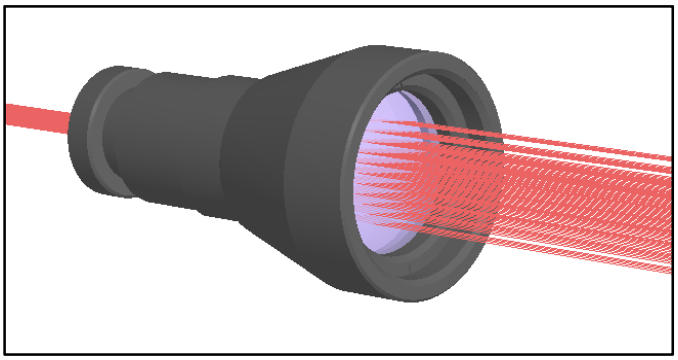
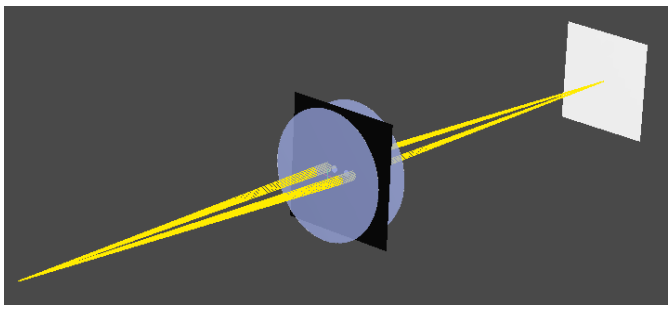
FRED® Application Note
Modeling Coherence

Photon ENGINEERING
Illuminating Ideas

This Application Note discusses aspects of modeling coherence in FRED and presents examples that demonstrate the tools available for coherent analyses.





FRED Tools for Modeling Coherence

- ✓ **Coherent Sources**
  FRED includes many default coherent sources, including: plane wave, point source, and laser beam. An alternate detailed source menu makes it easy and convenient to define custom sources.
- ✓ **Gaussian Ray Size Spot Diagram**
  Gaussian beams can be plotted on any plane showing each base ray and its $1/e^2$ ellipse, facilitating analysis and troubleshooting of beam divergence and sampling.
- ✓ **Ray Status Summary**
  Displays the status of every ray, making it simple to troubleshoot and diagnose ray errors.
- ✓ **Coherent Field Resampling**
  Spatially resample a scalar field to overcome coherent ray errors and under-sampling of a surface.
- ✓ **Coherent Field Analysis**
  Displays plots of amplitude, energy, phase, and wavefront for scalar or polarized vector fields.
- ✓ **Wavefront Calculation**
  Wavefront analysis and plotting with Zernike decomposition capability.
- ✓ **Partial Coherence**
  For specific applications, partially coherent sources and analyses can be performed.

Contents

440 S. Williams Blvd. Suite 106, Tucson, AZ 85711 – www.photonengr.com – sales@photonengr.com

Introduction

Modeling the coherent properties of propagating light is non-trivial. When modeling coherent systems in FRED, the user should have a general understanding of the way FRED performs coherent calculations using a generalized form of Gaussian Beam Decomposition (GBD). This Application Note describes some basic coherent modeling methods and considerations that should be made when using FRED, followed by an example of the *Coherent Field Resampling* feature as applied to a ThorLabs beam expander, and finally a diffractometer is used to demonstrate a model of partial coherence.

FRED Basics: Modeling Coherence

FRED uses a generalized form of Gaussian Beam Decomposition (GBD) to propagate coherent fields. The premise behind GBD, as first proposed by Arnaud in 1969, is that an arbitrary wavefront can be synthesized from a basis set of Gaussian beams that are propagated by raytracing. Conventional GBD methods limit the synthesis to one of two extreme conditions: a spatial decomposition using beamlets arranged on an evenly-spaced grid, or a Fourier decomposition, based on the spatial frequency content, into beamlets at a single spatial position with different phases and directions. An extension of Arnaud's method, developed by Gabor and implemented in FRED, allows these two methods to be utilized together in a flexible approach that is adaptable to a broader range of conditions.

In FRED, the optical field is represented by a superposition of Gaussian beamlets that are described and propagated using rays. A central "base" ray represents the trajectory of the beamlet and additional secondary "waist" and "divergence" rays track the evolution of the beamlet parameters. The relationship between a beamlet and its corresponding rays is shown in Figure 1. The rays fully describe the beamlet characteristics as they undergo refraction, reflection, and diffraction. This process is referred to as "complex raytracing." At any plane in the system, the coherent field can be calculated by determining the contribution of each beamlet at each position on the analysis surface and accounting for the phases.

Because they are Gaussian, the beamlets obey the well-known equation relating far-field divergence half-angle θ and minimum waist radius ω_0 (where λ is the wavelength and n is the index of refraction of the medium):

$$\tan\theta = \frac{\lambda}{\pi n \omega_0} \tag{1}$$

The beamlet's radius, $\omega_{beamlet}$, changes as it propagates and is related to the waist ray height, h_w, and divergence ray height, h_d, by:

$$\omega_{beamlet}^2 = h_w^2 + h_d^2 \tag{2}$$

where ray heights are measured relative to the base ray.

In order for the model to be accurate, the beamlets must remain Gaussian and perform best when they obey the paraxial approximation. This is perhaps the most important consideration in properly implementing

coherent raytracing. Attempting to operate outside this paraxial limit negates the ability of Gaussian beamlets to accurately sample optical components as they propagate. Failure of secondary rays to remain well-correlated with their base ray may lead to coherent ray errors and erroneous irradiance calculations.

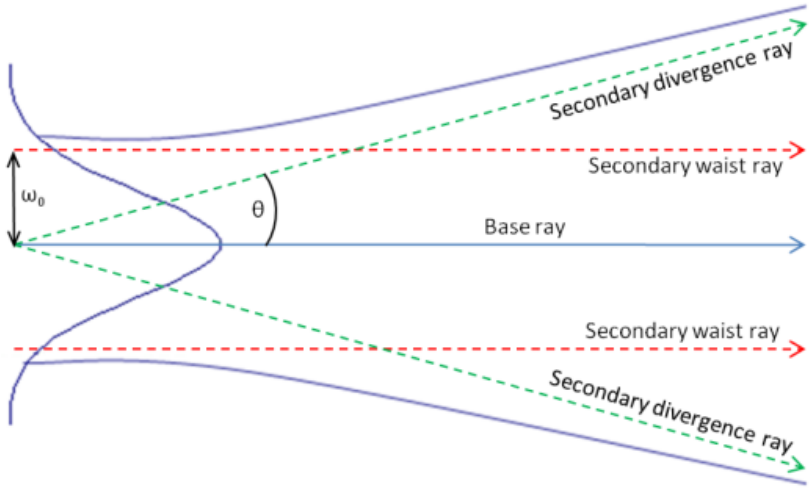


Figure 1. Gaussian beamlet with corresponding rays.

While not precisely defined, the paraxial approximation can be expressed in at least two forms: $\tan\theta \approx \theta$ and $\theta << \pi$. In both cases, a reasonable choice for θ is 0.1 radians, or about 6 degrees. The most obvious implication of the paraxial approximation, resulting from the Equation 1, is that the beamlet waist radius ω_o must be greater than or equal to about 3λ. In practice, the user should consider operating with some amount of margin, maybe 5-10λ. FRED keeps track of the beamlet's phase by tracing the secondary rays and calculating the optical invariant H'*n*U – H*n*U' (where H is the ray height, *n* is the material index, and U is $\tan\theta$). If the invariant for a given coherent ray exceeds $\lambda/4$, the ray is declared invalid for calculation purposes. The invariant serves as an internal measure of how well-behaved the beamlets are.

When a coherent source is defined, FRED sets up a ray grid and uses the grid spacing, $\Lambda_{x,y}$ (which is the width of the grid, $W_{x,y}$, divided by the number of rays across the grid, $N_{x,y}$), and a beam overlap factor (OF) to set the beamlet waist radius, ω_0, shown in Equations 3 and 4. Random ray grids should not be used for coherent sources because the grid spacing is uneven and beamlets of varying waists will be generated. There is independent control of the x and y values for the grid width and number of rays. The beamlet waist radius at the $1/e^{\pi/2}$ point where FRED traces rays is,

$$\omega_{0x,y}^{e^{-\pi/2}} = \frac{\Lambda_{x,y} OF}{2} = \frac{W_{x,y} OF}{2N_{x,y}} \tag{3}$$

Using the more common $1/e^2$ waist radius convention adds an extra factor,

$$\omega_{0x,y}^{e^{-2}} = \frac{W_{x,y} OF}{2N_{x,y}} \sqrt{\frac{4}{\pi}} \tag{4}$$

The overlap factor is the fractional overlap between adjacent beamlets on the grid, and has a default value of 1.5 (which should rarely be changed). Thus, to stay within the paraxial limit of $\omega_0 \geq 3\lambda$, for visible light (λ=0.5µm) the grid spacing should be $\Lambda \geq (2/1.5)\cdot 3\cdot 0.5\mu m = 2\mu m$; a more reasonable limit being closer to 5-10µm.

Focus on a FRED Feature: Coherent Field Resampling

There are certain instances when using a properly defined coherent source can still result in coherent ray errors. In this example, a ThorLabs 5x beam expander (BE05M) is used to illustrate FRED's *Coherent Field Resampling* feature as well as some other useful tools.

The beam expander is modeled in FRED by importing the CAD file provided from ThorLabs and replacing the CAD representation of the lens elements with native FRED lens elements that have the correct optical properties. Figure 2 shows the system layout in FRED using the *3D Cutaway view*.

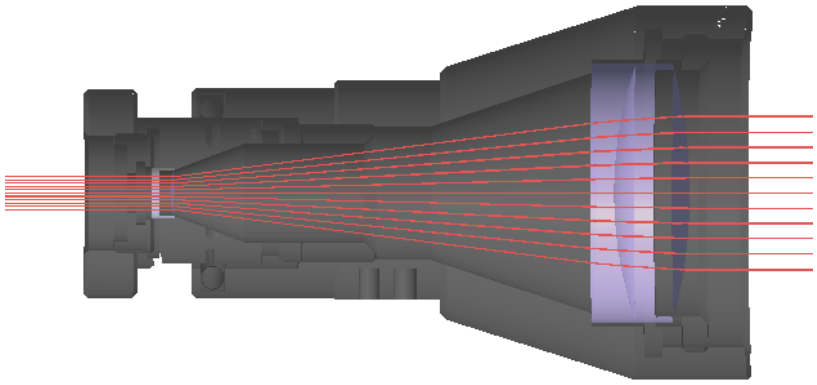


Figure 2. FRED model of a ThorLabs 5x beam expander. FRED's *3D Cutaway View* option is used to show the inside of the housing.

Coherent Source Definition

There are a number of default sources in FRED including a collimated source, point source, Gaussian TEM00 mode laser beam, and laser diode beam. A coherent Gaussian HeNe laser beam is used for this example. The inputs for a Gaussian laser beam are the beam size (waist semi-aperture), grid size (semi-aperture of the sample plane at the waist), and the number of points across the plane. A good rule of thumb is to set the beam size (beam waist) as half of the grid size. For this example, the beam is defined as circular with a beam waist radius of 0.5 mm (1mm diameter) on a 2mm x 2mm grid (W) with 41 rays across in each direction. This is a perfectly valid coherent source definition. It has a beamlet waist radius (ω_0,) of about 41.7 µm, well above the 10λ margin of 6.328 µm, and divergence angle (θ) of 0.28°, much below the paraxial limit of 6°:

The beamlet waist radius at the $1/e^2$ point (Equation 4) is

$$\omega_0 = \frac{W\,(OF)}{2N}\sqrt{\frac{4}{\pi}} = \frac{(2\text{ mm})\,(1.5)}{2(41)}\sqrt{\frac{4}{\pi}} = \mathbf{41.3\ \mu m}$$

And the beamlet divergence angle is

$$\tan\theta = \frac{\lambda}{\pi n \omega_0} = \frac{0.6328\ \mu m}{\pi(1)(41.3\ \mu m)} = 0.00488$$

$$\theta = \tan^{-1}(0.00488) = \mathbf{0.28}^\circ$$

Gaussian Ray Size Spot Diagram Tool

FRED's *Gaussian Ray Size Spot Diagram* analysis tool is extremely useful for examining Gaussian beamlet characteristics, visualizing secondary ray positions, and diagnosing coherent ray errors. The tool plots base rays with their corresponding $1/e^2$ ellipse. Although not explicitly plotted, secondary waist rays lie along this ellipse – typically four, in the ±x and ±y directions. FRED draws a $1/e^2$ secondary waist ray ellipse in the *Gaussian Ray Size Spot Diagram*, but uses secondary rays along a slightly smaller $1/e^{\pi/2}$ ellipse in the raytrace. Figure 3 show Gaussian spot diagrams of the laser source at two locations: (a) at the source and (b) 650 mm downstream. The base rays are perfectly collimated, but the laser beam itself is divergent, which can be observed by noting that the $1/e^2$ ellipses (circles in this case) have increased in size after propagating 650 mm.

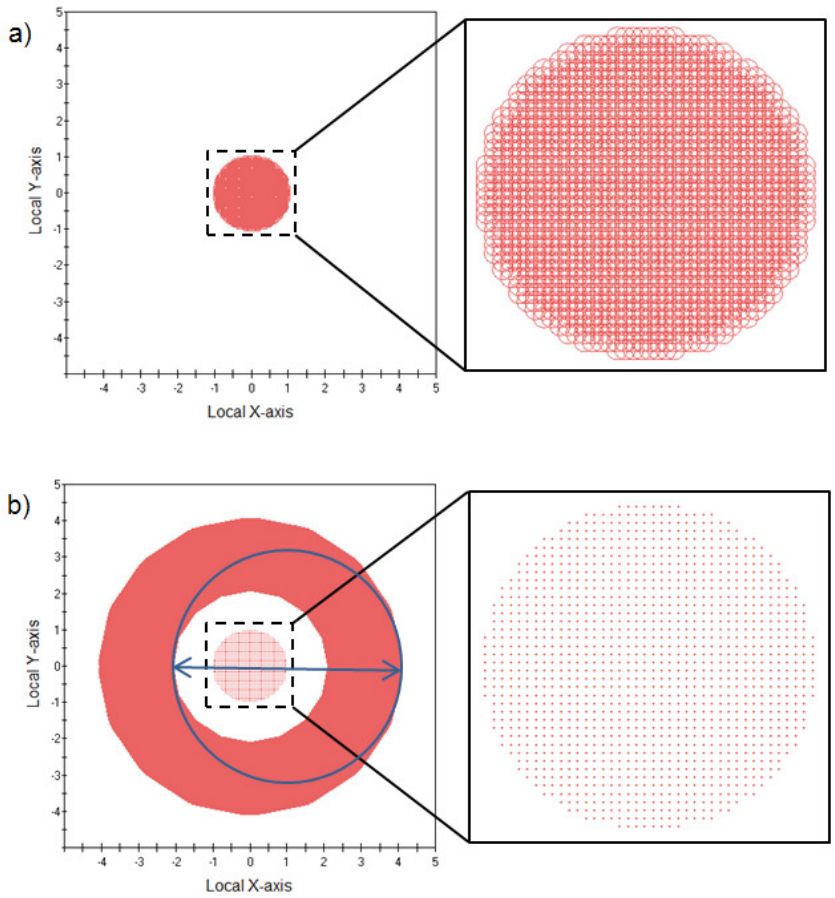


Figure 3. Gaussian Ray Size Spot Diagram with center enlarged to show detail. (a) At the source location. (b) After the beam has propagated 650 mm, with the right most beamlet outlined. The base rays are collimated and the secondary rays diverge.

Suppose the beam expander is part of an optical system that requires its first surface to be 650 mm from the laser source. One example is a Mach-Zehnder interferometer with different beam sizes in its arms, as shown in Figure 4.

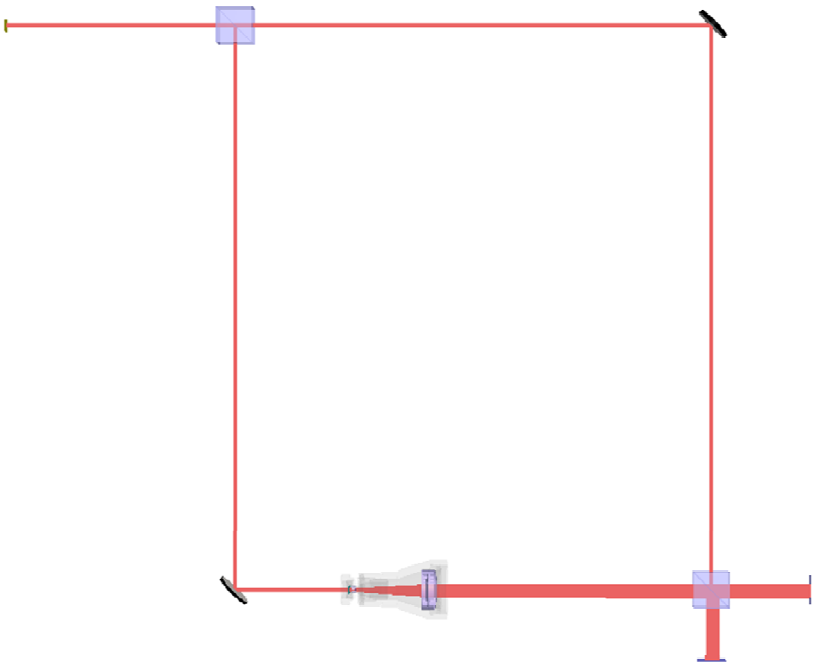


Figure 4. FRED model of Mach-Zehnder interferometer with a beam expander in one arm.

A closer look at Figure 3b shows that the outlined beamlet diameter is around 6 mm. The diverging lens of the beam expander has a diameter of only 5 mm. It may thus seem reasonable to conclude that the secondary rays are clipped, but this is not the case. One of the fundamental rules of complex raytracing is: *If the base ray intersects a surface, then all of its secondary rays must intersect the same surface.* FRED enforces this rule by mathematically extending the surface to intersect each secondary ray, as shown in Figure 5. When a raytrace is performed, all of the rays are traced through the system.

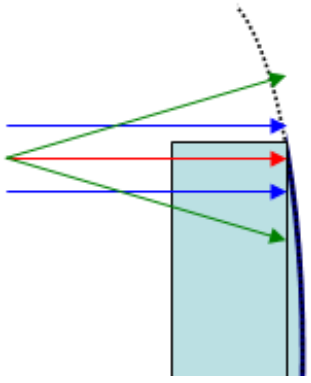


Figure 5. Mathematical extension of the optical surface is used for intersecting secondary rays that do not intersect the real surface.

There are three types of coherent ray errors (discussed in the following section), and FRED only shows a warning after the raytrace if it was unable to propagate the rays properly, as would occur if it was unable to mathematically extend a surface when necessary. The other two coherent ray errors produce warnings only once an analysis is attempted. In the case of the beam expander, the rays are traced with no errors or warnings because the spherical surface of the lens is easily extended, but issues do arise when an analysis is attempted because the secondary rays are no longer well-correlated with the base rays and the beamlets deviate from perfect Gaussians and become too divergent.

Ray Status

FRED's *Ray Status* tool is very handy in situations such as this beam expander model where there is a problem but the details and cause are unclear. The *Ray Status* outputs the status of all the rays currently in the system, as shown in Figure 6. There are three types of coherent ray errors:

1. *Coherent secondary ray raytrace errors*: This signifies that something happened during the raytrace that prevented all the rays from being properly traced. A warning that describes the specific issue is displayed in the Output Window in the *Raytrace Summary* after the raytrace is completed. For example, if a base ray travels through a ball lens but the secondary rays do not intersect the lens, it is impossible to extend the optical surface to force the secondary rays to make the intersection, and FRED outputs the following message: "Rays halted because unable to complete coherent secondary ray intersection (warn: 18)."
2. *Coherent secondary ray invariant violations*: This error occurs when the beamlet has deviated too far from a Gaussian. There are no errors or warnings when the raytrace is completed, only once an analysis is attempted.
3. *Coherent ray Gaussian exponential decay violations*: This error is very similar to the previous one, but indicates that the beamlet has become too divergent. This also does not report any errors or warnings until an analysis is attempted.

In the case of the beam expander, all 1313 rays violated the secondary ray invariant and Gaussian exponential decay and are thus invalid. As shown in Figure 6, the summary displays all errors that might have occurred during the raytrace and is very useful as a starting point when troubleshooting a system.

RAY STATUS SUMMARY:		(ThorLabsBeamExpander)
1313	total ray slots	
1313	total live rays	
1313	active	
0	traceable	
1313	specular	
0	scatter	
0	absorbed	
0	polarized	
1313	coherent	
1313	stopped by a surface's raytrace control specification	
0	fell below absolute transmitted power threshold	
0	fell below absolute reflected power threshold	
0	fell below relative transmitted power threshold	
0	fell below relative reflected threshold	
0	could not resolve material ambiguity	
0	could not resolve transmitted glue material ambiguity	
0	could not resolve reflected glue material ambiguity	
0	exceeded total intersection count	
0	exceeded consecutive intersection count	
0	exceeded specular ancestry threshold	
0	exceeded scatter ancestry threshold	
0	had coherent secondary ray raytrace errors	
1313	had coherent secondary ray invariant violations	
1313	had coherent ray Gaussian exponential decay violations	
0	are evanescent	
0	halted because total internal reflection not allowed	
0	acquired a bad position	
0	acquired a bad direction	
0	exceeded allowed number of steps in an inhomogenuous material	

Figure 6. Output window showing result from *Ray Status Summary* containing coherent ray errors.

Coherent Field Resampling

The *Coherent Field Resampling* feature can be used to resolve coherent ray errors by computing and generating a new set of coherent rays that reproduce the current scalar field over an extended spatial area. The new rays sum up to generate the same field, but the freshly synthesized beamlets have redefined waists and divergence angles. This is analogous to generating a brand new source with well-behaved beamlets. The area, pixel size, and location of the resampling grid are specified by an Analysis Surface entity. These parameters, just like the source creation grid, determine the beamlet properties. Therefore, the same considerations should be made concerning the pixel size and spacing as when defining a source. *Coherent Field Resampling* does the following: the wavefront is computed, any spherical and tilt terms are removed, the field is resampled, the spherical and tilt terms are reincorporated, and a new beamlet is created at the center of each pixel on the Analysis Surface to reproduce the original optical field. The *Coherent Field Resampling* dialog box is shown in Figure 7.

In this beam expander example, the field is resampled just in front of the first (diverging) lens with the same beam parameters as the original source, since it was already determined that the beamlet waist and divergence were reasonable. An Analysis Surface a little smaller than the size of the diverging lens (2 mm semi-aperture) is created, and number of divisions adjusted to result in the same beamlet waist as the original source (73). This adjustment is necessary because the original source was defined on a circular grid whereas the new field is defined on a square Analysis Surface.

The rays are traced from the source to the first lens by using the *Advanced Raytrace* feature, which allows precise control over the raytrace, including stopping the raytrace on a specific surface. The field can then be resampled (*Raytrace* > *Spatially Resample Scalar Field…*), deleting the existing rays and replacing them with

the newly defined ones. The resampled field should be identical to the original field, with the only difference being the rays used to represent it. The new rays are then traced through the rest of the system by using the *Trace Existing Rays* or *Trace and Render Existing* command.

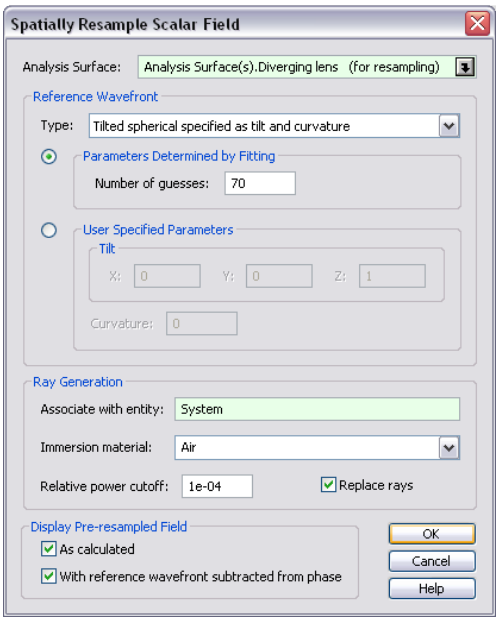


Figure 7. The *Field Resampling* utility dialog.

In addition to resolving coherent ray errors, *Coherent Field Resampling* can also be used in cases where a surface is being under-sampled. For example, if the beam expander's track length was much larger, the second lens would be overfilled and under-sampled, as shown in Figure 8. Because of the minimum grid size restraints described previously, increasing the number of source rays is not an appropriate solution. The *Coherent Field Resampling* feature can be used to reproduce the field at the second lens, and synthesize a new ray grid with sufficient spatial sampling of the lens.

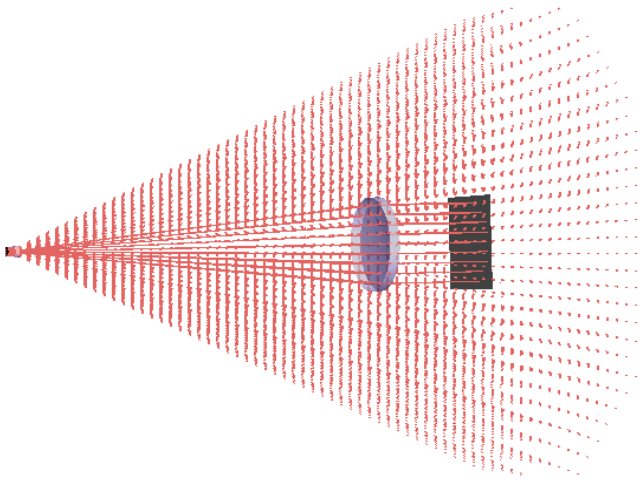


Figure 8. Large magnification afocal telescope resulting in under-sampling of the second lens.

Coherent Scalar Field Analysis

After the new rays have been synthesized they can be traced through the system without coherent ray errors, and the resulting field can be analyzed. The energy, phase, and wavefront of the coherent field can be investigated with the *Coherent Scalar Field* analysis tool. Figure 9 shows the scalar field menu with the available plot options boxed in red. It also shows the options for displaying and exporting plots, scaling data, smoothing and modifying plot data, displaying plot statistics, and performing a Fourier Transform. Figure 10 shows the field energy, phase, and wavefront at the output of the beam expander.

Figure 9. The *Scalar Field Analysis* menu is displayed when right-clicking the plot. Red box shows available plots.

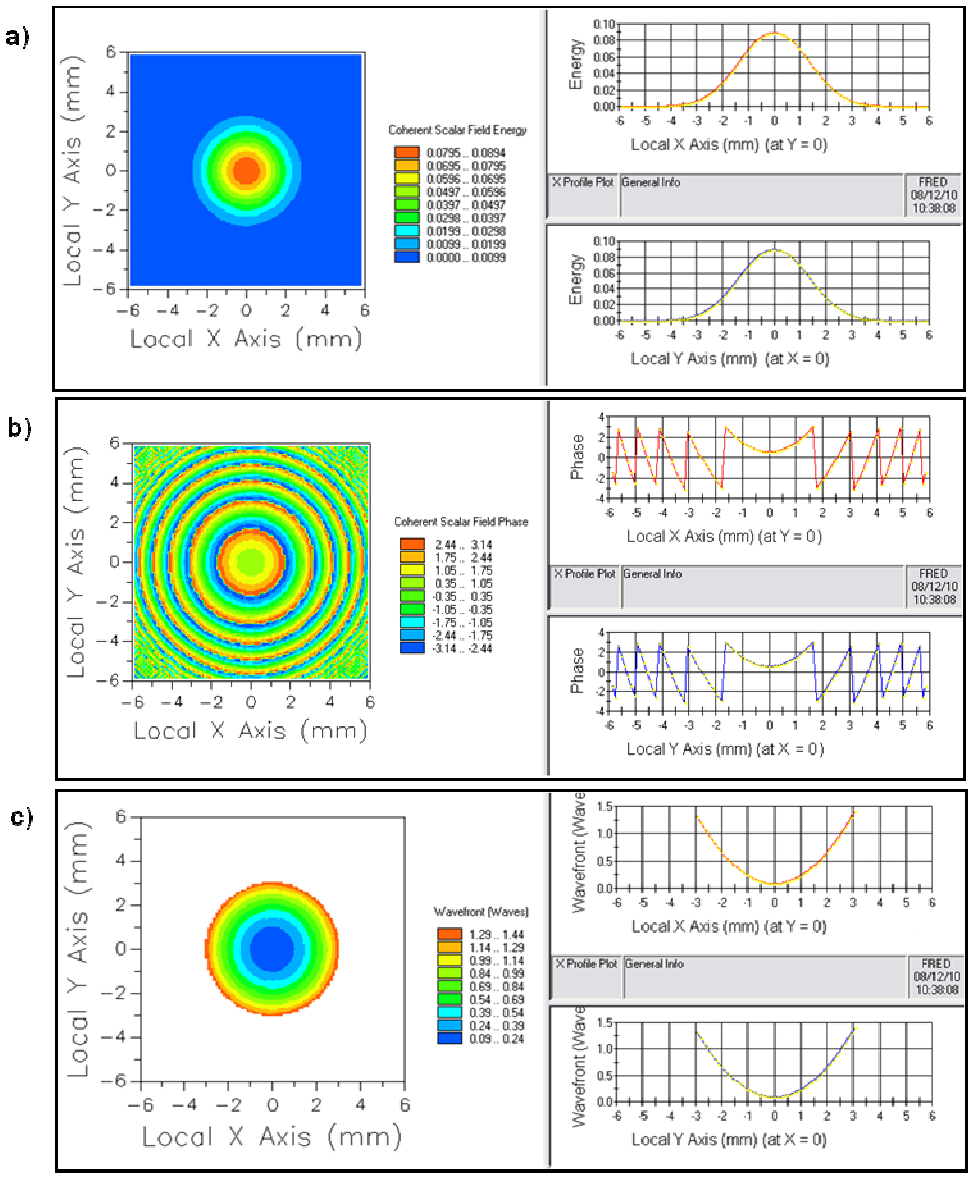


Figure 10. FRED output plots showing a) Field Energy, b) Field Phase, c) Wavefront.

Example of Partial Coherence in FRED: The Diffractometer

A partially coherent source can be modeled in FRED by a collection of coherent point sources, each with a different spatial position and wavelength. Modeling of partial coherence in FRED is limited to special cases where such a definition is valid.

One such case is the diffractometer, an interferometer that can be used to measure the spatial coherence of a source. This example is based on the setup described by Thompson and Wolf [1], shown in Figure 11. An extended, incoherent source, σ_0, is imaged by a lens, L_0, onto a pinhole, σ_1. The light emerging from σ_1 is collimated by L_1 and focused by L_2 onto the plane $\mathscr{F}$. An opaque screen, $\mathscr{A}$, containing two apertures, P_1 and P_2,

is placed between L_1 and L_2. The apertures can be of arbitrary size and shape, and be placed anywhere on plane $\mathscr{A}$.

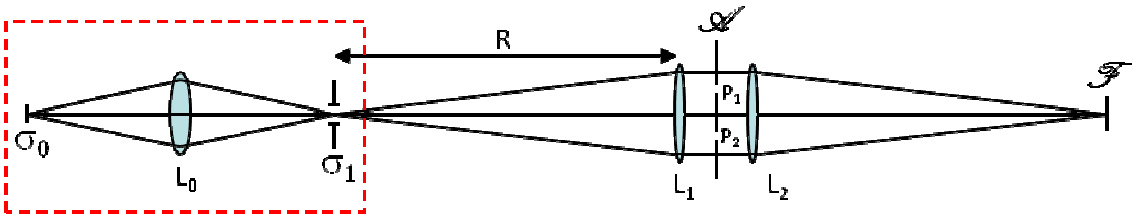


Figure 11. Diffractometer (after Thompson and Wolf).

The outlined section in Figure 11 can be modeled in FRED by a collection of point sources of differing wavelengths randomly positioned within the pinhole σ_1. This collection of sources meets the definition of a quasi-monochromatic source given by Born & Wolfe [2], if the wavelengths fall within a small bandwidth. For this example, wavelengths within 0.579 ± 0.002 μm are used. Each wavelength component of the source independently produces an interference pattern on the plane $\mathscr{F}$. FRED sums equal wavelengths coherently and different wavelengths incoherently. Thus, the total irradiance pattern on plane $\mathscr{F}$ is an incoherent sum of the individual coherent components from each wavelength.

Since only a small fraction of the rays pass through pinholes P_1 and P_2, the raytrace described above is very inefficient. A more efficient way to achieve the same result is to define two circular grids of rays just before the lens L_1. The ray grids should be analogous to the two pinholes (same spacing and location in the x-y plane), but slightly larger in size to ensure that the rays overfill the pinholes. The ray directions of the two circular sources are specified as originating from a random position within the σ_1 aperture. FRED has an option for specifying ray directions, called *Focus to/from a point*, which generates rays that focus to or from a point defined by the user. These two sources are created many times over at slightly different wavelengths, each set of grids defined as coming from a random position on the σ_1 aperture. Figure 12 shows two sets of sources with the rays extended back to the aperture and a limited number of rays to help demonstrate the source creation. To sum up the process, a random position within the aperture σ_1 is chosen. Two sources, one for each pinhole P_1 and P_2 (making up the source "set"), are defined just in front of the first lens. For both sources the ray direction is specified as having originated from the chosen random position. Then another position is chosen and two more corresponding sources are created. This process is repeated for a number of sources. In this example 75 such sets of sources are created using an embedded script. The goal is to simulate many point sources at random positions and wavelengths, within a small bandwidth, that have propagated to two regions just in front of L_1 such that they overfill the pinholes P_1 and P_2. Figure 13 shows the FRED model of the diffractometer. The rays were created in front of the first lens, but they have been extended back to σ_1 for the visual representation. This option is called *Post-Creation Ray Propagation* and can be useful in situations like this one, when it is helpful to visualize rays where they don't actually exist.

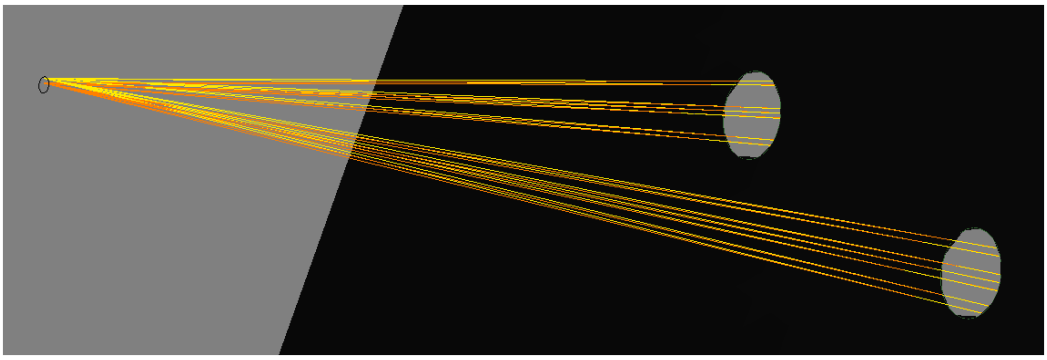


Figure 12. Two sets of sources being created analogous to the two pinholes, with the ray directions coming from a random position in the aperture on the left. The lenses are not displayed to prevent obstruction of the pinholes.

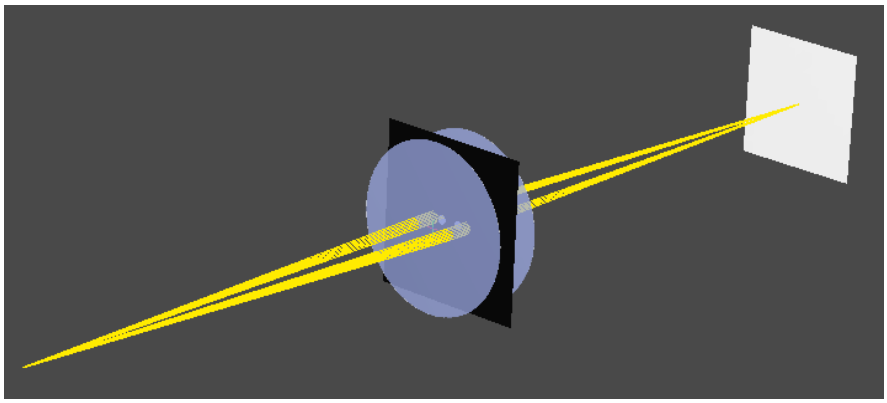


Figure 13. FRED model of diffractometer. System has been scaled down in the z-direction to provide a view of the entire system via Anamorphic 3D View.

According to the van Cittert-Zernike theorem, developed independently by P.H. van Cittert in 1934 and later by F. Zernike in 1938, the source collection at σ_1 gives rise to a correlation between the field at any two points P_1 and P_2 on the screen $\mathscr{A}$. The van Cittert-Zernike theorem establishes the complex degree of partial coherence as

$$\mu_{12} = \frac{2J_1(\mathrm{v})}{\mathrm{v}} e^{i\psi} \tag{4}$$

where

$$\mathrm{v} = \frac{2\pi}{\lambda_m}\frac{\rho d}{R}, \qquad \psi = \frac{2\pi}{\lambda_m}\left(\frac{r_1^2 - r_2^2}{2R}\right) \tag{5, 6}$$

and J_1 is the Bessel function of the first kind of order 1, ρ is the radius of the pinhole σ_1, d is the center-to-center distance between P_1 and P_2, R is the focal length of L_1, r_1 and r_2 are the distances of P_1 and P_2 offsets from the optical axis, and λ_m is the mean wavelength. The amplitude of μ_{12} is referred to as the degree of partial coherence, $|\mu_{12}|$.

The effect of d, the pinhole separation, on the degree of partial coherence, $|\mu_{12}|$, is examined. The following values are used in the simulation: ρ = 0.045 mm, R = 1505.6 mm, $r_1 = r_2 = 0$ (on axis), and λ_m = 0.579 μm. The dependence of $|\mu_{12}|$ on d is oscillatory because of the Bessel function, as shown in Figure 14. The interference patterns on the plane $\mathscr{F}$, as modeled in FRED, for four pinhole separations (which are marked by dots in Figure 14) are displayed in Figure 15. The FRED simulation results are in good agreement with Thompson and Wolf.

When the intensity at the two pinholes is equal and temporal coherence can be ignored, as is the case here, the fringe visibility is equal to the degree of partial coherence [3]. Fringe visibility is defined by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{7}$$

where I_{max} and I_{min} are the maximum and minimum values of the fringe irradiance. To compare the FRED model to the theoretical degree of partial coherence values, the fringe visibility was estimated for each of the four pinhole separations by using the maximum and minimum irradiance values of the central fringe. The red X's in Figure 14 correspond to the estimated fringe visibility based on the interference patterns shown in Figure 15. As demonstrated in Figures 14 and 15, the approximations based on the FRED model are reasonably close to the theoretical values.

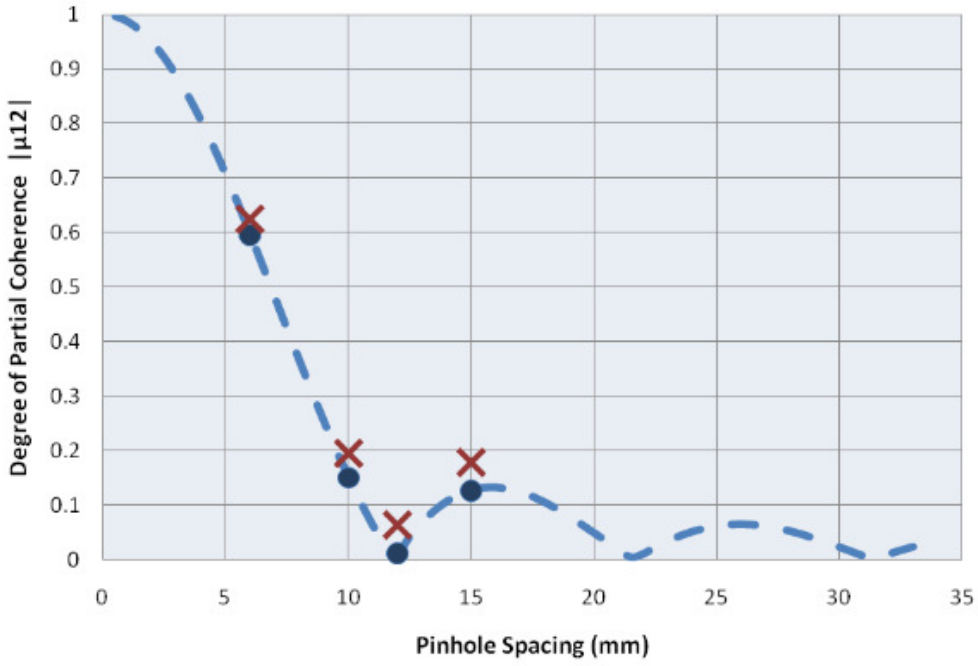
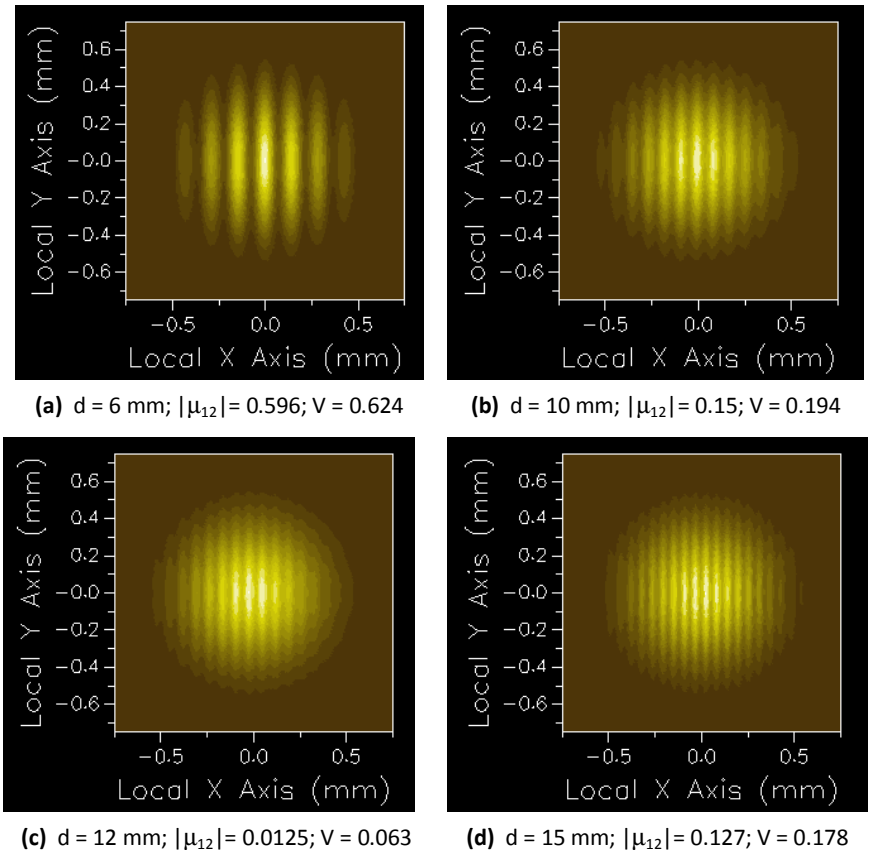


Figure 14. Diffractometer degree of partial coherence vs. pinhole separation. Dots represent theoretical |μ12| at the settings used in the FRED model. Red X's correspond to fringe visibility (which is equivalent to |μ12|) calculated based on FRED plots (shown in Figure 15).



(a) d = 6 mm; $|\mu_{12}|$= 0.596; V = 0.624

(b) d = 10 mm; $|\mu_{12}|$= 0.15; V = 0.194

(c) d = 12 mm; $|\mu_{12}|$= 0.0125; V = 0.063

(d) d = 15 mm; $|\mu_{12}|$= 0.127; V = 0.178

Figure 15. Fringe patterns at the plane $\mathscr{F}$ for four different pinhole spacings, d, with corresponding degree of partial coherence, $|\mu_{12}|$, and estimated fringe visibility, V.

References

1. B.J. Thompson & E. Wolf, *J. Opt. Soc. Amer.*, **47** (1957), p.895.

2. Born & Wolfe. Principle of Optics (6th Ed), Pergamon Press, Ch. 10, Sec. 4.3, p. 513

3. Hecht, Eugene. Optics (4th Ed), Addison Wesley, Ch 12.3, p. 571

For More Information…

On this Application Note and others

Visit our website at http://www.photonengr.com

Email us at support@photonengr.com

On purchasing a copy of FRED or getting a free Demo version

Email us at sales@photonengr.com

On our many tutorials and short courses

Visit our website at http://www.photonengr.com

Email us at tutorials@photonengr.com

Photon Engineering, LLC
440 South Williams Blvd., Suite # 106
Tucson, Arizona 85711
Tel: +1-520-733-9557
Fax +1-520-733-9609
http://www.photonengr.com