The physical concept of diffraction limit

1 Diffraction of water waves at seashore

It is a common sight of the sea waves surging back and forth against the seashore. When these waves happen to collide with a break water, only those waves which can take a round-about way are possible to propagate within the breakwater. This phenomenon is generally attributed as the diffraction of water waves. As the oscillation of the waves do not stop abruptly at the geometrical boundaries, the propagation of the water waves is practically continuous and hence the waves inside the breakwater do not have a sharp shadow.

2 Diffraction of light

Diffraction of light is generally formed as the encouragement of light on any obstacle and is attributed to the wavelength of light. Huygens' theory of light states that as a point source is responsible for a spherical wavefront, any point propagated by a light wave is solely responsible for the origin of secondary spherical waves that spread out in all directions. By adding the concept of interference to Huygens' theory, Fresnel stated that the complex amplitude of a light wave beyond the wavefront can be superimposed to that of all elementary waves which propagate from each point of the wavefront to the observed point. The complex amplitude at an arbitrary point on the wavefront is given by the equation,

$$\Phi = \frac{A}{r_0} e^{ikr_0}, \tag{2.1}$$

where A is the amplitude, r_0 is the radius and k is the propagation constant of the plane wave. Thus, if the distance from a point on the wavefront to an observed point is r, the infinitesimal area $d\phi$ between the observed point and the elementary waves is given by,

$$d\phi = AK(\theta)\frac{e^{ik(r+r_0)}}{rr_0}, \tag{2.2}$$

where $K(\theta)$ is termed as the obliquity factor such that $0 \leq K(\theta) \leq M$, where $K(\theta) \to 0$ as $\theta \to \frac{\pi}{2}$ and tends to a maximum value M, when $\theta \to 0$. θ is the angle subtended by a line between a point of the wavefront to the observed point to a line normal to the wavefront in the amplitude of the elementary waves in the plane wave. When there is no obstacle, the process of light propagation is given by the equation

$$\phi = \frac{Ae^{ikr_0}}{r_0} \int\int \frac{e^{ikr}}{r} K(\theta) d\sigma, \tag{2.3}$$

where $d\sigma$ is the infinitesimal area. The propagation behavior of spherical waves can be explained by assuming $K(\theta) = \frac{i}{\lambda}$ and hence can be shown that the amplitude of the secondary elementary waves decays by a factor equal to the increase of the light wavelength with a quarter phase delay period with respect to the primary incident wave. In the presence of an obstacle, the phenomenon of diffraction can be explained by assuming concentric circles based on the function $f_n = r + \frac{n}{2}\lambda$ and drawn around a point on the wavefront. Zones of concentric circles on a wave plane arc termed as Fresnel zones. When the distance between the obstacles is smaller than the wavelength of light, bending of light or encroachment of light takes place, leading to diffraction. The Fresnel- Kirchhoff equation is given as,

$$\phi = \frac{iA\cos\delta}{\lambda} \int\int_S \frac{e^{ik(r+r_0)}}{r_0 r} dS, \tag{2.4}$$

where $\cos\delta = \frac{\cos\theta_0 - \cos\theta}{2}$. For $\delta \ll 1$,

$$\phi = \frac{iA}{\lambda} \int\int_S \frac{e^{ik(r+r_0)}}{rr_0} dS. \tag{2.5}$$

The obliquity factor in the Fresnel- Kirchhoff equation is written as

$$K(\theta) = \frac{(\cos\theta_0 - \cos\theta)}{2\lambda} \exp(ik\pi/2). \tag{2.6}$$

Equation (2.6) implies that the amplitude of the secondary waves is inversely proportional to the wavelength of light and has phase delays of $\frac{\pi}{2}$. In diffraction grating experiment, the equation pertaining to the intensity of light can be derived from Eq. (2.4) and is given as

$$I = \frac{I_0}{N^2} \frac{\sin^2(l_s z)\sin^2(N l_{slit} z)}{(l_s z)^2 \sin^2(N l_{slit})}, \tag{2.7}$$

where I_0 is the intensity of incident light, l_s is the slit width, l_{slit} is the space of slits and N is the number of slits. Also $z = \frac{\pi\sin\phi}{\lambda}$, where ϕ is the diffraction angle. The maximum intensity of light is given by, $I_{max} = \frac{I_o \sin^2(l_s z)}{(l_s z)^2}$, where $\sin^2(l_{slit} z) = 0$. The diffraction grating formula is then given as $l_{slit}\sin\phi = n\lambda$, $n = 0, \pm 1, \pm 2.....$ I_{max} does not depend on N. I_{max} increases as l_s decreases. Hence one can obtain light of narrow wavelength distribution on employing diffraction grating.

3 The concept of diffraction limit

Diffraction limit is specialization of Heisenberg's uncertainty principle with respect to position $\triangle x$ and canonical momentum $p_x = \hbar k_x$ of a photon which is given as

$$\triangle x |p_x| \geq \frac{\hbar}{2}. \tag{3.1}$$

When λ_j is the wavelength of light in medium j, n_i is the refractive index of the medium and k_0 is the propagation constant in vacuum, the propagation constant k_i of the medium 'i' is given as $k_i = \frac{2\pi}{\lambda_i} = n_i|k_0|$. Also, $k_i^2 = k_x^2 + k_y^2 + k_z^2$. From the Heisenberg's uncertainty relation, one can infer that conventional propagation light cannot be confined to linear dimensions much smaller than the wavelength of light. Diffraction limit is very secure in the far-field optics or classical optics, where freely propagating photons are the important carriers. Free propagation of the photon is always characterized by a propagation vector **k**, whose three components $k_x, k_y.k_z$ are all real. As $|\mathbf{k}|^2 = k_i^2 = k_x^2 + k_y^2 + k_z^2$, none of k_x, k_y and k_z can happen to be greater than k. Hence in a freely propagating beam of light, the Heisenberg's uncertainty relation can be written as

$$\triangle x_{classical}|p_x| \geq \frac{\hbar}{2}. \tag{3.2}$$

But, $|p_x| = \hbar k_x = \dfrac{2\pi\hbar}{\lambda}$, where k_x and λ_x are the propagation constant and the wavelength of light in the given medium, respectively. On substituting these in Eq. (3.2), one gets

$$\begin{aligned} \triangle x_{classical}\frac{2\pi\hbar}{\lambda_x} &\geq \frac{\hbar}{2}, \qquad \text{or} \\ \triangle x_{classical} &\geq \frac{\lambda_x}{2\pi}. \end{aligned} \tag{3.3}$$

Thus, as $\dfrac{\lambda_x}{2\pi}$ is the minimum cross-section, the physical dimensions of the photonic device cannot be reduced beyond the wavelength of the light in the given medium when one considers conventional propagation light. This limiting case becomes detrimental when on wants to focus the spot size of light beyond a certain point and when one employs photolithography. Let l denote a factor that depends on the intensity distribution of light beam falling on the objective and let NA denote the numerical aperture of the focusing lens. Then the critical dimension can be roughly given as $CD = \dfrac{l\lambda}{NA}$. In modern photolithography, for optimized ring shaped illumination, the value of l can be as small as 0.36. Thus with $NA = 0.9$, near-ultraviolet light having the wavelength of $\lambda = 400nm$ can be focused down to $CD \simeq 130nm$. For best commercially available immersion objective, $NA = 1.4$. Hence for $l = 0.36, CD = 100nm$. Thus at the borderline of nanometer size dimension, the resolution right of classical optics fails miserably.

When evanescent waves are used instead of free wave propagation, diffraction limit becomes less of a deterrent and this constitutes the near-field optics. For evanescent waves, the amplitudes decay rapidly at least in one of the spacial dimensions, thereby making the respective propagation constant complex in general. Thus the propagation constants in the transverse directions, become much larger than the $|\mathbf{k}|$ and hence the uncertainty in the position can be made comparably smaller than the diffraction limiting case of $\dfrac{\lambda_i}{2\pi}$. As the evanescent waves are excited at the boundary of two different media, they have dominance only close to the interface. Therefore, evanescent fields are called near-fields. Near field optics finds applications mainly in scanning near-field optical microscopy.

An evanescent wave is generally produced as a result primary excitation of electronic dipoles induced near the materiel surface. These electric dipoles are found to align themselves in a periodic manner depending on the spatial phase of the incident light. On the other hand, the guided wave propagating through a sub-wavelength sized cross-sectional waveguide is generated by electronic dipoles induced along the one-dimensional waveguide axis. The corresponding electronic dipoles also align in a periodic manner depending on the spatial phase of the incident light. Example for the former include silicon and metallic wave guides used for silicon photonics and those for the latter include plasmonics. As the both the two-dimensional evanescent waves and the one-dimensional guided waves are produced solely due to the periodic alignment of electric dipoles that in turn depend on the spatial phase of the incident propagating light, they fall under the category of diffraction limited light waves. Even though an evanescent wave has complex or imaginary wave vector components at least along one of the spatial dimensions, the other two spatial dimensions of the real wave vector components, thereby coming under the category of diffraction limited light waves.

4 Additional reading and references

1. C. T. Cohen, J. R. Dupont and G. Grynberg, Photons and Atoms-Introduction to Quantum Electrodynamics (John Wiley and Sons, New York, 1989).

2. C. Klingshirn, Semiconductor Optics (Springer-Verlag, Germany, 2005).