Ultrashort Pulses in Focal Regions

by

Aaron Webster

A Master's Thesis

Institute of Optics, Information and Photonics
University of Erlangen-Nuremberg
June 2011

Contents

Abstract

The spatial and temporal behavior of ultrashort pulses in the vicinity of the focal region of a complex optical system is investigated numerically using an extended Debye approximation to the Rayleigh-Sommerfeld diffraction integrals. Data for the evaluation of such is obtained through raytracing using a coherent superposition of monochromatic fields to describe a pulse. Various techniques regarding methods for this numerical implementation are discussed and compared with existing literature.

Introduction

This thesis concerns the numerical simulation of ultrashort pulses in the focal regions of complex optical systems. The implementation uses a combination of raytracing and diffraction theory. The figure below schematizes the two realms of this method:
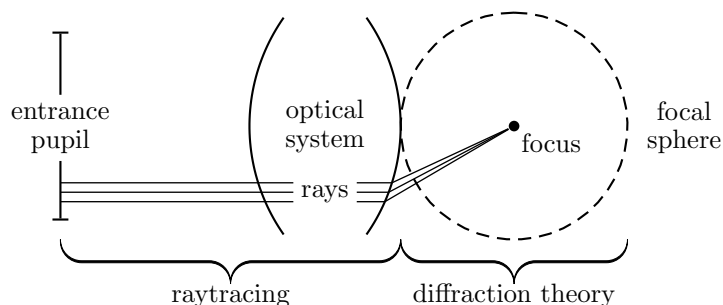
To the left is assumed a focusing optical system, fairly well corrected for aberrations. The system itself can be any arrangement of optical elements whose response can be, up to the final optical surface, determined by raytracing. This allows the investigation of so-called "real" optical systems - ones including effects such as material dispersion, tilt, misalignment, and monochromatic aberrations.

To the right of the optical system we define a surface known as the *focal sphere* – a spherical surface centered around the focus with a radius equal to the *back focal length*.[1] The information of the rays intersecting the focal sphere are then used to determine the structure of the electromagnetic field in the focal region using what is known as the *Debye approximation* applied to the Rayleigh-Sommerfeld diffraction integrals.

In total, the numerical simulation makes some basic assumptions:

- The optical system is fairly well corrected for aberrations. The optical path difference on the focal sphere should be at most a few wavelengths.

- Rays on the focal sphere propagate to a single focus.

- The radial boundary of the area of interest near focus is small compared to the radius of the focal sphere.

- Interference and diffraction on the focal sphere can be neglected.

- The volume within the focal sphere contains a material which is both linear and homogeneous.

Furthermore, because the Debye approximation uses information obtained through raytracing, the local permittivity and permeability of the materials in the optical system must vary with frequency only; there can be (to a reasonable approximation) no nonlinear intensity dependent effects. This puts some limits on the optical intensities for which our model would produce good results. We have estimated an upper limit on this to be around 10^9 W/cm^2. [2] Though this seems prohibitive, many systems which focus ultrashort pulses assume a vacuum behind the optical system, and only in the direct vicinity of the focus is there a material (e.g. atom or a molecule). In such a case our methods are perfectly suited to

---

[1] The distance from the focus to the last optical surface.
[2] Section 1.3.1

be combined with, for example, finite-difference time-domain (FDTD) methods. Therefore, this upper limit is important only within the optical system itself: a region in which the light is mostly expanded anyway.

This work will proceed as follows: Chapter 1 introduces the mathematical theories from which our treatment derives. We begin with a derivation of the plane wave solutions to Maxwell's equations – the basis for both raytracing and the Debye approximation. Proceeding, we apply the (scalar) Debye approximation for monochromatic light to the Rayleigh-Sommerfeld diffraction integrals. This is then extended to a vectorial theory. Finally, the formalism of pulsed light behavior is introduced and included into the theory.

Chapter 2 concerns the actual numerical implementation of the theory discussed in Chapter 1. This includes description and analysis of computer codes produced for this work.

Chapter 3 presents the main predictions of our work and makes comparisons with existing literature. In doing so our intent is to validate our tool and show the domains where this validity holds. With this in hand, further investigations are made possible.

This work uses a particular notation which may differ from that in the referenced literature. The reader is therefore directed to Appendix B, which describes the variable conventions which have been taken. Additionally, where not stated explicitly parameters for numerical computation of applicable figures are presented in Table B.2.

Chapter 1

Theory

1.1 The Debye Approximation

1.1.1 Maxwell's Equations

Foundational to our numerical calculation are plane wave solutions to Maxwell's equations. We begin with a derivation of such waves, starting with the differential form of Maxwell's equations

$$\nabla \cdot \mathbf{D} = 0 \qquad \text{Gauss's law} \qquad (1.1)$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{Gauss's law for magnetism} \qquad (1.2)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{Faraday's law of induction} \qquad (1.3)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \qquad \text{Ampere's law with Maxwell's correction} \qquad (1.4)$$

Where $\mathbf{E}$ is the electric field in $\mathrm{V\,m^{-1}}$, $\mathbf{B}$ is the magnetic field in Teslas, $\mathbf{D}$ is the electric displacement field in $\mathrm{C\,m^{-2}}$, and $\mathbf{H}$ is the so-called auxiliary magnetic field, in units of $\mathrm{A\,m^{-1}}$. Here the notation is such that vectors are in **bold**. Thus, for example, in Cartesian coordinates the electric field

$$\mathbf{E} = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \qquad (1.5)$$

In formulating Maxwell's equations it is assumed all electromagnetic propagation takes place in a *simple dielectric*, that is, the following assumptions are made:

1. The polarization density $\mathbf{P}$ is linear with the electric field $\mathbf{E}$.

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \qquad (1.6)$$

where χ_e is the electric susceptibility. From this and the definition of $\mathbf{D}$

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \qquad (1.7)$$

Where ϵ_0 is the permittivity of free space, one can simplify the $\mathbf{D}$-field as

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \qquad (1.8)$$

$$= \epsilon_0 \mathbf{E} + \epsilon_0 \chi_e \mathbf{E} \qquad (1.9)$$

$$= \epsilon_0 (1 + \chi_e) \mathbf{E} \qquad (1.10)$$

$$= \epsilon \mathbf{E} \qquad (1.11)$$

Here ϵ has been defined to be the permittivity of the dielectric. The convention in this work is usually that variables with a subscript 0 can always be assumed to be that variable *in vacuo*, while their material dependent counterparts have no subscript.

2. The magnetization $\mathbf{M}$, related to the auxiliary magnetic field by is linear in $\mathbf{H}$

$$\mathbf{M} = \mu_0 \chi_m \mathbf{H} \tag{1.12}$$

where χ_m is the volume magnetic susceptibility. This allows the magnetic field to be rewritten according to its definition

$$\mathbf{B} = \mu_0 \left(\mathbf{H} + \mathbf{M} \right) \tag{1.13}$$
$$= \mu_0 \left(1 + \chi_m \right) \mathbf{H} \tag{1.14}$$
$$= \mu \mathbf{H} \tag{1.15}$$

where μ_0 is the vacuum permeability of free space. In practice, most dielectrics have nearly zero magnetic response, so $\mathbf{M} = 0$ and $\mu_0 = \mu$. Nevertheless for the sake of generality it is treated as a material dependent parameter.

3. The material is homogeneous.

4. The material is isotropic.

Taking these assumptions into account, plane wave solutions for the electric field can be derived. Beginning with Faraday's law

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.16}$$

and taking the curl of both sides

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right) \tag{1.17}$$

$$= -\frac{\partial}{\partial t} \left(\nabla \times \mathbf{B} \right) \tag{1.18}$$

The left side can be expanded and simplified using the vector identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \tag{1.19}$$

in combination with Gauss's law, $\nabla \cdot \mathbf{E} = 0$. The right hand side is modified using Ampere's law with $\mathbf{B} = \mu \mathbf{H}$.

$$\nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu \frac{\partial}{\partial t} \left(\nabla \times \mathbf{H} \right) \tag{1.20}$$

$$-\nabla^2 \mathbf{E} = -\mu \frac{\partial^2 \mathbf{D}}{\partial t^2} \tag{1.21}$$

$$-\nabla^2 \mathbf{E} = -\mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{1.22}$$

$$\left(\nabla^2 - \mu \epsilon \frac{\partial^2}{\partial t^2} \right) \mathbf{E} = 0 \tag{1.23}$$

This is the electromagnetic wave equation in terms of the electric field. Choosing to solve this equation by separation of variables the following plane wave solutions can be obtained

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \, e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \tag{1.24}$$

where $k = \omega/c = \omega \sqrt{\epsilon \mu}$. In this notation, $\mathbf{k}$ is the (material) vectorial wavenumber, $\mathbf{r}$ is the spatial position, ω is angular frequency, and t is the dimension of time (see Chapter B of the Appendix). The initial value is chosen with the vectorial constant $\mathbf{E}_0$.

Equation 1.24 which, when applied to Equation 1.23, obeys the *superposition principle*

$$f(\phi_\alpha) + f(\phi_\beta) = f(\phi_\alpha + \phi_\beta) \tag{1.25}$$

such that if ϕ_α is a solution and ϕ_β is a solution, then $\phi_\alpha + \phi_\beta$ is also a solution. The proof of this is readily available by the substitution of any linear superposition of Equation 1.24 into Equation 1.23. As a simple example we can choose the one dimensional case of a superposition of just two waves with spatial frequencies k_α, k_β and temporal frequencies ω_α, ω_β. In terms of the electric field E,

$$E(x,t) = e^{i(k_\alpha x - \omega_\alpha t)} + e^{i(k_\beta x - \omega_\beta t)} \tag{1.26}$$

first by computing the partial derivatives

$$\nabla^2 E = -k_\alpha^2 e^{i(k_\alpha x - \omega_\alpha t)} - k_\beta^2 e^{i(k_\beta x - \omega_\beta t)} \tag{1.27}$$

$$\frac{\partial^2}{\partial t^2} E = -\omega_\alpha^2 e^{i(k_\alpha x - \omega_\alpha t)} - \omega_\beta^2 e^{i(k_\beta x - \omega_\beta t)} \tag{1.28}$$

using the condition that Equation 1.28 must be valid at each point when substituted into Equation 1.23, a separation ansatz can be made yielding

$$-k_\alpha^2 + \mu\epsilon\omega_\alpha^2 = 0 \tag{1.29}$$

and

$$-k_\beta^2 + \mu\epsilon\omega_\beta^2 = 0 \tag{1.30}$$

whereby equating the two yields

$$\left(-k_\alpha^2 - k_\beta^2\right) - \mu\epsilon\left(-\omega_\alpha^2 - \omega_\beta^2\right) = 0 \tag{1.31}$$

Using the definitions for k and ω,

$$k = \frac{2\pi}{\lambda}, \qquad \omega = \frac{2\pi c}{\lambda} \tag{1.32}$$

where λ is the wavelength, allows us to see the parameter c present in ω. This coupled with the definition of the speed of light in a medium c

$$c = \frac{1}{\sqrt{\mu\epsilon}} \tag{1.33}$$

allows the factor of $\mu\epsilon$ to cancel out, and so do the rest of the terms. The equation is satisfied. The extension of this to the vectorial case is as simple as repeating the procedure for each component. The superposition principle is fundamental to our work. Through it we can represent any solution to Maxwell's equations (as presented) by a linear superposition of plane waves of the form of Equation 1.24.

So far we have only considered solutions for the electric field, $\mathbf{E}$. For most systems, the magnetic field $\mathbf{B}$ is fully described by the electric field and the local direction of wave propagation (i.e. the wavevector $\mathbf{k}$). This can easily be verified by repeating the derivation starting with Ampere's law and using Gauss's law as a simplification on the left hand side and Faraday's law as a substitution on the right. There are, of course, nuances in computing such but for simplicity in this discussion we will deal with $\mathbf{E}$ only.

1.1.2 Rayleigh-Sommerfeld Diffraction Theory and the Debye Approximation

One of the fundamental results of diffraction theory is known as the *Kirchhoff-Helmholtz integral theorem*. This theorem relates the behavior of a scalar field within a volume to the field on a surface enclosing it. Simply stated, the electromagnetic field within a region is completely defined by the flux of the enclosing field. This is the statement of the Rayleigh-Sommerfeld diffraction equation

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{i\lambda} \iint_S \mathbf{E}(\mathbf{r}',t) \frac{e^{-ikR}}{R} \cos\theta \, ds \tag{1.34}$$

The geometry and variables for this equation are schematized in Figure 1.1.

Light, represented by rays, originate on the entrance pupil and are traced through an optical system. The system focuses light to its focal region. The Gaussian focus itself is located at the local (Cartesian)
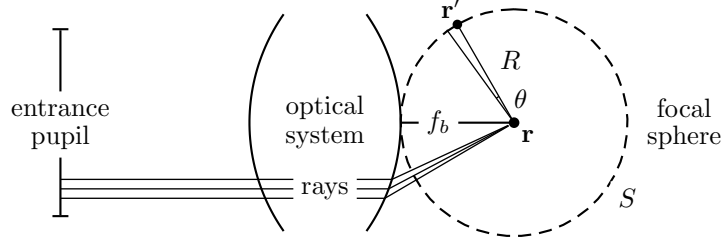
Figure 1.1: Definition of terms. Rays distributed on the entrance pupil pass through an (arbitrary) optical system and are focused to its Gaussian focus. The focal sphere with surface S is centered around the Gaussian focus with radius f_b: typically the minimum distance from said focus and the vertex of the last optical surface. The electric field $\mathbf{E}(\mathbf{r}', t)$ is evaluated at each point $\mathbf{r}'$ on the focal sphere by raytracing. The region of interest is a small volume near the Gaussian focus with local coordinates $\mathbf{r}$.

coordinates $\mathbf{r} = (0, 0, 0)$. Concentric to the focus is the *focal sphere* – a sphere centered around the focus with radius f_b equal to the back focal distance. The electric field on the focal sphere $\mathbf{E}(\mathbf{r}', t)$ is determined by the ray intersections with it at coordinates $\mathbf{r}' = (x', y', z')$.

The Rayleigh-Sommerfeld equation is an integration over the electric field $\mathbf{E}(\mathbf{r}', t)$ on the focal sphere of surface S with spherical waves of the form e^{-ikR}/R oblique to said sphere with an angle θ; here

$$R = \sqrt{|\mathbf{r}' - \mathbf{r}|^2} \tag{1.35}$$

is the radial distance from a point near the focus to the focal sphere. This integral can be simplified considerably to our advantage. First, as we are only interested in small volumes around the focus, the radial coordinate $R \approx f_b$. The first consequence of this is the replacement of the term

$$\frac{1}{R} \Rightarrow \frac{1}{f_b} \tag{1.36}$$

To make this approximation of R in the exponential, first we note the radial distance R can be written according to its definition in terms of the difference of radial distance from the focus $\mathbf{r}$ and the radial distance to focal sphere $|\mathbf{r}'| = f_b$.

$$R = \sqrt{|\mathbf{r}' - \mathbf{r}|^2} \tag{1.37}$$

$$= \sqrt{\mathbf{r}'^2 + \mathbf{r}^2 - 2\mathbf{r} \cdot \mathbf{r}'} \tag{1.38}$$

$$= \sqrt{f_b^2 + (\mathbf{r}^2 - 2\mathbf{r} \cdot \mathbf{r}')} \tag{1.39}$$

Assuming small values of $|\mathbf{r}|$ compared to f_b, the expression of Equation 1.39 can be approximated by the well known series approximation

$$\sqrt{1 + \alpha} \approx 1 + \frac{\alpha}{2} \quad \text{for} \quad \alpha \ll 1 \tag{1.40}$$

then

$$R = \sqrt{f_b^2 + (\mathbf{r}^2 - 2\mathbf{r} \cdot \mathbf{r}')} \tag{1.41}$$

$$= f_b \sqrt{1 + \frac{\mathbf{r}^2 - 2\mathbf{r} \cdot \mathbf{r}'}{f_b^2}} \tag{1.42}$$

$$\approx f_b \left(1 + \frac{\mathbf{r}^2 - 2\mathbf{r} \cdot \mathbf{r}'}{2f_b^2} \right) \tag{1.43}$$

$$\approx f_b + \frac{\mathbf{r}^2}{2f_b} - \frac{\mathbf{r} \cdot \mathbf{r}'}{f_b} \tag{1.44}$$

6

As with any vector, $\mathbf{r}'$ can be expressed in terms of its magnitude and its direction unit vector $\hat{\mathbf{n}}$

$$\mathbf{r}' = |\mathbf{r}'|\hat{\mathbf{n}} = -f_b\hat{\mathbf{n}} \tag{1.45}$$

Here $\hat{\mathbf{n}}$ points from the focal sphere to the focus, and is parallel to the wavevector $\mathbf{k}$. Using this information to continue reducing Equation 1.44,

$$R = f_b + \frac{\mathbf{r}^2}{2f_b} - \frac{\mathbf{r}\cdot\mathbf{r}'}{f_b} \tag{1.46}$$

$$= f_b + \frac{\mathbf{r}^2}{2f_b} + \hat{\mathbf{n}}\cdot\mathbf{r} \tag{1.47}$$

$$= f_b + \frac{\mathbf{r}^2}{2f_b} + \frac{\mathbf{k}\cdot\mathbf{r}}{|\mathbf{k}|} \tag{1.48}$$

Collecting these approximations to Equation 1.34 yields

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{\mathrm{i}\lambda f_b}\iint_S \mathbf{E}(\mathbf{r}',t)\,\mathrm{e}^{-\mathrm{i}kf_b}\,\mathrm{e}^{-\mathrm{i}kr^2/(2f_b)}\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\cos\theta\,\mathrm{d}s \tag{1.49}$$

If plane waves are obtained about the focal sphere and propagate nearly uniformly to the focus $\mathbf{r} = (0,0,0)$, the wavevector will be perpendicular to the focal sphere and the obliquity factor $\theta \approx 0$, $\cos\theta \approx 1$, can be removed from the equation.

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{\mathrm{i}\lambda f_b}\iint_S \mathbf{E}(\mathbf{r}',t)\,\mathrm{e}^{-\mathrm{i}kf_b}\,\mathrm{e}^{-\mathrm{i}kr^2/(2f_b)}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{r})}\,\mathrm{d}s \tag{1.50}$$

The term $\mathrm{e}^{-\mathrm{i}kr^2/(2f_b)}$ can be further approximated under the assumption that the volume of interest is small. Using $k = 2\pi/\lambda$,

$$\mathrm{e}^{-\mathrm{i}kr^2/(2f_b)} = \mathrm{e}^{-\mathrm{i}\pi r^2/(\lambda f_b)} \tag{1.51}$$

If $f_b \gg r^2/\lambda$, $\mathrm{e}^{-\mathrm{i}\pi r^2/(\lambda f_b)} \approx 1$ and can be neglected from the equation. Furthermore, since the term $\mathrm{e}^{-\mathrm{i}kf_b}$ represents a constant phase factor over the entire focal sphere, it can be neglected from the integral. The net result of these approximations leaves the Debye approximation for monochromatic light.

$$\boxed{\mathbf{E}(\mathbf{r},t) = \frac{1}{\mathrm{i}\lambda f_b}\iint_S \mathbf{E}(\mathbf{r}',t)\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{r})}\,\mathrm{d}s} \tag{1.52}$$

1.2 Ultrashort Pulses

So far we have been concerned with radiation of the form

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)} \tag{1.53}$$

(i.e. monochromatic light with angular frequency ω). As mentioned in the introduction, we are most interested in the behavior of *pulses* of light. Pulses are characterized not by one frequency, but a continuum of frequencies

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\mathbf{E}}(\omega)\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{r}-\omega t)}\,\mathrm{d}\omega \tag{1.54}$$

Here $\tilde{\mathbf{E}}(\omega)$ represents the *envelope* of the pulse in its angular frequency domain.

The particular formulation of Equation 1.54 is that of the Fourier transform. A Fourier transform enables the representation of an arbitrary pulse by a linear superposition of simple oscillatory functions - here mutually orthogonal complex exponential functions which satisfy the electromagnetic wave equation. The Fourier transform and its inverse are defined by

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{f}(\omega)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega = \mathscr{F}^{-}\big(\tilde{f}(\omega)\big) \tag{1.55}$$
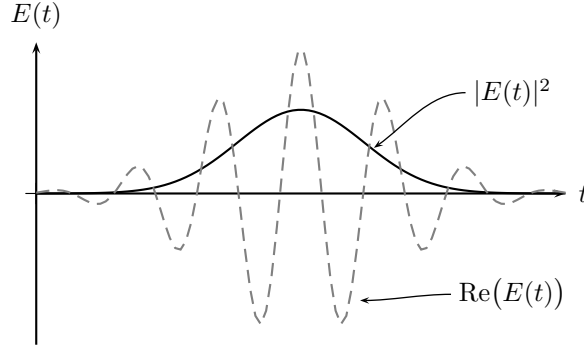
Figure 1.2: The envelope of a pulse is defined as the magnitude squared of its complex valued electric field. Since the function f in the derivations has no oscillatory component, the relation between f and E is given by $|f|^2 = |E|^2$.

and

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} f(t)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}t = \mathscr{F}^+\big(f(t)\big) \tag{1.56}$$

Where the convention $\mathscr{F}^+$ is the Fourier transform, and $\mathscr{F}^-$ is its inverse. In the same way, pulses can be represented in both the time and angular frequency domain as a Fourier transform pair

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{E}(\omega)\, \mathrm{e}^{-\mathrm{i}\omega t}\, \mathrm{d}\omega \tag{1.57}$$

$$\tilde{E}(\omega) = \int_{-\infty}^{\infty} E(t)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}t \tag{1.58}$$

$$\mathscr{F}^+\big(E(t)\big) = \tilde{E}(\omega) \tag{1.59}$$

$$\mathscr{F}^-\big(\tilde{E}(\omega)\big) = E(t) \tag{1.60}$$

1.2.1 Envelope Functions

The mathematical description of the pulse envelope is carried out with variables as designated in Figure 1.3. The envelopes are described in either the time or angular frequency domain by $f(t)$ or $\tilde{f}(\omega)$, respectively. The variables δt and $\delta \omega$ represent the *full width at half maximum* (FWHM) – the width of $|f(t)|^2$ or $|\tilde{f}(\omega)|^2$ at half of its maximum value. The product of the two, $\delta t \delta \omega$, is the *time bandwidth product*. As a consequence of the Fourier relationship, as the width of the pulse increases in the time domain, the spectral width decreases in the frequency domain and vice-versa. In terms of the actual electric field E, which contains a fast oscillating real and imaginary part, the function f is related by $|f|^2 = |E|^2$ (f has no oscillatory component). This is detailed in Figure 1.2.

For all pulse types we have normalized them so that the temporal integral

$$\int_{-\infty}^{\infty} \big|f(t)\big|^2\, \mathrm{d}t = 1 \tag{1.61}$$

and the integral over angular frequency

$$\int_{-\infty}^{\infty} \big|\tilde{f}(\omega)\big|^2\, \mathrm{d}\omega = 2\pi \tag{1.62}$$

For convince, this section treats envelope functions in a purely mathematical sense; the variables themselves dimensionless. Some common pulse shapes for modern femtosecond lasers are plotted in Figure 1.4. They are the Gaussian, the sech, and the Lorentzian. These specific pulse shapes are important in part because they are so called *transform limited*: for a given spectrum they represent the shortest pulse duration possible and the time-bandwidth product $\delta t \delta \omega$ is minimized for that pulse type.
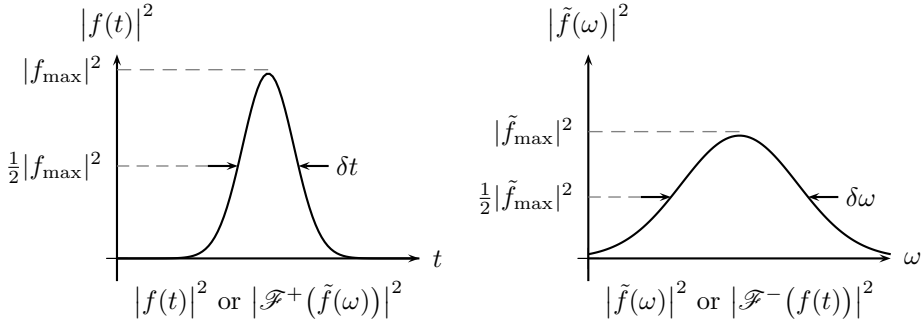
8

Figure 1.3: Definition of terms. A pulse can be represented in either the time $f(t)$ or angular frequency $\tilde{f}(\omega)$ domain through a Fourier transform relationship. The parameters δt or $\delta \omega$ are the width of the function at half its maximum value. As the width of the pulse increases in the time domain, the spectral width decreases in the frequency domain and vice-versa.

1.2.2 Gaussian Pulse

A Gaussian pulse is defined by the temporal envelope

$$f(t, \alpha) = A\, e^{-t^2/(2\alpha^2)} \tag{1.63}$$

Where α is some, yet to be determined parameter (controlling the width of the pulse) and A is a normalization constant. The first step is to normalize Equation 1.63 and find A as per the condition described in Equation 1.61. To find do so, the well known Gaussian integral is computed

$$\int_{-\infty}^{\infty} \left| e^{-t^2/(2\alpha^2)} \right|^2 dt = A^2 \alpha \sqrt{\pi} \tag{1.64}$$

$$A = \sqrt{\frac{1}{\alpha\sqrt{\pi}}} \tag{1.65}$$

Substitution of A into Equation 1.63, the following is obtained

$$f(t, \alpha) = \left(\alpha\sqrt{\pi}\right)^{-1/2} e^{-t^2/(2\alpha^2)} \tag{1.66}$$

In its current form using the variable α in effect controls the width of the envelope. Since we are interested in characterizing pulses by their FWHM, it is prudent to replace α with δt, such that δt represents the FWHM – the width of the magnitude squared envelope at half the maximum value.

$$|f(t, \alpha)|^2 = \frac{1}{2} |f(0)|^2 \tag{1.67}$$

$$\left| e^{-t^2/(2\alpha^2)} \right|^2 = \frac{1}{2} e^0 \tag{1.68}$$

$$e^{-t^2/(2\alpha^2)} = \frac{1}{\sqrt{2}} \tag{1.69}$$

$$-\frac{t^2}{2\alpha^2} = \log\left(\frac{1}{\sqrt{2}}\right) \tag{1.70}$$

$$\frac{t^2}{2\alpha^2} = \frac{\log 2}{2} \tag{1.71}$$

$$t^2 = \alpha^2 \log 2 \tag{1.72}$$

$$t = \pm\alpha\sqrt{\log 2} \tag{1.73}$$

$$\delta t = 2\alpha\sqrt{\log 2} \quad , \quad \alpha = \frac{\delta t}{2\sqrt{\log 2}} \tag{1.74}$$

Note that the solution in terms of t is symmetric about $t = 0$. The parameter δt is then $2|t|$, fitting with the definition of the FWHM (Figure 1.3).

9

The variable α can then be inserted into Equation 1.66, obtaining our final expression.

$$f(t, \delta t) = \left(\frac{2\sqrt{\log 2}}{\sqrt{\pi}\delta t} \right)^{1/2} e^{-2t^2 \log 2/\delta t^2} \tag{1.75}$$

We now have an expression where δt describes the full width at half maximum of the magnitude squared envelope. That is to say,

$$\left| f(t, \delta t) \right|^2 = \frac{1}{2} \left| f(0, \delta t) \right|^2 \tag{1.76}$$

The envelope is also normalized. By taking the Fourier transform of Equation 1.75, a complementary equation $\tilde{f}(\omega, \delta t)$, in terms of angular frequency (and the same parameter δt) can be obtained

$$\mathscr{F}^+ \big(f(t, \delta t) \big) = \tilde{f}(\omega, \delta t) \tag{1.77}$$

$$= \int_{-\infty}^{\infty} \left(\frac{2\sqrt{\log 2}}{\sqrt{\pi}\delta t} \right)^{1/2} e^{-2t^2 \log 2/\delta t^2} \, e^{\mathrm{i}\omega t} \, \mathrm{d}t \tag{1.78}$$

From a table of indefinite integrals[18], an integral of the same form is found to be

$$\int_{-\infty}^{\infty} e^{-at^2} \, e^{-2bt} \, \mathrm{d}t = \sqrt{\frac{\pi}{a}} \, e^{b^2/a} \tag{1.79}$$

Equating terms in the exponents from Equation 1.78 we obtain

$$a = \frac{2\log 2}{\delta t^2} \quad , \quad b = \frac{-\mathrm{i}\omega}{2} \tag{1.80}$$

The integral can then be evaluated

$$\tilde{f}(\omega, \delta t) = \int_{-\infty}^{\infty} \left(\frac{2\sqrt{\log 2}}{\sqrt{\pi}\delta t} \right)^{1/2} e^{-2t^2 \log 2/\delta t^2} \, e^{\mathrm{i}\omega t} \, \mathrm{d}t \tag{1.81}$$

$$= \left(\frac{\pi \delta t^2}{2\log 2} \right)^{1/2} \left(\frac{2\sqrt{\log 2}}{\sqrt{\pi}\delta t} \right)^{1/2} e^{-\omega^2 \delta t^2/(8\log 2)} \tag{1.82}$$

therefore

$$\tilde{f}(\omega, \delta t) = \left(\frac{\sqrt{\pi}\delta t}{\sqrt{\log 2}} \right)^{1/2} e^{-\omega^2 \delta t^2/(8\log 2)} \tag{1.83}$$

Equation 1.83 is also normalized. Note that

$$\int_{-\infty}^{\infty} \left| \tilde{f}(\omega) \right|^2 = 2\pi \tag{1.84}$$

This is consistent with the Fourier transformed pulse defined in units of angular frequency.

At this point it is important to point out that the pulse envelope in both the time and angular frequency domain have been characterized entirely in terms of the (temporal) FWHM δt. This is not a mistake. The intention here is to simplify the (eventual) numerical computation by having one parameter controlling the width of the pulse. For example, if a $40\,\mathrm{fs}$ pulse was to be obtained by sampling in angular frequency space, the integral

$$f(t, 40\,\mathrm{fs}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega, 40\,\mathrm{fs}) \, e^{-\mathrm{i}\omega t} \, \mathrm{d}\omega \tag{1.85}$$

would be computed to obtain the pulse (in units of femtoseconds). Of course, the actual FWHM of the pulse $\tilde{f}(\omega, \delta t)$ is not δt. This will be addressed shortly.

The literature often quotes what is referred to as the *time bandwidth product*, $\delta t \delta \omega$. As written, Equation 1.75 has a FWHM of δt – the magnitude squared of the function decreases to half its maximum value for $t = \pm \delta t / 2$. The FWHM of Equation 1.83 can be found in an analogous way by solving the following equation with width δt

$$\left| \tilde{f}(\omega, \delta t) \right|^2 = \frac{1}{2} \left| \tilde{f}(0, \delta t) \right|^2 \tag{1.86}$$

$$\left| e^{-\omega^2 \delta t^2 / (8 \log 2)} \right|^2 = \frac{1}{2} e^0 \tag{1.87}$$

$$e^{-\omega^2 \delta t^2 / (8 \log 2)} = \frac{1}{\sqrt{2}} \tag{1.88}$$

$$\frac{\omega^2 \delta t^2}{8 \log 2} = \frac{\log 2}{2} \tag{1.89}$$

$$\omega^2 = \left(\frac{2 \log 2}{\delta t} \right)^2 \tag{1.90}$$

$$\omega = \pm \frac{2 \log 2}{\delta t} \tag{1.91}$$

$$\delta \omega = \frac{4 \log 2}{\delta t} \tag{1.92}$$

Note in the last step we have defined $\delta \omega$ to be the FWHM of $\tilde{f}(\omega)$. Multiplying them together yields the time bandwidth product

$$\delta t \delta \omega = \delta t \frac{4 \log 2}{\delta t} \tag{1.93}$$

$$= 4 \log 2 \approx 2.7726 \tag{1.94}$$

Because it is in terms of angular frequency, this result differs from the time bandwidth product, often quoted in literature in terms of ordinary frequency, $\delta t \delta \nu$, by a factor of $(2\pi)^{-1}$

$$\delta t \delta \nu = \frac{\delta t \delta \omega}{2\pi} = \frac{2 \log 2}{\pi} \approx 0.4413 \tag{1.95}$$

Gaussian functions of this form have a useful expression for their definite integrals. This can be taken advantage of to determine what integration boundaries must be chosen to encompass a certain proportion of the total pulse energy.

For Gaussians symmetric about the origin, this integral is expressed in terms of the error function erf

$$\int_0^a e^{-q^2 t^2} \, \mathrm{d}t = \frac{\sqrt{\pi}}{2q} \operatorname{erf}(qa) \tag{1.96}$$

Because the Gaussian is symmetric about the origin $t = 0$, this can be trivially rewritten

$$\int_{-a}^a e^{-q^2 t^2} \, \mathrm{d}t = \frac{\sqrt{\pi}}{q} \operatorname{erf}(qa) \tag{1.97}$$

Using this form to evaluate Equation 1.75 with $q = 2\sqrt{\log 2}/\delta t$,

$$\int_{-a}^a |f(t, \delta t)|^2 \, \mathrm{d}t = \int_{-a}^a \frac{2\sqrt{\log 2}}{\sqrt{\pi} \delta t} e^{-4t^2 \log 2 / \delta t^2} \, \mathrm{d}t \tag{1.98}$$

$$= \frac{2\sqrt{\log 2}}{\sqrt{\pi} \delta t} \frac{\sqrt{\pi} \delta t}{2\sqrt{\log 2}} \operatorname{erf}\left(\frac{2a\sqrt{\log 2}}{\delta t} \right) \tag{1.99}$$

$$= \operatorname{erf}\left(\frac{2a\sqrt{\log 2}}{\delta t} \right) \tag{1.100}$$

11

Solving for a, we introduce a variable X such that X represents a proportion of the total pulse energy

$$\mathrm{erf}\left(\frac{2a\sqrt{\log 2}}{\delta t}\right) = X \tag{1.101}$$

$$\frac{2a\sqrt{\log 2}}{\delta t} = \mathrm{inverf}(X) \tag{1.102}$$

$$a = \frac{\delta t}{2\sqrt{\log 2}}\,\mathrm{inverf}(X) \tag{1.103}$$

Having found a, let $\zeta(\delta t, X)$ be defined as the function returning a,

$$\zeta(\delta t, X) = \frac{\delta t}{2\sqrt{\log 2}}\,\mathrm{inverf}(X) \tag{1.104}$$

Where $\mathrm{inverf}(X)$ is the inverse error function. This function expresses the bounds required to obtain a pulse with an energy of X of the total energy of an ideal Gaussian pulse. For example, the window for a 40 fs pulse encompassing 99 % of the total pulse energy would have bounds

$$\zeta(40\,\mathrm{fs}, 0.99) = \frac{40\,\mathrm{fs}}{2\sqrt{\log 2}}\,\mathrm{inverf}(0.99) \approx 43.75\,\mathrm{fs} \tag{1.105}$$

Therefore the window should have bounds at $\pm\zeta(40\,\mathrm{fs}, 0.99)$.

An equivalent set of boundaries for the Fourier transformed pulse can be found by the same procedure. Let $\tilde{\zeta}((\delta t, X)$ be defined as the analog to Equation 1.104: a function which returns the symmetric bounds in angular frequency space for which the integral of the magnitude squared envelope returns the proportion X of the total pulse energy.

$$\tilde{\zeta}(\delta t, X) = \frac{2\sqrt{\log 2}}{\delta t}\,\mathrm{inverf}(X) \tag{1.106}$$

1.2.3   Sech Pulse

In its simplest form, the sech pulse is defined

$$f(t, \alpha) = A\,\mathrm{sech}\left(\frac{2t}{\alpha}\right) \tag{1.107}$$

$$\tag{1.108}$$

To solve the integral

$$\int_{-\infty}^{\infty} |f(t, \alpha)|^2 \, \mathrm{d}t \tag{1.109}$$

and obtain the normalizing constant A, we use an integral of the form[1]

$$\int \frac{\mathrm{d}x}{\cosh^n(ax)} = \frac{\sinh(ax)}{a\,(n-1)\cosh^{n-1}(ax)} + \frac{n-2}{n-1}\int \frac{\mathrm{d}x}{\cosh^{n-2}(ax)} \tag{1.110}$$

noting $\mathrm{sech}\,x = \cosh^{-1} x$. With $n = 2$, and $a = 2/\alpha$ this becomes

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}t}{\cosh^2\left(\frac{2t}{\alpha}\right)} = \left.\frac{\alpha\sinh\left(\frac{2t}{\alpha}\right)}{2\cosh\left(\frac{2t}{\alpha}\right)}\right|_{-\infty}^{\infty} + 0 \tag{1.111}$$

$$= \left.\frac{\alpha}{2}\tanh\left(\frac{2t}{\alpha}\right)\right|_{-\infty}^{\infty} \tag{1.112}$$

$$= \frac{\alpha}{2}\left(\lim_{t\to\infty}\tanh\left(\frac{2t}{\alpha}\right) - \lim_{t\to-\infty}\tanh\left(\frac{2t}{\alpha}\right)\right) \tag{1.113}$$

12

The second limit can be found by rewriting the hyperbolic tangent in terms of exponentials

$$\tanh\left(\frac{2t}{\alpha}\right) = \frac{e^{2t/\alpha} - e^{-2t/\alpha}}{e^{2t/\alpha} + e^{-2t/\alpha}} \cdot \frac{e^{2t/\alpha}}{e^{2t/\alpha}} \tag{1.114}$$

$$= \frac{e^{4t/\alpha} - 1}{e^{4t/\alpha} + 1} \tag{1.115}$$

$$\lim_{t \to -\infty} \frac{e^{4t/\alpha} - 1}{e^{4t/\alpha} + 1} = -1 \tag{1.116}$$

and the first through a similar argument but by multiplication with a different unity function

$$\tanh\left(\frac{2t}{\alpha}\right) = \frac{e^{2t/\alpha} - e^{-2t/\alpha}}{e^{2t/\alpha} + e^{-2t/\alpha}} \cdot \frac{e^{-2t/\alpha}}{e^{-2t/\alpha}} \tag{1.117}$$

$$= \frac{1 - e^{-4t/\alpha}}{1 + e^{-4t/\alpha}} \tag{1.118}$$

$$\lim_{t \to \infty} \frac{1 - e^{-4t/\alpha}}{1 + e^{-4t/\alpha}} = 1 \tag{1.119}$$

In terms of the original integral this evaluates to

$$\int_{-\infty}^{\infty} A \, \frac{\mathrm{d}t}{\cosh^2\left(\frac{2t}{\alpha}\right)} = 1 \tag{1.120}$$

$$A^2 \frac{\alpha}{2} \left(1 - (-1)\right) = 1 \tag{1.121}$$

$$A = \frac{1}{\sqrt{\alpha}} \tag{1.122}$$

which is plugged into the original equation, normalizing it and obtaining

$$f(t, \alpha) = \frac{1}{\sqrt{\alpha}} \operatorname{sech}\left(\frac{2t}{\alpha}\right) \tag{1.123}$$

Solving for the parameter α

$$|f(t, \alpha)|^2 = \frac{1}{2} |f(0, \alpha)|^2 \tag{1.124}$$

$$\operatorname{sech}\left(\frac{2t}{\alpha}\right) = \frac{1}{\sqrt{2}} \tag{1.125}$$

$$\frac{2t}{\alpha} = \operatorname{arcsech}\left(\frac{1}{\sqrt{2}}\right) \tag{1.126}$$

$$t = \pm\frac{\alpha}{2} \operatorname{arcsech}\left(\frac{1}{\sqrt{2}}\right) \tag{1.127}$$

$$\delta t = \alpha \operatorname{arcsech}\left(\frac{1}{\sqrt{2}}\right) \quad, \quad \alpha = \frac{\delta t}{\operatorname{arcsech}\left(\frac{1}{\sqrt{2}}\right)} \tag{1.128}$$

using the identity

$$\operatorname{arcsech}(x) = \log \frac{1 + \sqrt{1 - x^2}}{x} \quad, \quad 0 < x \leq 1 \tag{1.129}$$

the function can be expressed in terms of exponentials. When α is substituted into Equation 1.123, the final form in the time domain is obtained

$$\boxed{f(t, \delta t) = \left(\frac{\log\left(1 + \sqrt{2}\right)}{\delta t}\right)^{1/2} \operatorname{sech}\left(\frac{2t \log\left(1 + \sqrt{2}\right)}{\delta t}\right)} \tag{1.130}$$

13

To take the Fourier transform of Equation 1.130, a table of integrals[18] is used to express the Fourier transform in terms of Gamma functions

$$\mathscr{F}^+\left(\mathrm{sech}^n\left(t/\alpha\right)\right) = \tilde{f}(\omega, \delta t) \tag{1.131}$$

$$= \frac{2^{n-1}}{\Gamma(n)}\Gamma\left(\frac{n + i\omega\alpha}{2}\right)\Gamma\left(\frac{n - i\omega\alpha}{2}\right) \tag{1.132}$$

along with the identity[1]

$$\Gamma\left(\frac{1}{2} + iy\right)\Gamma\left(\frac{1}{2} - iy\right) = \left|\Gamma\left(\frac{1}{2} + iy\right)\right|^2 \tag{1.133}$$

$$= \frac{\pi}{\cosh(\pi y)} \tag{1.134}$$

to obtain the Fourier transformed equation

$$\tilde{f}(\omega, \delta t) = \pi\left(\frac{\delta t}{4\log\left(1 + \sqrt{2}\right)}\right)^{1/2}\mathrm{sech}\left(\frac{\pi\omega\,\delta t}{4\log\left(1 + \sqrt{2}\right)}\right) \tag{1.135}$$

As in the case of the Gaussian pulse, the FWHM of the sech pulse in the time domain is the parameter δt. Solving for the width of the pulse in terms of $\tilde{f}(\omega, \delta t)$

$$\left|\tilde{f}(\omega, \delta t)\right|^2 = \frac{1}{2}\left|\tilde{f}(0, \delta t)\right|^2 \tag{1.136}$$

$$\mathrm{sech}\left(\frac{\pi\omega\,\delta t}{4\log\left(1 + \sqrt{2}\right)}\right) = \frac{1}{\sqrt{2}} \tag{1.137}$$

$$\frac{\pi\omega\,\delta t}{4\log\left(1 + \sqrt{2}\right)} = \mathrm{arcsech}\left(\frac{1}{\sqrt{2}}\right) \tag{1.138}$$

$$\omega = \pm\frac{4\log^2\left(1 + \sqrt{2}\right)}{\pi\,\delta t} \tag{1.139}$$

$$\delta\omega = \frac{8\log^2\left(1 + \sqrt{2}\right)}{\pi\,\delta t} \tag{1.140}$$

$$\tag{1.141}$$

and by multiplication of δt with $\delta\omega$ the time bandwidth product is obtained

$$\delta t\delta\omega = \delta t\,\frac{8\log^2\left(1 + \sqrt{2}\right)}{\pi\,\delta t} \tag{1.142}$$

$$= \frac{8\log^2\left(1 + \sqrt{2}\right)}{\pi} \approx 1.9782 \tag{1.143}$$

In terms of ordinary frequency

$$\delta t\delta\nu = \frac{\delta t\delta\omega}{2\pi} = \left(\frac{2\log\left(1 + \sqrt{2}\right)}{\pi}\right)^2 \approx 0.3148 \tag{1.144}$$

The sech pulse also has a useful formula for calculating the total energy contained within a region. Beginning with the definite integral of Equation 1.112, and following the same procedure as was taken for the Gaussian pulse

$$\int_{-a}^{a}\mathrm{sech}^2\left(\frac{2t}{\alpha}\right) = \frac{\alpha}{2}\tanh\left(\frac{2t}{\alpha}\right)\Big|_{-a}^{+a} \tag{1.145}$$

$$= \frac{1}{2}\left(\tanh\left(\frac{2a\log\left(1 + \sqrt{2}\right)}{\delta t}\right) - \tanh\left(\frac{-2a\log\left(1 + \sqrt{2}\right)}{\delta t}\right)\right) \tag{1.146}$$

$$= \tanh\left(\frac{2a\log\left(1 + \sqrt{2}\right)}{\delta t}\right) \tag{1.147}$$

14

then, solving a for a proportion of the pulse energy X the following is obtained

$$\tanh\left(\frac{2a\log\left(1+\sqrt{2}\right)}{\delta t}\right) = X \tag{1.148}$$

$$\frac{2a\log\left(1+\sqrt{2}\right)}{\delta t} = \operatorname{arctanh}(X) \tag{1.149}$$

$$a = \frac{\delta t}{2}\frac{\operatorname{arctanh}(X)}{\log\left(1+\sqrt{2}\right)} \tag{1.150}$$

And in the same regards as Equation 1.104,

$$\zeta(\delta t, X) = \frac{\delta t}{2}\frac{\operatorname{arctanh}(X)}{\log\left(1+\sqrt{2}\right)} \tag{1.151}$$

or, in terms of exponential functions

$$\zeta(\delta t, X) = \frac{\delta t}{4}\frac{\log\left(\frac{1+X}{1-X}\right)}{\log\left(1+\sqrt{2}\right)} \tag{1.152}$$

The same procedure can be used to obtain the parameter $\tilde{\zeta}(\delta t, X)$,

$$\tilde{\zeta}(\delta t, X) = \frac{4\operatorname{arctanh}(X)\log\left(1+\sqrt{2}\right)}{\pi\delta t} \tag{1.153}$$

$$= \frac{2\log\left(\frac{1+x}{1-x}\right)\log\left(1+\sqrt{2}\right)}{\pi\delta t} \tag{1.154}$$

1.2.4   Lorentzian Pulses

Beginning in the time domain with a basic form of a Lorentzian envelope function,

$$f(t, \alpha) = \frac{A}{1+\alpha^2 t^2} \tag{1.155}$$

Solving for the normalizing coefficient A using a table of definite integrals[18]

$$\int_{-\infty}^{\infty}|f(t,\alpha)|^2\,\mathrm{d}t = 1 \tag{1.156}$$

$$\frac{A^2\pi}{2\alpha} = 1 \tag{1.157}$$

$$A = \sqrt{\frac{2\alpha}{\pi}} \tag{1.158}$$

and inserting it into Equation 1.155, the following is obtained

$$f(t, \alpha) = \sqrt{\frac{2\alpha}{\pi}}\frac{1}{1+\alpha^2 t^2} \tag{1.159}$$

from which α can be determined

$$|f(t, \alpha)|^2 = \frac{1}{2} |f(0, \alpha)|^2 \tag{1.160}$$

$$\frac{1}{1 + \alpha^2 t^2} = \frac{1}{\sqrt{2}} \tag{1.161}$$

$$\sqrt{2} = 1 + \alpha^2 t^2 \tag{1.162}$$

$$\sqrt{2} - 1 = \alpha^2 t^2 \tag{1.163}$$

$$t^2 = \frac{\sqrt{2} - 1}{\alpha^2} \tag{1.164}$$

$$t = \pm \left(\frac{\sqrt{2} - 1}{\alpha^2} \right)^{1/2} \tag{1.165}$$

$$\delta t = 2 \frac{\sqrt{\sqrt{2} - 1}}{\alpha} \quad , \quad \alpha = 2 \frac{\sqrt{\sqrt{2} - 1}}{\delta t} \tag{1.166}$$

whereby substitution into Equation 1.159 yields the expression for the Lorentzian pulse in the time domain

$$f(t, \delta t) = \left(\frac{4\sqrt{\sqrt{2} - 1}}{\pi \delta t} \right)^{1/2} \frac{1}{1 + 4 \left(\sqrt{2} - 1 \right) t^2 / \delta t^2} \tag{1.167}$$

The Fourier transform of Equation 1.167 is given by

$$\mathscr{F}\big(f(t, \delta t)\big) = \tilde{f}(\omega, \delta t) = \int_{-\infty}^{\infty} f(t, \delta t) \, \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t \tag{1.168}$$

Using a similar integral form from the table[18]

$$\int_{-\infty}^{\infty} \frac{1}{1 + at^2} \, \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t = \frac{\pi}{\sqrt{a}} \, \mathrm{e}^{-|\omega|/\sqrt{a}} \tag{1.169}$$

Substitution yields

$$\tilde{f}(\omega, \delta t) = \left(\frac{\pi \delta t}{\sqrt{\sqrt{2} - 1}} \right)^{1/2} \exp\left(\frac{-|\omega|\delta t}{2\sqrt{\sqrt{2} - 1}} \right) \tag{1.170}$$

Finally, determining the time bandwidth product of $\tilde{f}(\omega, \delta t)$,

$$\left| \tilde{f}(\omega, \delta t) \right|^2 = \frac{1}{2} \left| \tilde{f}(0, \delta t) \right|^2 \tag{1.171}$$

$$\exp\left(\frac{-|\omega|\delta t}{2\sqrt{\sqrt{2} - 1}} \right) = \frac{1}{\sqrt{2}} \tag{1.172}$$

$$\frac{|\omega|\delta t}{2\sqrt{\sqrt{2} - 1}} = \frac{\log 2}{2} \tag{1.173}$$

$$|\omega| = \frac{2 \log 2 \, \sqrt{\sqrt{2} - 1}}{2\delta t} \tag{1.174}$$

$$\omega = \pm \frac{2 \log 2 \, \sqrt{\sqrt{2} - 1}}{2\delta t} \tag{1.175}$$

$$\delta\omega = \frac{2 \log 2 \, \sqrt{\sqrt{2} - 1}}{\delta t} \tag{1.176}$$

leaves

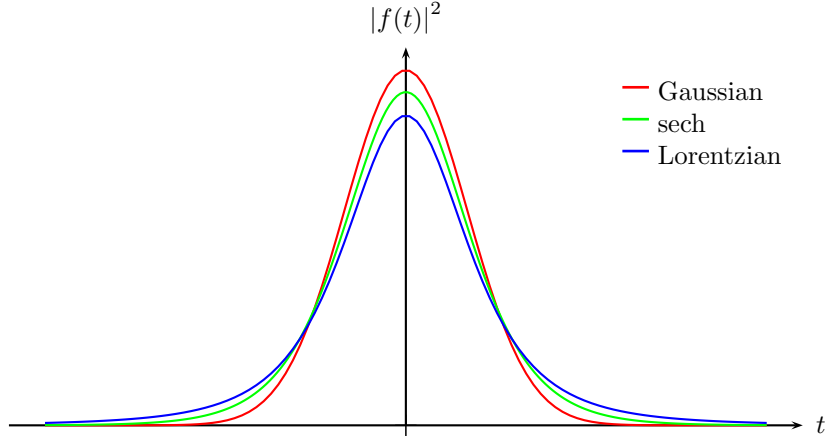$$\delta t \delta\omega = 2 \log 2 \, \sqrt{\sqrt{2} - 1} \approx 0.8922 \tag{1.177}$$

16

Figure 1.4: Plot of $|f(t)|^2$ for the Gaussian, sech, and Lorentzian pulse envelopes. Functions are normalized such that the total area under their curve is unity. Units are arbitrary. ($\delta t = 4.0$)

And in terms or ordinary frequency

$$\delta t \delta \nu = \frac{\delta t \delta \omega}{2\pi} = \frac{\log 2 \sqrt{\sqrt{2} - 1}}{\pi} \approx 0.142 \tag{1.178}$$

The Lorentzian pulse envelope in the time domain does not have a functional form for the proportional energy, as did the Gaussian and the sech pulses. It is therefore left to be evaluated numerically if needed by using the equation

$$\int_{-a}^{a} |f(t, \delta t)|^2 \, \mathrm{d}t = \frac{2}{\pi} \left(\frac{a\alpha}{1 + a^2\alpha^2} + \arctan(a\,\alpha) \right) = X \tag{1.179}$$

In terms of angular frequency however, a closed form can be found to be

$$\int_{-\tilde{a}}^{\tilde{a}} \left| \tilde{f}(\omega, \delta t) \right|^2 \, \mathrm{d}\omega = 2\pi \left(1 - \exp\left(-\sqrt{1 + \sqrt{2}}\, a\, \delta t \right) \right) = 2\pi X \tag{1.180}$$

with a solution

$$\tilde{\zeta}(\delta t, X) = \frac{-\log(1 - X)}{\delta t \sqrt{1 + \sqrt{2}}} \tag{1.181}$$

1.3 Dispersion

During propagation to the focal sphere the plane waves under the integrand of Equation 1.52: local plane waves associated with a light ray, are expected to propagate through an optical system as depicted in Figure 1.1. Naturally, this system may include dispersive optical elements. For such elements, the propagation speed c of the electromagnetic wave is no longer equal to its vacuum speed c_0, rather it is modified by the wavelength dependent refractive index $n(\lambda_0)$ of the material.

$$c = \frac{c_0}{n(\lambda_0)} \tag{1.182}$$

Where λ_0 is the wavelength of light measured in a vacuum. Using the relationships

$$k = \frac{2\pi}{\lambda} = \frac{2\pi\, n(\lambda_0)}{\lambda_0} \tag{1.183}$$

$$k_0 = \frac{2\pi}{\lambda_0} \tag{1.184}$$

$$\lambda = \frac{\lambda_0}{n(\lambda_0)} = k_0 c_0 \tag{1.185}$$

17

and

$$\omega = \frac{2\pi c}{\lambda} = \frac{2\pi c_0}{\lambda_0} \tag{1.186}$$

Note that as presented ω does not depend on the medium through which the component propagates. Along with $k = |\mathbf{k}|$, the wavevector can be factored from Equation 1.54

$$\mathbf{E}(\mathbf{r}', t) = \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\omega) \, e^{i(\mathbf{k} \cdot \mathbf{r}' - \omega t)} \, d\omega \tag{1.187}$$

$$= \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\omega) \, e^{i\,\mathbf{k} \cdot \mathbf{r}'} \, e^{-i\omega t} \, d\omega \tag{1.188}$$

$$= \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\omega) \, e^{i\,n(\omega)\omega/c_0} \, e^{\hat{\mathbf{n}} \cdot \mathbf{r}'} \, e^{-i\omega t} \, d\omega \tag{1.189}$$

In the last step, $\hat{\mathbf{n}}$ is the unit vector along the local ray/plane wave. $n(\lambda_0)$ then accounts for the phase relationship between the frequency components in the integrand. Note that all variables have been shifted to their respective values in a vacuum.

Equation 1.189 governs the pulse behavior through space and time. Since $n(\lambda_0)$ is available for each dielectric, usually given as an experimentally tabulated series, interpolation of such data can be made to numerically determine the effect it has on the pulse structure. Many authors use a series expansion of the dispersion relation to account for this, but we will not. It is fruitful, however, to understand the general effect of the first derivative $dn/d\lambda_0$[42].

- $dn/d\lambda_0 = 0$. All frequency components travel with the same speed. No effects on the pulse are observed.

- $dn/d\lambda_0 < 0$. Known as *normal dispersion*. The index of refraction decreases with wavelength; higher frequencies experience a larger index of refraction than lower frequencies, causing the former to travel more slowly than the later. The consequence of this is a temporal broadening of the pulse.

- $dn/d\lambda_0 > 0$. The opposite of normal dispersion is called *anomalous dispersion*. In such a situation lower frequencies experience a larger index of refraction than higher frequencies.

Figure 1.5 shows the evolution of the structure of a paragon pulse in a dispersive medium. The effects of dispersion can be seen quite clearly: there is a temporal broadening of the pulse width. It can also be observed that the pulse acquires a *chirp* - the spatial separation of the frequency components.

Higher order derivative terms of $n(\lambda_0)$ cause not only temporal broadening, but pulse distortion, as in Figure 1.5. This particular example shows an effect called *self-steepening*, where the frequency components travel at nonlinear speeds in relation to another. This is typically observed for very short pulses, but can also be observed if the pulse travels a great enough distance in a medium.

1.3.1 Nonlinear Regions

Real materials exhibit a permittivity ϵ and permeability μ that vary not only with frequency of the incident electromagnetic field, but also with intensity.

Nonlinearity is typically manifests at electric fields around or greater than $10 \times 10^8 \, \mathrm{V\,m^{-1}}$ – these are easily produced by modern pulsed lasers. This is not surprising: nonlinearity is the operating mechanism behind most ultrafast lasers. Although for all simulations we have assumed the focus is in a vacuum, it is important that we understand these effects and their contribution to the validity of our work.

Nonlinearity in glass is usually estimated by a perturbation of the index of refraction $n(\lambda_0)$ taking into account the intensity of light.

$$n(I, \lambda_0) = n(\lambda_0) + \gamma I \tag{1.190}$$

where γ is some material dependent constant in $\mathrm{m^2/W}$, such that when multiplied by I it remains dimensionless. γ can be stated in terms of a nonlinear refractive index n_2

$$\gamma(\mathrm{m^2/W}) = \frac{40\pi}{c_0 n_0} n_2(\mathrm{esu}) \tag{1.191}$$
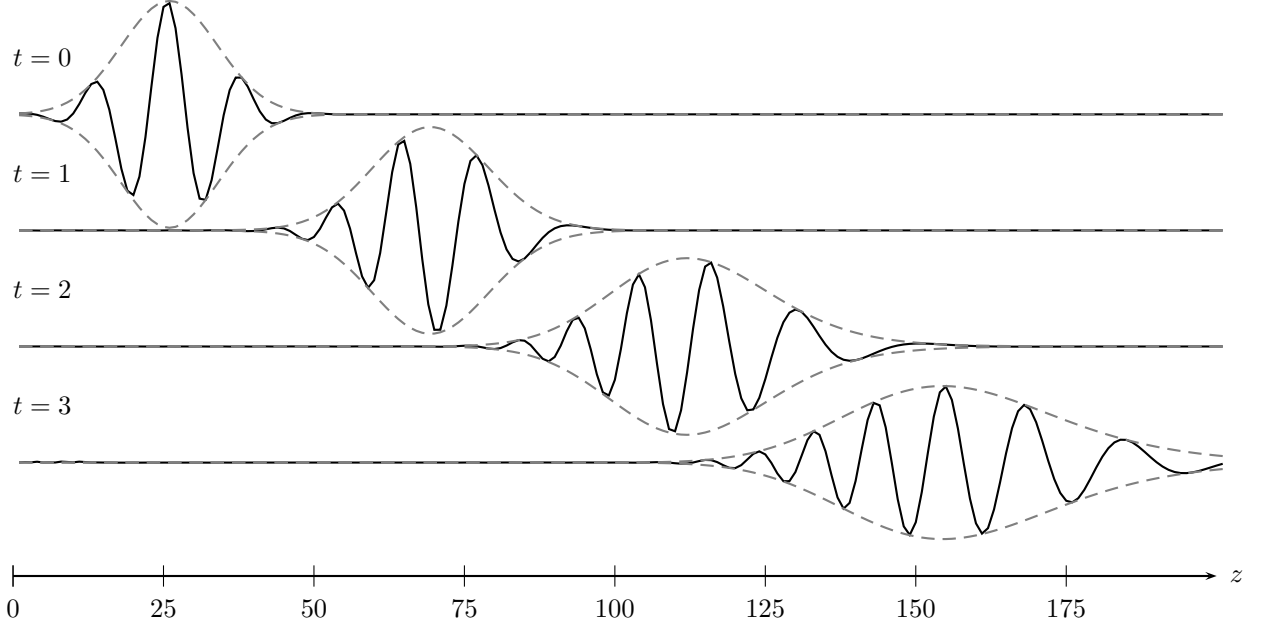
Figure 1.5: Example of pulse evolution in a dispersive medium (BK7) for different times t. Horizontal axis represents distance, and the vertical axis the electric field E. Units are arbitrary. Because BK7 has fairly low higher order dispersion terms, for $t = 0$ to 1 only first order dispersion effects are prevalent. At the extremities however, second order effects are present. Second order dispersion causes not only a broadening of the pulse but a chirp which distorts the overall pulse shape.

where n_2 is given in electrostatic units by the following, somewhat complicated formula [13]

$$n_2(\text{esu}) = \frac{68(n_d - 1)(n_d^2 + 2)^2}{\nu_d \sqrt{1.52 + \frac{(n_d^2 + 2)(n_d + 1)\nu_d}{6 n_d}}} \times 10^{-13}\, \text{esu} \tag{1.192}$$

where n_d and v_d are the refractive index and Abbe number of the material at 587.6 nm, c is the speed of light, n_0 is the refractive index of the material for the vacuum wavelength, and esu is the electrostatic unit, also known as a statcouloumb. Note that I is proportional to the time-averaged electric field, typically

$$I = \frac{c_0 n(\lambda_0) \epsilon_0}{2} |\mathbf{E}|^2 \tag{1.193}$$

The conversion between the units esu and m^2/W in Equation 1.191 is not dimensionally consistent, but it is nonetheless exact. This arises from the definition of esu as a derived unit in the cgs system of measurement.

The validity of our model depends on $\gamma \approx 0$, but it is difficult to state explicitly its error introduced into an arbitrary system. In general it will depend heavily on both the geometry of the optical system and the distribution of intensity within the pulse. Nonetheless, since typical values of γ for common glass types are on the order of $4 \times 10^{-20}\, \text{m}^2\, \text{W}^{-1}$, good results should be expected if the simulations are restricted to $I < 10^{14}\, \text{W}/\text{m}^2$ [13] when working with refractive optics.[1]

1.4 Strength Factor

In a raytracing simulation, the input electromagnetic field $\mathbf{E}_{\text{in}}(\mathbf{r}, t)$ passing through a surface S_{in} (the entrance pupil) is represented by a discrete sampling of rays. After propagation through an optical

---

[1]The change in n using this restriction is less than 10^{-6}.

system, these rays can be used to reconstruct the output field $\mathbf{E}_{\text{out}}(\mathbf{r}, t)$ on its surface S_{out} (the focal sphere).

Quite general to the subject of electromagnetic wave propagation is the concept of polarization. This arises because the plane of oscillation of the electric (and magnetic) field is orthogonal to the direction of wave propagation. In the above this is accounted for by noting that the electric field $\mathbf{E}$ is no longer proportional to what is denoted the polarization vector, $\mathbf{P}_0$ (not to be confused with the polarization density $\mathbf{P}$ from Maxwell's equations). The polarization vector $\mathbf{P}_0$ is a (generally complex) vector connected to each ray/local plane wave in the Debye integral. Similar to a Jones vector, it represents the local polarization state. However, it is the surface density of $\mathbf{P}_0$ which represents the local electric field; by changing the number of rays which sample the wave, the value of the electric field will not change, but the value of the polarization vector will. Accordingly, $\mathbf{E}$ is equal to the polarization vector per surface element $\mathrm{d}s$.

$$\mathbf{E} = \frac{\mathbf{P}_0}{\mathrm{d}s} \tag{1.194}$$

This comes from the property that in a numerical simulation, where the incident field is sampled by a discrete number of local plane waves (rays), the spacing between these points will change during refraction in the optical system. Since these local plane waves are superimposed at the focus, the electric field of a local plane wave itself is proportional to the density of the polarization vector of the associated ray. Since the light power $\mathrm{d}W = I\,\mathrm{d}s$ transported by each ray is constant[2]

$$I\,\mathrm{d}s = \text{constant} \tag{1.195}$$

and

$$I = |\mathbf{E}|^2 = \frac{|\mathbf{P}_0|^2}{(\mathrm{d}s)^2} \tag{1.196}$$

the proportionality can be accounted for by introducing a dimensionless "strength factor" g, sometimes called the *apodization*, defined as the ratio of the modulus squared of the polarization vector of a ray exiting the optical system, $|\mathbf{P}_{\text{out}}|^2$ to the modulus squared of the polarization vector entering it, $|\mathbf{P}_{\text{in}}|^2$

$$g^2 = \frac{|\mathbf{P}_{\text{out}}|^2}{|\mathbf{P}_{\text{in}}|^2} = \frac{\mathrm{d}s_{\text{out}}}{\mathrm{d}s_{\text{in}}} \tag{1.197}$$

There are several ways one can determine g:

- If the principle radii of curvature of the wavefronts R_1 and R_2 assigned to each ray are known, then

$$g^2 = \frac{R_{1,\text{out}}\, R_{2,\text{out}}}{R_{1,\text{in}}\, R_{2,\text{in}}} \tag{1.198}$$

  This can be achieved through differential raytracing.

- By constructing a Voronoi tessellation of the rays on the surfaces S_{in} and S_{out} and calculating the areas of the resultant polygons.

- For simple systems, the strength factor can be calculated analytically.

1.4.1 Analytic Derivation of the Strength Factor for Simple Systems

As mentioned in the previous section,

$$g = \sqrt{\frac{\mathrm{d}s_{\text{out}}}{\mathrm{d}s_{\text{in}}}} \tag{1.199}$$

---

[2]This is true assuming no absorption and no losses by e.g. Fresnel reflection, which have to be taken into account separately.
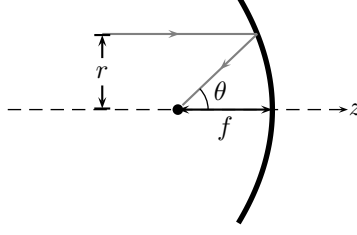
Figure 1.6: Schematic representation of the geometry for a parabolic mirror, radially symmetric about the optical axis (z). A ray, incident from the left at distance r from the optical axis is reflected by the mirror and propagates to the focus with focal length f, making an angle θ with the optical axis.

For a two dimensional circular aperture of rays incident on a plane, the surface area element is, in polar coordinates, $\mathrm{d}s_\mathrm{in} = 2\pi r\,\mathrm{d}r$. On the focal sphere the surface area element is given in terms of spherical coordinates by

$$\mathrm{d}s_r = r^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \tag{1.200}$$

If rays are traced to the focal sphere with surface S_out as in Figure 1.1, the surface area element on the focal sphere is, by integrating over the azimuthal angle ϕ

$$\mathrm{d}s_\mathrm{out} = 2\pi f^2 \sin\theta\,\mathrm{d}\theta \tag{1.201}$$

We now proceed to derive g on the focal sphere S_out for a circular aperture of rays on the entrance pupil S_in. Three simple optical elements will be addressed, from which we can make assessments on the validity of numerically computed values. Each of these are shown comparatively in Figure 1.9

Parabolic Mirror

The geometry for the parabolic mirror is shown in Figure 1.6, where z is the optical axis. Assume a ray which is incident onto the parabolic mirror, parallel to the optical axis with a radial coordinate r. Then, the angle θ of the reflected ray with respect to the optical axis fulfills the equation

$$r = 2f \tan\left(\frac{\theta}{2}\right) \tag{1.202}$$

and

$$\mathrm{d}r = f \sec^2\left(\frac{\theta}{2}\right)\mathrm{d}\theta \tag{1.203}$$

then

$$g^2 = \frac{\mathrm{d}s_\mathrm{out}}{\mathrm{d}s_\mathrm{in}} \tag{1.204}$$

$$= \frac{2\pi f^2 \sin\theta\,\mathrm{d}\theta}{2\pi r\,\mathrm{d}r} \tag{1.205}$$

$$= \frac{2\pi f^2 \sin\theta\,\mathrm{d}\theta}{2\pi \left(2f \tan\left(\frac{\theta}{2}\right)\right)\left(f \sec^2\left(\frac{\theta}{2}\right)\mathrm{d}\theta\right)} \tag{1.206}$$

$$= \frac{\sin\theta}{2 \tan\left(\frac{\theta}{2}\right) \sec^2\left(\frac{\theta}{2}\right)} \tag{1.207}$$

$$\tag{1.208}$$

First, noting the identity $\sec^2\alpha = 1/\cos^2\alpha$, g^2 becomes

$$g^2 = \frac{\sin\theta \cos^2\left(\frac{\theta}{2}\right)}{2 \tan\left(\frac{\theta}{2}\right)} \tag{1.209}$$
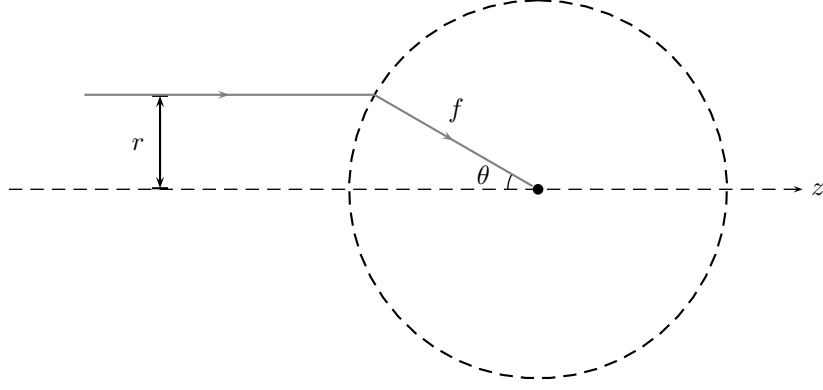
21

Figure 1.7: Schematic representation of an aplanatic system, radially symmetric about the optical axis z. The system's exit pupil is a sphere around the focus. A ray incident from the left at a distance r from the optical axis is deflected by the system, producing a spherical exit pupil around the focus. The focal length is denoted by f and the angle the ray makes with the optical axis θ.

Applying the double angle formula $\sin\alpha = 2\sin\left(\frac{\alpha}{2}\right)\cos\left(\frac{\alpha}{2}\right)$ to the numerator and the identity $\tan\alpha = \sin\alpha/\cos\alpha$ to the denominator,

$$g^2 = \frac{2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\cos^2\left(\frac{\theta}{2}\right)}{2\tan\left(\frac{\theta}{2}\right)} \tag{1.210}$$

$$= \frac{2\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\cos^2\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)}{2\sin\left(\frac{\theta}{2}\right)} \tag{1.211}$$

$$= \cos^4\left(\frac{\theta}{2}\right) \tag{1.212}$$

$$g = \frac{1+\cos\theta}{2} \tag{1.213}$$

Aplanatic System

An aplanatic system is general to any system whose exit pupil is a sphere around the focus. Consequently, such a system is assumed to follow the Abbe sine condition. Such a system is depicted in Figure 1.7 In this case, the relationship in r

$$r = f\sin\theta \tag{1.214}$$

and

$$\mathrm{d}r = f\cos\theta\,\mathrm{d}\theta \tag{1.215}$$

are such that

$$g^2 = \frac{\mathrm{d}s_{\mathrm{out}}}{\mathrm{d}s_{\mathrm{in}}} \tag{1.216}$$

$$= \frac{2\pi f^2\sin\theta\,\mathrm{d}\theta}{2\pi r\,\mathrm{d}r} \tag{1.217}$$

$$= \frac{2\pi f^2\sin\theta\,\mathrm{d}\theta}{2\pi\left(f\sin\theta\right)\left(f\cos\theta\,\mathrm{d}\theta\right)} \tag{1.218}$$

$$= \frac{\sin\theta}{\sin\theta\cos\theta} \tag{1.219}$$

$$g = \sqrt{\sec\theta} \tag{1.220}$$

Flat Diffractive Element

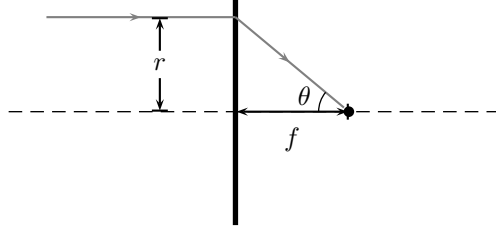A flat diffractive element is described geometrical arrangement in Figure 1.8. In such an system, light

Figure 1.8: Schematic representation of a flat diffractive element, radially symmetric about the optical axis z. A ray originating from the left with distance r from the optical axis is deflected towards the focus, making an angle θ with said axis. The focal length of the system is denoted by f.
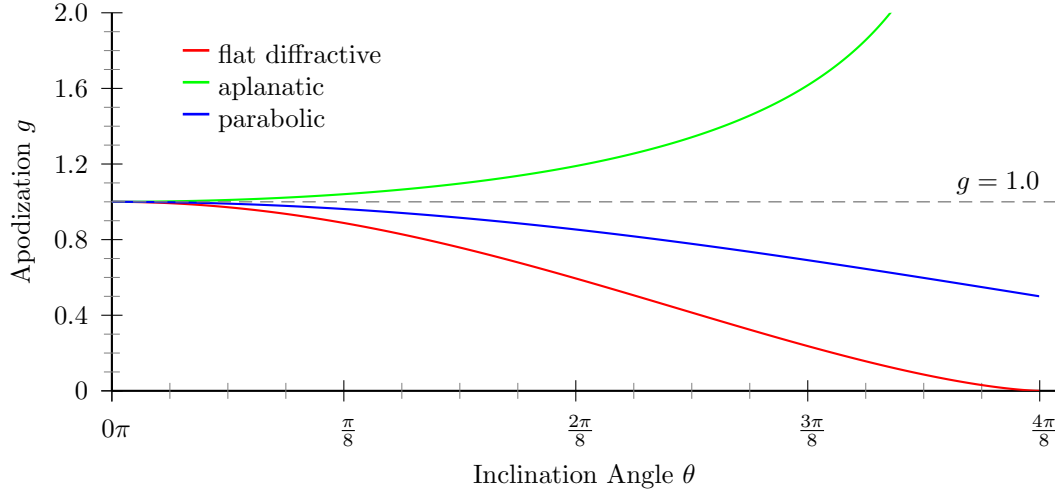

Figure 1.9: Comparison of the apodization factor g on the focal sphere for different optical elements illuminated by a circular aperture (Equations 1.228, 1.213, and 1.220).

from the plane of the element is focused to a point, so

$$r = f \tan \theta \tag{1.221}$$

$$\mathrm{d}r = f \sec^2 \theta \, \mathrm{d}\theta \tag{1.222}$$

then

$$g^2 = \frac{\mathrm{d}s_{\text{out}}}{\mathrm{d}s_{\text{in}}} \tag{1.223}$$

$$= \frac{2\pi f^2 \sin \theta d\theta}{2\pi r \, \mathrm{d}r} \tag{1.224}$$

$$= \frac{2\pi f^2 \sin \theta d\theta}{2\pi \left(f \tan \theta \right) \left(f \sec^2 \theta \, \mathrm{d}\theta \right)} \tag{1.225}$$

$$= \frac{\sin \theta}{\tan \theta \sec^2 \theta} \tag{1.226}$$

$$= \cos^3 \theta \tag{1.227}$$

$$g = \cos^{3/2} \theta \tag{1.228}$$

1.5   Raytracing

We are now in a position where we can express the Debye integral in a form appropriate for numerical calculation. This is done through a reformulation of such as a discrete sum over rays on the focal sphere. Recall the Debye integral of Equation 1.52

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{\mathrm{i}\lambda f_b} \iint_S \mathbf{E}(\mathbf{r}',t)\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{r})}\,\mathrm{d}s \tag{1.229}$$

where each plane wave on the focal sphere $\mathbf{E}(\mathbf{r}',t)$ was of the form

$$\mathbf{E}(\mathbf{r}',t) = \tilde{\mathbf{E}}(\omega)\,\mathrm{e}^{\mathrm{i}\left(\mathbf{k}\cdot\mathbf{r}'-\omega t\right)} \tag{1.230}$$

In terms of rays, note that the vector $\mathbf{k}$ includes information of the local ray direction.

$$\mathbf{k} = \frac{2\pi}{\lambda}\hat{\mathbf{n}} \tag{1.231}$$

Where $\hat{\mathbf{n}}$ is the normal unit vector to the plane wave. To include pulse behavior as discussed in Section 1.2, Equation 1.52 is substituted into Equation 1.54.

$$\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{\infty} \frac{1}{\mathrm{i}\lambda f_b} \iint_S \mathbf{E}(\mathbf{r}',t)\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{r})}\,\mathrm{d}s\,\mathrm{d}\omega \tag{1.232}$$

From the Debye integral it can be seen that the exciting field $\mathbf{E}(\mathbf{r}',t)$ is calculated from the optical path length (OPL) of each ray (computed according to Snell's law), and by taking into account the time-dependent term $\exp(-\mathrm{i}\omega t)$. Including these explicitly in $\mathbf{E}(\mathbf{r}',t)$, the Debye integral for pulses becomes

$$\boxed{\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\omega)\,\frac{n(\lambda_0)}{\mathrm{i}\lambda_0 f_b} \iint_S \mathrm{e}^{\mathrm{i}\,n(\lambda_0)\,(\mathbf{k}_0\cdot(\mathbf{r}'-\mathbf{r})-k_0\cdot\mathrm{OPL})}\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}s\,\mathrm{d}\omega} \tag{1.233}$$

(Debye integral for pulses)

Equation 1.233 can be further expressed as a discrete sum suitable for numerical computation. The most basic way to accomplish this is through a definition of the definite integral

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{N\to\infty} \sum_{j=1}^{N} f(x_j^*)\Delta x \tag{1.234}$$

where a and b are the limits of integration, $f(x_j^*)$ is the function evaluated at a discrete value x_j^*, and Δx is the step size. In this light Equation 1.233 becomes [3]

$$\mathbf{E}(\mathbf{r},t) = \lim_{N\to\infty} \sum_{j=1}^{N} \left\{ \frac{n(\lambda_0)}{\mathrm{i}\lambda_0 f_b}\,\tilde{\mathbf{E}}(\omega)\,\mathrm{e}^{\mathrm{i}\,n(\lambda_0)\,(\mathbf{k}_0\cdot(\mathbf{r}'-\mathbf{r})-k_0\cdot\mathrm{OPL})}\,\mathrm{e}^{-\mathrm{i}\omega t}\,\Delta s\,\Delta\omega \right\}_j \tag{1.235}$$

The final modification to the Debye summation is to include the conservation of energy consequence of ray propagation, the apodization factor g of Equation 1.199. Including this factor and rewriting in terms of the polarization vector $\mathbf{P}_0$ (Equation 1.194), the final form of our summation becomes

$$\boxed{\mathbf{E}(\mathbf{r},t) = \lim_{N\to\infty} \sum_{j=1}^{N} \left\{ \mathbf{P}_0\,\frac{g\,n(\lambda_0)}{\mathrm{i}\lambda_0 f_b}\,\mathrm{e}^{\mathrm{i}\,n(\lambda_0)\,(\mathbf{k}_0\cdot(\mathbf{r}'-\mathbf{r})-k_0\cdot\mathrm{OPL})}\,\mathrm{e}^{-\mathrm{i}\omega t}\,\Delta\omega \right\}_j} \tag{1.236}$$

(Debye sum for pulses)

If certain conditions are met, it is possible to greatly speed up the numerical computation of Equation 1.236 by using a *fast Fourier transform* (FFT). This is discussed in Section 2.1.

---

[3]Here we have used the bracket and subscript to consolidate the indices. For example, $\alpha_j \cdot \beta_j = \{\alpha \cdot \beta\}_j$ in this notation.

Chapter 2

Implementation

2.1 The Fast Fourier Transform

In Section 1.5 we expressed the Debye integral as a discrete sum of rays on the focal sphere. In this chapter it is our aim to discuss how the Debye sum of Equation 1.236 can be computed numerically with a computer.

For a finite number of sampling points N, Equation 1.236 takes the form of a discrete Fourier transform (DFT).

$$Y_k = \sum_{j=1}^{N} X_j \, e^{2\pi jk\sqrt{-1}/N} \tag{2.1}$$

Where the variables j and k are indices (not variables of a function) and we have written $\sqrt{-1} = \mathrm{i}$ explicitly to avoid confusion. The DFT maps a set of discrete complex input data $X = \{X_1 \ldots X_j : X_j \in \mathbb{Z}\}$ onto a set of discrete complex output data $Y = \{Y_1 \ldots Y_k : Y_k \in \mathbb{Z}\}$. In the context of Equation 1.236 this represents a mapping (within our approximation) of ray information on the focal sphere to the electric field in a small volume near the focus.

Direct computation of a DFT is of $O(N^2)$ complexity for N sampling points. However, because the DFT data must be equidistantly and linearly sampled with respect to its variables, a fast Fourier transform (FFT) can be used. The FFT is able to compute sums of the form of Equation 2.1 as a $O(N \log N)$ complexity problem. At the beginning of this work, it was thought that an FFT would be the natural course of action. However, certain limitations of the method soon became readily apparent. Two programs, `fftdebye` and `directdebye` were developed in tandem operating off the same data. The former to compute the fast Fourier transform and the later to do direct computation. Though the requirements for `directdebye` seem a priori to be computationally more demanding[1], it was chosen as the method used to compute nearly all of the examples in this work. Note that when use the term "equidistant sampling" we refer to linear sampling in each of the dimensions, independent of one another.

2.1.1 Memory Considerations

The electric field in the immediate vicinity of the focus is represented in four dimensional parameter space: three Cartesian and one time (x, y, z, and t). The minimum amount of memory required to fully represent this region with a 16 B complex double precision floating point number[2]

$$N_{\text{total}} = N_x \cdot N_y \cdot N_z \cdot N_t \cdot 6 \cdot 16 \text{ B} \tag{2.2}$$

where N_d is the number of samples along the respective dimension $\{d : x, y, z, t\}$ and the factor of 6 comes from the three electric field components E_x, E_y, and E_z, which each require both an input and output array. In place fast Fourier transforms are only known for certain[3] array sizes.

---

[1] This sentence is a tiny bit misleading. Refer to the discussion at the end of Subsection 2.2.2
[2] B is the unit of bytes, 8 bits = 1 byte.
[3] `directdebye` was written using the FFTW [14] library, which can only perform in-place transforms for sizes of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 32 and 64 along each dimension.

Equation 2.2 could potentially represent an enormous amount of memory on present day computers. For the modest case of $N_d = 128$, $128^4 \cdot 6 = 24$ GB. While not beyond modern computing capabilities, this number can be reduced a bit. First, if one is interested in only one component of the electric field vector, say $|E_x|$, $|E_y|$, or $|E_z|$, the factor of six can be reduced to a factor of two. Additionally, for the calculation of $\mathbf{E}$, in many cases only slices perpendicular or parallel to the optical axis are considered, allowing three dimensional slices in (x, y, t) for specific values along the z-axis (or for a fixed x or y value) to be computed. In this case, Equation 2.2 can be amended to $N_x \cdot N_y \cdot N_t \cdot 2 \cdot 16$ B, reducing the memory requirements for $N_{\text{total}} = 128$ to $128^3 \cdot 2 \cdot 16$ B $= 64$ MB per slice – a considerable improvement!

The choice of computing specific z slices in this reduction was not arbitrary. For many optical systems, the total illumination about the focal sphere will be less than or equal to 2π solid angle, or one half of the sphere. From this constraint, each ray $\tilde{\mathbf{E}}(x', y', z' > 0, t)$ then has a unique representation in terms of their spatial coordinates x' and y'. This also implies each ray has a unique representation in terms of its k-vector. Since we defined

$$\mathbf{k}_0 = \frac{2\pi}{\lambda_0}\hat{\mathbf{n}} \tag{2.3}$$

and each ray propagates from its intersection $\mathbf{r}' = (x', y', z')$, $z' = \sqrt{f^2 - (x'^2 + y'^2)}$, with the focal sphere to a point close to $\mathbf{r} = (0, 0, 0)$, geometrically equidistant sampling in terms of x' and y' is equivalent to sampling k_x and k_y equidistantly, albeit multiplied by a constant equal for all rays of the same wavelength.

2.1.2 Units in the Fast Fourier Transform

As discussed, the fast Fourier transform makes a map on the set of numbers $X_j \mapsto Y_k$ and vice versa. There are a symmetric set of equations that define the relationship more succinctly. Let's say you take N samples of the spatial coordinate Y with bounds Y_a and Y_b: $\{Y \in \mathbb{R} : Y_a \leq Y \leq Y_b\}$. The spacing Y_Δ between Y_k and Y_{k+1} is just the range $Y_{\text{range}} = Y_b - Y_a$ divided by the number of samples N.

$$Y_\Delta = \frac{Y_{\text{range}}}{N} \tag{2.4}$$

The resultant transform in terms of X has units inverse to that of Y

$$X_\Delta = \frac{1}{Y_{\text{range}}} \tag{2.5}$$

and

$$X_{\text{range}} = \frac{1}{Y_\Delta} = \frac{N}{Y_{\text{range}}} \tag{2.6}$$

Restated, these relationships are

$$Y_\Delta = \frac{Y_{\text{range}}}{N} \qquad Y_{\text{range}} = \frac{1}{X_\Delta} \tag{2.7}$$

$$X_\Delta = \frac{1}{Y_{\text{range}}} \qquad X_{\text{range}} = \frac{1}{Y_\Delta} \tag{2.8}$$

And in terms of the ranges of the data

$$Y_\Delta = \frac{Y_b - Y_a}{N} \qquad Y_{\text{range}} = Y_b - Y_a \tag{2.9}$$

$$X_\Delta = \frac{1}{Y_b - Y_a} \qquad X_b - X_a = \frac{N}{Y_b - Y_a} \tag{2.10}$$

There is also a kind of uncertainty relationship between them which can be obtained by rewriting Equation 2.8

$$Y_{\text{range}} X_{\text{range}} = N \qquad Y_\Delta X_\Delta = \frac{1}{N} \tag{2.11}$$

In the context of the DFT, the variables ν and t and the function values f and $\tilde{f}$ are indexed by k and j to discrete samples ν_j, t_k, F_k, and $\tilde{f}_j$. Since

$$t_k = k \, \Delta t \quad \text{and} \quad \nu_j = j \, \Delta \nu \tag{2.12}$$

$$t_k \nu_j = k \, \Delta t \, j \, \Delta \nu \tag{2.13}$$

$$= jk \frac{\tilde{f}_{\text{range}}}{N} \frac{1}{\tilde{f}_{\text{range}}} \tag{2.14}$$

$$= \frac{jk}{N} \quad , \quad \Delta t \Delta \nu = \frac{1}{N} \tag{2.15}$$

Which is exactly the condition imposed by the DFT.

Note that only such sampling allows the DFT of Equation 2.1 to be evaluated numerically. Take for example, the Fourier transform of unitary and ordinary frequency

$$f(t) = \int_{-\infty}^{\infty} \tilde{f}(\nu) \, e^{-i2\pi\nu t} \, d\nu \tag{2.16}$$

As a sum, this integral would be, according to Equation 1.234,

$$f(t) = \lim_{N \to \infty} \sum_{j=1}^{N} \tilde{f}(\nu_j) \, e^{-i2\pi\nu_j t} \Delta \nu \tag{2.17}$$

2.1.3 Sampling Condition for the DFT

Because the DFT composes functions using a finite summation of complex exponentials, there is an unavoidable periodicity in its approximation. This is known as *aliasing*. In order to avoid such effects, proper consideration must be taken when sampling the variables $\mathbf{k}$ and ω on the focal sphere.

Spatial Domain

The Fourier transform of the one dimensional spatial coordinate r is given by

$$f(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) \, e^{ikr} \, dk \tag{2.18}$$

Or, written as a discrete sum over N components,

$$f(r) = \frac{\Delta k}{2\pi} \sum_{j=1}^{N} \tilde{f}(k_j) \, e^{ik_j r} \tag{2.19}$$

$$= \frac{\Delta k}{2\pi} \sum_{j=1}^{N} \tilde{f}(j\Delta k) \, e^{i(j\Delta k)r} \tag{2.20}$$

where $i = \sqrt{-1}$. Note that the complex exponential will repeat itself when

$$\Delta k \, r = 2\pi \tag{2.21}$$

The value of Δk is just the range of k divided by the number of samples taken. Or, if the focal sphere is rotationally symmetric about the optical axis

$$\Delta k = \frac{2\bar{k}}{N} \tag{2.22}$$

where $\bar{k}$ has been used to indicate the maximal value of k: $\bar{k} = \max(k)$. There is a useful relationship between $\bar{k}$ and the numerical aperture of a system. Let k and r now be the three dimensional vectors $\mathbf{k} = (k_x, k_y, k_z)$ and $\mathbf{r} = (r_x, r_y, r_z)$ on the focal sphere. Assume the focal sphere is centered about the

origin of the coordinate system, and furthermore that it is symmetric with regards to the azimuthal coordinate ϕ. The maximal values of k_x and k_y are

$$\bar{k}_x = \frac{2\pi}{\lambda} \sin\theta \qquad\qquad \bar{k}_y = \frac{2\pi}{\lambda} \sin\theta \qquad (2.23)$$

then, by Equation 2.22

$$\Delta k_x = \frac{4\pi \sin\theta}{\lambda N_x} \qquad\qquad \Delta k_y = \frac{4\pi \sin\theta}{\lambda N_y} \qquad (2.24)$$

Note that the components of the k-vector are related to each other by

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \left(\frac{2\pi}{\lambda}\right)^2 \qquad (2.25)$$

As such, the sampling in k_z will be neglected because it is automatically determined for a fixed k_x and k_y[4]. To determine how many samples must be taken to avoid aliasing, the condition of Equation 2.21 is imposed on each of the components

$$\Delta k_x \bar{r}_x = 2\pi \qquad\qquad \Delta k_y \bar{r}_y = 2\pi \qquad (2.26)$$

$$N_x = \frac{4\pi\, \bar{r}_x \sin\theta}{2\pi\lambda} \qquad\qquad N_y = \frac{4\pi\, \bar{r}_y \sin\theta}{2\pi\lambda} \qquad (2.27)$$

$$= \frac{2\, \bar{r}_x \sin\theta}{\lambda} \qquad\qquad = \frac{2\, \bar{r}_y \sin\theta}{\lambda} \qquad (2.28)$$

Alternatively, since $\mathrm{NA} = \sin\theta$, the equations can be rewritten

$$N_x = \frac{2\, \bar{r}_x \,\mathrm{NA}}{\lambda} \qquad\qquad N_y = \frac{2\, \bar{r}_y \,\mathrm{NA}}{\lambda} \qquad (2.29)$$

As an example, to compute a $200 \times 200\,\mu\mathrm{m}$ plane along the r_x and r_y with $r_z = 0$, assuming $\mathrm{NA} = 0.5$ and $\lambda = 800\,\mathrm{nm}$, Equation 2.28 predicts 250×250 samples on the focal sphere to avoid aliasing. This is of course limited by the same condition used to approximate Equation 1.52

$$f \gg \frac{r^2}{\lambda} = \frac{\sqrt{2} \cdot 200\,\mu\mathrm{m}^2}{0.8\,\mu\mathrm{m}} \approx 35\,\mathrm{mm} \qquad (2.30)$$

Where the factor of $\sqrt{2}$ was used to take into account the diagonal of the quadratic field.

Time Domain

In the time domain the sampling requirements follows the same rubric, namely aliasing will occur if $\Delta\omega\,\bar{t} = 2\pi$. To compute the minimum number of sampling points, it is first required to impose boundary values for the pulse. Let $\tilde{\zeta}(\delta t, X)$ be defined as a function which returns the symmetric bounds for which the infinite integral over the magnitude squared electric field pulse envelope in the angular frequency domain would return the value X

$$\int_{-\tilde{\zeta}(\delta t, X)}^{\tilde{\zeta}(\delta t, X)} |E(\omega)|^2 \; \mathrm{d}\omega = X \qquad (2.31)$$

This assumes the integral of $|E(\omega)|^2$ over the entire domain is unity. As an example, for a Gaussian pulse this function would be

$$\tilde{\zeta}(\delta t, X) = \frac{2\sqrt{\log 2}}{\delta t} \,\mathrm{inverf}(X) \qquad (2.32)$$

---

[4]For focusing systems with 4π solid angle, the sign of k_z will be different on each half sphere. This is handled by treating the spheres separately, adjusting the sign of k_z as appropriate

The value returned by $\tilde{\zeta}(\delta t, X)$ then represents half the maximum value on the sampling domain

$$\Delta\omega = \frac{2\,\tilde{\zeta}(\delta t, X)}{N_t} \tag{2.33}$$

Solving for N_t,

$$\Delta\omega\,\bar{t} = 2\pi \tag{2.34}$$

$$\frac{2\tilde{\zeta}(\delta t, X)}{N_t}\,\bar{t} = 2\pi \tag{2.35}$$

$$N_t = \frac{\tilde{\zeta}(\delta t, X)\,\bar{t}}{\pi} \tag{2.36}$$

As an example, Equation 2.36 will be used to compute N_t for a 40 fs pulse in a rectangular region of $200 \times 400\,\mu\mathrm{m}$. The maximal distance diagonally across this region is $\sqrt{5} \cdot 200\,\mu\mathrm{m}$. In a vacuum this represents a maximal traversal time of

$$\bar{t} = \frac{\sqrt{5} \cdot 200\,\mu\mathrm{m}}{c_0} \approx 1.4917 \times 10^{-12}\,\mathrm{s} \tag{2.37}$$

Using $X = 0.99999$, the equation predicts about 62 samples required along the bounds set by $\tilde{\zeta}(\delta t, X)$.

2.2 Interpolation

As stated in the beginning of the chapter, for the evaluation of the FFT the input data must be equidistantly sampled with respect to $\mathbf{k}$ and ω (or k_x, k_y, and ω (or consequently, x', y' and ω)). This represents a significant challenge. For most optical systems, equidistant sampling of rays on the entrance pupil produces non-equidistant ray intersections on the focal sphere which propagate to the focus. This is demonstrated schematically in Figure 2.1[5]. This means the values of $\mathbf{k}$ on the focal sphere will not be linearly spaced as required for computation of the fast Fourier transform.

In general, there are three methods able to address these issues which have been studied:

- By "choosing" the rays on the entrance pupil such that they intercept the focal sphere as required.

- With interpolation of the ray data.

- For aplanatic systems this requirement is fulfilled automatically. In the (two dimensional) entrance pupil of such a system, the rays are given by the coordinates

$$r = f \sin\theta \tag{2.38}$$

$$x = r \cos\phi = f \cos\phi \sin\theta \tag{2.39}$$

$$y = r \sin\phi = f \sin\phi \sin\theta \tag{2.40}$$

$$\tag{2.41}$$

and in terms of their spatial frequencies

$$k_r = \frac{2\pi}{\lambda} \sin\theta \tag{2.42}$$

$$k_x = k_r \cos\phi = \frac{2\pi}{\lambda} \cos\phi \sin\theta \tag{2.43}$$

$$k_y = k_r \sin\phi = \frac{2\pi}{\lambda} \sin\phi \sin\theta \tag{2.44}$$

Therefore via the sine condition, if the incident field is equidistant in spatial coordinates, the field on the focal sphere is also equidistant in spatial frequencies.

These methods will be the subject of our present discussion.

---

[5]Note the convention that when a small globe is placed in the lower left-hand corner, that figure is projection of a three dimensional object. Two dimensional objects are represented by a circle.
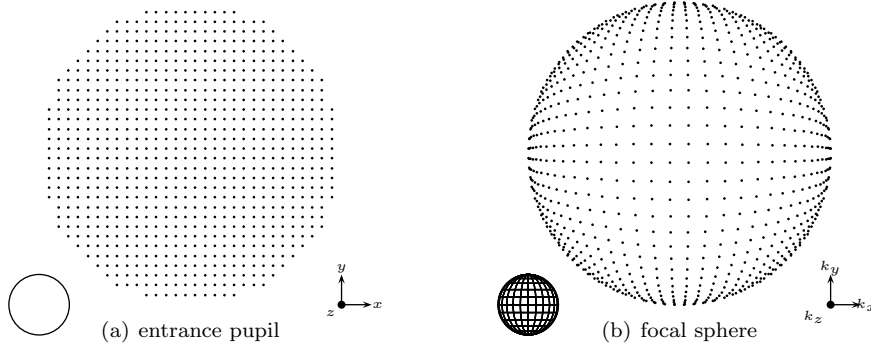
Figure 2.1: Statement of problem: Equidistantly spaced rays on the two dimensional planar entrance pupil (left) usually produce non-equidistantly spaced rays on the three dimensional focal sphere (right). Both are three dimensional projections along the same plane, but while the entrance pupil has its rays distributed in an essentially two dimensional plane, the focal sphere has its rays distributed in three dimensions. What appears to be wavefront warping on the focal sphere is an artifact of this particular projection. (Parabolic mirror, NA = 0.5).

Algorithm	CPU Time (s)	Iterations
Nelder-Mead[21]	0.086	191.636
Subplex[35]	0.147	310.438
Principal Axis[8]	0.148	334.142
COBYLA[29]	0.162	187.053
BOBYQA[30]	0.166	109.45
NEWUOA[28]	0.220	100.989
CRS2[21]	1.132	2234.08
DIRECT-L[16]	5.328	1177.91
ISRES[38]	22.397	23478.4

Table 2.1: Performance of various algorithms for the ray optimization problem. 1024 points were chosen to optimize. The column "Iterations" is the number of rays that were traced and consequently, iterations of the algorithm, required before the stopping criteria was met. Accordingly, the first row of the table would be interpreted as "The Nelder-Mead algorithm took a total of 0.086 s to optimize 1024 points, averaging 191.636 iterations per point, and approximately $191.636 \times 1024 \approx 196235$ rays had to be traced in total."

2.2.1 Optimization of Initial Rays

Although the mapping $S_{\text{in}} \to S_{\text{out}}$ can in some cases be derived analytically, for the sake of generality we will not attempt to do so. Instead, for each desired sampling point on the focal sphere we search for the ray which most closely intersects the focal sphere at that point. This minimization process was done using algorithms available through the nlopt library [20]. Table 2.1 shows the performance of several derivative-free algorithms available through the library for 1024 rays in a 32×32 grid focused by a parabolic mirror, sorted by CPU time. The last column, "Iterations" describes the mean number iterations (and consequently, the number of rays that had to be traced) for a given point before the stopping criteria was met. For simple cases such as the parabolic mirror, we expect the CPU time to be dominated by the overhead associated with the optimization algorithm. For more complicated systems, it is possible that the raytracing routines would dominate. This might effect the choice of algorithm.

One could imagine a system where two rays on the entrance pupil mapped to the same location on the focal sphere. We have not studied such systems, nevertheless to mitigate such effects it would be possible to introduce a penalty in the optimization algorithm's fitness function to account for this - a reduction in the fitness for choosing a ray in a location too close to a ray which has already had its
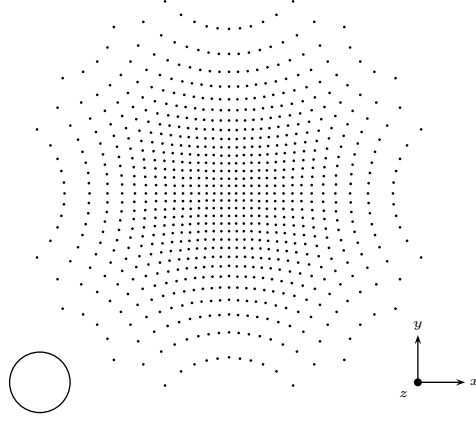
Figure 2.2: Predicted two dimensional distribution of rays in the entrance pupil in order to obtain equidistant sampling in k_x and k_y on the focal sphere for a parabolic mirror (Figure 3.1, $f = 50\,\text{mm}$, $\text{NA} = 0.5$, $\lambda_0 = 800\,\text{nm}$). For the above plot: Mean error $6.9084 \times 10^{-16}\,\text{mm}$, standard deviation $2.1166 \times 10^{-16}\,\text{mm}$, maximum error $9.9922 \times 10^{-16}\,\text{mm}$, and minimum error $1.1311 \times 10^{-16}\,\text{mm}$.

position determined. It could also be argued that such systems are of no practical consequence to this study as this would physically represent wavefront warping on the focal sphere. In such a case, the Debye approximation breaks down and the numerical calculation should be suspended. In Figure 2.2 we have used the Nelder-Mead algorithm to predict the distribution of rays on the entrance pupil for a parabolic mirror[6][7].

2.2.2 Interpolation

A more obvious method used to facilitate the FFT is an interpolation of ray data on the focal sphere. For this two methods have been studied. The first is based on tension cubic spline interpolation on a unit sphere[33] [34], made available part of the `ngmath` library using `cssgrid` (cubic spline sphere gridder). These packages were originally developed for interpolation of atmospheric data, and as such have coefficients which account for the earth's deviations from a spherical shape. In the numerical implementation, these coefficients have been changed or ignored to assure a true unit sphere.

The second is interpolation using radial basis functions (RBFs)[9]. RBFs are able to operate on scattered 3D datasets (point cloud data). Our investigation was done using a `c` interface to the FORTRAN library FITPACK[11]. As many systems do not have more than 2π total illumination, each ray position has a unique representation in x' and y' for one of either $z' < 0$ or $z' > 0$. This can be used to our advantage; RBFs can be used on a two dimensional surface in Cartesian coordinates (though at a cost of some accuracy at the rim for high numerical aperture systems).

Interpolation of ray data requires separate consideration of each component of ray information. To do this, it is assumed each ray can be uniquely defined by a set of Cartesian coordinates, x' and y' being that rays's intersection with the focal sphere. The points on the focal sphere of which to obtain values are then determined. Each ray becomes a 2-dimensional data point in a function $f(x', y', var)$, where var is one of the 9 variables in Table 2.2. Accordingly, each point on the focal sphere requires 9 separate interpolations. This table assumes the complex polarization vector $\mathbf{P}_0$ contains the combined information from itself, the apodization factor, and any material absorption.

Complex Interpolation

Interpolation of complex numbers such as $\mathbf{P}_0$ has a slight nuance due to the different representations complex numbers can take. Consider two such numbers in polar form, p_1 and p_2, specified by a radial

---

[6]Section 3.1

[7]While it doesn't affect the merit of what has been presented thus far, it should be added that the actual implementation and testing of the ray optimization code was in a very narrow scope. This was done with software developed by [4] for several simple optical systems, obtaining results consistent with Figure 2.2. However, due to limitations in the code it was not combined with our implementation of `fftdebye`.

variable	description	cmp
$\hat{\mathbf{n}}$	ray direction unit vector	2
$\mathbf{P}_0$	complex polarization vector	6
OPL	optical path length	1
	total	9

Table 2.2: Each ray carries 9 different pieces of information which must be interpolated in order to facilitate a fast Fourier transform. Above are listed the different variables and their description. The column "cmp" is how many vectorial components a certain variable contains. This table is equivalent to the information contained in Equation 1.236.

and angular component

$$p_1 = r_1 \mathrm{e}^{i\theta_1} \tag{2.45}$$

$$p_2 = r_2 \mathrm{e}^{i\theta_2} \tag{2.46}$$

A straightforward average $\bar{\theta}$ of the angles θ_1 and θ_2

$$\bar{\theta} = \frac{1}{2}(\theta_1 + \theta_2) \tag{2.47}$$

is not sufficient due to the periodicity of an angle θ $\{\theta \in \mathbb{R} : 0 \le \theta \le 2\pi\}$. For example, if $\theta_1 = \pi/6$ and $\theta_2 = 11\pi/6$

$$\bar{\theta} = \frac{1}{2}\left(\frac{\pi}{6} + \frac{11\pi}{6}\right) \tag{2.48}$$

$$= \pi \tag{2.49}$$

This result is incorrect. A cursory look at a unit circle will show this average should be 0 (or 2π). To avoid such mistakes, the polar representation is cast into a Cartesian one having orthogonal real and imaginary components

$$r\mathrm{e}^{i\theta} = r(\cos\theta + i\sin\theta) \tag{2.50}$$

$$= r\cos\theta + ir\sin\theta \tag{2.51}$$

The average of N angles θ_j is then defined according to the angle between the independently computed averages of the real and imaginary parts

$$\bar{\theta} = \mathrm{arctan2}\left(\frac{\frac{1}{N}\sum_{j=1}^{N}\sin\theta_j}{\frac{1}{N}\sum_{j=1}^{N}\cos\theta_j}\right) \tag{2.52}$$

$$= \mathrm{arctan2}\left(\frac{\sum_{j=1}^{N}\sin\theta_j}{\sum_{j=1}^{N}\cos\theta_j}\right) \tag{2.53}$$

As for interpolation about the radial component in polar representation, this is straightforward because

$$\left|r\mathrm{e}^{i\theta}\right| = \sqrt{(r\cos\theta)^2 + (r\sin\theta)^2} \tag{2.54}$$

$$= \sqrt{r^2(\cos^2\theta + \sin^2\theta)} \tag{2.55}$$

$$= r \tag{2.56}$$

and therefore for N radial components

$$\bar{r} = \frac{1}{N}\sum_{j=1}^{N}r_j \tag{2.57}$$
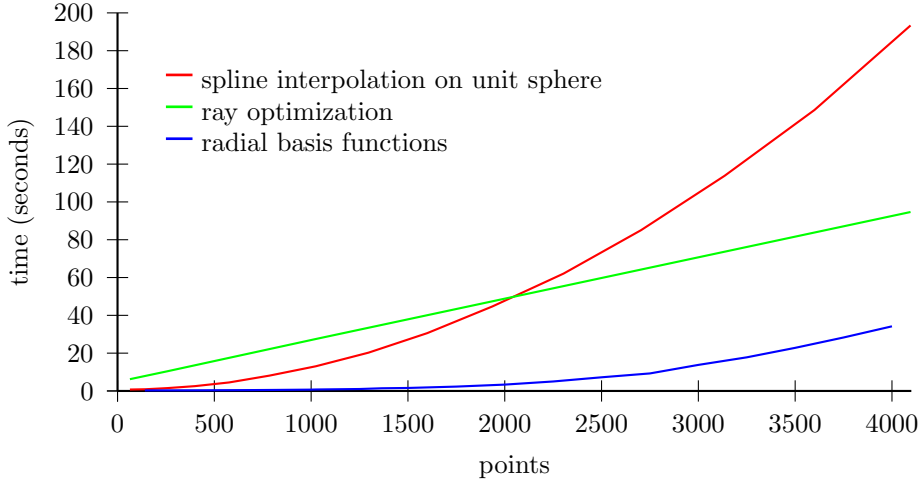
Figure 2.3: Comparison of running times for interpolation and optimization of ray data on the focal sphere. The domain of the data, points, is in terms of the total number of points under interpolation. For each interpolation run, there were the same number of input data points as were points to be interpolated.

Complex numbers in a Cartesian representation are much simpler to interpolate. In this form, a number c consists of an orthogonal real and imaginary component

$$c = a + \mathrm{i}b \tag{2.58}$$

This is essentially equivalent to Equation 2.51, and the averages are found likewise by computing the averages of the real and imaginary part independently.

Evaluation

One of the main issues with either interpolation method is the complexity of the problem. For both tension spline interpolation on the unit sphere and radial basis functions the complexity of the algorithms is $O(N^2)$ [32] [2] (though in some cases RBFs can be evaluated more quickly).

The running times for all three algorithms have been experimentally determined and are shown in Figure 2.3.

The running times for the tension spline method matches very well what is predicted in the literature. For the ray optimization method, we obtain $O(N)$ complexity. This is logical because the operations behind each subsequent ray optimization is independent of all others. Finally, the running time for the RBF implementation we tested differs from the literature values: it seems to be best described by $O(N^3)$ complexity. The reason for this is unknown at the moment, though the values in [2] suggest that the complete solution to the problem requires different steps in the algorithm with different complexity classes.

In addition to the above, because the formalism of Equation 2.1 requires equidistant sampling of k_x and k_y in Cartesian coordinates, sets of points in **k** will in general **not** lie on great arcs in spherical coordinates. The result of this is when computing the knots for the splines, an input of N points requires N distinct spline constructions. This is opposed to spline interpolation where both input and output points are in Cartesian coordinates; in such a case once the interpolating spline has been initialized it can be reused for subsequent evaluations. This increases the running time for this method.
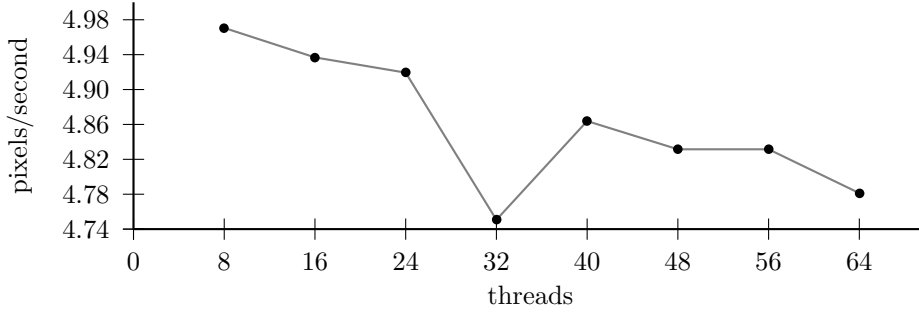
Figure 2.4: Computing performance: threads versus pixels/second/thread. "pixels" refers to the computation of one pixel in a region near the focus using a total number of 1290432 rays on the focal sphere. The speed of `directdebye` is almost independent of the number of threads involved in the computation. Computation carried out on a cluster with 3.0 GHz Intel Xeon 5400 CPUs.

2.3   Direct Integration

Though two programs were developed for the calculations seen in this work, all the figures to be presented were done with the `directdebye` program. `directdebye` computes the sum in Equation 1.236 directly. In this way the computation is greatly simplified: the sampling of the electric field need not be equidistant. The amount of memory required can also be reduced as the number of points to be determined in the focus is independent of the number of points sampled on the focal sphere.[8] We can also choose exactly the points which we wish to sample; an advantage not possible using an FFT. Furthermore, preliminary comparisons between the two methods indicated that the cost of interpolation was much greater than the speed gained through FFT methods.

Numerical evaluation of a `directdebye` is known as *embarrassingly parallel*. That is to say, computation of each point of the electric field $\mathbf{E}(\mathbf{r}, t)$ in the focal region can be done independently of every other point. To take advantage of this, `directdebye` was programmed using the Message Passing Interface (MPI) specification for parallel computing.[9]

The current implementation of `directdebye` averages about 4.86 pixels/s/thread working on 1290432 rays (20164 rays/wavelength for 64 wavelengths[10]). This is almost independent of the number of CPUs, as seen in Figure 2.4. To get a rough feel for how long such a computation might take, values typical for figures presented in Chapter 3 are 500×1000 pixel slices using 20163 rays/wavelength. Operating with 4 compute nodes, each running 4 threads each the total computation time is about 105 minutes.

Figure 2.4 confirms the performance measurements made in combination with different numbers of threads. Even with the overhead to manage many nodes, the performance scales almost perfectly linearly with the number of cores involved. This suggests that a future implementation using a more parallel computing architecture, such as a graphics processing unit, could lead to a substantial speed improvement.

2.3.1   Ideas for Improving Direct Computation

A number of strategies for speeding up direct calculation have been suggested, but as of the time of this writing have not been implemented.

- Assume real polarization. In the case of linear polarization, the polarization vectors can be assumed to be real. This would lead to the calculation of a sine or a cosine rather than a complex exponential. This would lead to a factor of two performance gain.

---

[8]That is to say, if we obey the sampling condition for the input data on the focal sphere we can sample the appropriate regions in the focus with arbitrary resolution. The sampling condition for the FFT and the DFT are the same.

[9]Regionales Rechenzentrum Erlangen, "Woody" cluster. 217 compute nodes, each with two Xeon 5160 "Woodcrest" chips (4 cores) running at 3.0 GHz with 4 MB Shared Level 2 Cache per dual core, 8 GB of RAM and 160 GB of local scratch disk.

[10]The calculation where we obtained this particular number of sampling points in the frequency domain was carried out in Section 2.1.2

- Adaptive integration. Rather than computing the Debye integral as a discrete sum through the trapezoidal rule, there are many other robust algorithms, such as adaptive Gaussian quadrature, which have been developed for this purpose[17]. This could work by incorporating raytracing algorithms via the "choosing the right rays" method into the integration function. An advantage of this method would be the inclusion of routines for estimating error in the output field.

- Symmetry. It is often the case that the input field is symmetric with respect to some coordinate system or axis. Ways of automatically including this symmetry could lead to improvements on the order of the dimensionality reduced.

- Ideal sampling. There should exist an ideal distribution of rays on the focal sphere matching the sampling condition. If the optical aberrations are known beforehand, it should be possible to determine such a distribution, obtaining a minimum amount of rays for a desired accuracy.

2.3.2 Data Communication in `directdebye`

The parallel evaluation of the Debye sum is implemented as follows. Our computing environment consists of one or more nodes. Each node (typically consisting of a single physical machine) has the ability to simultaneously execute one or more threads. Each thread is assumed to have access to a complete set of ray data on the focal sphere. The threads are given a list of coordinates for which they are to compute the Debye sum for said coordinate using the ray data. After each thread has finished computation of all its coordinates, it sends its data to the *root node*. The root node assembles all the data collected from all the threads and writes this information to the disk.

For many systems, especially those with high numerical aperture or a broad spectral range, the number of sampling points on the focal sphere, and concurrently the number of rays which must be traced, can be very large - on the order of 10^5 rays per spectral component. This is, of course, strongly dependent on the optical system and observation region. Nonetheless, this condition can easily lead to binary ray data of several gigabytes. In a cluster computing environment, the simultaneous demand for a large dataset on a network with finite bandwidth is problematic.

If we assume that at the beginning of the simulation one node of the cluster has the data[11], the time to transfer to all other nodes in the computation grows linearly with the number of nodes involved: if it takes t seconds to transfer the file to one of n nodes, it takes $n \cdot t$ seconds for either the sequential or concurrent transfer to take place. This time can quickly grow to consume more time than the computation of the focused field! (For example, transfer of a 100 MB file takes 10 s with 64 nodes, the transfer time is 600 s).

A way to overcome this problem is to transfer the data in a binary tree fashion. Consider the network depicted schematically in Figure 2.5. Assume all nodes are connected to each other with some equal bandwidth connection, and furthermore that the bandwidth between any two nodes is independent of the network activity of the whole. In practice this is a valid assumption for most cluster environments.

Two trees have been constructed: the "haves" and the "have nots". The former have the data in question, the later do not. At each level the number of concurrent transfers (represented by lines with arrows) double. The elegance of this system is the simplicity of the mathematical arrangement. It can be seen from Figure 2.5 that each level contains 2^n entries and 2^n simultaneous transfers. The levels for the "haves" always ranges from $1 \ldots 2^n$, and the "have nots" range from $2^n + 1 \ldots 2^{n+1}$, where n is the level. The time savings is of order $\mathrm{ceil}(\log_2(n))$.

2.3.3 Distribution of Rays on the Entrance Pupil

In computing the Debye integral, the particular distribution of rays on the entrance pupil was found to have a significant effect on the evaluation of the focused field. The three distributions tested are shown in Figure 2.6 (to the left of the images). They are the geometric, the square lattice, and the random distribution. For each, the rays placed within the perimeter of the circular entrance pupil with radius r.

The random distribution is a uniform random distribution[12]. It provides N rays in a circular area of πr^2.

---

[11] Furthermore, we assume local access of the data by the threads associated with each node have no performance penalty. In practice, this has been observed to be a good assumption

[12] In particular this specific implementation from RAYTRACE [26] is a linear congruential generator, from Numerical Recipes in C [31], Chapter 7, "`ran1`."

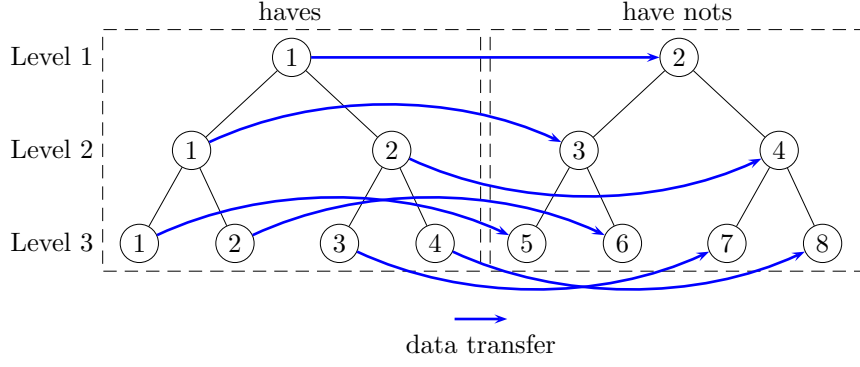haves · have nots

Level 1 · Level 2 · Level 3

data transfer

Figure 2.5: Distribution of a single large dataset in a cluster environment. Circles indicate nodes. Arrows indicate direction and location of data transfer for each Level. The figure can be read as follows: At Level 1, node number 1 has the data and transfers it to node 2. At Level 2, both node 1 and node 2 now have the data, and transfer it to node 3 and 4, respectively. On Level three, nodes 1 to 4 now have the data and begin transfer to nodes 5 to 8. This process continues down the tree until every node has the data.

The square lattice distribution produces rays situated on a square lattice. The conventional way to produce such a distribution is to create an array of $N_x \times N_y = N$ rays on a square and remove those outside the entrance pupil radius. The total number of rays N within this radius is, at least in the limit of infinitely number rays, equal to the ratio between the areas of the unit square and unit circle, or $\pi/4$ [13]. Done this way, the spacing between rays in each dimension is simply r/N_x and r/N_y. This can be a little bit confusing when trying to compare this to other distributions because the total number of rays differ. To augment this, distributions of this type had their total number of rays modified by $N \mapsto N \times 4/\pi$.

The geometric distribution is based on a set of concentric polygons. It is produced as follows: For an entrance pupil of radius r_p, divide that radius into R intervals with index $i = 0..R-1$ of length $dr = r_p/R$. The radius at each interval $r_i = (i + 1/2)\, dr$ for $i \geq 1$ and $r_0 = 0$. This distribution has the property that if N total rays are desired, there will be $R = \sqrt{N}$ intervals on the radius. This is due to the fact that on each sub-interval there are $(2i + 1)$ generated points. The sum of all these points for n sub-intervals is

$$\sum_{i=0}^{n} (2i + 1) = n^2 \tag{2.59}$$

Pseudocode for producing such a distribution in Cartesian coordinates is as follows

```
radius := (desired radius)
R      := sqrt(number of desired output points, a power of two)
dr     := 1/R
FOR i = 0 to R
 r = (i + 0.5) * dr * radius
 FOR j = 0 to (2*i+1)
  theta = j * 2 * PI / (2*i + 1) + PI / (2*i + 1)
  PRINT r*cos(theta), r*sin(theta)
 END FOR
END FOR
```

Figure 2.6 shows the differences between the distributions in the focus field for sampling on the focal sphere with 20164 rays/wavelength and 64 wavelengths, for a total of $N = 1290496$ rays. The optical system chosen was an ideal parabolic mirror (Section 3.1), NA = 0.5 and focal length of $f_b = 50\,\text{mm}$ [14].

---

[13] Area of a Box $A_\square$ with sides of length $2r$ is $A_\square = (2r)^2$. Area of a circle $A_\bigcirc$ with radius r is $A_\bigcirc = \pi r^2$. The ratio of the two is $A_\square/A_\bigcirc = \pi/4$

[14] The corresponding radius of the entrance pupil for these parameters is $\approx 26.7949\,\text{mm}$. See Equation 3.13

The geometric distribution seems to produce the best result with the fewest number of sampling rays. The square lattice distribution shows aliasing effects beginning around 150 μm from the focus. The random distribution shows no aliasing effects, but the random sampling causes severe distortion in the resultant image. This is most likely explained by noting that the geometric and square lattice distributions are uniformly distributed. The random however, is not. In this case each ray would need a special factor taking into account its surface element on the focal sphere: something that was not done in this simulation.

It is interesting to note that the geometric distribution seems to produce the best results. For $N = 1024$ rays and a radius $r = 1$, the mean nearest $\bar{r}_{nn}$ neighbor for each point in the geometric distribution was calculated to be 0.0404457 (standard deviation 0.007647). In comparison for the square lattice that was $\bar{r}_{nn} = 0.0553$ (standard deviation 0), and for the random distribution $\bar{r}_{nn} = 0.0278623$ (standard deviation 0.0145428). A priori assumptions would lead one to believe the increase in aliasing for the square lattice versus the geometric would be due to the fact that the geometric distribution is closer to the ideal packing for spheres in two dimensions. However, when the density of the square lattice is increased to match that of the ideal packing[15], it still exhibits higher aliasing than the corresponding geometric distribution. This suggests an interesting question, and is one we will not visit in this work: "What is the ideal sampling for rays on the focal sphere for this type of numerical integration?"

2.4 Calculation of the Strength Factor

For an arbitrary optical system the strength factor, g, has no analytical expression. One must therefore calculate it numerically. Besides differential raytracing, two additional methods have been studied for this purpose. They are the subject of our present discussion.

2.4.1 Calculation by Voronoi Tessellation

A Voronoi tessellation partitions (tessellates) a space into convex polytopes based upon a set of generating points. Each polytope contains exactly one generating point, and the set of points within each polytope is closer to its generating point than any other generating point. Voronoi tessellations in $\mathbb{R}^2$ are called Voronoi diagrams, and the properties of these we will attempt to take advantage of.

At this point it is important to reassert our desire to know the surface element which each ray represents. Take for example a regular square lattice of rays, size $R_x \times R_y$ with $N_x \times N_y$ rays. The surface element is simply

$$\mathrm{d}s = \frac{R_x R_y}{N_x N_y} \tag{2.60}$$

An interesting property of the Voronoi diagram is that regular lattice of generating points gives a regular tessellation of polygons in two dimensions. For the case of a square lattice, it is a regular tessellation of squares with equal areas. The areas of these squares is equivalent to that found with 2.60. In this same vein we can trace rays up to the focal sphere and apply a Voronoi tessellation of these rays constrained to the surface of that sphere. The polygons are now spherical polygons and their areas [25] [10] should represent ds_{out}. Using this method also requires pruning of the polygons that form the border to the main polygon cluster, as in Figure 2.7. For optics such as aspheres, parabolic mirrors, and spherical lenses this is as simple as removing all polygons whose vertices contain the point $(r, -\pi/2, 0)$[16] in spherical coordinates. For many focusing optics the main polygon cluster will center around $(r, \pi/2, 0)$ which means all polygons on the unit sphere which contain the vertex $(r, -\pi/2, 0)$ bound that cluster. More complicated structures can have these polygons determined by, for example, computing the convex hull of the points on the focal sphere, but we have not investigated such systems.

It is worth pointing out that there is no strong mathematical motivation for this idea that we are aware of. Although the method works well for the examples calculated, the reader is warned that there may be many systems for which this may not apply. It is, however, fruitful. Most commercial raytracing codes do not include an implementation of differential raytracing, so this gives this method somewhat

---

[15]Ideal packing for spheres in $\mathbb{R}^2$ is the "hexagonal lattice".

[16]The tuplet here is in spherical coordinates (r, θ, ϕ) where r is the radius, $\{\theta \in \mathbb{R} : -\pi/2 \leq \theta \leq \pi/2\}$ is the inclination, and $\{\phi \in \mathbb{R} : -\pi \leq \phi \leq \pi\}$ is the azimuth.

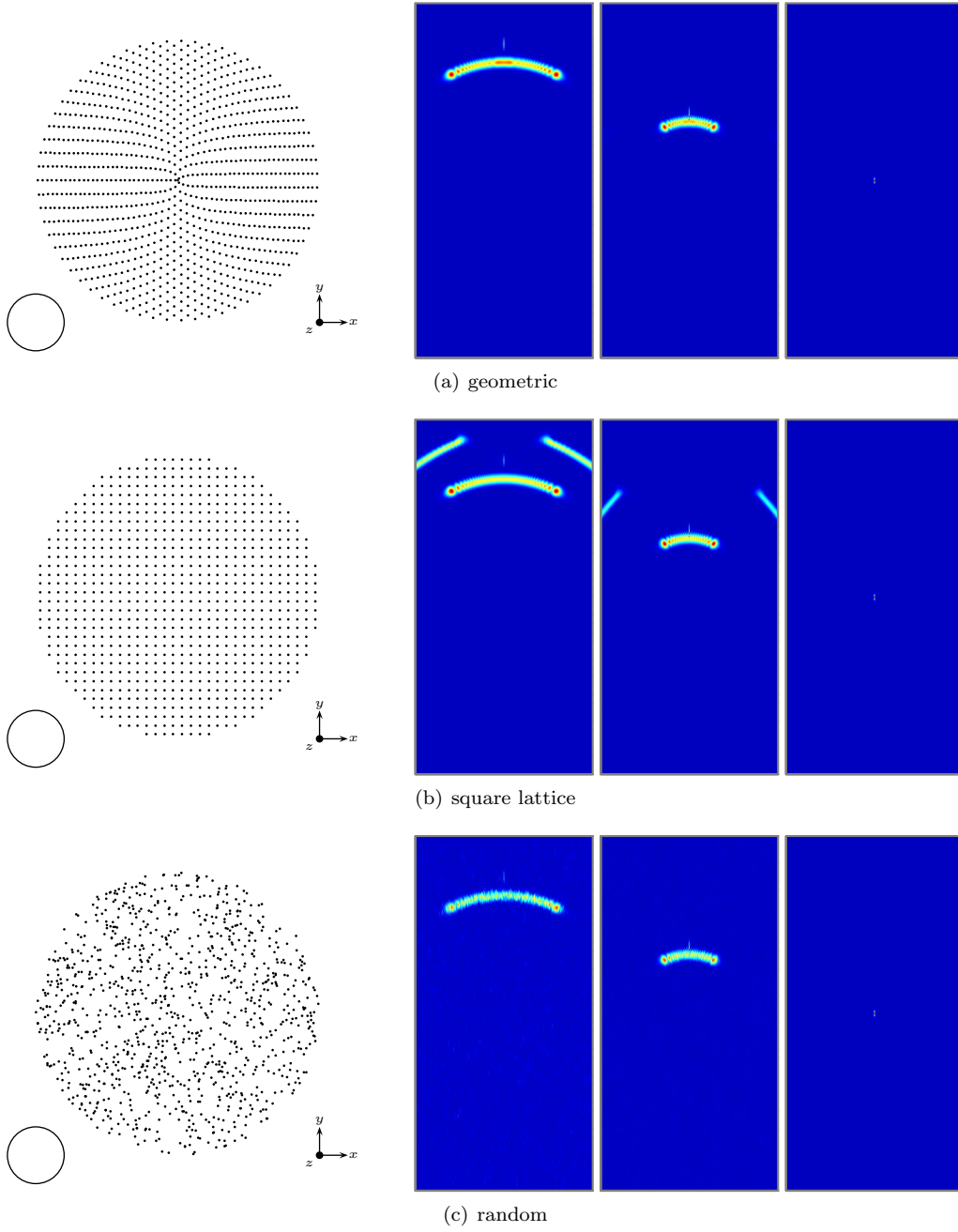(a) geometric

(b) square lattice

(c) random

Figure 2.6: Comparison of different ray distributions on the entrance pupil and their effect on the field structure at the focus. Given the same number of rays, the circular distribution produces the most accurate results. Discussion in text. The distributions have $N = 1024$ rays each. The resultant fields were calculated using 20164 rays/wavelength with 64 wavelengths, for a total of $N = 1290496$ rays with a 40 fs Gaussian pulse at $\lambda_0 = 800$ nm. Spatial and temporal units for the three electric field images are the same as for the first three timesteps from the left in Figure 3.6.
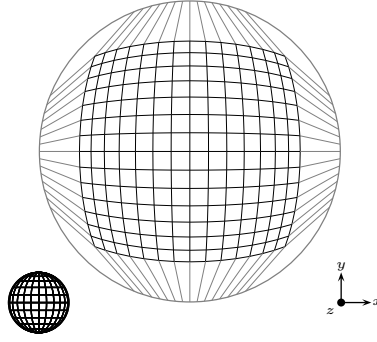
Figure 2.7: Bounding polygons produced by Voronoi tessellation have areas not applicable to the problem. They must be pruned. The pruned polygons are shown in gray for polygons distributed on the focal sphere of a parabolic mirror. This figure represents a two dimensional projection of spherical polygons on a three dimensional sphere. The wire-frame globe in the lower left indicates this is a two dimensional projection of a three dimensional object.

(a) Entrance pupil   (b) Focal sphere - parabola   (c) Focal sphere - asphere
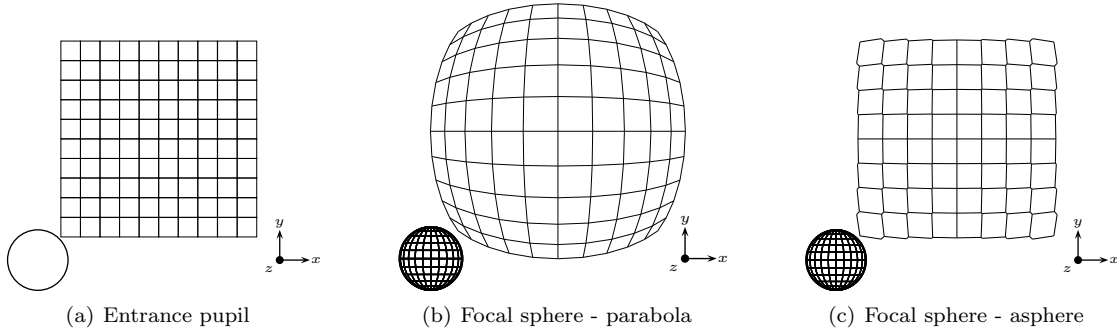
Figure 2.8: Voronoi tessellation of a square lattice of rays on the focal sphere (left), that same set of rays on the focal sphere after being focused by a parabolic mirror and an aspheric lens (center, right). The lattice on the entrance pupil is in a two dimensional plane (indicated by the bottom left circle), while the lattice on the focal spheres are two dimensional projections of a three dimensional surface (indicated by the "globe" in the bottom left).

of an advantage: one can determine the apodization factor from the intersections of the rays on the entrance pupil and the focal sphere alone.

Figure 2.9 shows the resultant tessellations for several simple systems. Both were computed for a square lattice of rays in the entrance pupil. The figure shows very good agreement between theoretical values of ds and those predicted with Voronoi diagrams. Again, the main advantage of this method is that it is able to calculate ds by only knowing the intersection points of a ray with a surface. Therefore no modification of existing raytracing code is necessary.

There are a limited number of systems that can be calculated exactly, so it is difficult to ascertain other systems where this might fail. Further comparisons with results obtained via differential raytracing are warranted.

2.4.2 Calculation by Nearest Neighbor

Another method, similar to Voronoi tessellation, was researched as a candidate to calculate the strength factor. The strength factor can be thought of as the ratios of the ray densities on the entrance pupil and the focal sphere. Numerically the ray density is determined by employing a fast kd-tree search[24] to find, for each desired ray intersection point on S_{out} and S_{in}, how many neighboring ray intersection points are formed within a certain, fixed, radius.

The results, plotted in Figure 2.10 are fairly consistent albeit unimpressive. The accuracy seems to
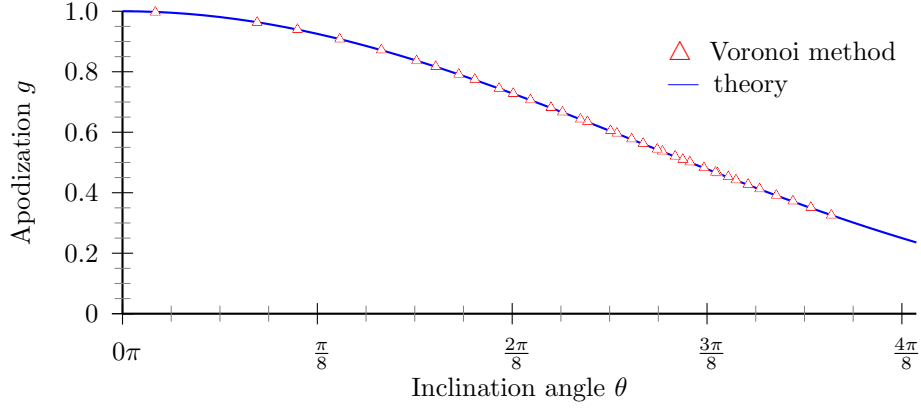
Figure 2.9: Comparison of g calculated using the Voronoi method and that predicted by the theory of Equation 1.213. System is an ideal parabolic mirror, NA = 1.0, $f = 50\,\text{mm}$, and entrance pupil radius of $100\,\text{mm}$. Horizontal axis is the inclination angle from the positive z axis.

hinge on having a fairly high ray density compared to the search radius. For aplanatic, or low numerical aperture systems, this works well, as the ray density is similarly distributed on the focal sphere. However, for parabolic or diffractive elements, obtaining a high sampling density near the optical axis means a very large ray density at the aperture edges.

This method also has a disadvantage in that pruning must be done for all points on the aperture edge which are within the length of one search radius from the boundary. Of course one can, for the price of precision, reduce this radius, but it remains that the strength factor can still not be determined for all rays in the focal sphere.
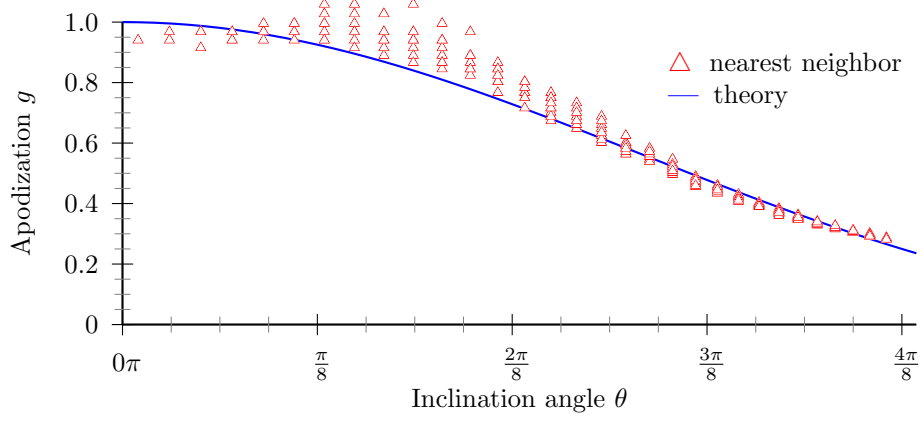
Figure 2.10: Comparison of the apodization factor calculated with the nearest neighbor method on the focal sphere for an ideal parabolic mirror, NA = 1.0, $f = 50\,\text{mm}$, and entrance pupil radius of $100\,\text{mm}$. The theoretical result was calculated using Equation 1.213. The nearest neighbor method tends to produce inconsistent results with high error. Because the nearest neighbor method tends to not be axially symmetric, different values for g appear along the domain.

Chapter 3

Analysis

3.1 Reflective Optics

Reflective optics are a natural place to begin investigations into focused ultrashort pulses. In this section, an *ideal* reflective surface is assumed, neglecting effects from the (usually both frequency and polarization dependent) material dispersion and absorption.

The raytracing was done with the software program RAYTRACE[26], and the Debye sum was computed from those rays using the method presented in Section 2.3 combined with the software program `directdebye`.

3.1.1 Parabolic Mirror - Monochromatic Light

Although this work is primarily interested with the focusing of pulses, it is nonetheless fruitful to look at stationary intensity distributions for monochromatic light. Stationary intensity predictions of `directdebye` for the case of a parabolic mirror were verified by comparison to another program `polfocus`[23], which has been extensively checked against itself and other peer-reviewed results[12] [39] [44] [45] [22]. In this section we present these comparisons to show that the results of `directdebye` are consistent.

Geometry

The optical system is set up as in Figure 3.1. It consists of an on-axis paraboloid of focal length f illuminated by a plane wave represented by rays within an entrance pupil of radius r. The rays are parallel to the optical axis. After reflecting off the paraboloid the rays propagate to the focus at the origin of the coordinate system $\mathbf{r} = (0, 0, 0)$.

In terms of the optical axis z and focal length f, the paraboloid of Figure 3.1 is described radially by
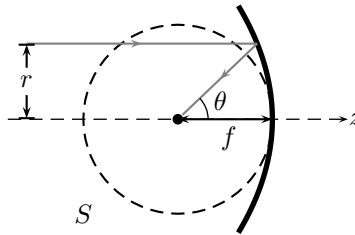
Figure 3.1: Two dimensional schematic representation of an on-axis paraboloid used for focusing. The paraboloid is centered and symmetric about the optical axis. Rays are incident parallel from the $-z$ direction within an entrance pupil of radius r, and reflect off the surface propagating to the focus at coordinates $\mathbf{r} = (0, 0, 0)$. The focal sphere of surface S is concentric with that point and represented by the dashed circle.

the equation

$$z(r) = -\frac{r^2}{4f} + f \tag{3.1}$$

where $r = \sqrt{x^2 + y^2}$ in Cartesian coordinates and the origin of the coordinate system ($r = 0$) is at the focal point.[1] The numerical aperture NA is controlled by adjusting the size of the entrance pupil. There is a direct relationship between the radius r of the entrance pupil and the numerical aperture. Recall Equation 1.202 which describes the relationship between the height r of the incident ray and the angle θ of the light ray after reflection by a parabolic surface

$$r = 2f \tan\left(\frac{\theta}{2}\right) \tag{3.2}$$

r is interpreted as being the radius of the entrance pupil. The numerical aperture of any system is described by the aperture angle θ via

$$\mathrm{NA} = \sin\theta \tag{3.3}$$

Here, the refractive index of the medium in which the light ray propagates is assumed to be 1. Solving Equation 3.2 for θ

$$\theta = 2\arctan\left(\frac{r}{2f}\right) \tag{3.4}$$

and substituting into Equation 3.3

$$\mathrm{NA} = \sin\left(2\arctan\frac{r}{2f}\right) \tag{3.5}$$

This can be simplified by using the double angle formula for sin

$$\sin(2x) = 2\sin x \, \cos x \tag{3.6}$$

and two trigonometric identities

$$\sin(\arctan x) = \frac{x}{\sqrt{1+x^2}} \tag{3.7}$$

$$\cos(\arctan x) = \frac{1}{\sqrt{1+x^2}} \tag{3.8}$$

Equation 3.5 becomes

$$\mathrm{NA} = \sin\left(2\arctan\left(\frac{r}{2f}\right)\right) \tag{3.9}$$

$$= 2\sin\left(\arctan\left(\frac{r}{2f}\right)\right)\cos\left(\arctan\left(\frac{r}{2f}\right)\right) \tag{3.10}$$

$$= 2\frac{(r/2f)}{\sqrt{1+(r/2f)^2}}\frac{1}{\sqrt{1+(r/2f)^2}} \tag{3.11}$$

$$= 2\frac{(r/2f)}{1+(r/2f)^2} \tag{3.12}$$

$$= \frac{4fr}{4f^2 + r^2} \tag{3.13}$$

As a specific numerical aperture is often desired, expression in terms of r can be re-derived. Note Equation 3.13 can be rewritten as an equation quadratic in r,

$$\mathrm{NA} = \frac{4fr}{4f^2 + r^2} \tag{3.14}$$

$$\mathrm{NA}\,r^2 - 4fr + 4\mathrm{NA}\,f^2 = 0 \tag{3.15}$$

---

[1]This geometry, of course, assumes symmetry with the z axis. However, the ray distribution on the focal sphere need not be symmetric in the same respect.

Using the quadratic formula with $a = \mathrm{NA}$, $b = -4f$, and $c = 4\,\mathrm{NA}\,f^2$,

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3.16}$$

$$= \frac{4f \pm \sqrt{16f^2 - 16f^2\mathrm{NA}^2}}{2\,\mathrm{NA}} \tag{3.17}$$

For $\mathrm{NA} = 0$, the positive root of this equation makes no physical sense, so the negative root is considered only, leaving

$$r = \frac{2f\left(1 - \sqrt{1 - \mathrm{NA}^2}\right)}{\mathrm{NA}} \tag{3.18}$$

The pulse fronts are also spherical. This is a consequence of the geometry of a parabola which produces, from the point of view of geometric optics, an ideal spherical wave given an incident plane wave parallel to the optical axis. Recall the form of Equation 1.202. The total optical path length from a plane wave entering from the $-z$ direction is composed of two parts: the optical path length from the plane wave's origin to the surface of the paraboloid, and from that surface to the focus. The former is a re-expression of Equation 1.202

$$z(r) = -\frac{r^2}{4f} + f \tag{3.19}$$

and the distance from this point to the focus is

$$\sqrt{r^2 + z(r)^2} \tag{3.20}$$

Therefore the total optical path length is

$$\mathrm{OPL} = \sqrt{r^2 + z(r)^2} + z(r) \tag{3.21}$$

$$= \sqrt{r^2 + \left(f - \frac{r^2}{4f}\right)^2} + f - \frac{r^2}{4f} \tag{3.22}$$

$$= \sqrt{r^2 + f^2 + \frac{r^4}{16f^2} - \frac{r^2}{2}} + f - \frac{r^2}{4f} \tag{3.23}$$

$$= \sqrt{\frac{r^2}{2} + f^2 + \frac{r^4}{16f^2}} + f - \frac{r^2}{4f} \tag{3.24}$$

$$= \sqrt{\left(f + \frac{r^2}{4f}\right)^2} + f - \frac{r^2}{4f} \tag{3.25}$$

$$= f + \frac{r^2}{4f} + f - \frac{r^2}{4f} \tag{3.26}$$

$$= 2f \tag{3.27}$$

This result is independent of r, confirming that the total optical path lengths of different rays associated with a planar wavefront and converging to the focus are equal. An equivalent statement is that that a paraboloid focuses a plane wave to a spherical wave and vice versa.

Electric Field Near the Focus

We now proceed to look at the electric field near the focus for this system. For all figures we assume a parabolic geometry with $\lambda_0 = 800\,\mathrm{nm}$ and $f = 50\,\mathrm{mm}$. Figure 3.2 shows the stationary intensity distribution in the $z = 0$ slice of the x-y focal plane (top), and the corresponding $x = 0$ slice of the y-z plane (bottom). The numerical aperture of this focusing paraboloid is $\mathrm{NA} = 0.5$. The colormaps of both figures have been logarithmically scaled to show structure.

(a) $z = 0$ slice of the x-y plane
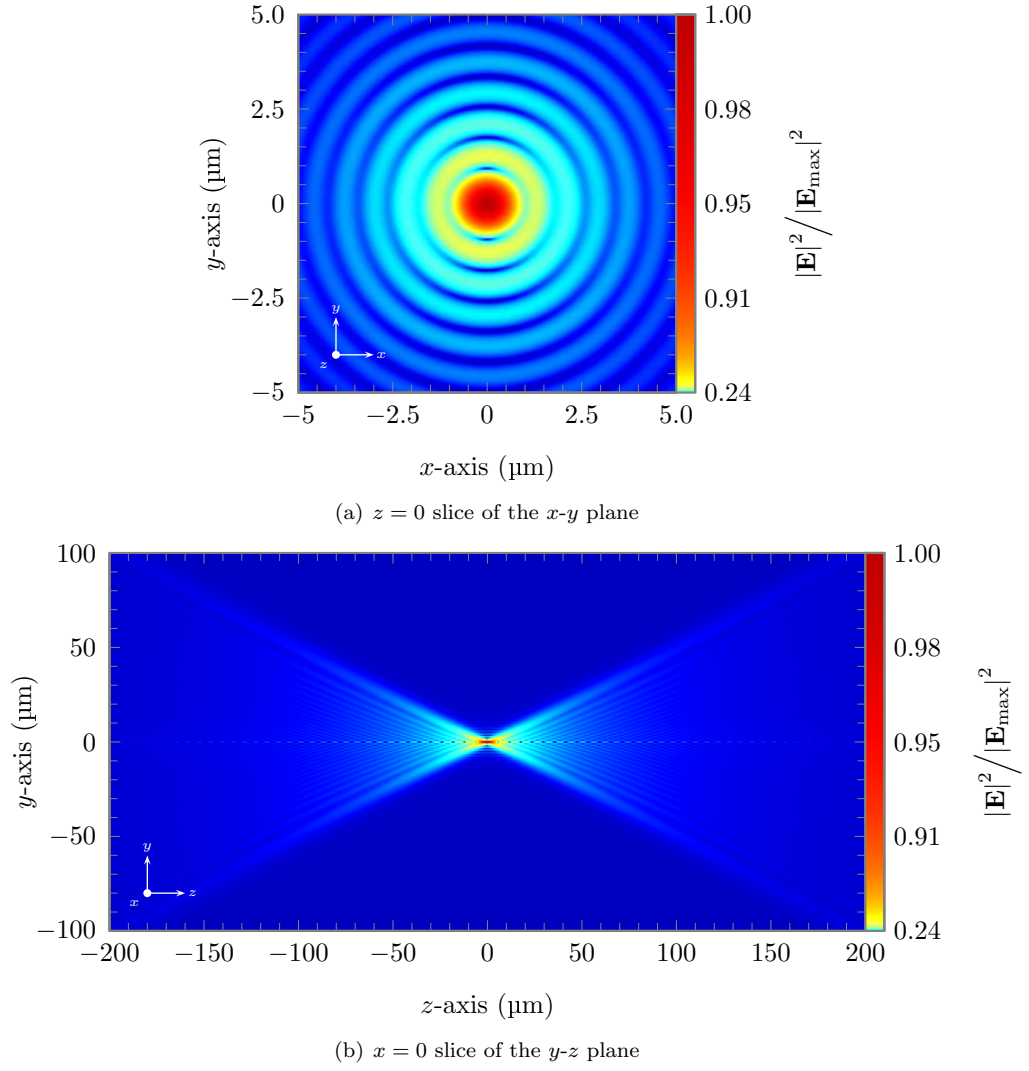

(b) $x = 0$ slice of the y-z plane

Figure 3.2: Stationary intensity distribution for monochromatic incident light in two different planes of the focus of a parabolic mirror, $\lambda_0 = 800\,\mu\text{m}$, $f = 50\,\text{mm}$, and NA $= 0.5$. Figures have been logarithmically scaled. Polarization of the incident field was linear along the x direction.

These results for the $z = 0$ slice of the x-y focal plane computed with `directdebye` agree almost perfectly both in structure and in the maximum predicted intensity with `polfocus`. This is shown in Figure 3.3 for different numerical apertures, the domain representing the radial distance from the focus ($x = y = z = 0$). Both were calculated using a sampling of 142×142 rays on the entrance pupil, however `directdebye` used the geometric distribution of rays while `polfocus` uses a square lattice.

While not particularly evident in Figure 3.3, as the numerical aperture of the system increases, polarization effects in the focus become particularly important[23] [41]. This effect is demonstrated in Figure 3.4 for different polarization components of the electric field, incident linear in the x direction. As the numerical aperture along the domain is increased, the spot becomes increasingly elongated along the polarization direction. These effects can be augmented by using an incident radial polarization instead of a linear one. Figure 3.5 shows the difference between an incident linear polarization in the x-direction, and for radial polarization. To the left of each field is shown schematically the polarization arrangement of rays on the entrance pupil, the vectors pointing parallel to the polarization direction. To the right is their corresponding field structure in the $z = 0$ slice of the x-y plane of the focus. As shown in [40], some of the tightest spots and highest field strengths can be achieved by the use of such a polarization distribution.

3.1.2 Parabolic Mirror - Pulsed Light

In this section the pulse behavior for the same ideal parabolic mirror is discussed. Polarization is included, but effects stemming from dispersion and absorption have been ignored. A more complete ray-tracing implementation including a wider array of effects is warranted for future investigation. Focusing paraboloids have the advantage of being widely studied in the literature [44] [45], giving a good basis on which to judge the predictions.

Figure 3.6 shows the intensity $|\mathbf{E}|^2$ in the $y = 0$ slice of the x-z focal plane for a 40 fs Gaussian[2] pulse at $\lambda_0 = 800$ nm. The figure is a composite of five frames separated[3] by $\Delta t \approx 2.22 \times 10^{-13}$ s, with the focus exactly at $t = 0$. The input polarization was linear in the x direction. The focal length of the parabolic mirror is 50 mm. Having chosen a numerical aperture of NA = 0.5, Equation 3.18 was used to compute the entrance pupil to be approximately 26.7949 mm.

It can be seen from Figure 3.6 that the pulse travels at the speed of light in a vacuum c_0. In four timesteps, $4 \cdot \Delta t \cdot c_0 \approx 267$ μm, which is approximately the distance which can be seen from the center of the leftmost pulse to the center of the rightmost pulse.

Note that since we are working with an ideal mirror, there is no group velocity dispersion and no pulse spreading; the width of the pulse does not vary radially around the pulse fronts. It can also be seen that the pulse fronts are perfectly symmetric around the focus. This is in contrast to the case for dispersive optics, as we will see in the next section.

The pulse front also exhibits a diffraction pattern. This is due to the entrance pupil having a uniform intensity profile - the diffraction pattern is just a result of the region in the focus being the Fourier transform of its aperture function. Pulses often have, in addition to a temporal profile, a Gaussian spatial intensity profile in that aperture. The consequence of this is the removal of the high frequency Fourier components, causing the diffraction pattern to be reduced. This is shown in Figure 3.7.

High numerical aperture systems offer the ability to achieve some of the highest possible focused intensities[3]. Specifically, the geometry of the paraboloid allows one to obtain very high numerical apertures by increasing the radius of the entrance pupil. Figure 3.8 shows consistent results for two high numeral aperture systems: NA = 0.75 and NA = 1.0.[4]

3.2 Boundary Wave Pulses

The pulse structure for the parabolic mirror of the previous section includes a small intensity spike directly on the optical axis, immediately before or after the main pulse. Though it may appear to be an artifact, this intensity spike is a real effect known as a *boundary wave pulse*.
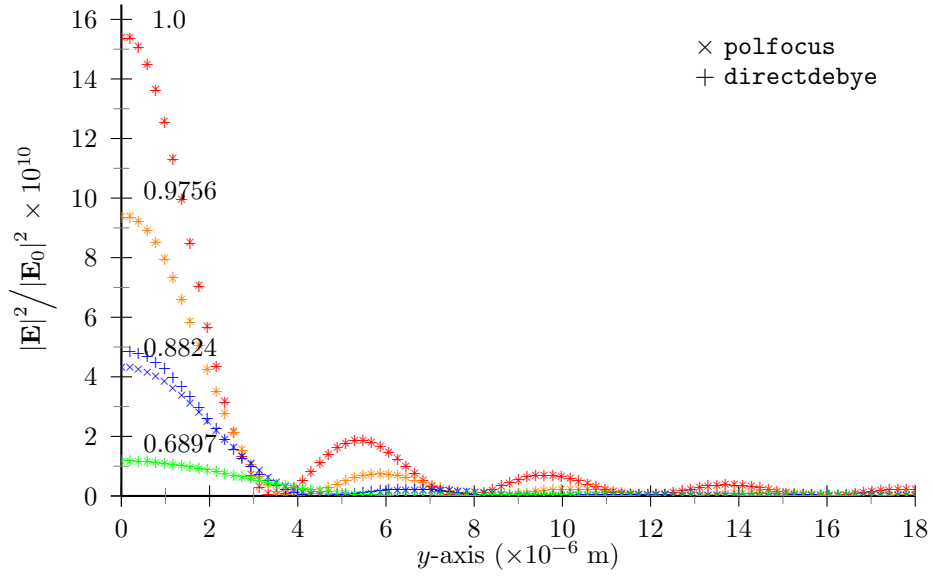
---

[2]The envelope function was computed using the definition of Equation 1.75 in Section 1.2.2.

[3]Exact spacings for this figure are $\Delta t = 400$ μm$/(6c_0)$

[4]As per Equation 3.18 with $f = 50$ mm, the corresponding radii for NA = 0.75 and NA = 1.0 are $r \approx 45.1416$ mm and $r = 100$ mm, respectively
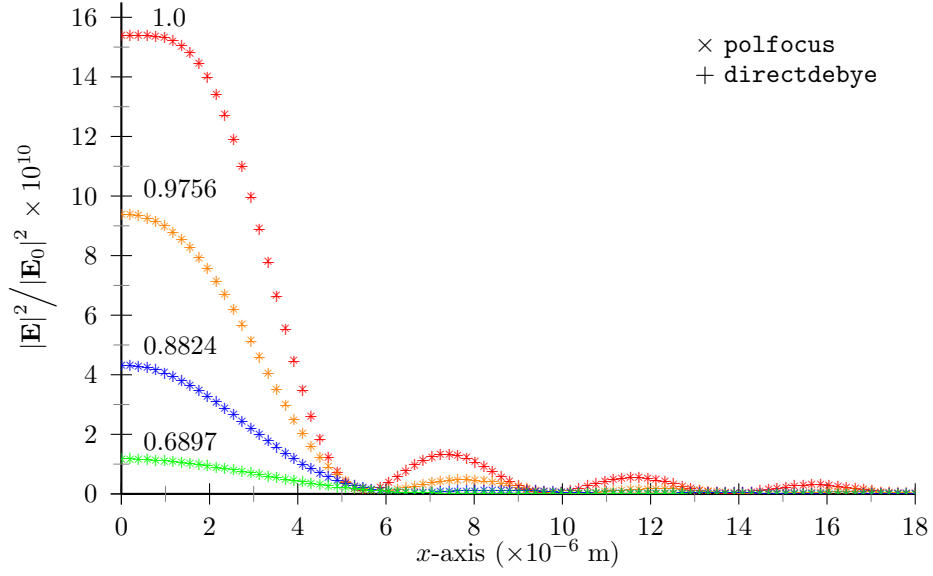
(a) x-axis comparison

(b) y-axis comparison

Figure 3.3: Comparison of $|\mathbf{E}|^2 / |\mathbf{E}_0|^2$ for 1 W incident light power using the `polfocus` and `directdebye` programs. Optical system is a parabolic mirror of varying numerical apertures which have been labeled above. For all plots, $\lambda_0 = 800$ nm, and $f = 50$ mm. Incident polarization is linear in the x-direction.
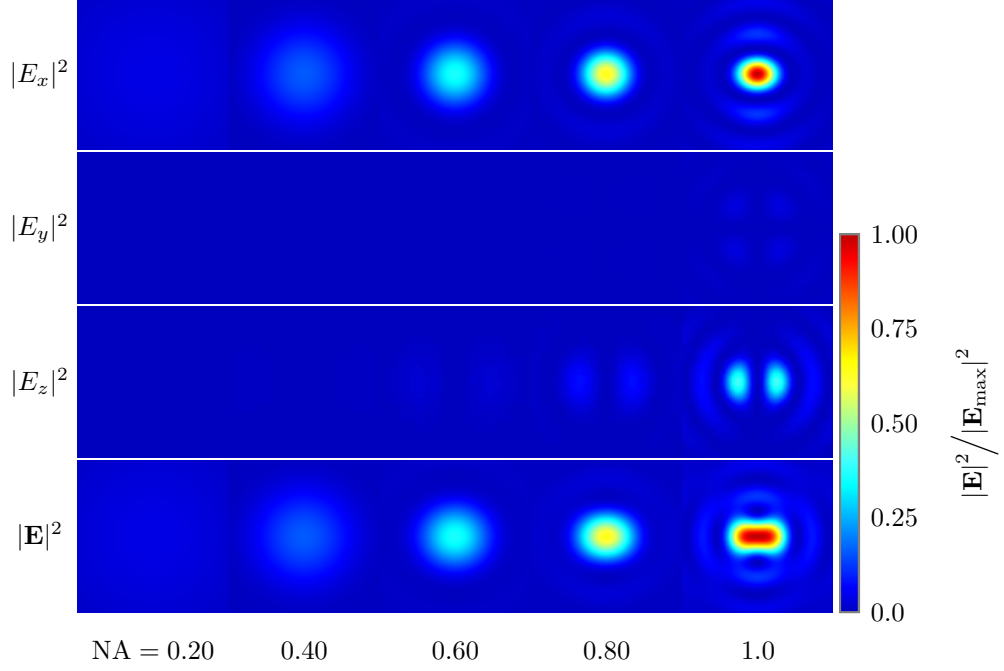
Figure 3.4: Polarization effects become increasingly important as the numerical aperture is increased. Here is shown the electric field in the focus ($z = 0$ slice of the x-y plane) for a parabolic mirror ($\lambda_0 = 800\,\text{nm}$, $f = 50\,\text{mm}$) for six different numerical apertures and each polarization component of the electric field. Each of the 24 corresponding slices represent a square area of $2 \times 2\,\mu\text{m}$, and have been normalized relative to each other.

Observe Figure 3.9, top. Shown here is single timestep for a $\lambda_0 = 800\,\text{nm}$ 40 fs Gaussian pulse in the vicinity of the focus for our familiar parabolic geometry, NA = 0.5 and $f = 50\,\text{mm}$. Directly behind the main pulse, shown in detail, is the boundary wave pulse. Below is shown for the same parameters a pulse with a Gaussian *spatial* profile radially about the entrance pupil: 1/e of $|\mathbf{E}|$ is at $r = 20.25\,\text{mm}$ in the entrance pupil. Note the following features of boundary wave pulses in Figure 3.9:

1. It does not appear for a Gaussian spatial distribution in the entrance pupil.

2. It is always coincident with the optical axis.

3. The shape of the boundary wave pulse does not change as it propagates through the focus.

Recent explanations for this phenomenon[36][37] are based on the assertion that diffraction can be thought of as the interference of two waves: boundary waves, those originating at the aperture edge, and geometrical waves originating within the aperture [5].

The explanation of the boundary wave as an interference effect originating at the aperture edge seems, a priori, to be well supported by the structure of the focused field, for if we replace the fully illuminated aperture with an incident Gaussian spatial profile the boundary wave disappears (Figure 3.9).

Among the first predictions available under these assumptions is if a boundary wave pulse is due to diffraction at the aperture edge, the diffraction pattern should be the Fourier transform of the edge of its its aperture function. First, consider a system with a circular exit pupil. The diffraction pattern of the boundary wave should be that of an infinitely thin circular annulus. To derive this pattern, first note the form of a Fourier transform of $f(r, \theta)$ in radially symmetric polar coordinates
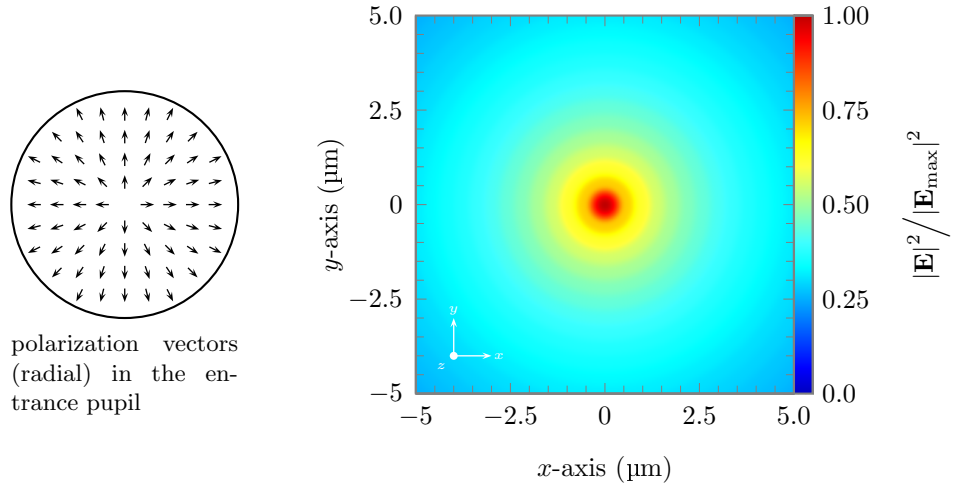
$$\mathscr{F}^+\big(f(r,\theta)\big) = \int_0^\infty \int_0^{2\pi} f(r,\theta)\, e^{i k_r r \cos\theta}\, r\, dr\, d\theta \tag{3.28}$$

For a circular annulus of radius a, $f(r, \theta)$ can be represented by a delta function in r only

$$f(r) = \delta(r - a) \tag{3.29}$$

48
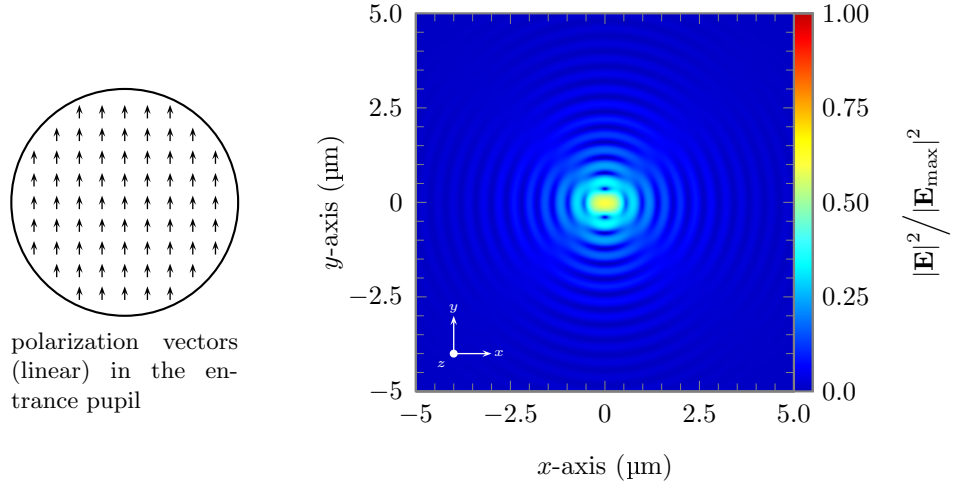
(a) linear polarization

(b) radial polarization

Figure 3.5: Polarization effects in the focus can be augmented using a radial polarization distribution. To the left is shown schematically the polarization vectors for rays in the entrance pupil, and to the right the corresponding electric field intensity $|\mathbf{E}|^2$ in the $z = 0$ slice of the x-y plane of the focus. System is a parabolic mirror, NA = 1.0, $f = 50\,\mathrm{mm}$, and $\lambda_0 = 800\,\mathrm{nm}$. Figures have been normalized relative to each other.
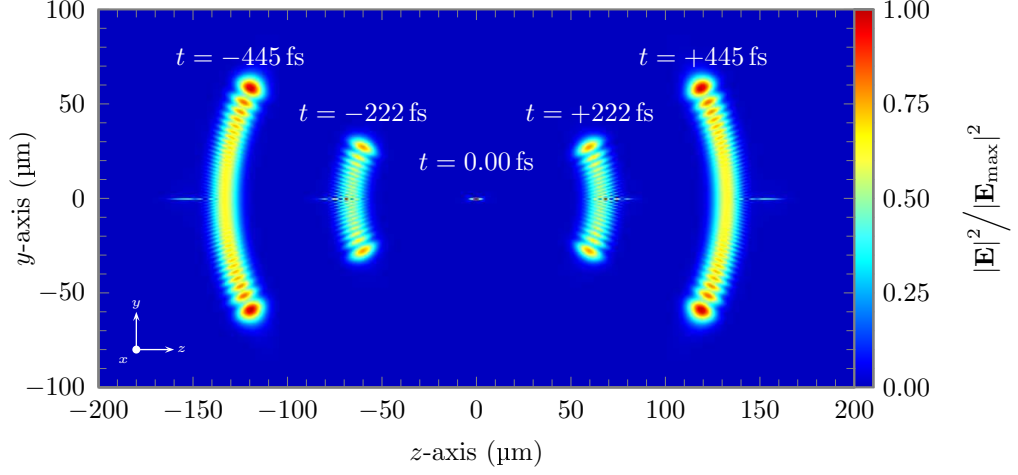
(a) each timestep normalized individually
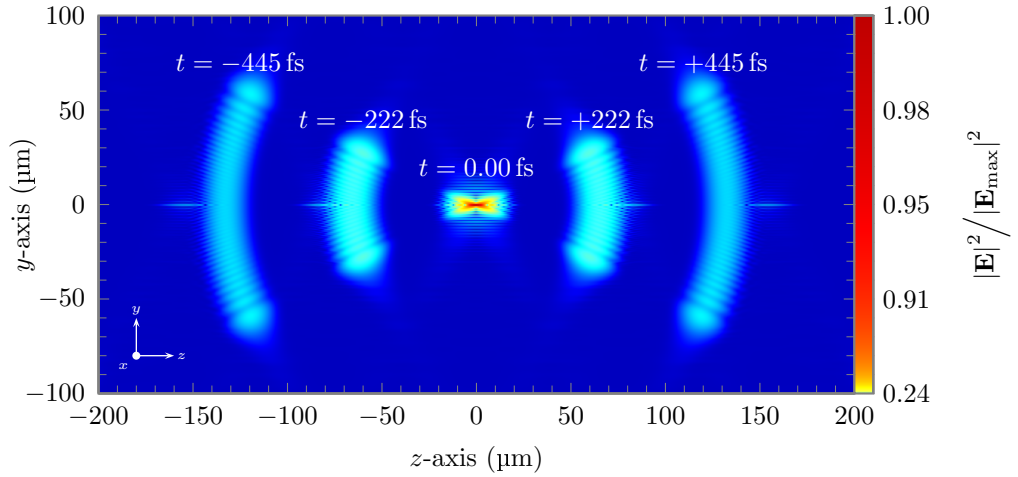
(b) timesteps normalized relative to each other

Figure 3.6: Intensity $|\mathbf{E}|^2$ of a 40 fs Gaussian pulse in the y-z plane of the focus of a parabolic mirror with focal length 50 mm, input polarization linear in the x-direction. The numerical aperture was chosen to be NA = 0.5. Units along the axis are in microns. The image is a composite of five frames separated by 2.2238×10^{-13} s, with the focus exactly at $t = 0$. The top figure shows the structure of the five frames with each frame normalized individually, while below the frames have been scaled relative to each other.
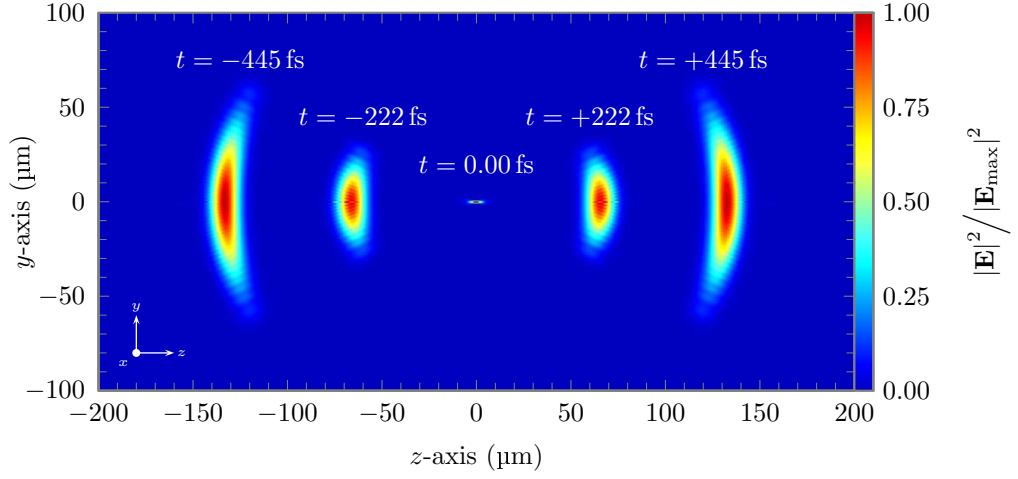
(a) each timestep normalized individually
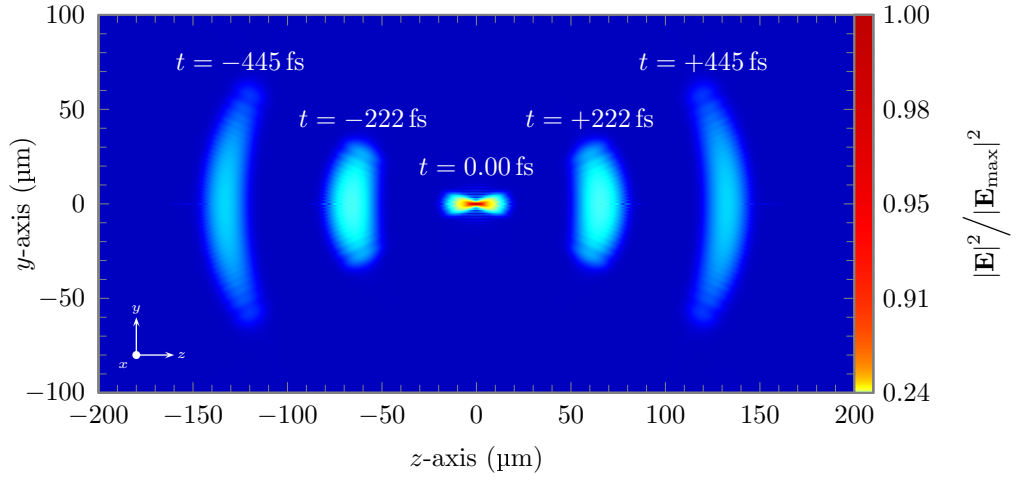
(b) timesteps normalized relative to each other

Figure 3.7: Complementary to Figure 3.6, this figure uses the same parameters but with a Gaussian spatial distribution in the entrance pupil. The spatial distribution was chosen such that $1/e$ of $|\mathbf{E}|$ was present at $r = 20.25\,\text{mm}$ in the entrance pupil.
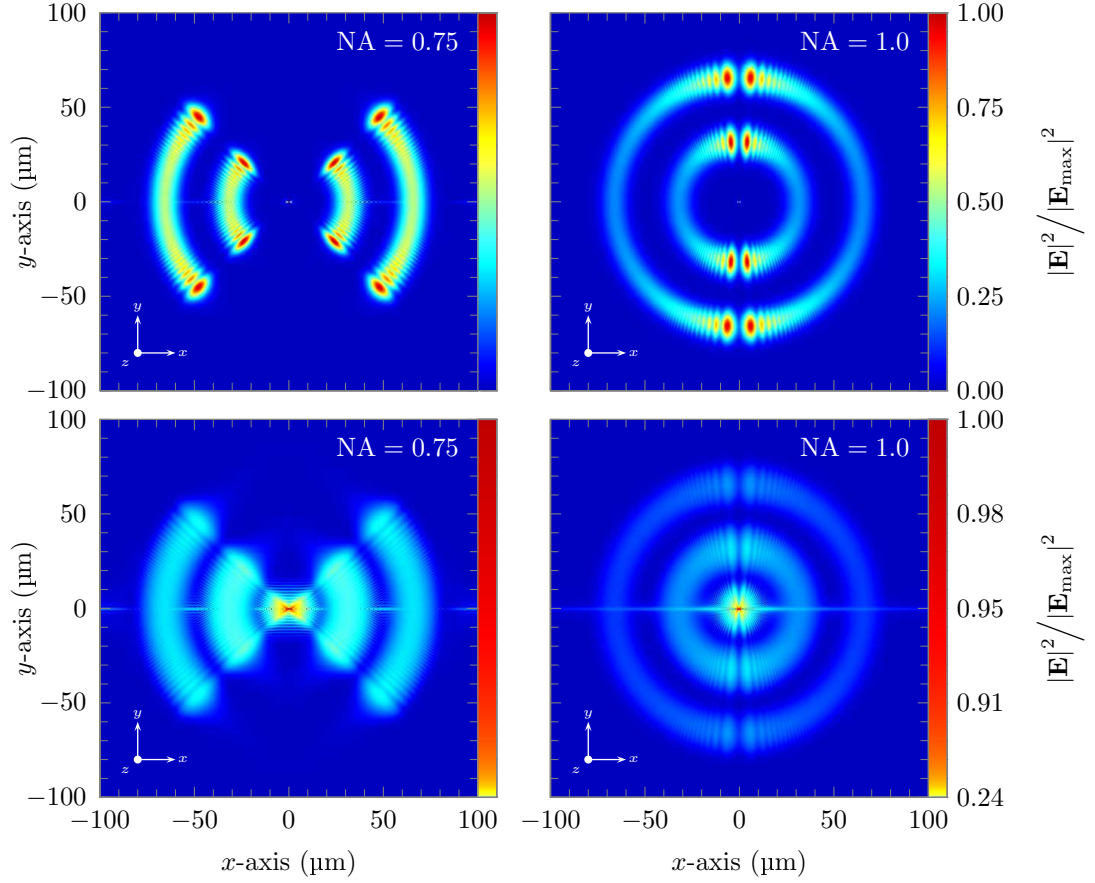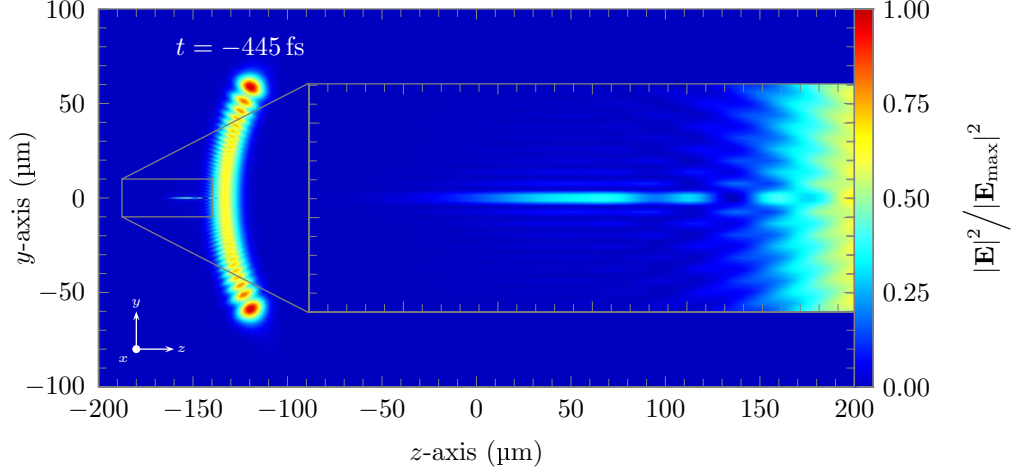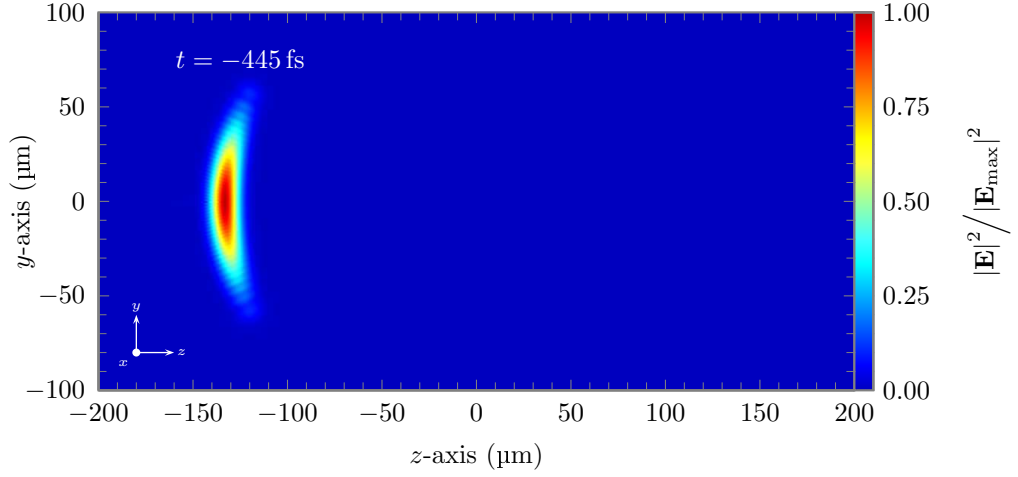
Figure 3.8: Electric field in the y-z plane of the focus for two high numerical aperture parabolic mirrors. These plots use the same geometry and pulse characteristics as in Figure 3.6, but the radius of the entrance pupil has been extended to increase the numerical aperture.

(a) uniform spatial intensity in the entrance pupil

(b) Gaussian spatial intensity distribution in the entrance pupil

Figure 3.9: Detail of a boundary wave pulse in the focal plane of a parabolic mirror (NA = 0.5, $f = 50\,\text{mm}$, and $\lambda_0 = 800\,\text{nm}$) for a particular timestep $t \approx 4.7346 \times 10^{-13}\,\text{s}$. The top figure shows the pulse for a uniformly illuminated entrance pupil. Within the top figure is a detail of the boundary wave pulse. The bottom figure shows this same system under the same parameters, but here we assume a Gaussian spatial profile radially about the entrance pupil with $1/e$ of $|\mathbf{E}|$ at a radius $20.25\,\text{mm}$ within the entrance pupil. Note the boundary wave pulse disappears with a Gaussian spatial distribution.
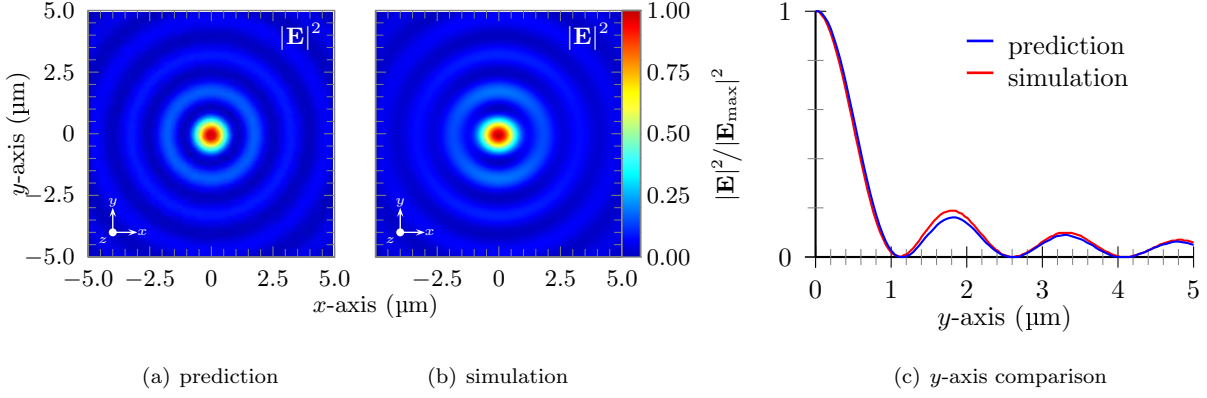
(a) prediction   (b) simulation   (c) y-axis comparison

Figure 3.10: x-y slices of the boundary wave pulses for $z = -153.96\,\mu\mathrm{m}$ and $t = -444.75\,\mathrm{fs}$ using the same geometry and optical setup found in Figure 3.6. The slight elongation of the spot in the simulation is due to incident linear polarization in the x-direction. The intensity of the boundary wave pulse here is approximately $73.29\,\%$ that of the main pulse.

The Fourier transform is then

$$\mathscr{F}^{+}\big(f(r)\big) = \int_0^\infty \int_0^{2\pi} f(r,\theta)\, \mathrm{e}^{\mathrm{i}k_r r \cos\theta}\, r\, \mathrm{d}r\, \mathrm{d}\theta \tag{3.30}$$

$$= \int_0^\infty f(r)\, J_0(k_r r)\, r\, \mathrm{d}r \tag{3.31}$$

$$= \int_0^\infty \delta(r-a)\, J_0(k_r r)\, r\, \mathrm{d}r \tag{3.32}$$

$$= a J_0(k_r a) \tag{3.33}$$

Where J_0 is the Bessel function of the first kind, order 0. The second step in the above took advantage of the fact that Fourier integrals of this type can reduced to a *Hankel transform*. The electric field of the boundary wave pulse should then be described by

$$\mathbf{E}(r) \propto \mathbf{E}_0 J_0(r) \tag{3.34}$$

A comparison between this predicted diffraction pattern and a simulation data consisting of an x-y slice of the boundary wave pulse shown in Figure 3.9 is presented in Figure 3.10 Despite the elongation of the spot due to polarization effects, the boundary wave pulse seems well described by Equation 3.34. Figure 3.10 presents both the predicted and observed structure of the boundary wave pulse for a circular entrance pupil using the same parabolic geometry of Section 3.1 (radius $23.7949\,\mathrm{mm}$)

Another interesting property of the boundary wave pulse is its velocity of propagation. Figure 3.18 shows the normalized intensity of different time slices along the z-axis for the same geometry as Figure 3.6. Following the main pulse in blue, it is seen to travel at c_0 The boundary wave, however, appears to travel

$$153\,\mu\mathrm{m}/4.4475\,\mathrm{s} \approx 1.1474 c_0 \tag{3.35}$$

It appears as if the pulse propagates at superluminal speed!

There is a geometrical argument which explains this apparent behavior, outlined in Figure 3.12. Consider a plane wave at point E on the aperture rim whose k-vector points to the focus at point F. Such an ideal plane wave is infinitely extended in the plane orthogonal to its direction of propagation, and intersects the optical axis at point B. If the plane wave is a distance $\|EF\|$ from the focus, the component which intersects with the optical axis is at a distance $\|BF\| = \|EF\| \sec\theta$ (where $\theta = \arcsin(\mathrm{NA})$). Consequently, if the main pulse propagates at velocity c, then the boundary wave pulse will appear to propagate at velocity[6]

$$c_{\mathrm{bw}} = c \sec\theta \tag{3.36}$$

(a) uniform spatial intensity
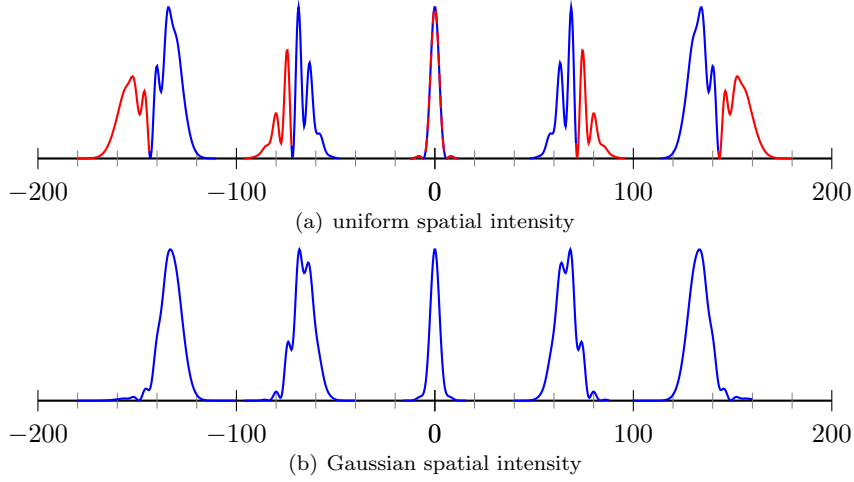

(b) Gaussian spatial intensity

Figure 3.11: Behavior of the boundary wave pulse along the z-axis for four discrete timesteps about the focus. The top figure is for a fully illuminated aperture and shows both the main pulse (blue) and the boundary pulse (red). The bottom figure shows a Gaussian spatial intensity in blue for the same timesteps. Each timestep has its pulses normalized independent of the other timesteps. The four discrete timesteps are separated by approximately 3.8×10^{-13} s. At $t = 0$ the pulses are coincident.

The delay between the two is similarly expressed. If the main pulse propagates at c, then at time t its position on the optical axis is ct (assuming $t = 0$ is coincident with the focus). The position of the boundary wave pulse is then $ct \sec \theta$, and the spatial delay Δz between the two is

$$\Delta z = ct(1 - \sec \theta) \tag{3.37}$$

In the limit of high numerical apertures, NA $\to 1.0$ or $\theta \to \pi/2$, the sec function diverges to infinity. In this case no boundary wave pulse is observed – the separation becomes infinite

$$\Delta z = \lim_{\theta \to \pi/2} (1 - \sec \theta) = +\infty \tag{3.38}$$

This is evidenced in Figure 3.8(b). In the limit of low numerical apertures the separation is

$$\Delta z = \lim_{\theta \to 0} (1 - \sec \theta) = 0 \tag{3.39}$$

meaning the main pulse and the boundary wave pulse are coincident. In general, it will be difficult to discern the pulses if their separation distance is less than the pulse's FWHM δt. Solving Equation 3.37 for the maximum distance from $z = 0$, $\bar{z}$ at which the separation distance between the main and boundary wave pulse is $\delta t/2$ gives the following useful relationship

$$\Delta z = \frac{c \, \delta t}{2} \tag{3.40}$$

$$ct(1 - \sec \theta) = \frac{c \, \delta t}{2} \tag{3.41}$$

$$\bar{z} = \frac{c \, \delta t}{2(1 - \sec \theta)} \tag{3.42}$$

As an example, for NA $= 0.5$ and $\delta t = 40$ fs, $\theta = \pi/6$ and $\bar{z} \approx 40 \, \mu$m. Conversely, for Δz to grow to a maximum of $c \, \delta t/2$ a distance of $200 \, \mu$m from the focus, the maximum numerical aperture is NA ≈ 0.25. In terms of the velocity of propagation for NA $= 0.5$, $c_{\text{bw}} \approx 1.1547$. This is within $1 \, \%$ of the value ≈ 1.1474 obtained from analysis of Figure 3.6.

Even though the boundary wave pulse travels at superluminal speed (and can be detected doing so [7]), it does not violate causality. The pulse is only an interference phenomenon; in the calculation of the propagation speed we have assumed that all waves originating from the annular aperture propagate at the speed of light.
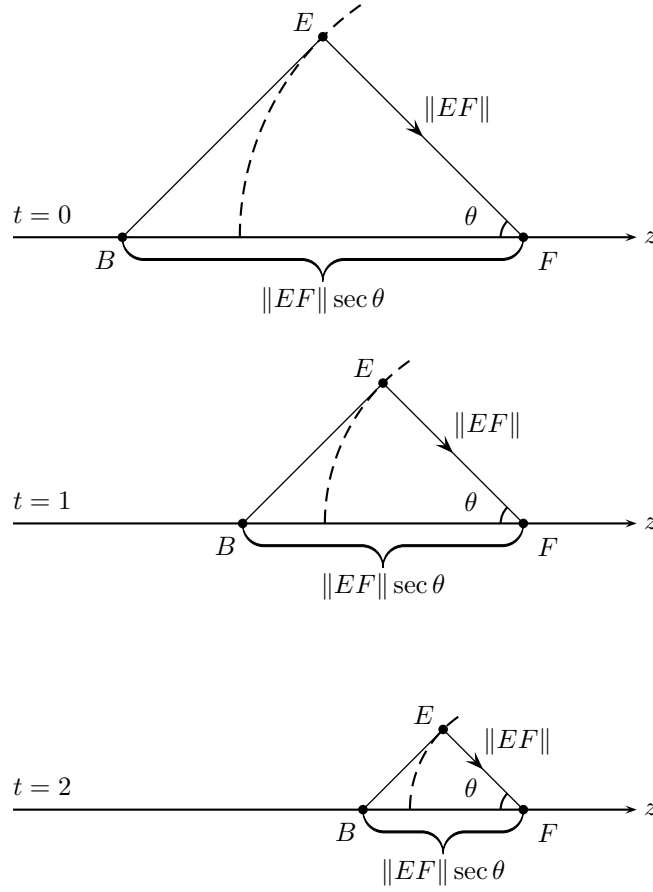
Figure 3.12: Definition of geometry for the boundary wave pulse. Imagine a plane wave, represented by a ray, propagating at the speed of light from a point E to the focus at F. Shown here is three example time steps. If the plane wave is infinitely extended orthogonal to its direction of propagation, it will intersect the optical axis at point B. The distance $\|BF\|$ is equal to $\|EF\| \sec \theta$. It can be seen from this arrangement that if a plane wave propagates from E to F at velocity v, the boundary wave pulse will propagate from B to F at velocity $v \sec \theta$.
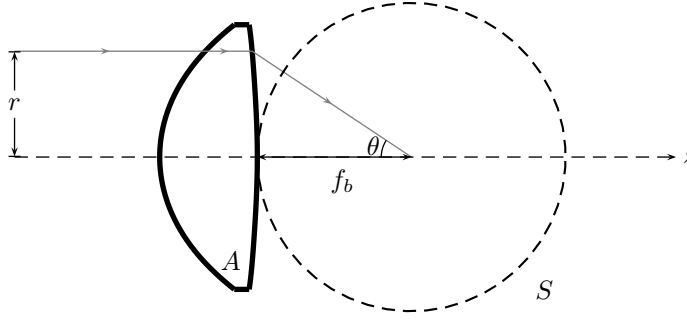
Figure 3.13: Aspheric test system under consideration consisting of a single refractive optic (Thorlabs 350240-B). The optic is illuminated by a plane wave sampled by rays along the entrance pupil of radius r. Light passes through the asphere, labeled A, whose best focus is at $\mathbf{r} = (0, 0, 0)$ (the center dot). The optic is made of ECO-550 glass. Assuming $\lambda_0 = 800\,\text{nm}$ the focal length $f \approx 8.00\,\text{mm}$ and back focal length $f_b \approx 5.825\,\text{mm}$. The numerical aperture is NA = 0.5. Figure is to scale.

	750 nm	800 nm	850 nm
Mean	1.6	0.34	1.5
RMS	0.81	0.12	0.79
P-V	2.8	0.46	2.7
Max	2.8	0.46	2.7
Min	0	0	0

Table 3.1: Numerical data for the wave aberrations plotted in Figure 3.14. "Mean" is the mean aberration, "RMS" is the root-mean-square statistical deviation from a perfect reference sphere averaged over the entire wavefront, "P-V" is the peak to valley distance. All numbers are in units of the wavelength they reference.

3.3 Refractive Optics

Many strategies for focusing pulses involve refractive optics. In contrast to the ideal reflective mirrors of the previous section, refractive optics exhibit dispersion. Effects from dispersion, whose theory was introduced in Section 1.3, will be important to this discussion.

3.3.1 Aspheric Lens

The first system under investigation consists of a single aspheric lens as depicted in Figure 3.13. This optic was chosen to facilitate comparisons both with the parabolic mirror from the previous sections and from results presented in [15]. Accordingly, the figures to follow continue to assume a 40 fs pulse at $\lambda_0 = 800\,\mu\text{m}$, input polarization linear in the x direction, and a numerical aperture of NA = 0.5.

The optic is fairly well corrected around the center pulse wavelength of 800 nm. Figure 3.14 details its wave aberrations on the focal sphere for 750, 800, and 850 nm. Pertinent information to this figure is cataloged in Table 3.1. At 800 nm the RMS wave aberrations are approximately $\lambda/8$. This is not quite diffraction limited: P-V wave aberrations are 0.46λ, where a diffraction limited optic would exhibit P-V $\leq 0.25\lambda$. However, as the spot radius predicted by RAYTRACE is $1.0693\,\mu\text{m}$, it is well within the tolerances imposed through the Debye approximation. In the derivation of Equation 1.52, it was assumed that the radius of the focal sphere f_b was much greater than the square of the radius of the focal volume divided by the wavelength.

$$f_b \gg \frac{r^2}{\lambda} \tag{3.43}$$

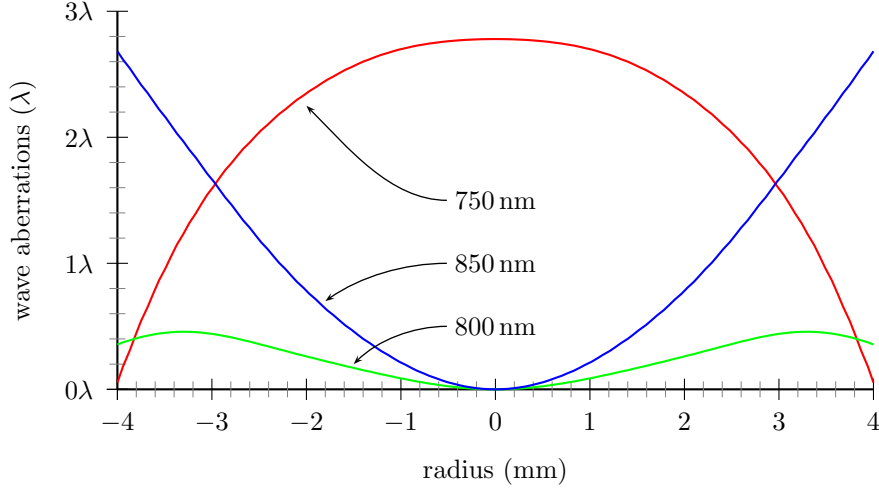For a spot radius of $1.0693\,\mu\text{m}$, $r^2/\lambda \approx 1.4293\,\mu\text{m} \ll f_b$ remains valid.

Figure 3.14: Wave aberrations on the focal sphere, radius equal to the back focal length $f_b \approx 5.825\,\text{mm}$, as computed by RAYTRACE for the aspheric optic of Figure 3.13. For the target wavelength of 800 nm this optic is not quite diffraction limited.
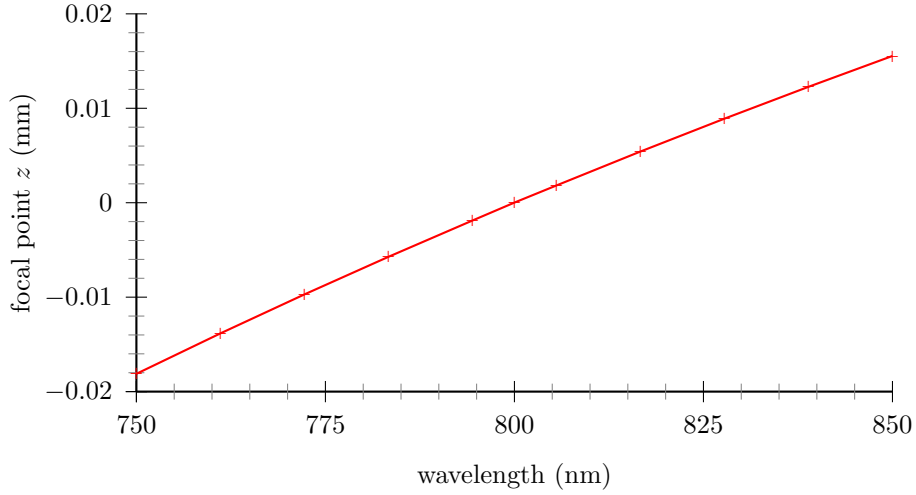

Figure 3.15: Focal point along the optical axis as a function of wavelength for the aspheric lens. The plot is shown such that the focal point for $\lambda = 800\,\text{nm}$ is $z = 0.0\,\text{mm}$.

To supplement the data in Figure 3.14, the focal point as a function of wavelength is given in Figure 3.15.

It is often convenient to estimate the effects of dispersion without computing the entire field in the focal plane. To accomplish this, a single ray is traced along the optical axis for spectral components centered around the target wavelength. The optical path lengths OPL of these rays are collected and divided by the speed of light to give their traversal time. Assuming the center wavelength traverses the optical system with time t_0, a slightly modified version of Equation 1.54 can be used to estimate the pulse structure in the time domain.

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{E}(\omega) \, e^{i\omega(t-t_0)} \, d\omega \qquad (3.44)$$

Tracing a single ray along the optical axis should represent the worst case scenario for the asphere in discussion, because the thickest part of the lens lies also on the optical axis. The effect of material dispersion this optic has on a 40 fs pulse is shown in 3.16 The result is a temporal broadening of about 41 %, with very little pulse front distortion. The low distortion is explained by noting that the second order dispersion of this glass (ECO-550) is approximately linear within the spectral range of this pulse:
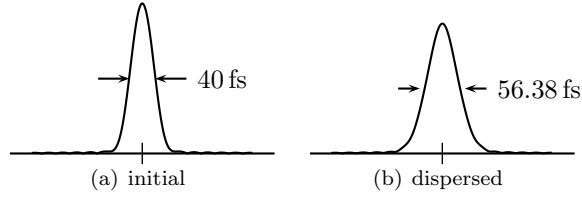
Figure 3.16: Effect of dispersion on the on-axis pulse width in the time domain for the aspheric optic of Figure 3.13.

$\Delta \, \mathrm{d}n/\, \mathrm{d}\lambda \approx 0.0172$ in the range of 700 nm to 800 nm. This is slightly worse than the common BK7, from which $\Delta \, \mathrm{d}n/\, \mathrm{d}\lambda \approx 0.0105$ in the same range.

The full electric field in the y-z plane is shown in Figure 3.17. Due to the dispersion imposed by the radially varying thickness of the asphere, the pulse shape itself likewise exhibits a radially varying temporal width. This delay has been estimated to be ≈ 16 fs on the optical axis and ≈ 5 fs at the rim as compared with a dispersion-free system.

The aspheric system produces a unique and interesting pulse front. This is, primarily, due to the presence of chromatic aberrations [7]: for an aberration free system such as an ideal focusing paraboloid, the pulse front would be symmetric about the focus and a flat front would be present in the focal plane ($z = 0$). Chromatic aberration has the effect of shifting the front, in this case to about 200 μm behind the focus.

Due to the fully illuminated aperture, the aspheric lens exhibits a boundary wave pulse similar to the parabolic mirror. Though the optical system is markedly different, the pulse travels at a velocity predicted by Equation 3.36, being dependent only on the numerical aperture. The main and boundary wave pulses on the optical axis have been plotted together in Figure 3.18 using the same data as in Figure 3.17.

3.3.2 Microscope Objective

A well corrected optical system desires to minimize all possible aberrations. For refractive systems, the paragon of this is commonly found in microscope objectives. This section will use as an example one such objective designed by Nikon Corporation. The construction of this optic is detailed in [27].

Keeping accordance with other optics in this work, the entrance pupil of the objective has been set such that the numerical aperture is NA = 0.5. The incident pulse continues to be 40 fs at a center wavelength of 800 nm. The focal length is $f = 4.02$ mm. Due to the high number of optical elements and specific construction parameters, the reader is directed to [27] for more detailed information. This citation also includes more detailed information on the performance characteristics of this optic.

Though designed for operation in the ultraviolet, the objective is corrected for operation at around 800 nm. Its wave aberrations on the focal sphere are shown in Figure 3.21. For the center wavelength, the RMS wave aberrations are 0.08λ. These characteristics, combined with a predicted spot radius of 4.9711×10^{-7} m, suggest that this optic is within the parameters set by the Debye approximation. To supplement the data in Figure 3.21, the focal point as a function of wavelength is given in Figure 3.20.

Since there are many optical elements involved, and consequentially the pulse spends a comparatively long time in the glass, the dispersive effects on the temporal width of the pulse are much greater than a single optic such as the asphere. Shown in Figure 3.22 is the on-axis effects of dispersion for an input pulse of 40 fs. The pulse duration has evolved to ≈ 134 fs - more than a three fold increase. Although the pulse has been significantly broadened, the system exhibits very low second order dispersion and distortion. This is exhibited in Figure 3.22. Here the pulse after propagation through the system has been compared with an otherwise unmodified Gaussian pulse of the same width. Note that the pulses structure is nearly identical, owing to this same low second order dispersion.

The full field structure for the $x = 0$ slice of the y-z plane is shown in Figure 3.23. In contrast to the aspheric lens presented in the previous section, there is no radially varying temporal width to the pulse. This is an example of a well corrected objective: it focuses all spectral components (within some range) nearly to the same focal point.
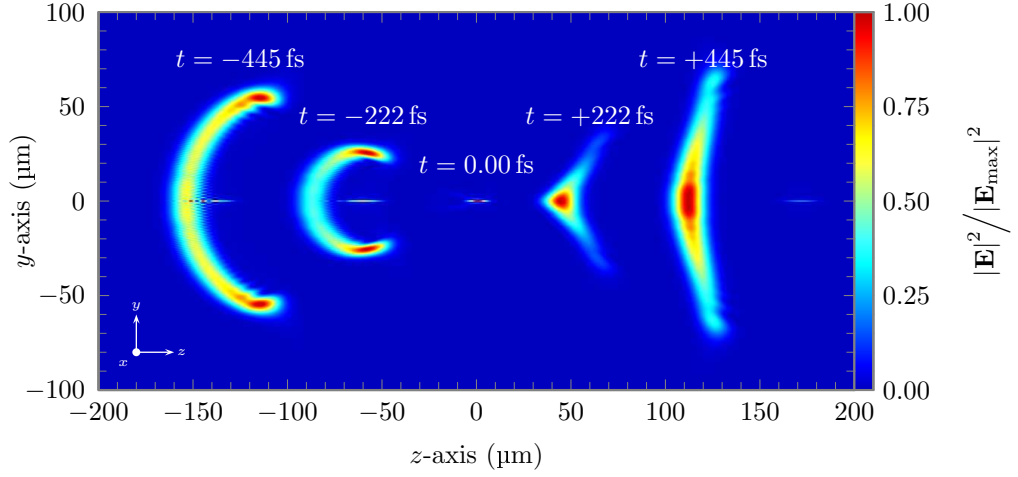
(a) each timestep normalized individually
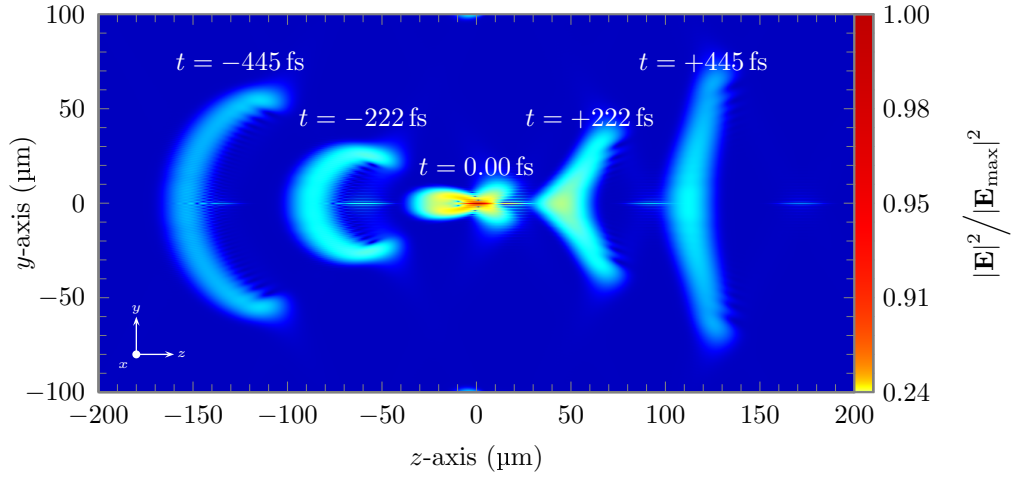
(b) timesteps normalized relative to each other

Figure 3.17: Intensity $|\mathbf{E}|^2$ 40 fs Gaussian pulse in the y-z plane of the focus of a aspheric lens (Thorlabs 350240-B) with back focal length $f_b \approx 5.825$ mm, input polarization linear in the x-direction. The numerical aperture was chosen to be NA = 0.5. Units along the axis are in microns. The image is a composite of five frames separated by 2.2238×10^{-13} s, with the focus exactly at $t = 0$. The top figure shows the structure of the five frames with each frame normalized individually, while below the frames have been scaled relative to each other.
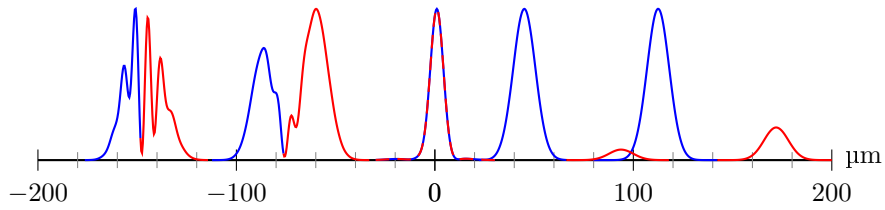

Figure 3.18: A boundary wave pulse also occurs for the aspheric lens. This figure details such behavior along the optical axis. The main pulse is shown in blue, while the boundary wave pulse is shown in red. Each pair has been normalized relative to each other, but not relative to other pulse pairs. Intensity units are arbitrary. The times for the pulses are the same as for Figure 3.17. At $t = 0$ the pulses are coincident.
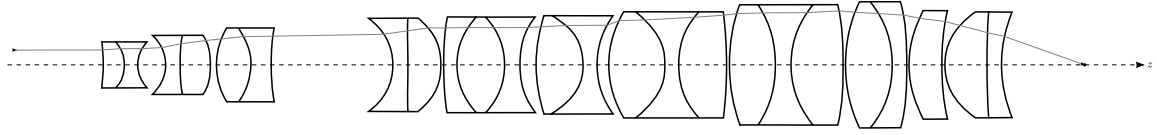
60

Figure 3.19: Schematic of the microscope objective used in the calculations of this section, detailed in [27]. The system consists of 34 optical surfaces, with the distance from the first to last surface being approximately 84 mm.
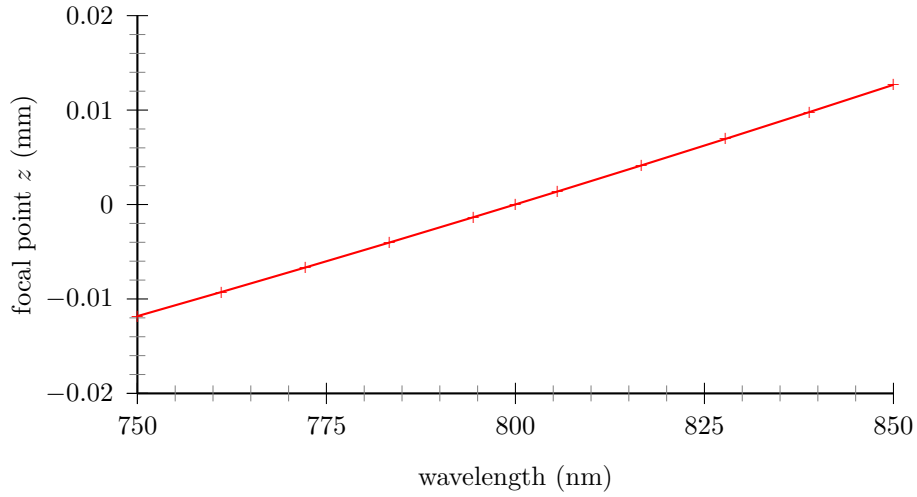

Figure 3.20: Focal point along the optical axis as a function of wavelength for the microscope objective. The plot is shown such that the focal point for $\lambda = 800$ nm is $z = 0.0$ mm.
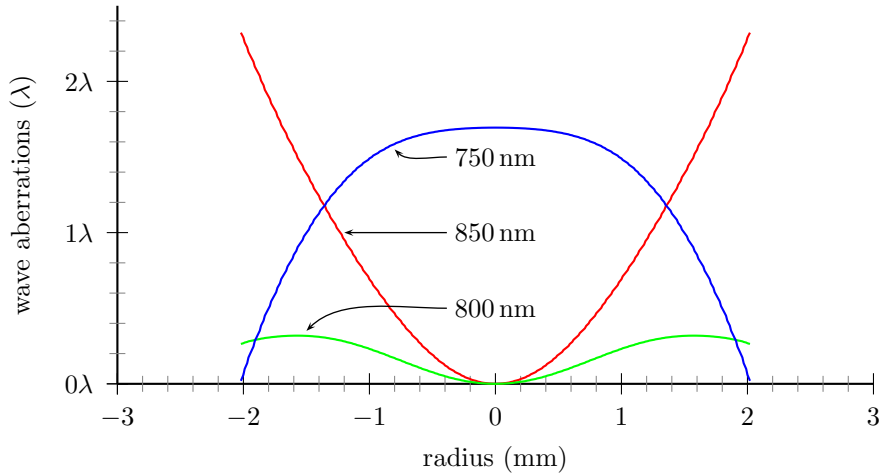

Figure 3.21: Wave aberrations on the focal sphere, radius equal to the back focal length $f_b \approx 7.4549$ mm, as computed by RAYTRACE for the microscope objective under discussion. For the target wavelength of 800 nm this optic is diffraction limited.
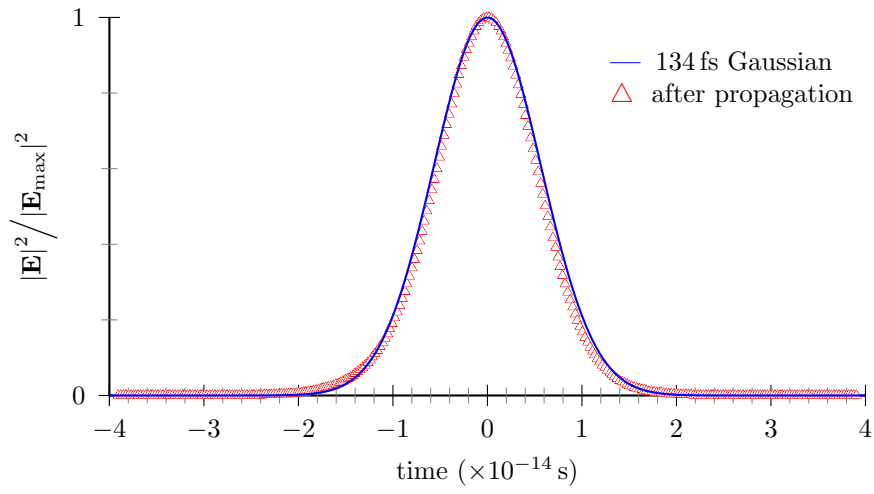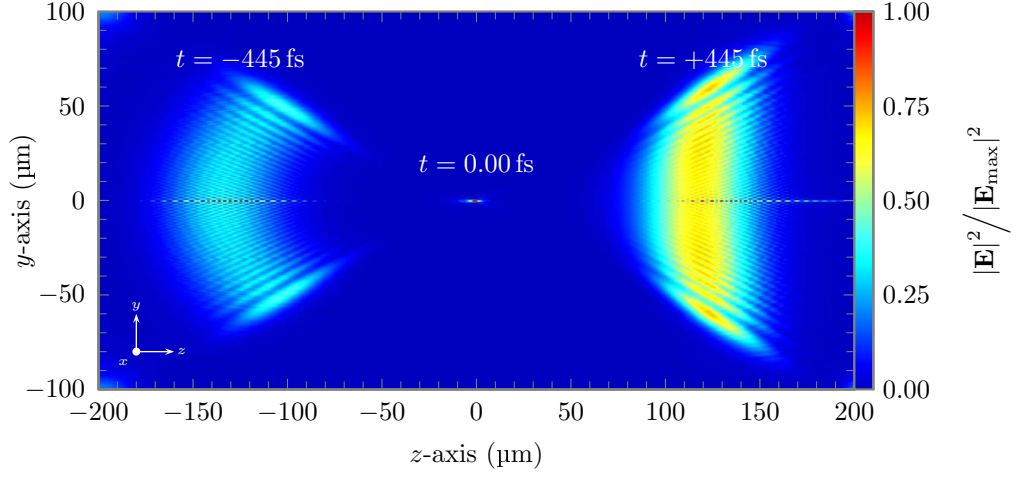
Figure 3.22: Temporal behavior of a 40 fs incident pulse in the focus of the microscope objective. Due to dispersion, the pulse has been broadened to nearly 134 fs. An otherwise unmodified 134 fs Gaussian pulse is plotted against this data to show that, while there is certainly temporal broadening, the overall pulse shape remains largely undistorted.

(a) each timestep normalized individually
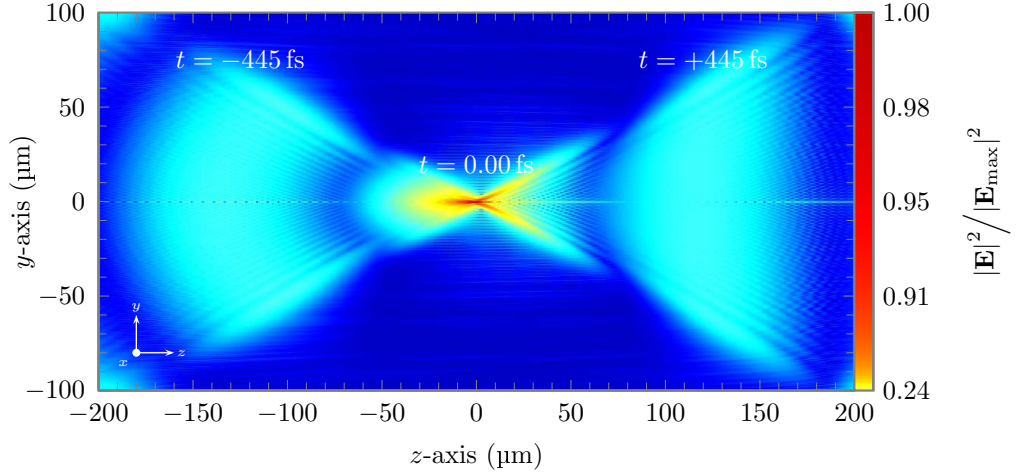
(b) timesteps normalized relative to each other

Figure 3.23: Intensity $|\mathbf{E}|^2$ in the $x = 0$ slice of the y-z plane of the focus of the microscope objective for an incident 40 fs Gaussian pulse at $\lambda_0 = 800$ nm. The objective has a back focal length $f_b \approx 7.4549$ mm and a numerical aperture of NA = 0.5. Input polarization is linear in the x-direction. Units along the axis are in microns. The image is a composite of three frames separated by 4.45×10^{-13} s, with the focus exactly at $t = 0$. The top figure shows the structure of the five frames with each frame normalized individually, while below the frames have been scaled relative to each other.

Chapter 4

Summary

The preceding chapters outlined a theoretical framework for the calculation of the electromagnetic field in the focus of certain complex optical systems under the scope of the Debye approximation. This was complemented by a numerical implementation, allowing said field to be determined with ray data on the focal sphere. Such ray data may be obtained by any number of raytracing software; in this regard the numerical implementation is agnostic. Different methods were also discussed on how to obtain certain parameters (such as the apodization factor), which may not be readily available in every raytracing environment.

A convenient characterization of pulses was laid out in Section 1.2. By using these conventions, comparing parameters from a simulation to those obtained through experiment or another independently written simulation is straightforward.

In Chapter 3 several optical systems were presented and their pulse focusing behavior analyzed. These systems are meant to be representative of a wider class of focusing geometries common to ultrashort pulses. The results show good agreement with other simulations of this nature.

There are several areas of further investigation warranted by the work presented thus far. In terms of the actual implementation, it may be possible to overcome some of the problems with interpolation and reformulate the Debye sum as a fast Fourier transform. This could greatly speed up computation and make the simulation more accessible. The numerical simulation could also benefit by being carefully checked with experimental data. Certain methods[7] exist which are able to resolve the pulse structure directly. This would give an idea of how well and under what regions the simulation models real world physics.

Many experiments assume a vacuum behind the optical system, and only in the direct vicinity of the focus a material (e.g. atom or a molecule). In such a case our methods are perfectly suited to be combined with, for example, finite-difference time-domain (FDTD) methods. This could be done by sampling the electromagnetic field for points on a small boundary around the focus using the Debye approximation, and then modeling light-matter interaction using FDTD. This is somewhat analogous to how the electromagnetic field was sampled on the focal sphere by rays and then propagated to the focal region through the Debye approximation. There are interesting methods which further apply to FDTD, such as the near to far field transform[43]. This may allow the scattered light to be reconstructed and possibly recycled into additional optical systems.

Bibliography

[1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover publications, 1964.

[2] E.A. Bashkov and V.S. Babkov. Research of rbf-algorithm and his modifications apply possibilities for the construction of shape computer models in medical practic. In *Proc Int. Conference "Simulation-2008".* Pukhov Institute for Modelling in Energy Engineering, 2008.

[3] IM Bassett. Limit to concentration by focusing. *Journal of Modern Optics*, 33(3):279–286, 1986.

[4] Alexandre Becoulet. Gnu optical design and simulation library, 2009.

[5] M. Bom and E. Wolf. Principles of optics. *Pergamon, New York*, 1965.

[6] Z. Bor and ZL Horvath. Distortion of femtosecond pulses in lenses. Wave optical description. *Optics Communications*, 94(4):249–258, 1992.

[7] P. Bowlan, P. Gabolde, and R. Trebino. Directly measuring the spatio-temporal electric field of focusing ultrashort pulses. *Optics Express*, 15(16):10219–10230, 2007.

[8] R.P. Brent. *Algorithms for minimization without derivatives.* Dover Pubns, 2002.

[9] JC Carr, RK Beatson, JB Cherrie, TJ Mitchell, WR Fright, BC McCallum, and TR Evans. Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 67–76. ACM, 2001.

[10] R. Chamberlain and W. Duquette. Some algorithms for polygons on a sphere, 2007.

[11] A.K. Cline. FITPACK-software package for curve and surface fitting employing splines under tension. *Department of Computer Sciences, University of Texas, Austin*, 1981.

[12] N. Davidson and N. Bokor. High-numerical-aperture focusing of radially polarized doughnut beams with a parabolic mirror and a flat diffractive lens. *Optics letters*, 29(12):1318–1320, 2004.

[13] I. Fanderlick. *Optical properties of glass.* Elsevier Science Pub., New York, NY, 1983.

[14] M. Frigo and S.G. Johnson. FFTW: An adaptive software architecture for the FFT. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 3, pages 1381–1384. IEEE, 1998.

[15] U. Fuchs, U. Zeitner, and A. Tünnermann. Ultra-short pulse propagation in complex optical systems. *Optics express*, 13(10):3852–3861, 2005.

[16] J.M. Gablonsky and C.T. Kelley. A locally-biased form of the DIRECT algorithm. *Journal of Global Optimization*, 21(1):27–37, 2001.

[17] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. *GNU scientific library.* Citeseer, 2002.

[18] I.S. Gradshteyn, I.M. Ryzhik, A. Jeffrey, D. Zwillinger, and S. Technica. *Table of integrals, series, and products*, volume 1980. Academic press New York, 1965.

[19] P.S. Heckbert. *Graphics gems IV*. Morgan Kaufmann, 1994.

[20] Steven G. Johnson. The nlopt nonlinear-optimization package.

[21] P. Kaelo and MM Ali. Some variants of the controlled random search algorithm for global optimization. *Journal of Optimization Theory and Applications*, 130(2):253–264, 2006.

[22] M. Lieb and A. Meixner. A high numerical aperture parabolic mirror as imaging device for confocal microscopy. *Optics Express*, 8(7):458–474, 2001.

[23] N. Lindlein, S. Quabis, U. Peschel, and G. Leuchs. High numerical aperture imaging with different polarization patterns. *Optics Express*, 15(9):5827–5842, 2007.

[24] S. Maneewongvatana and D.M. Mount. Its okay to be skinny, if your friends are fat. In *Center for Geometric Computing 4th Annual Workshop on Computational Geometry*. Citeseer, 1999.

[25] Robert D. Miller. *Computing the Area of a Spherical Polygon*, pages 132–137. In [19], 1994.

[26] M. Lano O. Stolz und A. Mitnacht N. Lindlein, F. Simon. Raytrace version 17, 2009.

[27] Kenji Ono. Microscope Objective Lenses, 2002. US Patent 6,392,814 B1.

[28] M. Powell. The NEWUOA software for unconstrained optimization without derivatives. *Large-Scale Nonlinear Optimization*, pages 255–297, 2006.

[29] M.J.D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation, Advances in Optimization and Numerical Analysis. In *Proceedings of the Sixth Workshop on Optimization and Numerical Analysis, Oaxaca, Mexico*, volume 275, pages 51–67, 1994.

[30] MJD Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, Reino Unido*, 2009.

[31] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. Numerical recipe in C. *Cambridge, New York*, 1992.

[32] R.J. Renka. Interpolation of data on the surface of a sphere. *ACM Transactions on Mathematical Software (TOMS)*, 10(4):417–436, 1984.

[33] R.J. Renka. Algorithm 772: STRIPACK: Delaunay triangulation and Voronoi diagram on the surface of a sphere. *ACM Transactions on Mathematical Software (TOMS)*, 23(3):434, 1997.

[34] R.J. Renka. Algorithm 773: SSRFPACK: interpolation of scattered data on the surface of a sphere with a surface under tension. *ACM Transactions on Mathematical Software (TOMS)*, 23(3):442, 1997.

[35] T. Rowan. Functional stability analysis of numerical algorithms. *Unpublished doctoral dissertation, University of Texas at Austin*, 1990.

[36] A. Rubinowicz. Thomas Young and the theory of diffraction. 1957.

[37] A. Rubinowicz. V The Miyamoto-Wolf Diffraction Wave. *Progress in Optics*, 4:199–240, 1965.

[38] T.P. Runarsson and X. Yao. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 35(2):233–243, 2005.

[39] CJR Sheppard, A. Choudhury, and J. Gannaway. Electromagnetic field near the focus of wide-angular lens and mirror systems. *Microwaves, Optics and Acoustics, IEE Journal on*, 1(4):129–132, 2009.

[40] M. Sondermann, N. Lindlein, and G. Leuchs. Maximizing the electric field strength in the foci of high numerical aperture optics. *Preprint at http://arxiv.org/abs/0811.2098*, 2008.

[41] J.J. Stamnes. *Waves in focal regions*. Hilger, 1986.

[42] Daniel A. Steck. *Classical and Modern Optics*. 2008.

[43] A. Taflove, S.C. Hagness, et al. *Computational electrodynamics: The finite-difference time-domain method*. Artech House Norwood, MA, 2000.

[44] P. Varga and P. Török. Focusing of electromagnetic waves by paraboloid mirrors. I. Theory. *JOSA A*, 17(11):2081–2089, 2000.

[45] P. Varga and P. Török. Focusing of electromagnetic waves by paraboloid mirrors. II. Numerical results. *JOSA A*, 17(11):2090–2095, 2000.

Appendix A

Command Line Program

`directdebye`

NAME

`directdebye` computes the discrete Debye integral from raytracing data.

SYNOPSIS

`directdebye` [*OPTIONS*]

DESCRIPTION

Computes the Debye integral as a discrete sum over input ray data using the trapezoidal rule. The program expects the ray data to be a tab separated file with the following fields:

0 to 2	ray position x y and z
3 to 5	normalized ray direction e_x e_y,and e_z
6 to 7	polarization in x (real, complex)
8 to 9	polarization in y (real, complex)
10 to 11	polarization in z (real, complex)
12	wavelength
13	apodization factor
14	ray intensity for the particular wavelength

The output is a four dimensional HDF file in x, y, z and t. `directdebye` uses MPI to distribute its calculation amongst multiple processors. Since direct computation of the Debye integral takes $O(n^2)$ operations, it is recommended the program be run in an environment which can take advantage of this.

OPTIONS

--efield-out *x:y:z:t*
>   Size of the output array in pixels.

--efield-ub *x:y:z:t*
>   Upper bounds, in millimeters, for each dimension of the output array.

--efield-ub *x:y:z:t*
>   Lower bounds, in millimeters, for each dimension of the output array.

--output-file *file*
>   Output will be written to the HDF file *file*. Default is 'efield-out.h5'.

--output-E*n*2
>   Output the complex vectorial component of the field $|\mathbf{E}|^2$. n can be x, y, or z. Default is to output all components.

--output-E*n*

Output the complex vectorial field E_n. n can be x, y, or z. Default is to output all components.

--output-phase-E*n*

Output the phase of the vectorial field E_n. n can be x, y, or z. Default is to output all components.

--temporal-offset *offset*

Temporal offset. Use this parameter to make t=0 at the center of the coordinate system. Default is 0.

--radius *radius*

Radius of the focal sphere. Not mandatory. If not given, the output will only be qualitatively correct. Default is 1.

--pulse-envelope $f(x)$

Specify the pulse envelope manually. Causes the 'ray intensity' field (column 14) of the input file to be multiplied by the function $f(x)$, where x is in units of $2\pi c_0/\lambda$; c_0 is the speed of light in meters per second and lambda is the parameter specified in the ray input file. The string given can be any dimensional function in x supported by libmatheval. If no envelope is given it is assumed the envelope is specified in terms of the 'ray intensity' field.

--config *file*

Read configuration parameters from *file*. Each line of the configuration file has the format:

option [value]

where option is one of the above options without the '--' or '='. Options on the command line are supersede options in the configuration file.

BUGS

Send bug reports to Aaron Webster `<aaron.webster@physik.uni-erlangen.de>`

AUTHORS

Written by Aaron Webster. Copyright (c) 2011 Max Planck Institute for the Science of Light and University of Erlangen Nuremberg, Erlangen, Germany

Appendix B

Notation and Figures

variable	name	page of definition
$\mathbf{k}$	spatial angular frequency	5
$\mathbf{r}$	position vector	4
$\hat{\mathbf{n}}$	unit normal vector	7
ω	angular frequency	4
ν	spatial frequency	27
c	speed of light	5
λ	wavelength	5
$\mathrm{d}s$	surface element	5
S	surface of integration	5
$\mathbf{E}$	electric field	3
$\mathbf{B}$	magnetic field	3
$\mathbf{P}_0$	polarization vector	20
g	strength factor, or apodization	20
δt	full width at half maximum in the time domain	8
$\delta\omega$	full width at half maximum in the angular frequency domain	8
f_b	back focal length	1
$n, n(\lambda_0)$	index of refraction	17
OPL	optical path length	24

Table B.1: Different variables used throughout this work, their names, and the page number where their definition can be found.

figure	ray distribution	N_x	N_y	N_ω
2.6(a)	geometric	142	142	64
2.6(b)	square lattice	142	142	64
2.6(c)	random	142	142	64
3.2(a)	square lattice	250	250	1
3.2(b)	square lattice	250	250	1
3.3	geometric	142	142	1
3.4	geometric	142	142	1
3.5	square lattice	500	500	1
3.6	square lattice	250	250	62
3.8(a)	square lattice	375	375	27
3.8(b)	square lattice	500	500	27
3.17	square lattice	250	250	62
3.18	square lattice	250	250	62
3.23	square lattice	250	250	62

Table B.2: Detail of the actual numerical sampling of rays on the focal sphere for various figures in the text. The sampling conditions are described in Section 2.1. The ray distributions are described in Subsection 2.3.3.

Appendix C

Declaration of Originality

I affirm that the work presented in this thesis is, to the best of my knowledge, original and my own, except as acknowledged in the text.

Signed ———————————— (Aaron Webster)

Date ————————————