Chapter 2

Literature Review

2.1: Geometrical Optics

In geometrical optics, when the short wavelength approximation $\lambda_0 \rightarrow 0$ is fulfilled, the propagation of light is governed by three empirical laws. The law of rectilinear propagation, the law of reflection and the law of refraction enable the propagation path to be determined geometrically. Even though geometrical optics only refers to the propagation properties of light and not the physical nature of light, it is accurate in the classical regime.

According to geometrical optics, the propagation paths are completely determined by the initial values i.e. position and angle of incidence from a reference plane. Using the law of rectilinear propagation, the law of refraction and the law of reflection, the position and the angle of incidence at the final plane can be obtained. By repeated applications of these laws, the path of light rays through any surface and medium can be computed.

This ray tracing method is carried out not only for a single ray but also for a number of suitably selected rays. For an optical system, the image can easily be constructed at the image plane with the aids of these selected rays. Therefore, the evaluation of these rays will give a good estimate of the optical performance of the system.

2.2: The Generalized Eikonal Equation

In a few recent studies, the classical eikonal equation has been generalized to include higher order terms[1, 2, 3]. The higher order terms, which consist of the second derivatives of spatial and temporal profiles of the wave leads to the time dependency and local inhomogeneity of the effective refractive index i.e. a homogeneous material is induced to become inhomogeneous. The generalized eikonal equation is able to show wave phenomena through these terms. The time dependent generalized eikonal equation is able to show the dispersion of a pulse in a linear and non-linear media. It is also able to give the same results of soliton propagation obtained by the inverse scattering method. At the same time, in the stationary case, this formalism can generate the phenomena of diffraction, self-trapping and the self-focusing effects.

The case of time dependent generalized eikonal equation has been derived in Quek et al[1] and the case of generalized eikonal equation has been further developed by Yap et al[2]. The following derivation is based on the textbook of Born and Wolf and further developed by Quek et al[1] and Yap et al[2].

The stationary-wave equation in a homogeneous medium is given by

$$\nabla^2\vec{E} + k_0^2 n^2\vec{E} = 0 \tag{2.1}$$

where n is the refractive index, and $k_0 = \frac{\omega}{c} = \frac{2\pi}{\lambda_0}$ represents the wave number in free space, ω being the angular frequency of the wave and λ_0 the free space wavelength.

When the field is polarized in the $\hat{e}$ direction transverse to the beam-propagation axis, the solution for the wave equation can be generally expressed in the following

$$\vec{E} = \phi(r)\exp[ik_0 L(r)]\hat{e} \tag{2.2}$$

where $\phi(r)$ is a slowly varying envelope function and $L(r)$ is a real scalar function of position. Both are assumed to be independent of k_0. The rapid variation of the optical field is represented by the exponential term.

By substituting equation (2.2) into equation (2.1), two equations from the real part and the imaginary part can be obtained. Both parts are separately equated to zero. The imaginary part gives the following equation

$$\nabla \cdot (\phi^2 \nabla L) = 0 \tag{2.3}$$

which is in the form of a continuity equation while the real part gives

$$(\nabla L)^2 = n^2 + \frac{1}{k_0^2 \phi} \nabla^2 \phi \tag{2.4}$$

In the limit of $\lambda_0 \rightarrow 0$ i.e. $k_0 \rightarrow \infty$, equation (2.4) is reduced to

$$(\nabla L)^2 = n^2 \tag{2.5}$$

This equation is called the classical eikonal equation. The quantity L is called the eikonal. However, retaining the second term in equation (2.4)

$$(\nabla L)^2 = n_G^2 \tag{2.6}$$

allows one to obtain wave phenomena. Equation (2.6) is termed as the generalized eikonal equation where n_G is the generalized index of refraction

$$n_G^2 = n^2 + \frac{1}{k_0^2 \phi} \nabla^2 \phi \tag{2.7}$$

2.3: Ray Tracing and Its Geometrical Interpretation

On account of the generalized eikonal equation, $\frac{\nabla L}{n_G}$ is a unit vector $\hat{s}$ given by

$$\hat{s} = \frac{\nabla L}{n_G} \tag{2.8}$$

$$= \frac{\nabla L}{|\nabla L|}$$

where $\hat{s}$ is in the direction of the time averaged Poynting vector. The geometrical light rays are then defined as the orthogonal trajectories of the geometrical wave fronts where L is constant.

The unit vector is represented by

$$\hat{s} = \frac{d\vec{r}}{ds} \tag{2.9}$$

where $\vec{r}(s)$ denotes the position vector of a point P on a ray and s is the arc length of the ray. From here, it was shown in Born and Wolf[7] that by manipulation, it then becomes

$$\frac{d}{ds}(n_G \frac{dr}{ds}) = \nabla n_G \tag{2.10a}$$

or

$$\frac{d}{ds}(\nabla L) = \nabla n_G \tag{2.10b}$$

This shows that the ray vector is tangent to the ray.

Using these properties, the ray can be constructed from the generalized eikonal equation

$$(\nabla L)^2 = n_G^2 .$$

By casting it into Cartesian coordinates[5, 23], the above equation becomes

$$\left(\frac{\partial L}{\partial x}\hat{x}+\frac{\partial L}{\partial y}\hat{y}+\frac{\partial L}{\partial z}\hat{z}\right)^2=n_G^2 \qquad (2.11).$$

These partial terms in x, y and z direction can be represented by

$$L_x=\frac{\partial L}{\partial x} \qquad (2.12a)$$

$$L_y=\frac{\partial L}{\partial y} \qquad (2.12b)$$

and

$$L_z=\frac{\partial L}{\partial z} \qquad (2.12c).$$

Since we only take into consideration two dimensions, the y term is then dropped from equation (2.11),

$$L_z^2=n_G^2-L_x^2 \qquad (2.13)$$

with z as the direction of propagation. Differentiation of equation (2.13) as a well-behaved, continuous function, in the x and z direction yields two separate equations: -

$$\frac{\partial L_x}{\partial z}=\frac{1}{2L_z}\frac{\partial n_G^2}{\partial x}-\frac{L_x}{L_z}\frac{\partial L_x}{\partial x} \qquad (2.14)$$

$$\frac{\partial L_z}{\partial z}=\frac{1}{2L_z}\frac{\partial n_G^2}{\partial z}-\frac{L_x}{L_z}\frac{\partial L_z}{\partial x} \qquad (2.15).$$

Casting the continuity equation, again in two dimensions in Cartesian coordinates, gives

$$\frac{\partial \phi}{\partial z}=-\frac{1}{2}\frac{\phi}{L_z}\left(\frac{\partial L_x}{\partial x}+\frac{\partial L_z}{\partial z}\right)-\frac{L_x}{L_z}\frac{\partial \phi}{\partial x} \qquad (2.16)$$

while equation (2.7) can be rewritten as

$$n_G^2 = n_0^2 + \frac{1}{k_0^2\phi}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\phi \tag{2.17}$$

Calculation of ϕ, L_x, L_z and n_G^2 can be carried out from these four differential equations enabling the construction of rays.

2.4: Wave Phenomena

In optics, the geometrical model only gives information on the physical properties of light propagation. There are certain regimes in which a simple geometrical model of energy propagation is inadequate. These regimes involve the description of wave effects on phenomena such as diffraction and interference. All of these cases involve beam size of the same order of magnitude with the wavelength.

2.4.1: Diffraction

Diffraction phenomena[8, 11, 14] refer to the spreading of waves into regions that are blocked by an opaque obstacle, out of the direct line of sight of all oncoming waves. Diffraction theory deals with the deviation from geometrical model in the immediate neighbourhood of the boundaries of shadows and the region where a large number of rays meet. These regions give an appearance of dark and bright fringes.

Wave theory, first proposed by Huygens, is sufficient to explain not only the rectilinear propagation of light but also the minute deviations from it – diffraction phenomenon. According to Huygens, every point of a wave front may be considered as a center of a secondary disturbance, which may give rise to spherical wavelets. From the above construction, Fresnel then accounted for diffraction with the postulate that the secondary wavelets mutually interfere.

The Huygens-Fresnel principle has been given a mathematical basis developed by Kirchhoff. He has expressed the solution of a homogeneous wave equation, at any arbitrary point in the field, in terms of the values of the solution and its first derivatives at all points on an arbitrary closed surface surrounding the point. However, the expression governing the contributions from different elements of the surface is more complicated than that assumed by Fresnel. In many cases, it can be approximated to the formulation of Fresnel. The following approximations, referred to by Born and Wolf[7], are the Kirchhoff's boundary conditions, making the following assumptions: -

(1) Over a surface B, which is far from the aperture (which lies just behind the opaque portion of the screen) the field U and its derivative are zero.

$$U = 0, \qquad \frac{\partial U}{\partial x} = 0$$

(2) On the surface A (surface exposed to oncoming waves), the field along its derivative has exactly the same value as it would have had if the screen has been absent. The field on aperture A is assumed to be $U = \left(\frac{A}{s}\right)\exp(-iks)$ for a point source illumination where s represents the distance of an arbitrary point on the aperture from the point source and A as the amplitude of the wave.

It is further assumed that the distance between the source and the aperture is large compared to that of wavelength. Under these approximations[7, 14], the integral is then reduced to

$$U(P) = \frac{1}{4\pi}\iint_{A}\left\{U\frac{\partial}{\partial n}\left(\frac{\exp(-iks)}{s}\right) - \frac{\exp(-iks)}{s}\frac{\partial U}{\partial n}\right\}dS \qquad (2.18),$$

which is known as the Kirchhoff integral, where s represents the distance of surface element dS to a chosen origin. The unit vector $\hat{n}$ is a unit normal to the surface of the aperture.

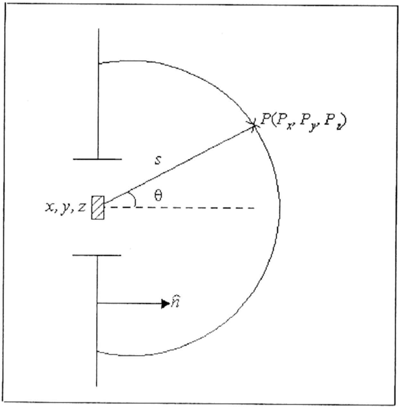

Figure 2.1

The Kirchhoff integral above, equation (2.18), can be reduced to the Fresnel integral by making a few assumptions. The source is assumed to be placed at the aperture instead of located at a distance from the aperture. From this point, the value $\frac{\partial U}{\partial n}$ is then equal to zero, leaving the integral theorem of Kirchoff with only one term

$$U(P) = \frac{1}{4\pi}\iint_A U \frac{\partial}{\partial n}\left(\frac{e^{-iks}}{s}\right) dS. \quad (2.19)$$

The differentiation of $\frac{e^{-iks}}{s}$ gives two terms and by considering k is large or λ is small. Therefore, only the leading term with constant k is maintained.

$$U(P) = \frac{1}{4\pi}\iint_A -ikU \frac{e^{-iks}}{s} \hat{n}.\hat{s}dS \quad (2.20)$$

The dot product of the unit vector of the distance between the aperture to the point P and that of $\hat{n}$ gives $\cos\theta$. The distance s is reduced by binomial expansion, and by taking in the first term only, to

$$s \approx P_z + \Delta x + \Delta y \quad (2.21)$$

with P_x, P_y and P_z as the location of point P in Cartesian coordinates while x, y and z refers to the location of the source at the aperture. Substitution of equation (2.21) into equation (2.20) reduces the equation to

$$U(P) = \frac{-ik}{4\pi}\iint_A U \frac{\exp(-ik(P_z + \Delta x + \Delta y))}{P_z} dS. \quad (2.22)$$

where only the first term of $\cos\theta$ expansion is taken into account. Then, it will reduce to the one dimensional Fresnel integral form

$$U(P) = \frac{-ik\exp(-ikP_z)}{4\pi P_z}\int U \exp(-ik\Delta x)dx. \quad (2.23)$$

It is then used to compute results for comparison purposes of different cases of initial conditions of U.

2.4.2: Interference

Another interesting and important wave phenomenon is the occurrence of interference. It happens when two or more light beams are superposed, the resultant disturbance can no longer be described in terms of rays. The intensity in the region of superposition is found to vary from point to point between a maximum and a minimum which may be zero. The two beams are said to have interfered. The maximum point is formed from constructive interference with the resultant intensity greater than that of the two separate intensities. The minimum point, formed from destructive interference, has resultant intensity less than the sum of the separate intensities. Overall, it does not violate the principle of conservation of energy.

For two monochromatic waves, $\mathbf{E_1}$ and $\mathbf{E_2}$, superposed at some point P, the total electric field at that point is

$$\mathbf{E} = \mathbf{E_1} + \mathbf{E_2}$$

and the total intensity is given by

$$\mathbf{E^2} = \mathbf{E_1}^2 + \mathbf{E_2}^2 + 2\mathbf{E_1}.\mathbf{E_2}.$$

giving an interference term of

$$\mathbf{J_{12}} = 2\langle \mathbf{E_1}.\mathbf{E_2} \rangle.$$

Taking into account the complex amplitudes of the two waves and the fact that the phases will produce a phase difference of δ, caused by the difference of the optical paths,

$$\mathbf{J_{12}} = (\mathbf{a_1 b_1} + \mathbf{a_2 b_2} + \mathbf{a_3 b_3}) \cos \delta$$

The interference term depends on the amplitude of the components (a_i, b_i; $i = 1, 2, 3$) and also the phase difference.

2.4.3: Diffraction Grating

Diffraction grating refers to any periodic arrangement of diffracting elements, which causes the incident wave to be periodic variation in amplitude or phase, or both. Elements of a grating, taken as narrow slits of the same width, act as sources of disturbances that radiate uniformly. The properties of diffraction grating are associated with the interference effects between disturbances from corresponding parts of the separate elements i.e. the periodicity of diffracting elements. It is regarded as an extension to that of a double slit for interference.

For a one-dimensional grating of N-parallel grooves of arbitrary profiles, with d as the separation of the elements in the grating, the light distribution is given by

$$
\begin{aligned}
U(P) &= U^{(0)}(p)\sum_{n=0}^{N-1} e^{-ikndp} \\
&= U^{(0)}(p)\frac{1-e^{-iNkdp}}{1-e^{-ikdp}}.
\end{aligned}
$$

It is the same as that of a set of coherent secondary sources, each characterized by the same amplitude function $|U^{(0)}(p)|$ with the same phase that differs from each other by an integral multiple of kdp.

The most striking modification[11] in pattern as the number of slit increases consists of the narrowing of the interference maxima. The sharpness of these principal maxima increases rapidly until they become narrow lines. There also exist secondary maxima between principal maxima but they are of less importance, as they are quite weak in intensity.

The three phenomena that were described above were only explainable by wave optics and not by geometrical optics. However, we will demonstrate that these effects can be reproduced using ray-tracing concepts by utilization of the generalized eikonal formalism.

2.5: Previous Work and Problems

Using the above method, the propagation of a Gaussian beam has been investigated by Quek et al[1]. By using some approximations, the propagation of a Gaussian beam has been analytically solved. It agrees with the previous work obtained by using the Kirchhoff integral using Fresnel approximation.

However, the propagation of other finite beams has no simple analytic solution. Therefore, a general numerical scheme is needed. The numerical scheme proposed by Quek and Yap et al uses the following algorithm.

1. Calculate n_G using initial values.
2. Calculate $\frac{\partial L_x}{\partial z}$, $\frac{\partial L_z}{\partial z}$.
3. Calculate $\frac{\partial \phi}{\partial z}$.
4. Obtain L_x, L_z and ϕ at $z = z_i + \delta z$.
5. Repeat 1 to 4 until $z = z_{final}$.

In order to test this algorithm, two programs have been developed. The first program uses fixed coordinates while the second is based on moving coördinates. Gaussian beam has been used to test the stability and accuracy of these two programs. It is found that both programs agree with the result obtained by Kirchoff integral to the accuracy of 10^{-4} at $z_{final} = 10$m.

The first program based on fixed grid coordinate by Quek et al[4] focussed on the formalism for the stationary beam propagation, time-dependent pulse propagation and wave propagation in non-linear medium. The above algorithm was first written in the canonical Hamiltonian form, which allows the characteristic equations to maintain its form after transformation to other curvilinear coordinates. It was used to investigate different types of envelope functions of finite beams in a homogeneous media.

The envelope function of an infinite plane wave was applied to the above algorithm. It reduces the generalized eikonal equation to the classical eikonal equation, giving the propagation behaviour exactly as predicted by geometrical optics. When the envelope function was changed to that of a Gaussian beam profile, it shows a spreading of the beam indicating the diffraction effect due to higher order terms. It is similar to the conventional diffraction results. This fact was verified by using an envelope function of a non-diffractive beam. This was obtained by using cylindrical coordinates where the lowest mode obtained is that of wave-guide solution in the core of a step function cylindrical optical fibre. From the above, it was concluded that the higher order term gives rise to the diffraction effect.

Further study was carried out on the time-dependent generalized eikonal equation written in the Hamiltonian-Jacobi form. It was found that the existence of the local acceleration causes a spreading of the local velocity, distorting the pulse thus leading to the dispersion effect. However, the space-time rays remain straight and parallel to each other if the group velocity equals the phase velocity, giving propagation in non-dispersive medium.

The generalized eikonal equation allows for the undertaking of the dispersive effect. The storing of energy in the dispersive media causes the group velocity to be different from the phase velocity. It then led to the reduction in the local velocity at the center, dispersing the pulse.

Thus far, the generalized eikonal equation had been shown to be able to handle diffraction and dispersion in a unified manner. It was shown that, while both effects have similarity in showing greater effect when a narrower beam was used, they are different in the aspect that the diffraction effect is not material dependent. Dispersion happens when the material used has a group velocity that differs from the phase velocity. Diffraction happens in all materials but can be modulated to achieve non-diffractive propagation as in the transverse modes in a cylindrical fibre.

The generalized eikonal equation was then extended[1, 2, 3] to include the propagation in a non-linear medium allowing the study of various effects such as self-trapping, self-focussing and solitonic propagation. Self-trapped effect was obtained by balancing the non-linear effect with the diffraction of a stationary beam causing the local curvatures to vanish. The 1-D soliton effect was, however, obtained by balancing the non-linear effect with the dispersion effect. This caused the local acceleration to vanish.

The numerical work used for the first program was based on the Lax-Wendroff scheme. It is only able to give an accuracy of 10^{-2} for a fourth order super Gaussian beam at z_{final} = 1m and 10^{-5} at z_{final} = 1m for a Gaussian beam.

The second program developed by Yap[5] was based on moving grid. It shows a stability problem when used for other type of profiles. This program was based on the algorithm that is extended from the one discussed in the previous section. Yap utilized the classical ray equation in the vector form and transforms it to other more suitable forms to enable implementation of numerical method. The method by Puchalski was adapted by replacing all refractive index with local refractive index. This method traces the trajectories of each individual ray. The value of the variables are calculated corresponding to the new position of the ray, creating a non-uniform computing grid that is self-generated by the propagation properties.

The cases studied by Yap[5] using this program include the propagation of a Gaussian beam in graded index medium and lens design. In tracing the Gaussian beam in a graded index medium, the higher order term in the generalized eikonal equation was first dropped from the local refractive index. This condition shows a ray path that is sinusoidal from the ray trace method. This result is in good agreement with the analytic solution. The higher order term was recovered, changing the ray trajectories. Energy that flows along these trajectories is focused to a narrow width, then the diffraction term takes over, a stage where focusing effect is compensated by the diffraction effect. The beam is then diverged out. The trajectories never cross each other for a finite beam. Again, the results from the calculation agree well with the analytical solution.

As for the numerical ray tracing in the lens design, the generalized eikonal equation represents a way of determining the initial condition and the surface properties that are required to produce certain final image. The diffraction effect prevents the beams from being focused to an infinitesimal spot. The numerical results show that beams are focused to a minimum spot size before starting to diverge again. The generalized eikonal equation has a great advantage in its ability to carry out ray tracing in lens design when the medium is non-linear. The inclusion of the non-linear effect in lenses had been shown to be able to focus a minimum spot size smaller than that set by diffraction effect.