Chapter 3 Foundations of Scalar Diffraction Theory

Prof. Hsuan-Ting Chang

September 10, 2019

Chapter 3
Foundations of Scalar Diffraction Theory

3.1 Historical Introduction
3.2 From a vector to a scalar theory
3.3 Some mathematical preliminaries
3.4 The Kirchhoff formulation of Diffraction by a planar screen
3.5 The Rayleigh-Sommerfeld formulation of diffraction
3.6 Comparison of the Kirchhoff and Rayleigh-Sommerfeld theories
3.7 Further discussion of the Huygens-Fresnel principle
3.8 Generalization of nonmonochromatic waves
3.9 Diffraction at boundaries
3.10 The angular spectrum of plane waves

1. *Refraction*: the bending of light rays that takes place when they pass through a region in which there is a gradient of the local velocity of propagation of the wave.
2. *Reflection*: light rays are bent at a metallic or dielectric interface. The angle of reflection is always equal to the angle of incidence.
3. **Diffraction**: any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction.
4. Diffraction is caused by the confinement of the lateral extent of a wave, and is most appreciable when that confinement is to sizes comparable with a wavelength of the radiation being used.

Figure 3.3

The transition from light to shadow was gradual rather than abrupt.

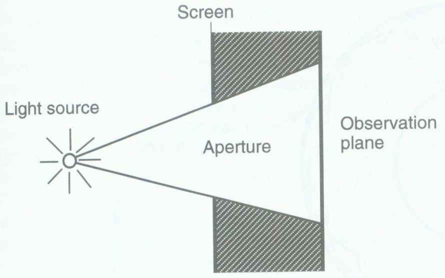

Figure 3.3 Arrangement used for observing diffraction of light.

Figure 3.4

If each point on the wavefront of a disturbance were considered to be a new source of a "secondary" spherical disturbance, then the wavefront at a later instant could be found by constructing the "envelope" of the secondary wavelets.

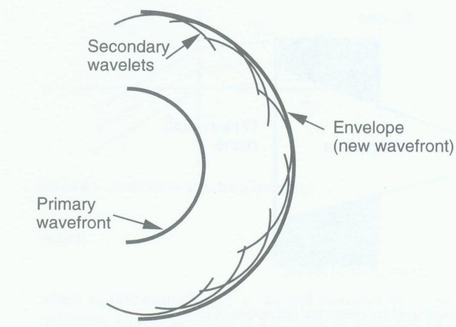

Figure 3.4 Huygens envelope construction.

3.2 From a Vector to a Scalar Theory – 1

1. Maxwell's Equations:

$$
\begin{aligned}
\nabla \times \vec{\mathcal{E}} &= -\mu \frac{\partial \vec{\mathcal{H}}}{\partial t} \\
\nabla \times \vec{\mathcal{H}} &= \epsilon \frac{\partial \vec{\mathcal{E}}}{\partial t} \\
\nabla \cdot \epsilon \vec{\mathcal{E}} &= 0 \\
\nabla \cdot \mu \vec{\mathcal{H}} &= 0
\end{aligned}
\qquad (3.2)
$$

2. $\vec{\mathcal{E}}$: electric field; $\vec{\mathcal{H}}$: magnetic filed; μ and ϵ are the permeability and permittivity, respectively, of the medium in which the wave is propagating.
3. $\vec{\mathcal{E}}$ and $\vec{\mathcal{H}}$ are functions of both the position P and time t.

4. The symbols $\times$ and $\cdot$ represent a vector cross product and a vector dot product, respectively, while

$$\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k},$$

where $\vec{i}, \vec{j}$ and $\vec{k}$ are unit vectors in the x, y and z direction, respectively.

5. Properties of the medium in which the wave is propagating:
isotropic – its properties are independent of the direction of ploarization of the wave,
homogeneous – the permittivity is constant throughout the region of propagation,
nondispersive – the permittivity is independent of wavelength over the wavelength region occupied by the propagation wave,
nonmagnetic – the magnetic permeability is always equal to μ_0, the vacuum permeability.

6. Apply the $\nabla\times$ operation to the left and right sides of the first equation for $\vec{\varepsilon}$, we make use of the vector identity

$$\nabla \times (\nabla \times \vec{\varepsilon}) = \nabla(\nabla \cdot \vec{\varepsilon}) - \nabla^2 \vec{\varepsilon}. \tag{3.3}$$

7. If the propagation medium is linear, isotropic, homogeneous (constant ϵ), and nondispersive, substitution of the two Maxwell's equations for $\vec{\varepsilon}$ in Eq. (3.3) yields

$$\nabla^2 \vec{\varepsilon} - \frac{n^2}{c^2} \frac{\partial^2 \vec{\varepsilon}}{\partial t^2} = 0, \tag{3.4}$$

where n is the refractive index of the medium.

8. The magnetic field satisfies an identical equation,

$$\nabla^2\vec{\mathcal{H}} - \frac{n^2}{c^2}\frac{\partial^2\vec{\mathcal{H}}}{\partial t^2} = 0.$$

9. Since the vector wave equation is obeyed by both $\vec{\varepsilon}$ and $\vec{\mathcal{H}}$, an identical scalar wave equation is obeyed by all components of those vectors.
10. Thus, for exampel, ε_X obeys the equation

$$\nabla^2\varepsilon_X - \frac{n^2}{c^2}\frac{\partial^2\varepsilon_X}{\partial t^2} = 0,$$

and similarly for $\varepsilon_Y, \varepsilon_Z, \mathcal{H}_\mathcal{X}, \mathcal{H}_\mathcal{Y}$, and $\mathcal{H}_\mathcal{Z}$.

11. Therefore, it is possible to summarize the behavior of all components of $\vec{\mathcal{E}}$ and $\vec{\mathcal{H}}$ through a single scalar wave equation,

$$\nabla^2 u(P,t) - \frac{n^2}{c^2}\frac{\partial^2 u(P,t)}{\partial t^2} = 0, \tag{3.7}$$

where $u(P,t)$ represents any of the scalar field components.
12. If the aperture has an area that is large compared with a wavelength, the *coupling effects of the boundary conditions* on the $\vec{\mathcal{E}}$ and $\vec{\mathcal{H}}$ fields will be small.
13. The scalar theory is accurate provided that the diffracting structures are *large* compared with the wavelength of light.

1. **3.3.1 Helmholtz Equation**
2. For a monochromatic wave, the scalar field is

$$u(P,t) = A(P)\cos[2\pi\nu t + \phi(P)], \qquad (3.9)$$

   where $A(P)$ and $\phi(P)$ are the amplitude and phase, respectively, of the wave at position P, while ν is the optical frequency.
3. A more compact form

$$u(P,t) = Re\{U(P)\exp(-j2\pi\nu t)\}, \qquad (3.10)$$

   where $U(P)$ is a complex function of position (*phasor*),

$$U(P) = A(P)\exp[-j\phi(P)].$$

4. If Eq. (3.10) is substituted in Eq. (3.12)=Eq. (3.7), it follows that U must obey the time-independent equation (*Helmholtz equation*):

$$(\bigtriangledown^2 + k^2)U = 0. \tag{3.13}$$

Here k is termed the *wave number* and is given by

$$k = 2\pi n \frac{\nu}{c} = \frac{2\pi}{\lambda},$$

and λ is the wavelength in the dielectric medium.

3.3.2 Green's Theorem

1. GOAL: Calculation of the complex disturbance U at an observation point in space.
2. **Let $U(P)$ and $G(P)$ be any two complex-valued functions of position, and let S be a closed surface surrounding a volume V. If U, G, and their first and second partial derivatives are single-valued and continuous within and on S, then we have**

$$\int\int\int_V (U \nabla^2 G - G \nabla^2 U) dv = \int\int_S (U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n}) ds$$

where $\partial/\partial n$ signifies a partial derivative in the *outward* **normal direction at each point on S.**

3. This theorem is the prime foundation of scalar diffraction theory.

3.3.3 The Integral Theorem of Helmholtz and Kirchhoff - 1

1. Let P_0 be the point of observation, S be an arbitrary closed surface surrounding P_0.
2. To express the optical disturbance at P_0 in terms of its values on the surface S, we follow Kirchhoff in applying Green's theorem and in choosing as an auxiliary function G, a unit-amplitude spherical wave expanding about P_0 (*free space* Green's function).
3. The value of G at an arbitrary point P_1 is

$$G(P_1) = \frac{\exp(jkr_{01})}{r_{01}},$$

where r_{01} is the length of the vector $\vec{r}_{01}$ pointing from P_0 to P_1.

Fig 3.5

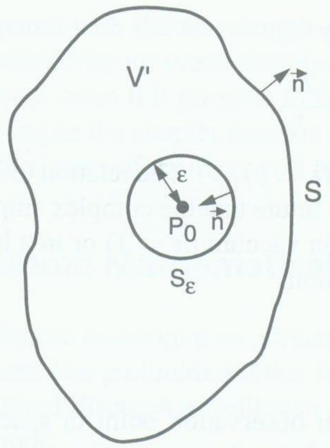

Figure 3.5 Surface of integration.

3.3.3 The Integral Theorem of Helmholtz and Kirchhoff - 2

4. To be legitimately used in Green's theorem, the function G must be continuous within the enclosed volume V. Therefore to exclude the discontinuity at P_0, a small spherical surface S_ϵ, of radius ϵ, is inserted about the point P_0.
5. The volume of integration V' being that volume lying between S and S_ϵ, and the surface of integration being the composite surface $S' = S + S_\epsilon$.

6. The disturbance G satisfies the Helmholtz equation

$$(\bigtriangledown^2 + k^2)G = 0 \qquad (3.18)$$

Then substituting Eqns. (3.13) and (3.18) in the left-hand side of Green's theorem,

$$\int\int\int_{V'}(U\bigtriangledown^2 G - G\bigtriangledown^2 U)dv = -\int\int\int_{V'}(UGk^2 - GUk^2)dv \equiv 0$$

The theorem reduces to

$$\int\int_{S'}(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n})ds = 0$$

or

$$-\int\int_{S_\epsilon}(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n})ds = \int\int_{S}(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n})ds.$$

7. For general point P_1 on S', ($\vec{n}$: outward normal)

$$G(P_1) = \frac{\exp(jkr_{01})}{r_{01}}$$

and

$$\frac{\partial G(P_1)}{\partial n} = \frac{\partial \vec{r}_{01}}{\partial \vec{n}} \frac{\partial G(P_1)}{\partial r_{01}} = \cos(\vec{n}, \vec{r}_{01})(jk - \frac{1}{\vec{r}_{01}}) \frac{\exp(jkr_{01})}{r_{01}}. \tag{3.20}$$

where $\frac{\partial \vec{r}_{01}}{\partial \vec{n}} = \cos(\vec{n}, \vec{r}_{01})$.

8. For the particular case of P_1 on S_ϵ, $\cos(\vec{n}, \vec{r}_{01}) = -1$,

$$G(P_1) = \frac{e^{jk\epsilon}}{\epsilon} \quad \text{and} \quad \frac{\partial G(P_1)}{\partial n} = \frac{e^{jk\epsilon}}{\epsilon}(\frac{1}{\epsilon} - jk).$$

9. Letting $\epsilon \to 0$, the continuity of U at P_0 allow us to write

$$\lim_{\epsilon \to 0} \int\int_{S_\epsilon} (U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n}) ds$$

$$= \lim_{\epsilon \to 0} 4\pi\epsilon^2 \left[U(P_0)\frac{\exp(jk\epsilon)}{\epsilon}(\frac{1}{\epsilon} - jk) - \frac{\partial U(P_0)}{\partial n}\frac{\exp(jk\epsilon)}{\epsilon} \right]$$

$$= 4\pi U(P_0).$$

10. Finally,

$$U(P_0) = \frac{1}{4\pi} \int\int_S \left\{ \frac{\partial U}{\partial n} \left[\frac{\exp(jkr_{01})}{r_{01}} \right] - U\frac{\partial}{\partial n} \left[\frac{\exp(jkr_{01})}{r_{01}} \right] \right\} ds. \quad (3.21)$$

The *integral theorem of Helmholtz and Kirchhoff*; it plays an important role in the development of the scalar theory of diffraction.

11. The field at any point P_0 can be expressed in terms of the "boundary values" of the wave on any closed surface surrounding that point.

Fig 3.6

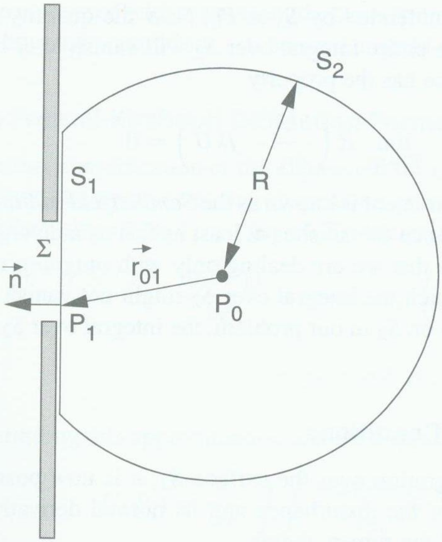

Figure 3.6 Kirchhoff formulation of diffraction by a plane screen.

3.4 The Kirchhoff Formulation of Diffraction by a Planar Screen - 1

1. **3.4.1 Application of the Integral Theorem (for S_2 plane)**
2. As shown in Figure 3.6, following the Kirchhoff, the closed surface S is chosen to consist of two parts. Let a plane surface, S_1, lying directly behind the diffraction screen, be joined and closed by a large spherical cap, S_2, of radius R and centered at the observation point P_0.
3. Since $S = S_1 + S_2$, then

$$U(P_0) = \frac{1}{4\pi} \int \int_{S_1+S_2} (G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n}) ds$$

where

$$G = \frac{\exp(jkr_{01})}{r_{01}}.$$

3.4.1 Application of the Integral Theorem (for S_2 plane) - 2

4. Since both U and G will fall off as $1/R$, the integrand will ultimately vanish, yielding a contribution of zero from the surface integral over S_2.
5. However, the area of integration increases as R^2, so this argument is incomplete. Evidently, a more careful investigation is required before the contribution from S_2 can be disposed of.
6. On S_2,

$$G = \frac{\exp(jkR)}{R}$$

and, from Eq. (3.20) and $\cos(\vec{n}, R) = 1$, we obtain

$$\frac{\partial G}{\partial n} = (jk - \frac{1}{R})\frac{\exp(jkR)}{R} \approx jkG \quad \text{(for large } R\text{)}.$$

3.4.1 Application of the Integral Theorem (for S_2 plane) - 3

7. The integration reduces to

$$\int\int_{S_2}[G\frac{\partial U}{\partial n} - U(jkG)]ds = \int_{\Omega} G(\frac{\partial U}{\partial n} - jkU)R^2 d\omega,$$

   where Ω is the solid angle subtended by S_2 at P_0.

8. Now $|RG| = |\exp(jkR)| = 1$ is uniformly bounded on S_2 and $Gds = RG \times \frac{1}{R}ds$. The entire integral over S_2 will vanish as R becomes arbitrary large, provided that

$$\lim_{R\to\infty}\int\int_{S_2}[G\frac{\partial U}{\partial n} - U(jkG)]ds = \lim_{R\to\infty}\int\int_{S_2}[\frac{\partial U}{\partial n} - U(jk)]Gds = 0,$$

   and this required that (see next page)

3.4.1 Application of the Integral Theorem (for S_2 plane) - 4

$$\lim_{R\to\infty} \frac{1}{R} \int\int_{S_2} (\frac{\partial U}{\partial n} - jkU) ds = 0$$

$$\lim_{R\to\infty} R(\frac{\partial U}{\partial n} - jkU) = 0 \tag{3.22}$$

9. This *requirement* is known as the **Sommerfeld radiation condition** (for S_2 plane) and is satisfied if U vanishes at least as fast as a diverging spherical wave. The integral over S_2 will yield a contribution of *precisely zero*.

3.4.2 The Kirchhoff Boundary Conditions (for S_1 plane) - 1

1. After considering Sommerfeld radiation condition, the disturbance at P_0 over the infinite plane S_1 is expressed by,

$$U(P_0) = \frac{1}{4\pi} \int \int_{S_1} (G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n}) ds. \qquad (3.23)$$

2. The major contribution to the integral (3.23) arises from the point of S_1 located within the aperture Σ, where we expect the integrand to be largest.
3. *Kirchhoff boundary conditions*:
   (1). Across the surface Σ, the field distribution U and its derivative $\partial U / \partial n$ are exactly the same as they would be in the absence of the screen.
   (2). Over the portion of S_1 that lies in the geometric shadow of the screen, the field distribution U and its derivative $\partial U / \partial n$ are identically zero.

4. Thus (3.23) is reduced to

$$U(P_0) = \frac{1}{4\pi} \int\int_{\Sigma} (G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}) ds. \qquad (3.24)$$

5. However, it is important to realize that neither can be exactly true. The screen will inevitably perturb the fields on Σ to some degree. In addition, fields will inevitably extend behind the screen for a distance of several wavelengths.
6. However, if the dimensions of the aperture are large compared with a wavelength, the fringing effects can be safely neglected, and the two boundary conditions can be used.

3.4.3 The Fresnel-Kirchhoff Diffraction Formula (for S_1 plane) - 1

1. A further simplification for $U(P_0)$ is obtained by noting $k \gg 1/r_{01}$, (3.20) becomes

$$\frac{\partial G(P_1)}{\partial n} = \cos(\vec{n}, \vec{r}_{01})(jk - \frac{1}{\vec{r}_{01}})\frac{\exp(jkr_{01})}{r_{01}}$$

$$\approx jk\cos(\vec{n}, \vec{r}_{01})\frac{\exp(jkr_{01})}{r_{01}}. \qquad (3.25)$$

2. Then

$$U(P_0) = \frac{1}{4\pi}\int\int_{\Sigma}\frac{\exp(jkr_{01})}{r_{01}}\left[\frac{\partial U}{\partial n} - jkU\cos(\vec{n}, \vec{r}_{01})\right]ds. \qquad (3.26)$$

Fig 3.7

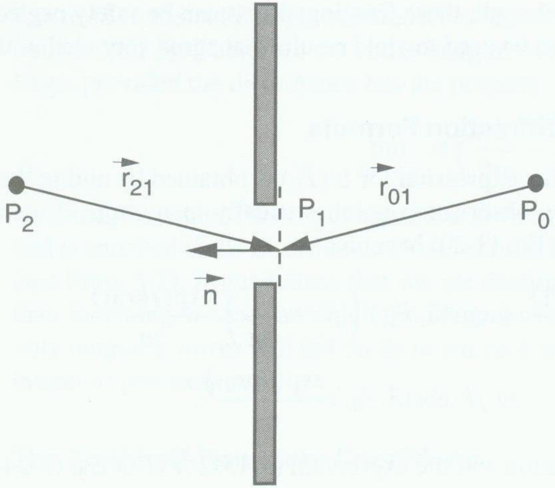

Figure 3.7 Point-source illumination of a plane screen.

3.4.3 The Fresnel-Kirchhoff Diffraction Formula (for S_1 plane) - 2

3. (Figure 3.7) Now suppose that the aperture is illuminated by a single spherical wave,

$$U(P_1) = \frac{A\exp(jkr_{21})}{r_{21}}$$

arising from a point source P_2, a distance r_{21} from P_1.

4. If r_{21} is many optical wavelengths, then

$$U(P_0) = \frac{A}{j\lambda}\int\int_{\Sigma} \frac{\exp[jk(r_{21}+r_{01})]}{r_{21}r_{01}}\left[\frac{\cos(\vec{n},\vec{r}_{01}) - \cos(\vec{n},\vec{r}_{21})}{2}\right]ds. \quad (3.27)$$

This result, which holds *only for an illumination consisting of a single point source*, is known as the *Fresnel-Kirchhoff diffraction formula.*

3.4.3 The Fresnel-Kirchhoff Diffraction Formula (for S_1 plane) - 3

5. The *reciprocity theorem of Helmholtz:* A point source at P_0 will produce at P_2 the same effect that a point source of equal intensity placed at P_2 will produce at P_0.
6. Finally, an interesting interpretation of the diffraction formula (3.27) is

$$U(P_0) = \int\int_{\Sigma} U'(P_1)\frac{\exp(jkr_{01})}{r_{01}}ds, \tag{3.28}$$

where

$$U'(P_1) =$$

$$\frac{1}{j\lambda}[\frac{A\exp(jkr_{21})}{r_{21}}][\frac{\cos(\vec{n},\vec{r}_{01}) - \cos(\vec{n},\vec{r}_{21})}{2}]. \tag{3.29}$$

3.4.3 The Fresnel-Kirchhoff Diffraction Formula (for S_1 plane) - 4

7. The light field at P_0 arises from an infinity of fictitious "secondary" point sources located within the aperture itself.
8. The secondary sources $U'(P_1)$ have amplitudes and phases, that are related to
   (1). The illumination wavefront (scaling factor $1/\lambda$)
   (2). The angles of illumination and observation (obliquity factor $-1 \leq \frac{\cos(\vec{n},\vec{r}_{01})-\cos(\vec{n},\vec{r}_{21})}{2} \leq 1$)
   (3). Phase lead (90^o) of the secondary source.

3.5 The Rayleigh-Sommerfeld Formulation of Diffraction - 1

1. There are certain internal inconsistencies in the theory which motivated a search for a more satisfactory mathematical development.
2. The difficulties of the Kirchhoff theory stem from the fact that boundary conditions must be imposed on *both* the field strength and its normal derivative.
3. *Potential theory:* if a 2-D potential function and its normal derivative vanish *together* along any finite curve segment, then that potential function *must vanish over the entire plane.*
4. Similarily, if a solution of the 3-D wave equation vanishes on any finite surface element, it must vanish in all space.

3.5 The Rayleigh-Sommerfeld Formulation of Diffraction - 2

5. The two Kirchhoff boundary conditions together imply that the field is zero everywhere behind the aperture, which contradicts the known physical situation.
6. The Fresnel-Kirchhoff diffraction formula can be shown to fail to reproduce the assumed boundary conditions as the observation point approaches the screen or aperture.
7. The inconsistencies of the Kirchhoff theory were removed by Sommerfeld, who eliminated the necessity of imposing boundary values on both the disturbance and its normal derivative simultaneously.

1. Consider again Eq. (3.23)

$$U(P_0) = \frac{1}{4\pi} \int \int_{S_1} (\frac{\partial U}{\partial n} G - U \frac{\partial G}{\partial n}) ds.$$

The conditions for validity are:
1. The scalar theory holds.
2. Both U and G satisfy the homogeneous scalar wave equation.
3. The Sommerfeld radiation condition, Eq. (3.22), is satisfied.

2. Suppose that the Green's function G of the Kirchhoff theory were modified either G or $\partial G/\partial n$ vanishes over the entire surface S_1. The inconsistencies of the Kirchhoff theory would be eliminated.
3. Suppose G is generated not only by a point source located at P_0, but also simultaneously by a second point source at a position $\tilde{P}_0$, which is a mirror image of P_0 on the opposite side of the screen. Let the source at $\tilde{P}_0$ be of the same wavelength as the source at P_0, and suppose that the two sources are oscillating with a 180^o phase difference.

Fig 3.8

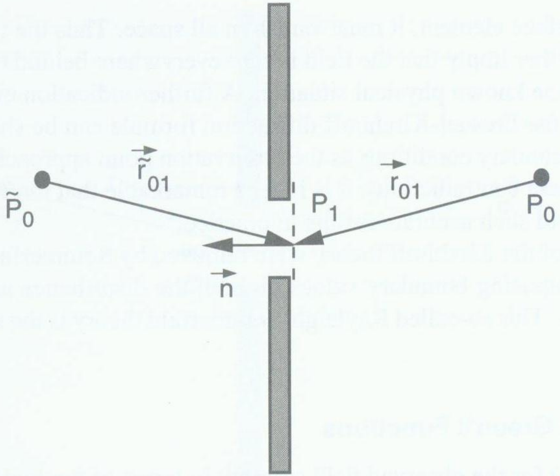

Figure 3.8 Rayleigh-Sommerfeld formulation of diffraction by a plane screen.

4. (Figure 3.8)

$$G_{-}(P_1) = \frac{\exp(jkr_{01})}{r_{01}} - \frac{\exp(jk\tilde{r}_{01})}{\tilde{r}_{01}}.$$

This function vanishes ($G_{-}(P_1) = 0$) on the plane aperture Σ,

$$U_I(P_0) = \frac{-1}{4\pi} \int\int_{\Sigma} U \frac{\partial G_{-}}{\partial n} ds.$$

We refer to this solution as the *first Rayleigh-Sommerfeld solution.*

5. Let $\tilde{r}_{01}$ be the distance from $\tilde{P}_0$ to P_1.

$$\frac{\partial G_{-}(P_1)}{\partial n} = \cos(\vec{n}, \vec{r}_{01})(jk - \frac{1}{r_{01}})\frac{\exp(jkr_{01})}{r_{01}}$$

$$- \cos(\vec{n}, \vec{\tilde{r}}_{01})(jk - \frac{1}{\tilde{r}_{01}})\frac{\exp(jk\tilde{r}_{01})}{\tilde{r}_{01}}.$$

3.5.1 Choice of Alternative Green's Functions - 4

6. For P_1 on S_1,

$$r_{01} = \tilde{r}_{01}, \quad \cos(\vec{n}, \vec{r}_{01}) = -\cos(\vec{n}, \vec{\tilde{r}}_{01}),$$

   and $r_{01} \gg \lambda$,

$$\frac{\partial G_-(P_1)}{\partial n} = 2jk\cos(\vec{n}, \vec{r}_{01})\frac{\exp(jkr_{01})}{r_{01}}, \qquad (3.35)$$

7.

$$\frac{\partial G_-(P_1)}{\partial n} = 2\frac{\partial G(P_1)}{\partial n} \text{ and } \quad U_I(P_0) = \frac{-1}{2\pi}\int\int_\Sigma U\frac{\partial G}{\partial n}ds.$$

8. An alternative and equally valid Green's function is found by allowing the two point sources to oscillate in phase,

$$G_{+}(P_1) = \frac{\exp(jkr_{01})}{r_{01}} + \frac{\exp(jk\tilde{r}_{01})}{\tilde{r}_{01}}. \qquad (3.37)$$

9. The normal derivative of this function vanishes across the screen and aperture, leading to the *second Rayleigh-Sommerfeld solution*,

$$U_{II}(P_0) = \frac{1}{4\pi} \int\int_{\Sigma} \frac{\partial U}{\partial n} G_{+} ds. \qquad (3.38)$$

10. On Σ and under the condition that $r_{01} \gg \lambda$, G_+ is twice the Kirchhoff Green's function G,

$$G_+ = 2G,$$

and

$$U_{II}(P_0) = \frac{1}{2\pi} \int\int_{\Sigma} \frac{\partial U}{\partial n} G ds.$$

3.5.2 The Rayleigh-Sommerfeld Diffraction Formula - 1

1. Substitute G_- for G in Eq. (3.23). Using (3.35),
$$U_I(P_0) = \frac{1}{j\lambda}\int\int_{S_1} U(P_1)\frac{\exp(jkr_{01})}{r_{01}}\cos(\vec{n},\vec{r}_{01})ds, \qquad (3.40)$$
(where $r_{01} \gg \lambda$).
2. The Kirchhoff boundary conditions may now be applied *to U alone*, yielding
$$U_I(P_0) = \frac{1}{j\lambda}\int\int_{\Sigma} U(P_1)\frac{\exp(jkr_{01})}{r_{01}}\cos(\vec{n},\vec{r}_{01})ds. \qquad (3.41)$$
Since no boundary conditions need be applied to $\partial U/\partial n$, the inconsistencies of the Kirchhoff theory have been removed.
3. If Eq. (3.37) is used,
$$U_{II}(P_0) = \frac{1}{2\pi}\int\int_{\Sigma} \frac{\partial U(P_1)}{\partial n}\frac{\exp(jkr_{01})}{r_{01}}ds. \qquad (3.42)$$

4. If the illumination of the aperture in all cases is a spherical wave diverging from a point source at position P_2:

$$U(P_1) = A\frac{\exp(jkr_{21})}{r_{21}}.$$

5. Using G_-, we obtain

$$U_I(P_0) = \frac{A}{j\lambda}\int\int_{\Sigma} \frac{\exp[jk(r_{21}+r_{01})]}{r_{21}r_{01}} \cos(\vec{n},\vec{r}_{01})ds. \qquad (3.43)$$

This result is known as the *Rayleigh-Sommerfeld diffraction formula*.

6. Using G_+, and assuming that $r_{21} \gg \lambda$,

$$U_{II}(P_0) = -\frac{A}{j\lambda}\int\int_{\Sigma} \frac{\exp[jk(r_{21}+r_{01})]}{r_{21}r_{01}} \cos(\vec{n},\vec{r}_{21})ds, \qquad (3.44)$$

where the angle between $\vec{n}$ and $\vec{r_{21}}$ is greater than 90^o.

3.6 Comparison of The Kirchhoff and Rayleigh-Sommerfeld Theorems - 1

1. Let
   G_K: the Green's function for Kirchhoff theory,
   G_- and G_+: the Green's functions for the two Rayleigh-Sommerfeld formulations
2. On the surface Σ, $G_+ = 2G_K$ and $\partial G_- / \partial n = 2\partial G_K / \partial n$.
3. *The Kirchhoff solution is the arithmetic average of the two Rayleigh-Smmerfeld solutions.*

$$U(P_0) = \frac{1}{2}[U_I(P_0) + U_{II}(P_0)].$$

4. The results from the Rayleigh-Sommerfeld theory differ from the Fresnel-Kirchhoff diffraction formula only through what is known as the *obliquity factor* ψ, which is the angular dependence introduced by the cosine terms.

3.6 Comparison of The Kirchhoff and Rayleigh-Sommerfeld Theorems - 2

5. For *all* cases we can write

$$U(P_0) = \frac{A}{j\lambda} \int \int_{\Sigma} \frac{\exp[jk(r_{21} + r_{01})]}{r_{21} r_{01}} \psi ds,$$

where

$$\psi = \begin{cases} \frac{1}{2}[\cos(\vec{n}, \vec{r}_{01}) - \cos(\vec{n}, \vec{r}_{21})], & \text{Kirchhoff theory (3.27)} \\ \cos(\vec{n}, \vec{r}_{01}), & \text{First Rayleigh-Sommerfeld solution (3.43)} \\ -\cos(\vec{n}, \vec{r}_{21}), & \text{Second Rayleigh-Sommerfeld solution (3.44)} \end{cases}$$

6. For the special case of an infinite distant point source producing normally incident plane wave illumination,

$$\psi = \begin{cases} \frac{1}{2}[1 + \cos\theta], & \text{Kirchhoff theory} \\ \cos\theta, & \text{First Rayleigh-Sommerfeld solution} \\ 1, & \text{Second Rayleigh-Sommerfeld solution} \end{cases}$$

where θ is the angle between the vectors $\vec{n}$ and $\vec{r}_{01}$.

3.6 Comparison of The Kirchhoff and Rayleigh-Sommerfeld Theorems - 3

7. Comments:
   - Kirchhoff solution and the two Rayleigh-Sommerfeld solutions to be essentially the same provided the aperture diameter is much greater than a wavelength.
   - When only small angles θ are involved, all three solutions are identical.
   - The Kirchhoff theory is more general than the Rayleigh-Sommerfeld theory since the latter requires that the diffracting screens be *planar*, while the former does not.
   - We generally use the first Rayleigh-Sommerfeld solution because of its simplicity.

3.7 Further Discussion of The Huygens-Fresnel Principle - 1

1. The Huygens-Fresnel principle, as predicted by the first Rayleigh-Sommerfeld solution (Eq. 3.41):

$$U(P_0) = \frac{1}{j\lambda} \int \int_{\Sigma} U(P_1) \frac{\exp(jkr_{01})}{r_{01}} \cos\theta ds. \qquad (3.51)$$

2. The observed field $U(P_0)$ is a superposition of diverging spherical waves $\exp(jkr_{01}/r_{01})$ originating from secondary sources located at each and every point P_1 within the aperture Σ.
3. The secondary source P_1 has the following properties:
    - Its complex amplitude is proportional to the amplitude of $U(P_0)$ at the corresponding point.
    - Its amplitude is inversely proportional to λ.
    - It has a phase lead 90^o than the incident wave.
    - Each secondary source has a directivity pattern $\cos\theta$.

4. The Huygens-Fresnel principle is nothing more than a *superposition integral*.

$$U(P_0) = \int \int_{\Sigma} h(P_0, P_1) U(P_1) ds,$$

where the impulse response $h(P_0, P_1)$ is

$$h(P_0, P_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos\theta.$$

We will find it is also *space-invariant* in Chapter 4.

3.8 Generalization to Nonmonochromatic Waves - 1

1. The more general case of a nonmonochromatic disturbance will now be considered briefly.
2. Consider the scalar disturbance $u(P_0, t)$ observed behind an aperture when a disturbance $u(P_1, t)$ is incident on that aperture.

3. The time functions may be expressed in terms of their inverse FTs:

$$u(P_1,t)=\int_{-\infty}^{\infty}U(P_1,\nu)\exp(j2\pi\nu t)d\nu$$

$$u(P_0,t)=\int_{-\infty}^{\infty}U(P_0,\nu)\exp(j2\pi\nu t)d\nu$$

4. By the exchange of variables $\nu'=-\nu$, yielding

$$\begin{aligned} u(P_1,t) &= \int_{-\infty}^{\infty}U(P_1,-\nu')\exp(-j2\pi\nu' t)d\nu' \\ u(P_0,t) &= \int_{-\infty}^{\infty}U(P_0,-\nu')\exp(-j2\pi\nu' t)d\nu' \quad (3.55) \end{aligned}$$

3.8 Generalization to Nonmonochromatic Waves - 3

5. The nonmonochromatic time functions $u(P_1, t)$ and $u(P_0, t)$ may be regarded as the linear combination of monochromatic time functions of the type represented in Eq. (3.10).
6. The monochromatic elementary functions are of various frequencies ν', the complex amplitudes of the disturbance at frequency ν' being simply $U(P_1, -\nu')$ and $U(P_0, -\nu')$.

7. Eq. (3.51) is used to write

$$U(P_0, -\nu') = -j\frac{\nu'}{v}\int\int_{\Sigma} U(P_1, -\nu')\frac{\exp(j2\pi\nu' r_{01}/v)}{r_{01}}\cos(\vec{n}, \vec{r}_{01})ds, \tag{3.56}$$

   where $v = c/n = \nu'\lambda$.

8. Subtitution of (3.56) into Eq. (3.55),

$$u(P_0, t) = \int\int_{\Sigma}\frac{\cos(\vec{n}, \vec{r}_{01})}{2\pi v r_{01}}\int_{-\infty}^{\infty} -j2\pi\nu' U(P_1, -\nu')\exp[-j2\pi\nu'(t-\frac{r_{01}}{v})]d\nu' ds.$$

9. Finally, the identity

$$\frac{d}{dt}u(P_1,t)$$

$$= \frac{d}{dt}\int_{-\infty}^{\infty} U(P_1,-\nu')\exp(-j2\pi\nu' t)d\nu'$$

$$= \int_{-\infty}^{\infty} -j2\pi\nu' U(P_1,-\nu')\exp(-j2\pi\nu' t)d\nu'$$

can be used to write

$$u(P_0,t)$$

$$= \int\int_{\Sigma} \frac{\cos(\vec{n},\vec{r}_{01})}{2\pi v r_{01}}\frac{d}{dt}u(P_1,t-\frac{r_{01}}{v})ds. \qquad (3.57)$$

10. The wave disturbance at point P_0 is linearly proportional to the *time derivative* of the disturbance at each P_1 on the aperture.
11. An understanding of diffraction of monochromatic waves can be used directly to synthesize the results for much more general nonmonochromatic waves.

3.9 Diffraction at Boundaries –1

1. Regard the observed field as consisting of a superposition of the incident wave transmitted through the aperture unperturbed, and a diffracted wave originating at the *rim* of the aperture.
2. Young's ideas: The field in the geometrical shadow of the screen has the form of a cylindrical wave originating on the rim of the screen.
3. In the directly illuminated region behind the plane of the screen the field is a superposition of this cylindrical wave with the directly transmitted wave.
4. The Kirchhoff diffraction formula can indeed be manipulated to yield a form that is equivalent to Young's ideas.

3.9 Diffraction at Boundaries –2

5. The field behind a diffracting obstacle is found by the principles of geometrical optics, modified by the inclusion of "diffracted rays" that originate at certain points on the obstacle itself.

3.10 The Angular Spectrum of Plane Waves

1. To formulate scalar diffraction theory in a framework that closely resembles the theory of linear, invariant systems.
2. If the complex field distribution of a monochromatic disturbance is Fourier-analyzed across any plane, the various spatial Fourier components can be identified as *plane waves travelling in different directions away from that plane*.

3.10.1 The Angular Spectrum and Its Physical Interpretation - 1

1. Let the complex field across that $z = 0$ plane be represented by $U(x, y, 0)$. Our objective is to calculate the resulting field $U(x, y, z)$ that appears across a second, parallel plane a distance z to the right of the first plane.
2. Across the $z = 0$ plane, the function U has a 2-D Fourier transform given by

$$A(f_X, f_Y, 0) = \int\int_{-\infty}^{\infty} U(x, y, 0) \exp[-j2\pi(f_X x + f_Y y)] dxdy.$$

3. Frequency decomposition: we write U as an inverse Fourier transform of its spectrum,

$$U(x, y, 0) = \int\int_{-\infty}^{\infty} A(f_X, f_Y, 0) \exp[j2\pi(f_X x + f_Y y)] df_X df_Y. \tag{3.59}$$

3.10.1 The Angular Spectrum and Its Physical Interpretation - 2

4. Consider a simple plane wave propagating with wave vector $\vec{k}$, where $\vec{k}$ has magnitude $2\pi/\lambda$ and has direction cosines (α, β, γ) (Fig. 3.9)

$$p(x, y, z; t) = \exp[j(\vec{k} \cdot \vec{r} - 2\pi\nu t)]$$

where $\vec{r} = x\hat{x} + y\hat{y} + z\hat{z}$ is a position vector, while $\vec{k} = \frac{2\pi}{\lambda}(\alpha\hat{x} + \beta\hat{y} + \gamma\hat{z})$.

5. Dropping the time dependence,

$$P(x, y, z) = \exp(j\vec{k} \cdot \vec{r}) = e^{j\frac{2\pi}{\lambda}(\alpha x + \beta y)} e^{j\frac{2\pi}{\lambda}\gamma z}.$$

Note that $\gamma = \sqrt{1 - \alpha^2 - \beta^2}$.

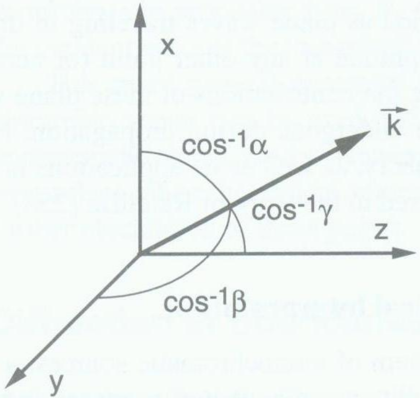

Figure 3.9 The wave vector $\vec{k}$.

3.10.1 The Angular Spectrum and Its Physical Interpretation - 3

6. Across the plane $z = 0$, $\exp[j2\pi(f_X x + f_Y y)]$ may be regarded as representing a plane wave propagating with direction cosines

$$\alpha = \lambda f_X, \quad \beta = \lambda f_Y, \quad \gamma = \sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}. \tag{3.62}$$

7. In the Fourier decomposition of U, the complex amplitude of the plane-wave component with spatial frequencies (f_X, f_Y) is simply $A(f_X, f_Y; 0)df_X df_Y$, evaluated at $(f_X = \alpha/\lambda, f_Y = \beta/\lambda)$.
8. For this reason, the function

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0) = \int\int_{-\infty}^{\infty} U(x, y, 0) \exp[-j2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)]dxdy \tag{3.63}$$

is called the *angular spectrum* of the disturbance $U(x, y, 0)$.

3.10.2 Propagation of the Angular Spectrum - 1

1. Consider the angular spectrum of U across a plane parallel to the (x, y) plane but at a distance z from it.

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = \int\int_{-\infty}^{\infty} U(x, y, z) \exp[-j2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)]dxdy. \tag{3.64}$$

2. Note that U can be written as

$$U(x, y, z) = \int\int_{-\infty}^{\infty} A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) \exp[j2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)]d\frac{\alpha}{\lambda}d\frac{\beta}{\lambda}. \tag{3.65}$$

3. Recall the Helmholtz equation $\bigtriangledown^2 U + k^2 U = 0$ at all source-free points. Then A must satisfy the differential equation

$$\frac{d^2}{dz^2}A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) + (\frac{2\pi}{\lambda})^2[1 - \alpha^2 - \beta^2]A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = 0.$$

3.10.2 Propagation of the Angular Spectrum - 2

4. An elementary solution:

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0) \exp(j\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z). \quad (3.66)$$

5. If the direction cosines (α, β) satisfy

$$\alpha^2 + \beta^2 < 1, \quad (3.67)$$

the effect of propagation over distance z is simply a change of the relative phases of the various components of the angular spectrum (relative phase delays are introduced).

3.10.2 Propagation of the Angular Spectrum - 3

6. If $\alpha^2 + \beta^2 > 1$, α and β are no longer interpretable as direction cosines. The equation is rewritten

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0) \exp(-\mu z), \qquad (3.68)$$

   where $\mu = \frac{2\pi}{\lambda}\sqrt{\alpha^2 + \beta^2 - 1}$.
7. Since μ is a positive real number, these wave components are rapidly attenuated by the propagation phenomenon. Such components are called *evanescent waves.*

8. Finally,

$$U(x,y,z) = \int\int_{-\infty}^{\infty} A(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0)\exp(j\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z)$$

$$\times \exp[j2\pi(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y)]d\frac{\alpha}{\lambda}d\frac{\beta}{\lambda}, \quad (3.69)$$

where the evanescent wave phenomenon will effectively limit the region with which Eq. (3.67) is satisfied.

9. No angular spectrum components beyond the evanescent wave cutoff contribute to $U(x,y,z)$.

3.10.3 Effects of a Diffracting Aperture on the Angular Spectrum - 1

1. Define the *amplitude transmittance function* of the aperture as the ratio of the transmitted field amplitude $U_t(x, y; 0)$ to the incident field amplitude $U_i(x, y; 0)$ at each (x, y) in the $z = 0$ plane,
$$t_A(x, y) = \frac{U_t(x, y; 0)}{U_i(x, y; 0)}. \tag{3.70}$$
2. Use the convolution theorem to relate the angular spectrum $A_i(\alpha/\lambda, \beta/\lambda)$ of the incident field and the angular spectrum $A_t(\alpha/\lambda, \beta/\lambda)$ of the transmitted field,
$$A_t(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}) = [A_i(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}) \otimes T(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda})], \tag{3.71}$$
where
$$T(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}) = \int\int_{-\infty}^{\infty} t_A(x, y) \exp[-j2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)]dxdy. \tag{3.72}$$

3.10.3 Effects of a Diffracting Aperture on the Angular Spectrum - 2

3. For the case of a unit amplitude plane wave illuminating the diffracting structure normally,

$$A_i(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}) = \delta(\frac{\alpha}{\lambda},\frac{\beta}{\lambda})$$

and

$$A_t(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}) = \delta(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}) \otimes T(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}) = T(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}).$$

4. The transmitted angular spectrum is found directly by Fourier transforming the amplitude transmittance function of the aperture. The smaller the aperture, the broader the angular spectrum behind the aperture.

3.10.4 The Propagation Phenomenon as a Linear Spatial Filter - 1

1. The propagation phenomenon acts as a linear space-invariant system and is characterized by a relatively simple transfer function.
2. It is now more convenient to leave the spectra as functions of spatial frequencies (f_X, f_Y) (Eq. 3.62).
3. Let the spatial spectrum of $U(x, y, z)$ again be represented by $A(f_X, f_Y; z)$.

$$U(x, y, z) = \int\int_{-\infty}^{\infty} A(f_X, f_Y; z) \exp[j2\pi(f_X x + f_Y y)] df_X df_Y.$$

3.10.4 The Propagation Phenomenon as a Linear Spatial Filter - 2

4. From Eq. (3.69)

$$U(x, y, z) = \int\int_{-\infty}^{\infty} A(f_X, f_Y; 0) \exp(j\frac{2\pi}{\lambda}\sqrt{1-(\lambda f_X)^2-(\lambda f_Y)^2}z)$$

$$\times \exp[j2\pi(f_X x + f_Y y)] df_X df_Y.$$

5. A comparison of the above two equations shows that

$$A(f_X, f_Y; z) = A(f_X, f_Y; 0) \times \exp(j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_X)^2-(\lambda f_Y)^2}). \tag{3.73}$$

3.10.4 The Propagation Phenomenon as a Linear Spatial Filter - 3

6. Finally, the transfer function of the wave propagation phenomenon is seen to be

$$H(f_X, f_Y) = \exp[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}] \qquad (3.74)$$

3.10.4 The Propagation Phenomenon as a Linear Spatial Filter - 3

7. The propagation phenomenon may be regarded as a linear, dispersive spatial filter with a finite bandwidth. The transmittance of the filter is zero outside a circular region of radius λ^{-1} in the frequency plane. Within the circular bandwidth, the modulus of the transfer function is unity but frequency-dependent phase shifts are introduced.
8. The phase dispersion is most significant at high spatial frequencies and vanishes as both f_X and f_Y approach zero. In addition, it increases as z increases.
9. The angular spectrum approach and the first Rayleigh-Sommerfeld solution yield identical predictions of diffracted field.