsymbol{\mathcal{L}}\left(\vec{\mathbf{r}}, \vec{\mathbf{r}}, \omega\right) \tag{8.10}$$

8.3 Light-Matter Interaction Operators

We begin with operators that change the polarization of the wave ensemble. Such operators can be borrowed directly from the *Jones calculus*. As before, let $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ be the basis under which an electromagnetic wave is decomposed into transverse orthogonal components, and let E_x, E_y be the transverse amplitudes. Then, the vector $[\, E_x \;\; E_y \,]^T$ is known as a *Jones vector* and light-matter interactions are modelled via *Jones matrices*—2×2 complex matrices that act upon Jones vectors.

Definition 8.5 (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

$$\mathscr{J}_{\boldsymbol{A}}\{\boldsymbol{\mathcal{L}}\} \triangleq \boldsymbol{A} \, \boldsymbol{\mathcal{L}} \, \boldsymbol{A}^\dagger \tag{8.11}$$

Note that while typical Jones calculus is generally applicable only to fully-coherent, full-polarized light propagating in an isotropic medium, this restriction is void under our formalism. For example, the matrix

$$\boldsymbol{P}_{\mathrm{linear},\vartheta} = \begin{bmatrix} \cos^2\vartheta & \cos\vartheta \sin\vartheta \\ \cos\vartheta \sin\vartheta & \sin^2\vartheta \end{bmatrix} \tag{8.12}$$

describes a linear polarizer (e.g., sunglasses), oriented at an angle of ϑ from the horizontal direction, and correctly acts upon a cross-spectral density matrix of any state of coherence and polarization.

Another useful example is a polarization rotation matrix, which can be used to rotate the transverse basis under which $\boldsymbol{\mathcal{L}}$ is decomposed:

$$\boldsymbol{Rot}_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{8.13}$$

Finally, a quarter-wave plate with slow vertical and fast horizontal axes, which serves to retard the phase of the portion of the wave ensemble that is polarized in the vertical direction:

$$\boldsymbol{W}_{\frac{\pi}{2}} = \begin{bmatrix} 1 & \\ & e^{-i\frac{\pi}{2}} \end{bmatrix} \tag{8.14}$$

Reflection and refraction operators. The only additional operators that involve light-matter interaction that we formalize are the simple reflection and refraction operators using the relations that were discussed in Subsection 7.5. Note that the Fresnel and Church reflection and refraction matrices that appear in Eqs. (7.34), (7.35) and (7.41) are nothing more than Jones matrices, and therefore the following definitions follow directly from the Jones operator, Definition 8.5.

Definition 8.6 (Fresnel Reflection Operator). Let $\boldsymbol{\mathcal{L}}$ be the RCSD impinging upon an interface between two media. Let $\hat{\mathbf{n}}$ be the normal of the interface and Ω_i the angle of incidence. Then, the reflected RCSD is

$$\mathscr{R}_{\mathrm{Fresnel}}\{\boldsymbol{\mathcal{L}}\} \triangleq \begin{bmatrix} r_s & \\ & r_p \end{bmatrix} \cdot \boldsymbol{\mathcal{L}} \cdot \begin{bmatrix} r_s & \\ & r_p \end{bmatrix}^\dagger \tag{8.15}$$

where $r_{s,p}$ are the Fresnel reflection coefficients, and we assume that the cross-spectral density matrix $\boldsymbol{\mathcal{L}}$ is formulated with respect to decomposition into s- and p-polarized incident waves, as discussed in Subsection 7.5.

Definition 8.7 (Fresnel Refraction Operator). Under identical context to the previous definition, the refracted RCSD is

$$\mathscr{T}_{\mathrm{Fresnel}}\{\boldsymbol{\mathcal{L}}\} \triangleq \begin{bmatrix} t_s & \\ & t_p \end{bmatrix} \cdot \boldsymbol{\mathcal{L}} \cdot \begin{bmatrix} t_s & \\ & t_p \end{bmatrix}^\dagger \tag{8.16}$$

where $t_{s,p}$ are the Fresnel refraction coefficients. Decomposition into s- and p-polarized waves is assumed as well.

Definition 8.8 (Church Reflection Operator). As before, the reflected RCSD under the Church reflection formulation is

$$\mathscr{R}_{\mathrm{Church}}\{\boldsymbol{\mathcal{L}}\} \triangleq \boldsymbol{Q} \, \boldsymbol{\mathcal{L}} \, \boldsymbol{Q}^\dagger \tag{8.17}$$

with $\boldsymbol{Q}$ being the Church reflection matrix, Eq. (7.40). Decomposition into s- and p-polarized waves is assumed.

9 APPLICATIONS

In this section we present one example application.

9.1 Rendering Diffraction on Reflection from a Surface

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

reflecting (or scattering) from the surface patch and arriving at an imaging device at position $\vec{\mathbf{e}}$.

Applying the SDTE (Eq. (8.5)), we are interested in the quantity

$$L(\vec{\mathbf{e}}) = \mathscr{L}\left\{\int_{\mathcal{S}^2} \mathrm{d}^2\Omega' \left|\Omega' \cdot \hat{\mathbf{n}}\right| \mathscr{D}_{\delta\vec{\mathbf{x}},\Omega'\to\Omega_o}\{\mathcal{S}_{\mathrm{inc}}\{\Lambda,\rho\}\}, \vec{\mathbf{e}}\right\} \tag{9.1}$$

where Ω_o is the direction of reflection of the surface patch and $\mathscr{D}$ is an abstract surface diffraction operator. Our focus is on deriving an analytic expression for $\mathscr{D}$ acting upon the sourced RCSD generated by $\mathcal{S}_{\mathrm{inc}}$. When the characteristic size of the surface patch is large compared to the height fluctuations described by the function h, the surface can be considered as an aperture. Similarly to the Harvey-Shack surface scatter theory [Krywonos 2006], we assume that the reflected field strength ratio is constant across the surface patch, hence we use the Church reflection matrix $\boldsymbol{Q}$ (Eq. (7.40)) and apply the reflection operator, Definition 8.8, to the incident sourced RCSD. Having $\boldsymbol{Q}$ be independent of the position on the surface patch simplifies the numeric calculations to complex scalars (instead of complex 2×2 matrices), but otherwise is done purely for simplicity and convenience.

Without loss of generality we center the surface patch at the origin, viz. $\vec{\mathbf{x}} = 0$. Then, $\vec{\mathbf{s}}$ is the vector from the patch to the source, $s = |\vec{\mathbf{s}}|$ is the mean distance from the surface patch to the source and $\hat{\mathbf{s}} = \vec{\mathbf{s}}/s$ is the direction to the source from the surface patch. And similar notation is employed for $\vec{\mathbf{e}}$. The sourced RCSD incident upon the surface patch is generated via Definition 8.1. We denote $\vec{\mathbf{x}}'$ and $\vec{\mathbf{x}}''$ as a pair of points on the surface and their projections onto the mean surface plane as $\vec{\mathbf{x}}'_\perp, \vec{\mathbf{x}}''_\perp$, likewise, we denote $\vec{\mathbf{s}}'$ as a point on the cross-section of the light source facing the surface patch. Making the typical far-field paraxial (small-angle) approximation, the distance between a point $\vec{\mathbf{x}}'$ on the surface to $\vec{\mathbf{s}}'$ on the source is

$$\left|\vec{\mathbf{x}}' - \vec{\mathbf{s}}'\right| = s + \frac{|\vec{\mathbf{x}}'| - 2\vec{\mathbf{s}}'\cdot\vec{\mathbf{x}}'}{2s'} + O\left(\frac{1}{s'^2}\right) \approx s' - \hat{\mathbf{s}}'\cdot\vec{\mathbf{x}}' \tag{9.2}$$

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

$$\left|\vec{\mathbf{x}}' - \vec{\mathbf{e}}\right| \approx e - \hat{\mathbf{e}}\cdot\vec{\mathbf{x}}' \tag{9.3}$$

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

$$\boldsymbol{\mathcal{L}}'\left(\vec{\mathbf{x}}'_\perp, \vec{\mathbf{x}}''_\perp, \omega\right) = \boldsymbol{Q}C e^{ik(\hat{\mathbf{s}}'+\hat{\mathbf{e}})\cdot(\vec{\mathbf{x}}'-\vec{\mathbf{x}}'')} \frac{\mathrm{J}_{\frac{3}{2}}\left[\rho k|\hat{\mathbf{x}}' - \hat{\mathbf{x}}''|\right]}{s'^2|\hat{\mathbf{x}}' - \hat{\mathbf{x}}''|^{\frac{3}{2}}} \boldsymbol{Q}^\dagger \tag{9.4}$$

where we applied Eqs. (9.2) and (9.3) and, for brevity, we denote C as the shorthand for the (position-independent) constants in Definition 8.1. The familiar wave propagation term, $\exp[ik(\hat{\mathbf{s}}' + \hat{\mathbf{e}})\cdot\vec{\mathbf{x}}']$, that appears in Eq. (9.4), can be regraded as the propagator that propagates the sourced RCSD to the "aperture"-plane (after reflection) taking the height fluctuations into account. We refer to it as the *frequency transmission function* [Born and Wolf 1999] that arises due to residual surface roughness. This expression can also be found in other computer graphics works that deal with rendering surface-induced diffractions [Holzschuch and Pacanowski 2017; Yan et al. 2018]. While the frequency transmission function describes the height fluctuations of the surface geometry, the scalar fraction that appears in Eq. (9.4) describes the coherence properties (second-order statistics) of the light incident upon the surface patch. We denote it as the *impulse response function*, $\mathcal{P}$, induced by the wave ensemble. Making the approximation

$$|\hat{\mathbf{x}}' - \hat{\mathbf{x}}''| \approx \frac{1}{s'}|\hat{\mathbf{s}}' \times (\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp)| \tag{9.5}$$

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

$$\mathcal{P}\left(\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp, \omega\right) \triangleq \frac{s^{\frac{3}{2}}\mathrm{J}_{\frac{3}{2}}\left[\frac{\rho k}{s'}\left|\hat{\mathbf{s}}' \times (\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp)\right|\right]}{\left|\hat{\mathbf{s}}' \times (\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp)\right|^{\frac{3}{2}}} \tag{9.6}$$

We now apply the diffraction operator, Definition 8.3, to Eq. (9.4) and substitute the resulting expression into Eq. (9.1) yields

$$L(\vec{\mathbf{e}}) = \frac{\Theta}{\sqrt{\rho}} \frac{1}{A_\Sigma|\hat{\mathbf{e}}\cdot\hat{\mathbf{n}}|} \int_0^\infty \frac{\mathrm{d}\omega}{\lambda^2} \operatorname{tr}\left[\boldsymbol{H}_1 \int_{\mathcal{S}^2_+} \mathrm{d}^2\hat{\mathbf{s}}'\, \hat{\mathbf{s}}'\cdot\hat{\mathbf{n}}\, \mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\}\boldsymbol{H}_2^T\right] \tag{9.7}$$

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

$$\Theta \triangleq 2\pi\left(1 - \sqrt{1 - \frac{\rho^2}{s^2}}\right) \tag{9.8}$$

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

$$\begin{aligned}\mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\} &= \int_\Sigma \mathrm{d}^2\vec{\mathbf{x}}'_\perp \int_\Sigma \mathrm{d}^2\vec{\mathbf{x}}''_\perp\, \boldsymbol{\mathcal{L}}'\left(\vec{\mathbf{x}}', \vec{\mathbf{x}}'', \omega\right) e^{-ik\left(\hat{\mathbf{e}}\cdot\vec{\mathbf{x}}'_\perp - \hat{\mathbf{e}}\cdot\vec{\mathbf{x}}''_\perp\right)} \\ &= \boldsymbol{Q}C\boldsymbol{Q}^\dagger \int_\Sigma \mathrm{d}^2\vec{\mathbf{x}}'_\perp\, e^{-ik\hat{\mathbf{e}}\cdot\vec{\mathbf{x}}'_\perp} \\ &\quad \times \left[\mathscr{F}_\Sigma\left\{e^{-ik(\hat{\mathbf{s}}'+\hat{\mathbf{e}})\cdot(\vec{\mathbf{x}}'-\vec{\mathbf{x}}'')}\right\} * \mathscr{F}_\Sigma\left\{\mathcal{P}\left(\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp, \omega\right)\right\}\right]\end{aligned} \tag{9.9}$$

where we applied the convolution theorem (Theorem 2.2). The $*$ denotes the two-dimensional convolution over the mean surface plane. Both the Fourier transforms in the expression above are over Σ and with respect to the double-primed integration variable $\vec{\mathbf{x}}''_\perp$. The convolution is evaluated at frequency $(-k\hat{\mathbf{e}})$. Note that by the Fourier transform shift identity and the variable change $\vec{\mathbf{y}}'_\perp = \vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp$ the following holds

$$\mathscr{F}\left\{\mathcal{P}(\vec{\mathbf{x}}'_\perp - \vec{\mathbf{x}}''_\perp, \omega)\right\}\left(\vec{\xi}\right) = -e^{-i\vec{\xi}\cdot\vec{\mathbf{x}}'_\perp}\, \mathscr{F}\left\{\mathcal{P}(\vec{\mathbf{y}}'_\perp, \omega)\right\}\left(-\vec{\xi}\right) \tag{9.10}$$

making the Fourier transform of the impulse response function independent of $\vec{\mathbf{x}}'_\perp$. Then, formally interchanging the orders of convolution and Fourier transform in Eq. (9.9) leads to

$$\begin{aligned}\mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\} = -\boldsymbol{Q}C\boldsymbol{Q}^\dagger \int_{\mathbb{R}^2} \mathrm{d}^2\vec{\xi}\, \mathscr{F}\left\{\mathcal{P}(\vec{\mathbf{x}}''_\perp, \omega)\right\}\left(-\vec{\xi}\right) \int_\Sigma \mathrm{d}^2\vec{\mathbf{x}}'_\perp\, e^{-i\vec{\xi}\cdot\vec{\mathbf{x}}'_\perp} \\ \times e^{-ik\hat{\mathbf{e}}\cdot\vec{\mathbf{x}}'_\perp} \int_\Sigma \mathrm{d}^2\vec{\mathbf{x}}''_\perp\, e^{i\left(k\hat{\mathbf{e}}+\vec{\xi}\right)\cdot\vec{\mathbf{x}}''_\perp} e^{-ik(\hat{\mathbf{s}}'+\hat{\mathbf{e}})\cdot(\vec{\mathbf{x}}'-\vec{\mathbf{x}}'')}\end{aligned} \tag{9.11}$$

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

$$\mathscr{F}^2_\Sigma\{\boldsymbol{\mathcal{L}}'\} = \boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q}^\dagger\left[\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\}\right](k\hat{\mathbf{e}}) \tag{9.12}$$

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

$$\mathcal{T}\left(\vec{\xi}\right) \triangleq \mathscr{F}_\Sigma\left\{e^{ik(\hat{\mathbf{s}}'+\hat{\mathbf{e}})\cdot\left[\vec{\mathbf{x}}'_\perp + \hat{\mathbf{n}}h(\vec{\mathbf{x}}'_\perp)\right]}\right\}(\vec{\xi}) \tag{9.13}$$

(where we write $\vec{\mathbf{x}}' = \vec{\mathbf{x}}'_\perp + \hat{\mathbf{n}}h(\vec{\mathbf{x}}'_\perp)$ to make the height function explicit) and $\mathcal{T}^2(\vec{\xi}) = \mathcal{T}(\vec{\xi})\mathcal{T}(\vec{\xi})^\star$ is the *frequency response function*. Eq. (9.12) is simply the two-dimensional convolution of the Fourier transform of the impulse response function with the frequency response function. Note that the impulse response function $\mathcal{P}$ is a symmetric real function, therefore its Fourier transform is real. Likewise, the frequency response function $\mathcal{T}^2$ is evidently real. Therefore, the double Fourier transform expression above is real as well, as expected seeing as we are computing the observed radiance.

Finally, substituting Eq. (9.12) into Eq. (9.7) and applying the paraxial approximation by settings $\vec{\mathbf{s}}' = \vec{\mathbf{s}}$, i.e. far-field with $s \gg \rho$, yields the simplified expression for the observed radiance of partially-coherent light reflected from an explicitly described surface:

$$L(\vec{\mathbf{e}}) = \frac{\Theta}{\sqrt{\rho}}\frac{1}{A_\Sigma|\hat{\mathbf{e}}\cdot\hat{\mathbf{n}}|}\int_0^\infty \frac{\mathrm{d}\omega}{\lambda^2}\Lambda(\omega) \times \operatorname{tr}\left[\boldsymbol{H}_1\boldsymbol{Q}(\boldsymbol{H}_2\boldsymbol{Q})^\dagger\right]\left(\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\}\right)(k\hat{\mathbf{e}}) \tag{9.14}$$

Numeric evaluation. The evaluation of Eq. (9.14) will typically be done numerically. To that end it is more suitable to rewrite the convolution as multiplication in Fourier space, viz.

$$\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} = \mathscr{F}\left\{\mathscr{F}^{-1}\{\mathcal{T}^2\}\cdot\mathcal{P}\right\} \tag{9.15}$$

Then, $\mathscr{F}^{-1}\{\mathcal{T}^2\}$ is evaluated via a pair of fast Fourier transforms (FFT). As we only need a single frequency of the last transform, the Goertzel algorithm (generalized to arbitrary frequencies [Sysel and Rajmic 2012]) is used as it is faster, compared to an FFT, and provides a sinc interpolated result. Overall, this implies a time-complexity of $O(n \log n)$.

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

Observe that for at the perfectly coherent limit, the impulse response function becomes (see Eq. (7.12)) a constant, viz. $\mathcal{P}(\vec{\mathbf{y}}, \omega) \propto k^2$. Then, its Fourier transform is $\mathscr{F}\{\mathcal{P}\} \propto 2\pi k^2\delta(\vec{\xi})$, i.e. proportional to a Dirac delta. Using the fact that the Dirac delta is the identity under convolution algebra, we get $\mathcal{T}^2 * \mathscr{F}\{\mathcal{P}\} \propto 2\pi k^2\mathcal{T}^2$. Hence, the observed intensity is proportional to the complex magnitude squared (i.e. power) of the Fourier transform of the amplitude distribution over the surface, as described by Eq. (6.26). That is, the problem reduces to a simple diffraction problem where we ignore coherence, as expected.

In this section we have derived an expression for the RCSD of light reflected of a surface under the context of a vectorized diffraction theory. Our primary result is showing that the reflected RCSD can be formulated as the convolution between a function describing the coherence properties of the incident wave ensemble and a function describing the surface-induced perturbations: That is, a convolution between the Fourier transform of the impulse response function $\mathcal{P}$, which is independent of the surface properties, and the frequency response function $\mathcal{T}^2$, which is the complex magnitude squared of the Fourier transforms of the surface's frequency transmission function and is independent of the properties of the incident wave ensemble. Some parallels can be drawn between this conclusion and known results in optics regarding transmission of mutual intensities through an optical system [Born and Wolf 1999]. See Figs. 5 to 8 for rendered BRDFs visualizing the scattering of some specific surfaces. Likewise, see Fig. 9 for a comparison with Yan et al. [2018].

Similar analysis can be performed to derive expressions for, e.g., the scattering of participating media, and other problems that can be formulated as a diffraction problem.

10 VALIDATION

To validate our formalism we compare solutions to the surface scattering problem (Subsection 9.1) rendered using Eq. (9.14) against a ground truth. To produce a ground truth we consider again our formal model of an (incoherent) light source that was used to express and derive the sourced cross-spectral density (Subsection 7.1): A large collection of (temporally invariant) discrete elementary radiators, each modelled as a point light source emitting polychromatic electromagnetic radiation with constant spectrum and strength and random polarization and phase. As discussed, this is a very good model for natural spontaneous emission light sources when the emitted radiation is observed over long periods. Because each elementary radiator is a point source, it produced perfectly spatially coherent radiation and we use the Smythe's diffraction formula (Eq. (6.26)) to evaluate the surface scattered radiation for a single elementary radiator.

Formally, given an elementary radiator positioned at $\vec{\mathbf{s}} \in \mathbb{R}^3$, the transverse components of its emitted radiation fields, propagated to a distance r from the source and at time t, can be written as

$${}^{\mathbf{s}}\vec{\mathbf{E}}(r,\omega) = \begin{bmatrix}{}^{\mathbf{s}}E_1(r,\omega)\\ {}^{\mathbf{s}}E_2(r,\omega)\end{bmatrix} = \sqrt{\Lambda(\omega)}\frac{e^{i(kr-\omega t+\varphi_{\mathbf{s}})}}{r}\begin{bmatrix}\cos\theta_{\mathbf{s}}\\ \sin\theta_{\mathbf{s}}\end{bmatrix} \tag{10.1}$$

where $\varphi_{\mathbf{s}}$ and $\theta_{\mathbf{s}}$ are the (mutually-independent) random initial phase and polarization rotation angle and Λ is the emitted power spectral density of a single elementary radiator. ${}^{\mathbf{s}}\vec{\mathbf{E}}(r,\omega)$ can be thought of as a realization of the wave ensemble comprising of contributions from all the elementary radiators, hence the notation. The surface scattered fields that arise due to radiation from the elementary radiator $\vec{\mathbf{s}}$ are denoted ${}^{\mathbf{s}}\vec{\mathbf{E}}'_{s,p}(\vec{\mathbf{e}},\omega)$, for the s- and p-polarized fields respectively, and those follow immediately from our surface reflection model (formalised via the Church polarization factors, Eqs. (7.36) to (7.39)):

$${}^{\mathbf{s}}E'_\zeta(\vec{\mathbf{r}},\omega) = \hat{\mathbf{e}}'_\zeta\cdot\mathscr{D}_{\vec{\mathbf{r}}}\left\{{}^{\mathbf{s}}E_1\cdot q_{s\zeta}\hat{\mathbf{e}}_s + {}^{\mathbf{s}}E_2\cdot q_{p\zeta}\hat{\mathbf{e}}_p\right\} \tag{10.2}$$

where $\zeta \in \{s,p\}$ indexes the transverse field component, $\{\hat{\mathbf{e}}_s, \hat{\mathbf{e}}_p\}$ and $\{\hat{\mathbf{e}}'_s, \hat{\mathbf{e}}'_p\}$ are the orthonormal bases for the incident and scattered wave's s- and p-polarization decomposition, respectively, q_{ss}, q_{sp},

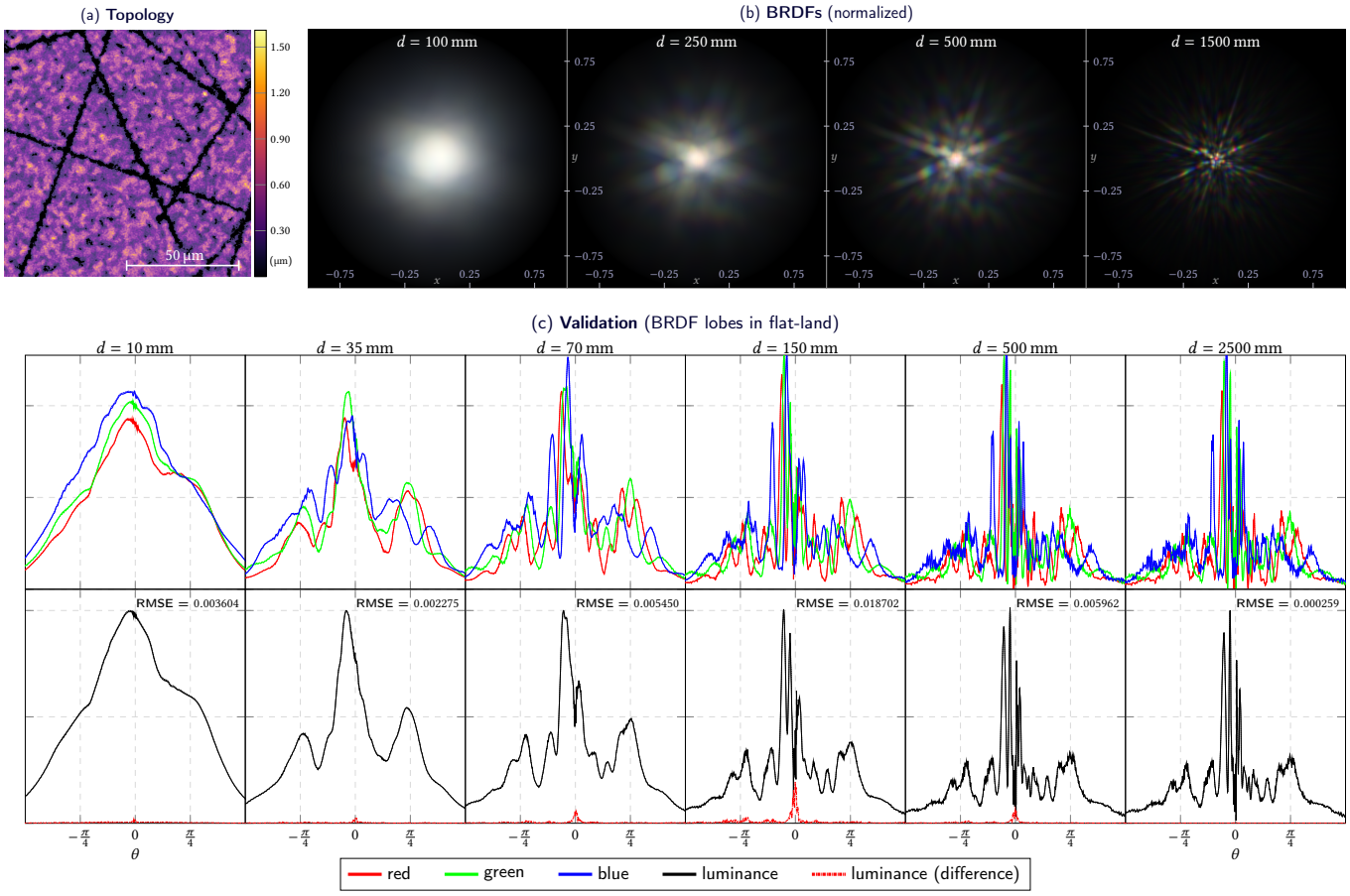

Fig. 5. (b) BRDF visualizations (normalized intensity) of the diffraction effects that arise when natural, partially-coherent light scatters of a surface were rendered using our method. A uniformly scratched aluminum surface (topography visualized in figure (a)) is illuminated by a D65 light source of radius 8 mm that is positioned directly above the surface. By varying the distance of the source from the surface we change the coherence properties of the incident wave ensemble, resulting in starkly different scattering behaviour. (c) Numeric validation of our method (solid plots) against the ground truth (dashed plots) in flat-land for the same scratched surface. The radius of the light source is 1 mm and the ground truth is generated using 49 087 elementary radiators. We compare (first row) each primary colour component as well as the (second row) total luminance. Absolute luminance difference is plotted as well (second row, dash-dotted red plot).

q_{ps} and q_{pp} are the Church reflection (polarization) factors and $\mathscr{D}$ is the diffraction operator, which is simply defined using the Smythe's diffraction formula:

$$\begin{aligned}\mathscr{D}_{\vec{\mathbf{r}}}\left\{\vec{\mathbf{E}}\right\} &\triangleq ik\frac{e^{ikr}}{2\pi r}\hat{\mathbf{r}}\times\int_{\Sigma}\mathrm{d}^2\vec{\mathbf{x}}'_{\perp}\left[\hat{\mathbf{n}}\times\vec{\mathbf{E}}(x'_{\perp},\omega)\right]e^{-ik\hat{\mathbf{r}}\cdot\vec{\mathbf{x}}'_{\perp}}\\ &= ik\frac{e^{ikr}}{2\pi r}\hat{\mathbf{r}}\times\mathscr{F}_{\Sigma}\left\{\hat{\mathbf{n}}\times\vec{\mathbf{E}}(x'_{\perp},\omega)\right\}(k\hat{\mathbf{r}})\end{aligned}\tag{10.3}$$

where we reuse the same notation as in Subsection 9.1: Σ is the mean surface plane that subtends the actual surface with $\vec{\mathbf{x}}'_{\perp}\in\Sigma$ being points on that plane and corresponding points on the surface are $\vec{\mathbf{x}}'=\vec{\mathbf{x}}'_{\perp}+\hat{\mathbf{n}}h(\vec{\mathbf{x}}'_{\perp})$. Finally, the intensity, observed by an observer positioned at $\vec{\mathbf{r}}$, is the time-averaged Poynting vector (Eq. (4.15)) of the super position of scattered contribution from all elementary radiators, viz.

$$I(\vec{\mathbf{r}},\omega)=\left\langle\left|\sum_{\mathbf{s}}{}^{\mathbf{s}}E'_{s}(\vec{\mathbf{r}},\omega)\right|^2+\left|\sum_{\mathbf{s}}{}^{\mathbf{s}}E'_{p}(\vec{\mathbf{r}},\omega)\right|^2\right\rangle_{\mathbf{t}}\tag{10.4}$$

The radiators are correlated only over a very short spatial distance. As discussed, this correlation distance serves to only scale the intensity of the produced radiation by a constant. Therefore, we can simply assume that this spatial distance is small and all elementary radiators are uncorrelated. This greatly simplifies the numerical calculations, as contributions can now be added on an incoherent basis:

$$\left\langle\left|\sum_{\mathbf{s}}{}^{\mathbf{s}}E'_{\zeta}(\vec{\mathbf{r}},\omega)\right|^2\right\rangle_{\mathbf{t}}=\sum_{\mathbf{s}}\left\langle\left|{}^{\mathbf{s}}E'_{\zeta}(\vec{\mathbf{r}},\omega)\right|^2\right\rangle_{\mathbf{t}}\tag{10.5}$$

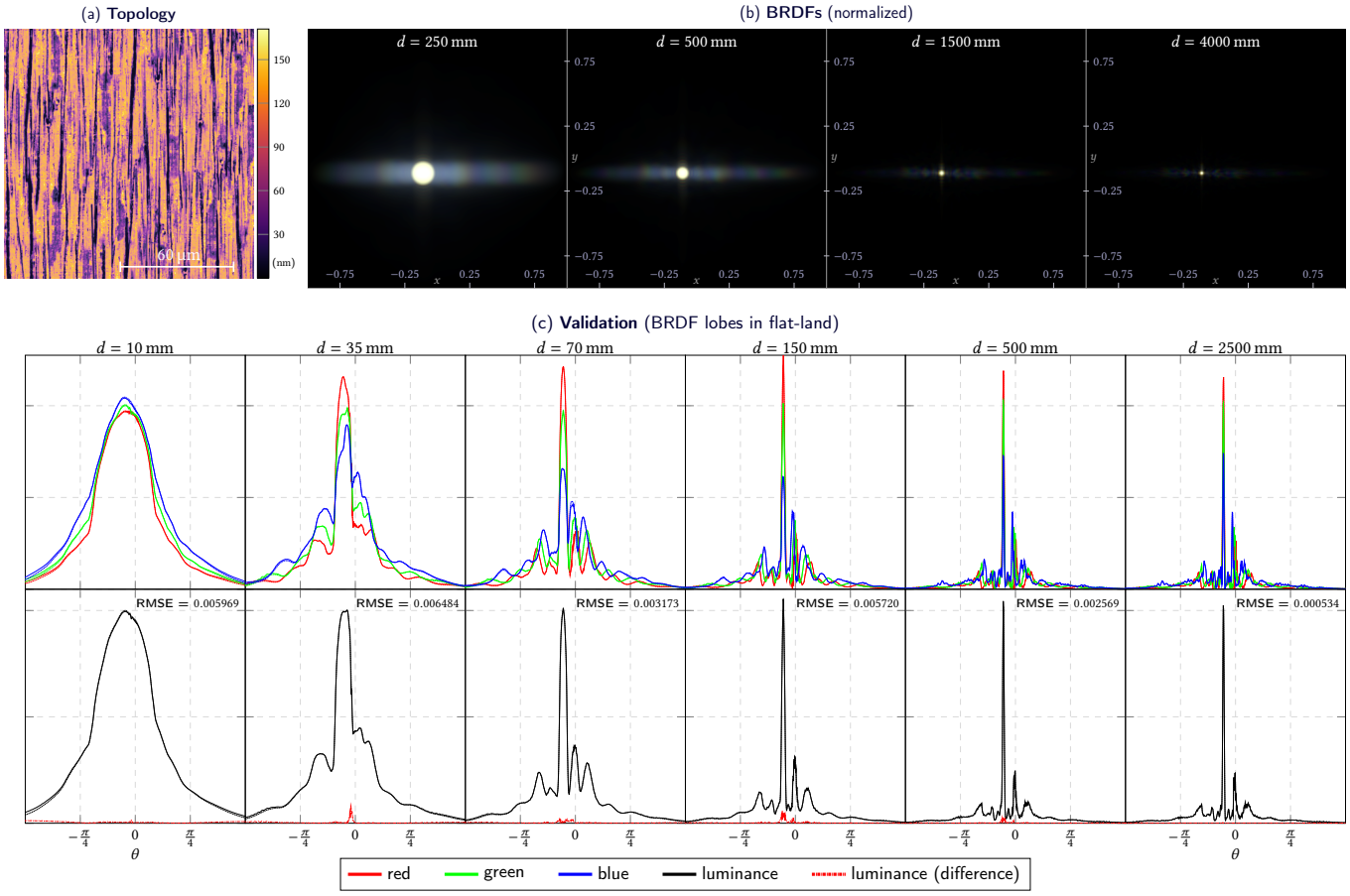

Fig. 6. (a) An aluminum surface that has undergone electric polish resulting in low intrinsic roughness and highly anisotropic details. (b) BRDF visualizations (normalized intensity) under partially-coherent illumination produced by a fluorescent F1 model light source. The anisotropic surface features give rise to a strongly directional scattering behaviour. (c) Validation in flat-land. For validation the radius of the light source is set to 1 mm and the ground truth is generated using 49 087 elementary radiators.

Moreover, with some algebra the time-averages above can be written as

$$\left\langle \left| {}^{\mathbf{s}}E'_{\zeta}(\vec{\mathbf{r}},\omega) \right|^2 \right\rangle_{\mathrm{t}} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{d}t \left| {}^{\mathbf{s}}E'_{\zeta}(\vec{\mathbf{r}},\omega) \right|^2$$
$$= \frac{k^2\Lambda(\omega)}{8\pi^2 r^2} \left[\left| q_{s\zeta} \hat{\mathbf{e}}'_{\zeta} \cdot (\hat{\mathbf{r}} \times (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_s)) \right|^2 + \left| q_{p\zeta} \hat{\mathbf{e}}'_{\zeta} \cdot (\hat{\mathbf{r}} \times (\hat{\mathbf{n}} \times \hat{\mathbf{e}}_p)) \right|^2 \right]$$
$$\times \left| \mathscr{F}_{\Sigma} \left\{ \frac{1}{|\vec{\mathbf{s}} - \vec{\mathbf{x}}'|} e^{ik[\hat{\mathbf{s}}(\vec{\mathbf{x}}') + \hat{\mathbf{r}}(\vec{\mathbf{x}}')] \cdot \vec{\mathbf{x}}'} \right\} (k\hat{\mathbf{r}}) \right|^2 \qquad (10.6)$$

(this analysis uses the fact that $\varphi_{\mathbf{s}}$ and $\theta_{\mathbf{s}}$ are statistically independent to carry out the integration separately) where the unit vectors $\hat{\mathbf{s}}$ and $\hat{\mathbf{r}}$ above are the unit vectors in direction of the elementary radiator and observer, respectively, at the integrated surface position, i.e.

$$\hat{\mathbf{s}}(\vec{\mathbf{x}}') \triangleq \frac{\vec{\mathbf{s}} - \vec{\mathbf{x}}'}{|\vec{\mathbf{s}} - \vec{\mathbf{x}}'|} \qquad \hat{\mathbf{r}}(\vec{\mathbf{x}}') \triangleq \frac{\vec{\mathbf{r}} - \vec{\mathbf{x}}'}{|\vec{\mathbf{r}} - \vec{\mathbf{x}}'|} \qquad (10.7)$$

Substituting Eqs. (10.5) and (10.6) into Eq. (10.4) and integrating over the spectrum yields the final formula for the observed intensity of light scattered of a surface:

$$I(\vec{\mathbf{r}}) = \int_0^{\infty} \frac{\mathrm{d}\omega\, k^2\Lambda(\omega)}{8\pi^2 r^2} \sum_{\mathbf{s}} \left| \mathscr{F}_{\Sigma} \left\{ \frac{1}{|\vec{\mathbf{s}} - \vec{\mathbf{x}}'|} e^{ik[\hat{\mathbf{s}}(\vec{\mathbf{x}}') + \hat{\mathbf{r}}(\vec{\mathbf{x}}')] \cdot \vec{\mathbf{x}}'} \right\} (k\hat{\mathbf{r}}) \right|^2$$
$$\times \sum_{\xi,\zeta \in \{s,p\}} \left| q_{\xi\zeta} \hat{\mathbf{e}}'_{\zeta} \cdot \left(\hat{\mathbf{r}} \times \left(\hat{\mathbf{n}} \times \hat{\mathbf{e}}_{\xi} \right) \right) \right|^2 \qquad (10.8)$$

The Fourier transform that appears above is essentially the same frequency transmission function that was encountered in Subsection 9.1, and the complex magnitude of this Fourier transform is then the frequency response function $\mathcal{T}^2$. This is altogether expected as the scatter formalism of our surface is unchanged. Note that the derivations in this section ignore the formulae that were developed in Sections 7 and 8 and are devoid of any optical coherence formalism. Indeed, we derive Eq. (10.8) simply by modelling the spontaneous emission light source as a large collection of independent elementary radiators, each a point light source giving rise to (spatially) coherent radiation, and apply well-founded results in optics: Namely the Smythe diffraction formulae which are well studied

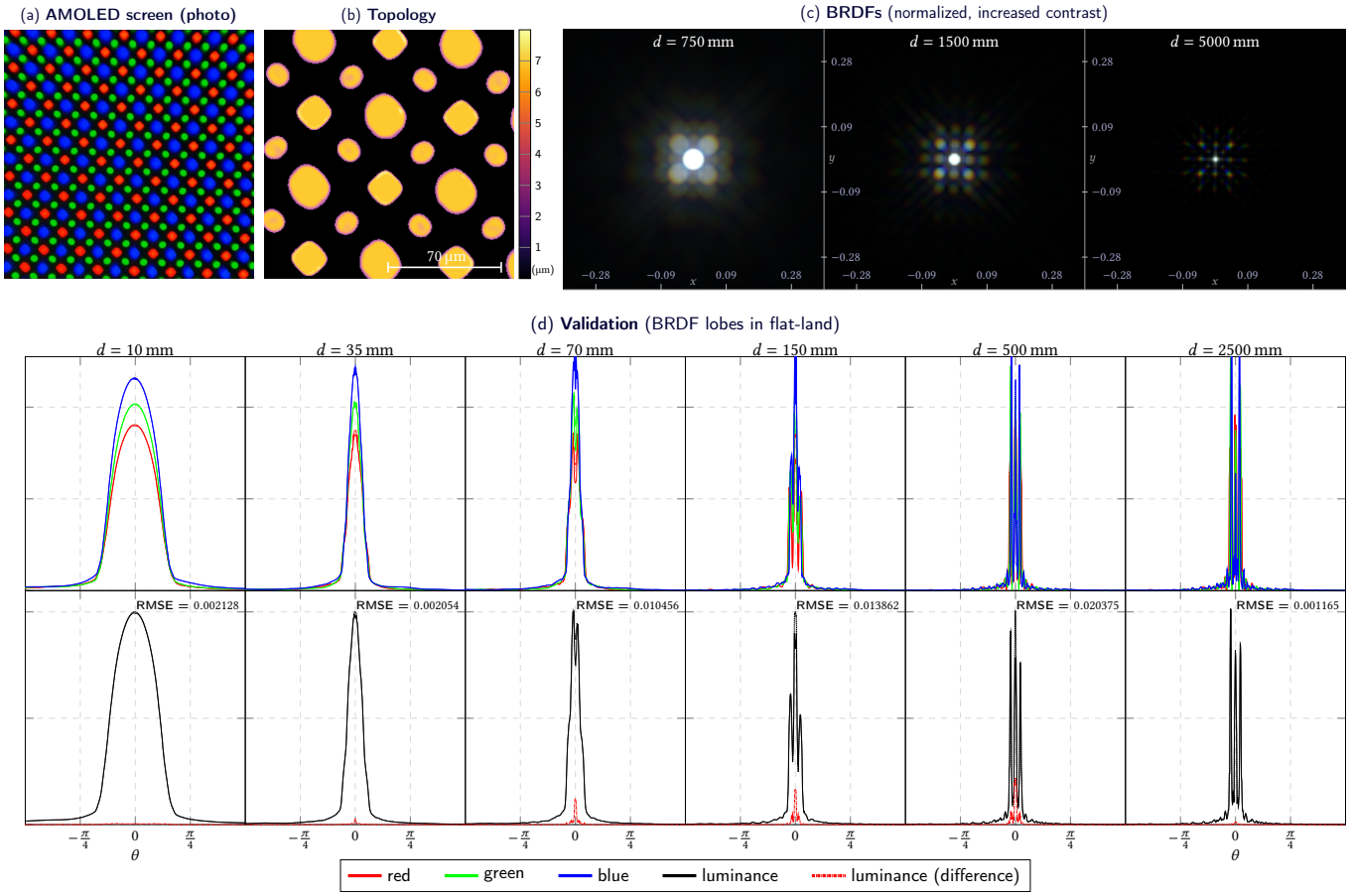

Fig. 7. (a) A microscopy photograph of an AMOLED screen with a diamond pitch pixel arrangement (used in some hand-held devices), with its (b) surface topology visualized. That topology admits an interesting 8-way symmetry in the induced diffraction patterns, which can be seen in the (c) rendered BRDFs (zoomed-in and with artificially increased contrast for visualization purposes). We used an F1 fluorescent light source with a radius of 8 mm positioned directly above the surface for the rendered BRDFs. The distances between the constructive-interference peaks is consistent with Bragg's diffraction law. (d) Validation in flat-land. For validation the radius of the light source is set to 1 mm and the ground truth is generated using 49 087 elementary radiators.

and known to be produce accurate solutions to Maxwell's equations in the Fraunhofer region [Zangwill 2013]. Assigning random phases to each elementary radiator can be thought of as applying a random phase screen to a coherent source [Xiao and Voelz 2006].

Finally, also observe that as the unit vectors $\hat{\mathbf{s}}(\vec{\mathbf{x}}')$ and $\hat{\mathbf{r}}(\vec{\mathbf{x}}')$ depend both on the surface position, i.e. the variable $\vec{\mathbf{x}}'$, as well as the position of the elementary radiator $\vec{\mathbf{s}}$. This greatly relaxes the paraxial approximations that are made in the derivation of the Smythe diffraction formulae (see Subsection 6.2), implying that Eq. (10.8) describes an accurate solution to Maxwell's equations when the distances to the source and observer are large compared to wavelength only, i.e. $r, s \gg \lambda$.

Results. We compare renderings of surface-induced diffraction effects achieved via Eq. (9.14) against the ground truth computed using Eq. (10.8). Because a high count of elementary radiators is required, numerical evaluation of the ground truth is only feasible in two-dimensional space. We use a light source with a radius of 1 mm and with elementary radiators spaced at 8 µm intervals, implying roughly 50 000 elementary radiators. Note that in $\mathbb{R}^2$ the derivations in Subsection 9.1 change very slightly: The source's geometry becomes a disk instead of a ball and thus its Fourier transform results in a Bessel function of a different order. The rest of the analysis remains the same.

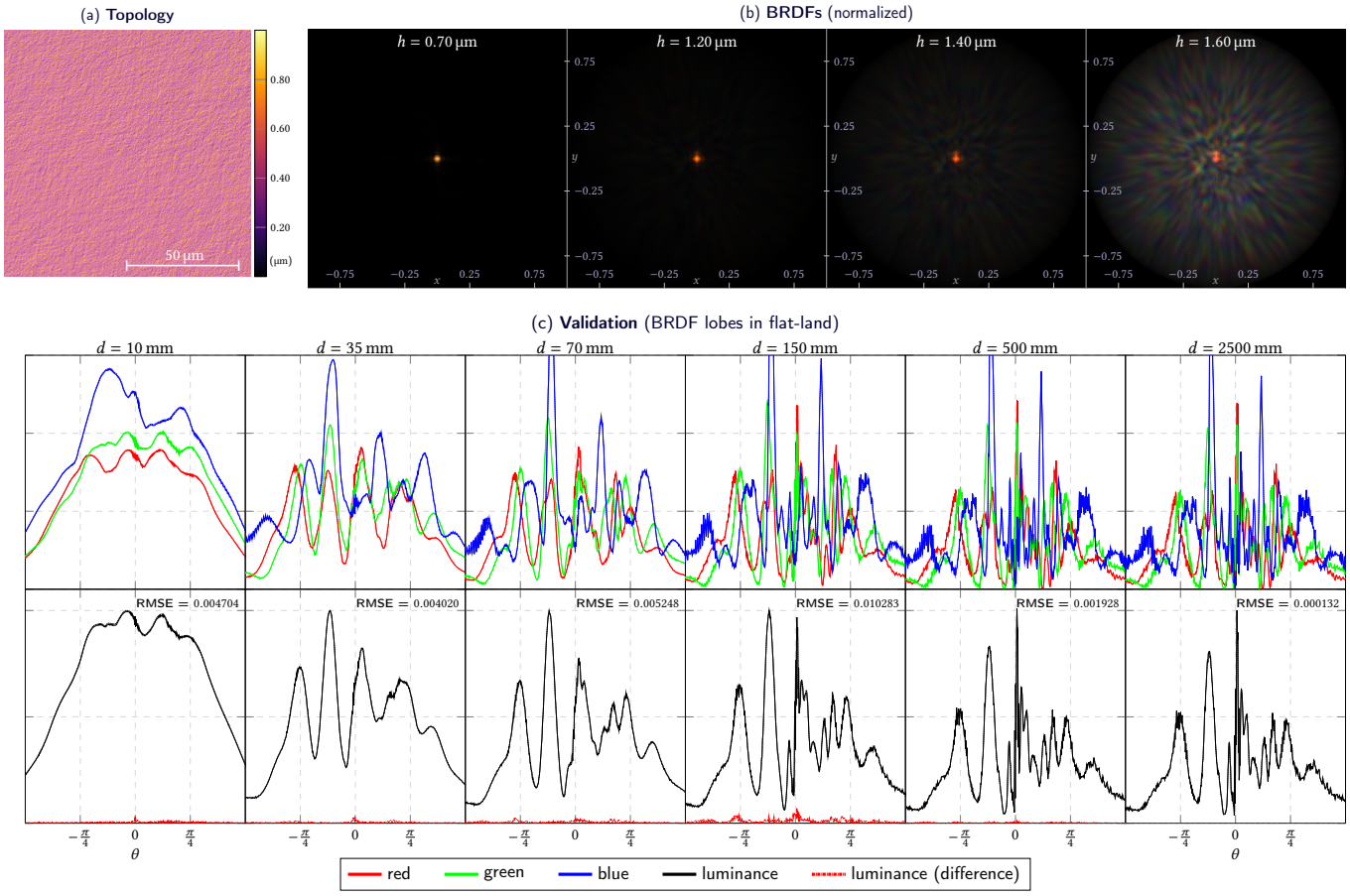

Fig. 8. (a) A sandpaper polished gold surface with isotropic statistics. (b) BRDF visualizations (normalized intensity) generated when the surface is illuminated by a D65 light source. We vary the height-scale of the surface, h, to compare the BRDFs between different levels of roughness: For a very smooth surface (on the left), the only visible details are the reflection of the light source, however, as we increase the roughness (left to right), surface diffractions effects appear isotropically. (c) Validation in flat-land. For validation the radius of the light source is set to 1 mm and the ground truth is generated using 49 087 elementary radiators.

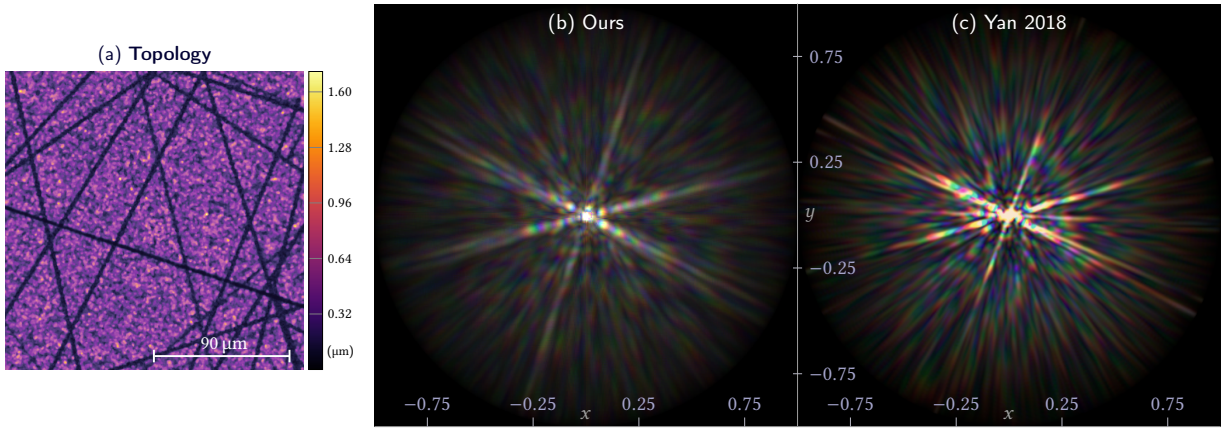

Fig. 9. Comparison of (b) our method against (c) Yan et al. [2018]. Both methods were employed on the (a) same surface topology. We used the implementation provided by Yan et al. [2018] to render the BRDF produced by their method. Some differences in colour temperature and white point are expected, however the produced patterns are similar. Both BRDFs were rendered with 32 spectral samples. The methods admit comparable rendering times.

Appendix

A FOURIER TRANSFORMS OF SPECIAL GEOMETRIES

A.1 Fourier Transform of a Ball

Let B^d be the unit ball in d-dimensions, i.e. $B^d = \{\vec{\mathbf{r}} \in \mathbb{R}^d : |\vec{\mathbf{r}}| \leq 1\}$. The Fourier transform of the ball is related to the Bessel function of the first kind of half-integer orders via [Gelfand 1964]

$$\mathscr{F}\left\{\mathbb{1}_{B^d}\right\}\left(\vec{\xi}\right) = \int_{B^d} \mathrm{d}^d\vec{\mathbf{r}}'\, e^{-i\vec{\mathbf{r}}'\cdot\vec{\xi}} = \frac{(2\pi)^\nu}{|\xi|^\nu}\mathrm{J}_\nu(|\xi|) \tag{A.1}$$

with $\nu = \frac{d}{2}$ and $\mathbb{1}_{B^d}$ the characteristic function of the ball. And in general for a ball of radius ρ:

$$\mathscr{F}\left\{\mathbb{1}_{\rho B^d}\right\}\left(\vec{\xi}\right) = \frac{(2\pi\rho)^\nu}{|\xi|^\nu}\mathrm{J}_\nu(\rho|\xi|) \tag{A.2}$$

A.2 Fourier Transform of a Cylinder

A cylinder in $\mathbb{R}^3$ centred around the origin with height h and radius ρ can be represented as

$$C(h,\rho) \triangleq \left\{(x,y,z) \in \mathbb{R}^3 : \sqrt{x^2+y^2} \leq \rho \wedge |z| \leq \frac{h}{2}\right\} \tag{A.3}$$

Let $\mathbb{1}_{C(h,\rho)}$ be its characteristic function and $\vec{\xi} = (x,y,z) \in \mathbb{R}^3$ an arbitrary frequency. Denote $\vec{\xi}_\perp = (x,y)$, the projection of $\vec{\xi}$ onto the xy-plane. Then, the Fourier transform of the 3-dimensional cylinder is evidently related to the transform of a 2-dimensional disk (with frequency $\vec{\xi}_\perp$), viz.

$$\begin{aligned}\mathscr{F}\left\{\mathbb{1}_{C(h,\rho)}\right\}\left(\vec{\xi}\right) &= \int_{C(h,\rho)} \mathrm{d}^3\vec{\mathbf{r}}'\, e^{-i\vec{\mathbf{r}}'\cdot\vec{\xi}} \\ &= \mathscr{F}\left\{\mathbb{1}_{\rho B^2}\right\}(\xi_\perp)\int_{-\frac{h}{2}}^{\frac{h}{2}} \mathrm{d}z'\, e^{-izz'} \\ &= \frac{2\pi h\rho}{|\xi_\perp|}\,\mathrm{sinc}\left(\frac{hz}{2}\right)\mathrm{J}_1(\rho|\xi_\perp|)\end{aligned} \tag{A.4}$$

A.3 Fourier Transform of a Polytope

The Fourier transform of a polygon in two-dimensional space was discussed by Shung-Wu Lee and Mittra [1983]. We present a slightly modified derivation that is applicable to higher-dimensions.

Let Σ be a d-dimensional simple (non self-intersecting) polytope and $\mathbb{1}_\Sigma$ its characteristic function. The Fourier transform is then

$$\begin{aligned}\mathscr{F}\{\mathbb{1}_\Sigma\}\left(\vec{\xi}\right) &= \int_\Sigma \mathrm{d}^d\vec{\mathbf{r}}'\, e^{-i\vec{\mathbf{r}}'\cdot\vec{\xi}} = -\frac{1}{|\xi|^2}\int_\Sigma \mathrm{d}^d\vec{\mathbf{r}}'\,\nabla^2 e^{-i\vec{\mathbf{r}}'\cdot\vec{\xi}} \\ &= -\frac{1}{|\xi|^2}\oint_{\partial\Sigma} \mathrm{d}^{d-1}\vec{\mathbf{s}}'\cdot\boldsymbol{\nabla} e^{-i\vec{\mathbf{s}}'\cdot\vec{\xi}} \\ &= \frac{i}{|\xi|^2}\oint_{\partial\Sigma} \mathrm{d}^{d-1}\vec{\mathbf{s}}'\left(\vec{\xi}\cdot\hat{\mathbf{n}}\right)e^{-i\vec{\mathbf{s}}'\cdot\vec{\xi}}\end{aligned} \tag{A.5}$$

where we used Green's first identity to go from the volume integral to the integral over the boundary of the polytope. $\hat{\mathbf{n}}$ is the (outwards) normal to the boundary at point $\vec{\mathbf{s}}$. Green's first identity can be repeatedly applied a further $d-2$ times until the boundary is a finite set of one dimensional line segments.

Given $d = 2$, the polygon consists of an ordered N-tuple of vertices denoted $(\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_N)$. We define $l_j = |\vec{\mathbf{x}}_{j+1} - \vec{\mathbf{x}}_j|$, the length of an edge; and $\hat{\mathbf{d}}_j = (\vec{\mathbf{x}}_{j+1} - \vec{\mathbf{x}}_j)/l_j$, the direction of an edge. Our indexing convention is modulo N, that is $\vec{\mathbf{x}}_{N+1} \equiv \vec{\mathbf{x}}_1$. Then, the last integral in Eq. (A.5) is a sum of line integrals:

$$\frac{i}{|\xi|^2}\oint_{\partial\Sigma} \mathrm{d}\vec{\mathbf{s}}'\, e^{-i\vec{\mathbf{s}}'\cdot\vec{\xi}}\left(\vec{\xi}\cdot\hat{\mathbf{n}}\right) = \frac{i}{|\xi|^2}\sum_{j=1}^{N}\left(\vec{\xi}\cdot\hat{\mathbf{n}}_j\right)\int_0^{l_j}\mathrm{d}\alpha\, e^{-i\left(\vec{\mathbf{x}}_j+\alpha\hat{\mathbf{d}}_j\right)\cdot\vec{\xi}} \tag{A.6}$$

where the integral admits a closed-form solution. Applying a few elementary simplifications we can conclude

$$\mathscr{F}\{\mathbb{1}_\Sigma\}\left(\vec{\xi}\right) = \frac{1}{|\xi|^2}\sum_{j=1}^{N}\frac{\vec{\xi}\cdot\hat{\mathbf{n}}_j}{\vec{\xi}\cdot\hat{\mathbf{d}}_j}\left[e^{-i\vec{\mathbf{x}}_j\cdot\vec{\xi}} - e^{-i\vec{\mathbf{x}}_{j+1}\cdot\vec{\xi}}\right] \tag{A.7}$$

for a two-dimensional polygon.

Transform of a rectangular aperture. The Fourier transform of an axis-aligned rectangular aperture with edge lengths a and b is a special case of the polygon transform. In this scenario $N = 4$; $\vec{\mathbf{x}}_1 = -\vec{\mathbf{x}}_3$ and $\vec{\mathbf{x}}_2 = -\vec{\mathbf{x}}_4$; $l_1 = l_3 = a$ and $l_2 = l_4 = b$; $\hat{\mathbf{d}}_1 = -\hat{\mathbf{d}}_3 = \hat{\mathbf{n}}_2 = -\hat{\mathbf{n}}_4 = +\hat{\mathbf{y}}$ and $\hat{\mathbf{d}}_2 = -\hat{\mathbf{d}}_4 = \hat{\mathbf{n}}_3 = -\hat{\mathbf{n}}_1 = +\hat{\mathbf{x}}$. Denote $\xi\cdot\hat{\mathbf{x}} = \cos\theta$, then, after a few simplifications the Fourier transform of the polygon (Eq. (A.7)) becomes

$$\begin{aligned}&\mathscr{F}\left\{\mathbb{1}_{\mathrm{rect}_{a,b}}\right\}\left(\vec{\xi}\right) \\ &= \frac{2}{|\xi|^2}\left\{\left[\cos\left(\xi_x\frac{b}{2} - \xi_y\frac{a}{2}\right) - \cos\left(\xi_x\frac{b}{2} + \xi_y\frac{a}{2}\right)\right](\tan\theta + \cot\theta)\right\} \\ &= \frac{4}{\xi_x\xi_y}\sin\left(\frac{b\xi_x}{2}\right)\sin\left(\frac{a\xi_y}{2}\right) = 2ab\,\mathrm{sinc}\left(\frac{b\xi_x}{2}\right)\mathrm{sinc}\left(\frac{a\xi_y}{2}\right)\end{aligned} \tag{A.8}$$

where we used trigonometric identities. Thus we have confirmed the well-known result that the transform of a rectangular aperture is proportional to the sinc function.

B HELMHOLTZ RECIPROCITY

Assume that a single source gives rise to the irradiance impinging on an aperture Σ from direction Ω_1, and using Eq. (8.5) the cross-spectral density can be written as (assuming quasi-homogeneous source, which serves a decent illustrative approximation in the far-field)

$$\boldsymbol{\mathcal{W}}\Big|_\Sigma = \int_{\mathcal{S}^2}\mathrm{d}^2\Omega'\left|\Omega'\cdot\hat{\mathbf{n}}\right|\boldsymbol{\mathcal{L}}^{(i)}_{\Omega'} = |\Omega_1||\Omega_1\cdot\hat{\mathbf{n}}|\,\boldsymbol{\mathcal{L}}^{(i)}_{\Omega_1} \tag{B.1}$$

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

$$\begin{aligned}\boldsymbol{\mathcal{L}}^{(s)}_{\Omega_1}\Big|_{\hat{\mathbf{r}}_1=\hat{\mathbf{r}}_2=\Omega_2} &= \mathscr{D}\left\{\boldsymbol{\mathcal{W}}\Big|_\Sigma, \Sigma\right\}\Big|_{\hat{\mathbf{r}}_1=\hat{\mathbf{r}}_2=\Omega_2} \\ &= \frac{|\Omega_1||\Omega_1\cdot\hat{\mathbf{n}}|e^{ik(r_1-r_2)}}{\lambda^2 A_\Sigma|\Omega_2\cdot\hat{\mathbf{n}}|}\boldsymbol{H}_{\Omega_1}\boldsymbol{\Psi}_{\Omega_1}\boldsymbol{H}^T_{\Omega_1}\end{aligned} \tag{B.2}$$

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

Ω_2 and observing the diffracted radiation in direction Ω_1, which analogously is

$$\boldsymbol{\mathcal{L}}^{(s)}_{\Omega_2}\Big|_{\hat{\mathbf{r}}_1=\hat{\mathbf{r}}_2=\Omega_1} = \frac{|\Omega_2||\Omega_2\cdot\hat{\mathbf{n}}|e^{ik(r_1-r_2)}}{\lambda^2 A_\Sigma|\Omega_1\cdot\hat{\mathbf{n}}|}\boldsymbol{H}_{\Omega_2}\boldsymbol{\Psi}_{\Omega_2}\boldsymbol{H}^T_{\Omega_2} \tag{B.3}$$

We choose our bases such that $\hat{\mathbf{e}}_1 = \hat{\mathbf{e}}'_1$ and both are perpendicular to the plane spanned by Ω_1 and Ω_2, and this fixes the second base components (up to a sign). In this case, with some vector algebra we deduce that $\boldsymbol{H}_{\Omega_1} = |\Omega_2\cdot\hat{\mathbf{n}}|\,\boldsymbol{I}$ and $\boldsymbol{H}_{\Omega_2} = |\Omega_1\cdot\hat{\mathbf{n}}|\,\boldsymbol{I}$. Hence, if $|\Omega_1| = |\Omega_2|$, then the "geometric" terms in the expressions Eqs. (B.2) and (B.3) (all terms except $\boldsymbol{\Psi}$) become equal. Nonetheless, clearly $\boldsymbol{\Psi}$ is not reciprocal in general.

REFERENCES

Milton Abramowitz and Irene Stegun. 1965. *Handbook of mathematical functions, with formulas, graphs, and mathematical tables.* Dover Publications, New York.

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

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

Isaac Freund. 2003. Polarization critical points in polychromatic optical fields. *Optics Communications* 227, 1 (2003), 61 – 71. https://doi.org/10.1016/j.optcom.2003.09.063

I. M. Gelfand. 1964. *Generalized functions.* Academic Press, New York.

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

J Hollas. 2004. *Modern spectroscopy.* Wiley, Chichester.

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

Konrad Jorgens. 1982. *Linear integral operators.* Pitman Advanced Pub. Program, Boston.

Olga Korotkova and Emil Wolf. 2005. Changes in the state of polarization of a random electromagnetic beam on propagation. *Optics Communications* 246, 1 (2005), 35 – 43. https://doi.org/10.1016/j.optcom.2004.10.078

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

Leonard Mandel and Emil Wolf. 1995. *Optical coherence and quantum optics.* Cambridge University Press, Cambridge.

Scott Miller. 2012. *Probability and Random Processes : With Applications to Signal Processing and Communications.* Elsevier Science, Burlington.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2016. *Physically Based Rendering: From Theory to Implementation (3rd ed.)* (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. 1266 pages.

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

Shung-Wu Lee and R. Mittra. 1983. Fourier transform of a polygonal shape function and its application in electromagnetics. *IEEE Transactions on Antennas and Propagation* 31, 1 (1983), 99–103. https://doi.org/10.1109/TAP.1983.1142981

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

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

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.

Xifeng Xiao and David Voelz. 2006. Wave optics simulation approach for partial spatially coherent beams. *Optics Express* 14, 16 (2006), 6986. https://doi.org/10.1364/oe.14.006986

Ling-Qi Yan, Miloš Hašan, Bruce Walter, Steve Marschner, and Ravi Ramamoorthi. 2018. Rendering Specular Microgeometry with Wave Optics. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2018)* 37, 4 (2018).

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