OPTICA DE FOURIER

Claudio Iemmi

laboratorio de procesado de imágenes
departamento de fisica - FCEyN - UBA

Departamento de Física – Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires – Argentina

Este documento no es un libro de texto sino el conjunto de algunos apuntes elaborados para la materia optativa y de postgrado *Óptica de Fourier* que se dicta en el Departamento de Física de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires. La materia tiene una duración cuatrimestral y además de la parte teórica que acá se describe, consta de experiencias demostrativas. Como es de esperar este curso evoluciona, de forma tal que en esta presente edición figuran temas que no se dictaron anteriormente y fueron eliminados otros que si eran tratados.
De todos modos debe recalcarse que el tratamiento de los temas aquí realizado es introductorio y sólo constituye un punto de partida para los interesados en las diversas áreas abordadas.
La bibliografía básica sugerida para el curso en general es
J.Goodman *Introduction to Fourier Optics* Mc Graw-Hill 2nd Ed. (1996)
M.Born & E.Wolf *Principles of Optics* Pergamon Press 6th Ed. (1980)
La bibliografía relacionada con temas específicos es citada dentro del texto

1er cuatrimestre de 2017 Claudio Iemmi

OPTICA DE FOURIER

INDICE

VII. MODULACIÓN DEL FRENTE DE ONDAS

VIII. HOLOGRAFÍA

IX. APLICACIONES DE LA HOLOGRAFÍA

I. TEORÍA ESCALAR DE LA DIFRACCIÓN

CLASE 1

INTRODUCCIÓN

El fenómeno conocido como difracción juega un rol fundamental en aquellas ramas de la física en donde se estudia la propagación de ondas. Así para comprender distintos procesos tales como la formación de imágenes, el tratamiento óptico de señales, etc. es esencial tener en cuenta este fenómeno. Comenzaremos entonces considerando los fundamentos de la teoría escalar de la difracción.

Sommerfeld la define como *cualquier desviación de los rayos de luz de su propagación rectilínea y que no puedan ser interpretados como reflexión o refracción*. El primero en reportar este fenómeno fue Grimaldi en una publicación del año 1665. El observó que cuando iluminaba una varilla con una fuente puntual, la sombra recogida sobre la pantalla poseía franjas más brillantes en su interior. Esto iba en contra de la propagación rectilínea de la luz postulada por la teoría corpuscular vigente en ese momento.

El paso inicial hacia la teoría que explicara este fenómeno fue realizado por Huygens en 1678. Este enunció de forma intuitiva que *cada punto de un frente de ondas primario sirve como fuente de onditas esféricas secundarias tales que el frente de ondas primario, un instante después, es la envolvente de dichas onditas. Además las onditas avanzan con una rapidez y frecuencia igual a la de la onda primaria en cada punto del espacio*. Sin embargo este principio por si solo no basta para explicar la difracción.

Durante los años 1801-1803 Young presentó su principio de interferencia y, en 1818, Fresnel unió las ideas intuitivas de Huygens con dicho principio, formulando lo que se conoce como principio de Huygens – Fresnel que establece que *cada punto sin obturación de un frente de ondas, en un instante de tiempo dado, sirve como una fuente de onditas esféricas secundarias, de la misma frecuencia de la onda primaria. La amplitud del campo óptico en cualquier punto adelante, es la superposición de todas estas onditas considerando sus amplitudes y fases relativas.* Fresnel logra calcular y predecir con mucha precisión diversas figuras de difracción, empleando este enunciado y adjuntando algunas hipótesis adicionales, aparentemente arbitrarias:

- La amplitud de las onditas secundarias difiere de la incidente en un factor λ^{-1}.

- La amplitud se halla también modulada por un factor llamado de oblicuidad (de hecho tales onditas no emiten hacia atrás).
- Existe un desfasaje de 90º entre la onda incidente y la emisión de la onda secundaria.

Fue Kirchhoff en 1882 quién mediante fundamentos matemáticos logra demostrar que estas hipótesis son resultados que surgen naturalmente de la aplicación de la teoría ondulatoria de la luz, es decir a partir de las ecuaciones de Maxwell (1864). Sin embargo Kirchhoff también hace dos suposiciones que se demostrarán incompatibles.
Sommerfeld en 1894 modifica lo planteado por Kirchhoff y logra así evitar esta incompatibilidad. Surge así la teoría de difracción llamada de Rayleigh – Sommerfeld.
Si bien tanto la teoría de Kirchhoff como la de Rayleigh – Sommerfeld se mostraron muy precisas en muchos casos, cabe destacar que no funcionan en otros ya que realizan algunas aproximaciones y simplificaciones de importancia. La más importante de ellas es que la luz es tratada como un fenómeno escalar. Es decir se supone que el campo eléctrico (en este caso interesa el campo eléctrico y no el magnético ya que es el vector óptico) tiene una sola componente ó, si tuviese dos, la otra puede ser tratada de igual forma. Esta aproximación desprecia el hecho de que las distintas componentes del campo eléctrico están acopladas por las ecuaciones de Maxwell y no pueden tratarse de forma independiente. Afortunadamente esta teoría da resultados muy precisos si se cumple que:

- La abertura difractora es grande en comparación con la longitud de onda.
- El campo difractado no se observa en un punto demasiado cerca de la abertura (no se puede estudiar que pasa a 4λ de la misma)

Está claro que estas condiciones no se cumplen en problemas de gran importancia como por ejemplo difracción por nano-estructuras, difracción cónica,etc. En estos casos debe utilizarse una teoría vectorial rigurosa para obtener resultados precisos. Estos problemas quedan excluidos de este curso. Esta teoría también falla en sistemas formadores de imágenes si se trabaja con ángulos muy grandes. Analicemos por ejemplo la Figura 1. Sobre el lado izquierdo de la lente no hay interferencia ya que los campos $\vec{E_1}$ y $\vec{E_2}$ son perpendiculares, mientras que sí lo hacen en el lado derecho. La teoría escalar no puede explicar que haya interferencia de un lado y no del otro.

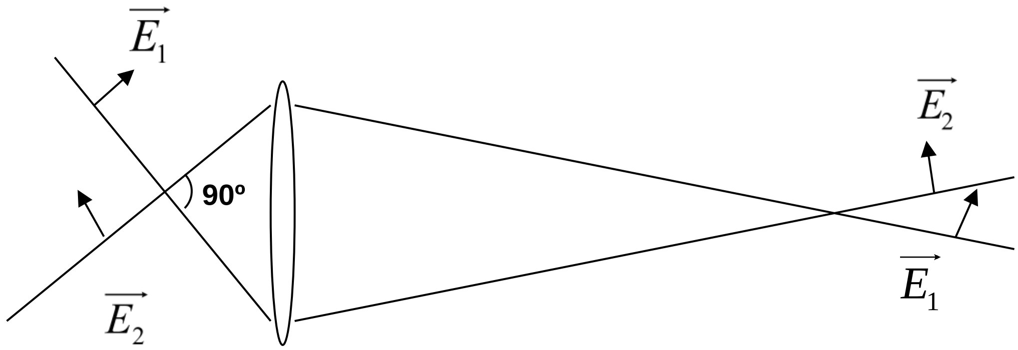


Figura 1: *Sobre el lado izquierdo de la lente los campos son perpendiculares y por lo tanto no interfieren, en cambio sí lo hacen sobre el lado derecho*

TEOREMA INTEGRAL DE HELMHOLTZ – KIRCHHOFF

Avancemos ahora en la formulación matemática de la teoría de difracción. Las ecuaciones de Maxwell en el sistema c.g.s. vienen dadas por:

$$\vec{\nabla} X\ \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}$$ Ley de Faraday

$$\vec{\nabla} X\ \vec{H} = \frac{4\pi}{c}\vec{J} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t}$$ Ley de Ampere

$$\vec{\nabla} . \vec{D} = 4\pi\rho$$ Ley de Coulomb

$$\vec{\nabla} . \vec{B} = 0$$ Ausencia de monopolos magnéticos

Donde $\vec{E}$ es el campo eléctrico, $\vec{B}$ el magnético, $\vec{D}$ es el vector desplazamiento, $\vec{H}$ es el vector intensidad magnética, $\vec{J}$ el vector densidad de corriente y ρ la densidad de carga. Por otra parte tenemos las relaciones constitutivas:

$$\vec{J} = \sigma\,\vec{E} \qquad \vec{B} = \mu\,\vec{H} \qquad \vec{D} = \varepsilon\,\vec{E}$$

Donde σ es la conductividad, μ es la permitividad y ε es la constante dieléctrica. Si además el medio es lineal, isótropo y homogéneo tenemos que:

$$\vec{\nabla}.\vec{D}=\varepsilon\,\vec{\nabla}.\vec{E} \qquad ; \qquad \vec{\nabla}X\ \vec{H}=\frac{1}{\mu}\vec{\nabla}X\ \vec{B}$$

Entonces las ecuaciones de Maxwell quedan

$$\vec{\nabla}X\ \vec{E}=-\frac{1}{c}\frac{\partial\vec{B}}{\partial t} \quad ; \quad \vec{\nabla}X\ \vec{B}=\frac{4\pi\mu}{c}\vec{J}+\frac{\mu\,\varepsilon}{c}\frac{\partial\vec{E}}{\partial t} \quad ; \quad \vec{\nabla}.\vec{E}=\frac{4\pi\,\rho}{\varepsilon} \quad ; \quad \vec{\nabla}.\vec{B}=0$$

Para llegar a la ecuación de ondas podemos, por ejemplo, tomar rotor en ambos miembros de la primera ecuación, tenemos así que

$$\vec{\nabla}\,X\,\vec{\nabla}\,X\ \vec{E}=\vec{\nabla}\left(\vec{\nabla}.\vec{E}\right)-\nabla^2\vec{E}=-\frac{1}{c}\frac{\partial\vec{\nabla}X\vec{B}}{\partial t}$$

Reemplazando la expresión para la segunda ecuación tenemos que

$$\vec{\nabla}\left(\vec{\nabla}.\vec{E}\right)-\nabla^2\vec{E}=-\frac{1}{c}\frac{\partial\left(\frac{4\pi\mu}{c}\vec{J}+\frac{\mu\,\varepsilon}{c}\frac{\partial\vec{E}}{\partial t}\right)}{\partial t} \qquad \Rightarrow$$

$$\nabla^2\vec{E}-\frac{\mu\,\varepsilon}{c^2}\frac{\partial^2\vec{E}}{\partial^2 t}=\underbrace{\frac{4\pi\mu}{c^2}\frac{\partial\vec{J}}{\partial t}+4\pi\,\vec{\nabla}\left(\frac{\rho}{\varepsilon}\right)}_{S}$$

Donde $\mu\ \varepsilon\ /\ c^2 = 1/\ v^2$ es la inversa de la velocidad de la luz en el medio y S describe las fuentes. Se obtiene una ecuación análoga para el campo magnético $\vec{B}$. En ausencia de fuentes S = 0 y recordando que nosotros vamos a trabajar con campos escalares una solución posible de la ecuación es:

$$E\left(\vec{r},t\right)=E_0\, e^{i\left(\vec{k}\,\vec{r}-\omega t\right)}$$

Si $E_0 = Cte$ esta expresión corresponderá a una onda plana, si $E_0 = {Cte}/{r}$ será una onda esférica y si $E_0 = {Cte}/{\sqrt{r}}$ describirá a una onda cilíndrica.

Dado que la ecuación de ondas es lineal y de coeficientes reales, la solución debe ser real, con lo cual la expresión que tenga sentido físico será $E\left(\vec{r},t\right)=\Re e\left[E_0\, e^{i\left(\vec{k}\,\vec{r}-\omega t\right)}\right]$.

Entonces si $E\left(\vec{r},t\right)$ es la perturbación óptica en $\vec{r}$ a tiempo t; asumiendo que es una función escalar (esto es linealmente polarizada) y **monocromática**, en ausencia de fuentes la ecuación de ondas es:

$$\nabla^2 E - \frac{1}{v^2}\frac{\partial^2 E}{\partial^2 t} = 0 \qquad \text{con} \qquad E\left(\vec{r},t\right)=E\left(\vec{r}\right)e^{-i\omega t}$$

Teniendo en cuenta la dependencia temporal y que $k = 2\pi\, {\nu}/{\mathrm{v}} = {\omega}/{\mathrm{v}} = {2\pi}/{\lambda}$ obtenemos la ecuación de Helmholtz independiente del tiempo

$$\left(\nabla^2 + k^2\right)E = 0 \tag{1}$$

Kirchhoff expresa la solución de esta ecuación en un punto de observación P_0 en términos del valor del campo E y de su derivada normal, sobre una superficie cerrada arbitraria S que rodea a dicho punto. (Esta solución había sido propuesta con anterioridad por Helmholtz para acústica).

Para encontrar esta solución utiliza el *Teorema de Green* que establece que si U_1 y U_2 son dos funciones complejas de la posición, **escalares**, con la primera y segunda derivadas univaluadas y contínuas dentro y sobre la superficie S entonces:

$$\iiint\limits_V \left(U_1 \nabla^2 U_2 - U_2 \nabla^2 U_1\right) dV = \oiint\limits_S \left(U_1 \vec{\nabla} U_2 - U_2 \vec{\nabla} U_1\right) \cdot n \; dS$$

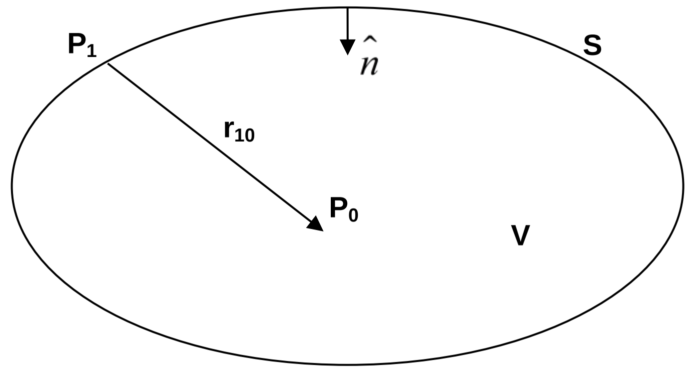


Figura 2: *Superficie y volumen de integración*

Está claro que si U_1 y U_2 son soluciones de la ecuación de Helmholtz, es decir

$$\nabla^2 U_1 + k^2 U_1 = 0 \quad \Rightarrow \quad U_2 \nabla^2 U_1 + U_2 k^2 U_1 = 0$$

$$\nabla^2 U_2 + k^2 U_2 = 0 \quad \Rightarrow \quad U_1 \nabla^2 U_2 + U_1 k^2 U_2 = 0$$

restando la segunda de la primera, tenemos que $U_1 \nabla^2 U_2 - U_2 \nabla^2 U_1 = 0$

por lo tanto

$$\oiint\limits_S \left(U_1 \vec{\nabla} U_2 - U_2 \vec{\nabla} U_1\right) \cdot n \; dS = 0 \tag{2}$$

Podemos elegir $U_1 = E$ y $U_2 = G\left(P_1, P_0\right) = \dfrac{e^{ikr_{10}}}{r_{10}}$ con

$$r_{10} = \left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2 + \left(z_0 - z_1\right)^2\right]^{\frac{1}{2}}$$

Esta función G es una onda esférica de amplitud 1 centrada en P_1. En general en los libros aparece $G\left(P_0, P_1\right) = \dfrac{e^{ikr_{01}}}{r_{01}}$ que es una onda centrada en P_0 (punto de observación). Nosotros lo tomamos al revés porque es más intuitivo, pero de todos modos debemos tener presente que esta función **no representa una onda real sino que es una herramienta matemática** para resolver el problema.

Ahora bien, si observamos el dibujo vemos que si queremos llevar P_1 hasta P_0 aparece una divergencia, entonces lo que se debe hacer es rodear a P_0 con una esfera de radio ε con el fin de excluir a P_0 de la región encerrada por S .

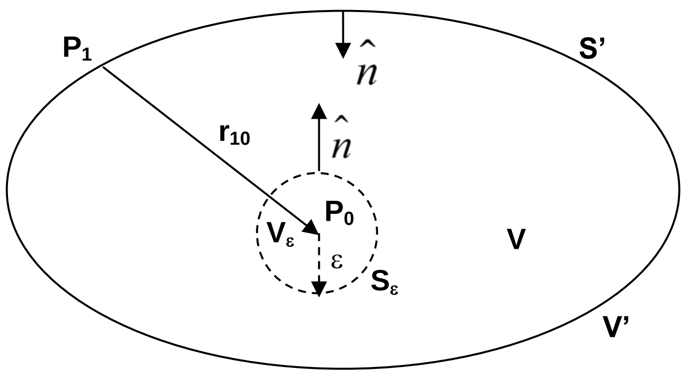


Figura 3: *Superficie y volumen de integración sin singularidades*

Ahora $S = S' + \lim_{\varepsilon \to 0} S_\varepsilon$; $V = V' - \lim_{\varepsilon \to 0} V_\varepsilon$

Entonces $\oiint_S = \oiint_{S'} + \lim_{\varepsilon \to 0} \oiint_{S_\varepsilon}$. Vamos a calcular la segunda integral teniendo en cuenta que el gradiente se toma sobre la variable 1 de modo que $\vec{\nabla}_1 = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial y_1}, \frac{\partial}{\partial z_1} \right)$

$$\vec{\nabla}_1 G = \vec{\nabla}_1 \left(\frac{e^{ikr_{10}}}{r_{10}} \right) = -\left(ik - \frac{1}{r_{10}} \right) \frac{e^{ikr_{10}}}{r_{10}} \hat{r}_{10} \quad \Rightarrow$$

$$\Rightarrow \quad \oiint_{S_\varepsilon} \left[-E \left(ik - \frac{1}{r_{10}} \right) \frac{e^{ikr_{10}}}{r_{10}} \hat{r}_{10} - \frac{e^{ikr_{10}}}{r_{10}} \vec{\nabla} E \right] \cdot n \, dS_\varepsilon =$$

Debemos recordar que $n \cdot \vec{\nabla} U = \frac{\partial U}{\partial n}$ con lo cual la ecuación anterior queda

$$\oiint_{S_\varepsilon} \frac{e^{ikr_{10}}}{r_{10}} \left[E\left(ik - \frac{1}{r_{10}} \right) - \frac{\partial E}{\partial n} \right] dS_\varepsilon$$ pero sobre $S_\varepsilon \quad r_{10} = \varepsilon \quad \Rightarrow$

$$\Rightarrow \quad = \oiint_{S_\varepsilon} \frac{e^{ik\varepsilon}}{\varepsilon} \left[E\left(ik - \frac{1}{\varepsilon} \right) - \frac{\partial E}{\partial n} \right] dS_\varepsilon$$ dado que $\varepsilon \to 0$ y recordando que E y G y sus derivadas son continuas

$$\lim_{\varepsilon \to 0} \oiint_{S_\varepsilon} f \, dS_\varepsilon \quad \to \quad \lim_{\varepsilon \to 0} f\left(P_0\right) 4\pi\varepsilon^2$$ La integral resulta

$$\lim_{\varepsilon \to 0} \left[\cancelto{0}{4\pi\varepsilon \, e^{ik\varepsilon} E \, ik} - 4\pi \, e^{ik\varepsilon} \, E - \cancelto{0}{4\pi\varepsilon \, e^{ik\varepsilon} \frac{\partial E}{\partial n}} \right] = -4\pi \, E\left(P_0\right)$$

Luego, retornando a la ecuación (2) tenemos que:

$$\boxed{E\left(P_0\right) = -\frac{1}{4\pi} \oiint_{S'} \left[\frac{e^{ikr_{10}}}{r_{10}} \frac{\partial E}{\partial n} - E \frac{\partial}{\partial n} \left(\frac{e^{ikr_{10}}}{r_{10}} \right) \right] dS'} \qquad (3)$$

Este resultado es conocido como el **Teorema integral de Helmholtz – Krichhoff**.

FORMULACIÓN DE KIRCHHOFF PARA LA DIFRACCIÓN POR UNA PANTALLA PLANA

Veamos ahora como aplicar el teorema integral al caso de difracción por una pantalla plana, infinita y opaca, con una abertura.

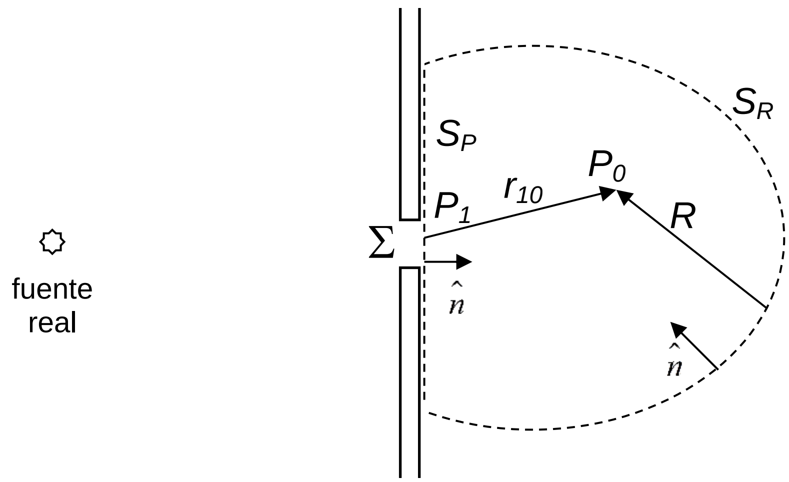


Figura 4: *Difracción por una pantalla plana según Kirchhoff*

Ahora $S = S_P + \Sigma + S_R \Rightarrow \oiint_S = \iint_{S_P} + \iint_{\Sigma} + \iint_{S_R}$

Se toma la superficie S_P plana porque sobre una superficie curva se complica terriblemente la función de Green G.

La fuente real que ilumina la pantalla está fuera del recinto S. Ahora vemos que la razón de tomar n hacia adentro (al revés que en el Goodman) es para que coincida con los z positivos sobre la abertura. Por otra parte se puede ver también la razón de tomar ondas emergentes de P_1 en vez de P_0 ya que ahora sobre Σ esas onditas coincidirán con las fuentes virtuales secundarias del *Principio de Huygens – Fresnel*.

Vamos a calcular primero la integral (3) en la superficie S_R.

$$-\frac{1}{4\pi}\iint_{S_R} \frac{\exp(ikR)}{R}\frac{\partial E}{\partial n} + \frac{\exp(ikR)}{R}\left(ik - \frac{1}{R}\right) E\,\hat{r}_{10}\cdot n\, dS_R$$

ya que

$$\frac{\partial}{\partial n}\left(\frac{\exp(ikr_{10})}{r_{10}}\right) = n\cdot\vec{\nabla}_1\left(\frac{\exp(ikr_{10})}{r_{10}}\right) = -\frac{\exp(ikR)}{R}\left(ik - \frac{1}{R}\right)\hat{r}_{10}\cdot n$$

De acuerdo a las hipótesis iniciales $R \gg \lambda \Rightarrow \frac{2\pi}{\lambda} \gg \frac{1}{R} \Rightarrow k \gg \frac{1}{R}$; además R se hace tender a ∞. Así la integral queda:

$$\iint_{S_R} = -\frac{1}{4\pi}\iint_{S_R} \frac{\exp(ikR)}{R}\left(\frac{\partial E}{\partial n} + ikE\, n\cdot\hat{r}_{10}\right) dS_R$$

Ahora bien, uno estaría tentado a decir que dado que *G* y *E* decaen cuando $R \to \infty \quad \Rightarrow \quad \iint \to 0$, pero no debemos olvidarnos que *S* crece como R^2.

Otro argumento sería que, dado que la radiación se propaga a la velocidad finita *c,* basta tomar S_R tan grande que las ondas aún no hayan llegado; pero existe el inconveniente que trabajamos con ondas monocromáticas y por lo tanto estas ondas, por definición, existieron siempre.

Obviamente en la práctica esto siempre es válido ya que en una experiencia siempre hay alguien que enciende la fuente, es decir estrictamente hablando, aunque usemos un láser las ondas son cuasi-monocromáticas.

Volviendo a nuestro problema, si escribimos $dS_R = R^2\, d\omega$ lo que debe exigirse es que

$$\lim_{R\to\infty} R\left(\frac{\partial E}{\partial n} + ikE\; n\cdot\hat{r}_{10}\right) = 0$$

Esta es la condición de radiación de Sommerfeld y es válida para campos que se atenúen por lo menos como $1/R$.

Ejercicio 1:

Probar que si la fuente real emite ondas esféricas, se verifica la condición de radiación

Dado que la perturbación que ilumina la apertura es una onda esférica, o una combinación lineal de ondas esféricas, esta condición siempre se cumple. Podemos pasar ahora a resolver las integrales sobre la pantalla.

Kirchhoff para ello realizó dos suposiciones:

- Sobre la superficie Σ la distribución de campo *E* y su derivada son exactamente iguales que en ausencia de la pantalla (esto es válido como una aproximación sólo si $\Sigma \gg \lambda$).

- Sobre la superficie S_p el campo E y $\dfrac{\partial E}{\partial n}$ son idénticamente cero.

Esta aproximación tampoco es estrictamente cierta ya que siempre hay una zona de penumbra cercana a la abertura. Por otra parte este requisito lleva a una inconsistencia, tal como lo veremos más adelante. Estas condiciones a menudo son conocidas como condiciones de contorno de Kirchhoff.

Así las integrales sobre $\sum$ y S_p se reducen a la integral sobre $\sum$

$$E\left(P_0\right) = -\frac{1}{4\pi}\iint_{\Sigma} \frac{\exp\left(ikr_{10}\right)}{r_{10}}\left[\frac{\partial E}{\partial n} + \left(ik - \frac{1}{r_{10}}\right) E\, n \cdot \hat{r}_{10}\right] dS$$

Como antes $r_{10} \gg \lambda \quad \Rightarrow \quad k \gg \dfrac{1}{r_{10}} \qquad \Rightarrow$

$$\boxed{E\left(P_0\right) = -\frac{1}{4\pi}\iint_{\Sigma} \frac{\exp\left(ikr_{10}\right)}{r_{10}}\left[\frac{\partial E}{\partial n} + ik\, E \cos\left(n \cdot \hat{r}_{10}\right)\right] dS} \qquad (4)$$

Esta es la ecuación de Kirchhoff para la difracción a través de una pantalla plana.

Veamos que sucede si iluminamos $\sum$ con una onda esférica proveniente de una fuente real ubicada en P_2. Ahora tenemos que

$$E\left(P_1\right) = A\frac{e^{ikr_{21}}}{r_{21}} \qquad \overrightarrow{\nabla}_1 E = A\left(ik - \frac{1}{r_{21}}\right)\frac{e^{ikr_{21}}}{r_{21}}\hat{r}_{21} \qquad \text{nuevamente } k \gg \frac{1}{r_{21}}$$

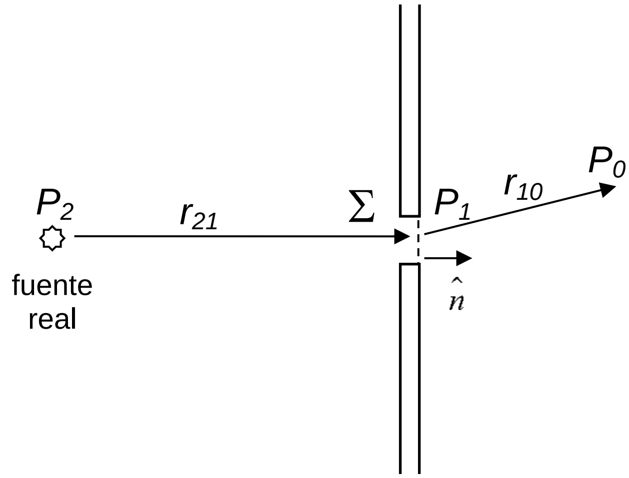


Figura 5: *Difracción por una pantalla plana iluminada con una onda esférica*

$$E(P_0) = -\frac{A}{4\pi} \iint_{\Sigma} \frac{\exp\left(ik\left(r_{21}+r_{10}\right)\right)}{r_{21}r_{10}}\ ik\left[\cos\left(n,\hat{r}_{21}\right)+\cos\left(n,\hat{r}_{10}\right)\right]dS$$

$$\boxed{E(P_0) = \frac{A}{i\lambda} \iint_{\Sigma} \frac{\exp\left(ik\left(r_{21}+r_{10}\right)\right)}{r_{21}r_{10}} \left[\frac{\cos\left(\hat{n},\hat{r}_{21}\right)+\cos\left(\hat{n},\hat{r}_{10}\right)}{2}\right]dS} \qquad (5)$$

Esta ecuación se conoce como la *Fórmula de difracción de Fresnel – Kirchhoff* y es válida para cuando se ilumina la apertura con una fuente puntual. Si la abertura tiene una cierta transmisión hay que introducir un factor T_{Σ}.

Vamos a escribir la expresión (5) de esta forma:

$$E(P_0) = \iint_{\Sigma} E'(P_1) \frac{\exp\left(ikr_{10}\right)}{r_{10}} dS \qquad \text{con}$$

$$E'(P_1) = \frac{A}{i\lambda} \frac{\exp\left(ikr_{21}\right)}{r_{21}} \left[\frac{\cos\left(n,\hat{r}_{21}\right)+\cos\left(n,\hat{r}_{10}\right)}{2}\right]$$

Vemos que podemos interpretar el campo en P_0 como el provisto por infinitas fuentes puntuales virtuales, ubicadas en los distintos P_1 de la abertura. Es decir llegamos a la formulación que intuitivamente habían hecho Huygens y Fresnel. Es más, vemos que la onda primaria que llega a P_1 tiene amplitud y fase $A\frac{e^{ikr_{21}}}{r_{21}}$ pero la onda emergente está desfasada con un adelanto de 90º y además está atenuada por el factor de oblicuidad [cos, cos], que vale 1 como máximo, y por la cantidad $1/\lambda$. Es decir que surgen de forma natural todos los requerimientos que Fresnel había impuesto en forma arbitraria para ajustar sus mediciones.

Obviamente estas fuentes virtuales no tienen un real significado físico, cabe entonces preguntarse como surge la alteración del campo al pasar por la abertura. Young (1802) sugirió que el campo observado correspondía a la interacción entre el campo de la onda incidente, directamente transmitido, con la onda difractada en los bordes de la abertura. Posteriores investigaciones confirmaron este punto de vista.

CLASE 2

FORMULACIÓN DE RAYLEIGH – SOMMERFELD PARA LA DIFRACCIÓN POR UNA PANTALLA PLANA

Si bien los resultados hallados por Kirchhoff coinciden muy bien con las experiencias, se presenta una inconsistencia en una de las hipótesis que utiliza. La misma está originada en el hecho de pedir condiciones simultáneas sobre el campo y sus derivadas.

Existe un teorema que dice que si una función potencial y su derivada normal se anulan simultáneamente a lo largo de un segmento de curva, entonces dicha función debe anularse en todo el plano (ver por ejemplo *Classical Electrodynamics* de Jackson).

Evidentemente esto no es lo que sucede, por lo tanto alguna hipótesis está mal planteada.

Tales inconsistencias fueron solucionadas por Sommerfeld quien eliminó la necesidad de imponer condiciones sobre E y $\frac{\partial E}{\partial n}$. Para ello eligió funciones de Green que fuesen combinaciones lineales de la anteriormente utilizada y tales que, o bien G ó $\frac{\partial G}{\partial n}$ se anulasen sobre $S_P+\Sigma$. De esta forma sólo se debe pedir condiciones sobre E ó $\frac{\partial E}{\partial n}$ en forma separada.

Cabe destacar que el hecho que una teoría sea autoconsistente y la otra no, no significa que una sea más precisa que la otra.

Sommerfeld plantea como soluciones de G

$$G_{\mp} = \frac{\exp(ikr_{10})}{r_{10}} \mp \frac{\exp(ikr_{10}^{'})}{r_{10}^{'}}$$

con $\vec{r}_{10}$ y $\vec{r}\,'_{10}$ tales que P'_0 es la imagen especular de P_0

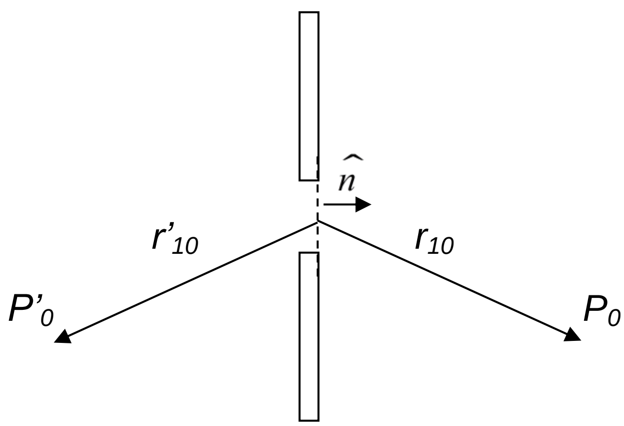


Figura 6: *Difracción por una pantalla plana en la formulación de Rayleigh-Sommerfeld*

Veamos cada solución por separado

$$\vec{\nabla}_1 G_- = -\frac{\exp(ikr_{10})}{r_{10}}\left(ik - \frac{1}{r_{10}}\right)\hat{r}_{10} + \frac{\exp(ikr_{10}^{'})}{r_{10}^{'}}\left(ik - \frac{1}{r_{10}^{'}}\right)\hat{r}_{10}^{'}$$

$$n \cdot \vec{\nabla}_1 G_- = \frac{\exp(ikr_{10})}{r_{10}}\left(ik - \frac{1}{r_{10}}\right)\left(-n \cdot \hat{r}_{10} + n \cdot \hat{r}_{10}^{'}\right) = -2\frac{\exp(ikr_{10})}{r_{10}}\left(ik - \frac{1}{r_{10}}\right) n \cdot \hat{r}_{10}$$

donde usamos que $\forall\, P_1 \in \Sigma, S_P \quad r_{10} = r_{10}^{'}$

Obviamente en estos puntos $G_- = 0$, entonces si recordamos que

$$E(P_0) = -\frac{1}{4\pi} \iint_{S_P+\Sigma} \left[G\frac{\partial E}{\partial n} - E\frac{\partial G}{\partial n} \right] dS$$

Basta ahora pedir que $E = 0$ sobre S_P, por lo tanto

$$E(P_0) = -\frac{1}{2\pi} \iint_{\Sigma} E(P_1) \frac{\exp(ikr_{10})}{r_{10}} \left(ik - \frac{1}{r_{10}} \right) n \cdot \hat{r}_{10}\, dS \qquad \text{como} \quad k \gg \frac{1}{r_{10}}$$

$$\boxed{E(P_0) = \frac{1}{i\lambda} \iint_{\Sigma} E(P_1) \frac{\exp(ikr_{10})}{r_{10}} \cos\left(n, \hat{r}_{10}\right) dS} \qquad (6)$$

Análogamente si usamos $G+$

$$n \cdot \overrightarrow{\nabla}_1 G_+ = \frac{\exp(ikr_{10})}{r_{10}} \left(ik - \frac{1}{r_{10}} \right) \left(n \cdot \hat{r}_{10} + n \cdot \hat{r}'_{10} \right) = 0$$; acá nuevamente usamos que

$\forall\, P_1 \in \sum, S_P \quad r_{10} = r'_{10}$

Entonces ahora sólo basta pedir que $\frac{\partial E}{\partial n} = 0$ sobre S_P, por lo tanto

$$\boxed{E(P_0) = -\frac{1}{2\pi} \iint_{\Sigma} \frac{\exp(ikr_{10})}{r_{10}}\; n \cdot \overrightarrow{\nabla}_1 E\, dS} \qquad (7)$$

Comparemos los resultados que se obtienen para una fuente puntual iluminando $\sum$

$$E\left(P_1\right) = A\frac{e^{ikr_{21}}}{r_{21}}$$

Para Kirchhoff habíamos obtenido la expresión (5)

$$E\left(P_0\right)=\frac{A}{i\lambda}\iint_{\Sigma}\frac{\exp\left(ik\left(r_{10}+r_{21}\right)\right)}{r_{21}r_{10}}\left[\frac{\cos\left(n,\hat{r}_{21}\right)+\cos\left(n,\hat{r}_{10}\right)}{2}\right]dS$$

Para *G_* la expresión (6) toma la forma

$$E\left(P_0\right)=\frac{A}{i\lambda}\iint_{\Sigma}\frac{\exp\left(ik\left(r_{10}+r_{21}\right)\right)}{r_{21}r_{10}}\cos\left(n,\hat{r}_{10}\right)dS$$

Para *G+* vamos a calcular previamente

$$n\cdot\overrightarrow{\nabla_1}E=A\frac{\exp\left(ikr_{21}\right)}{r_{21}}\left(ik-\frac{1}{r_{21}}\right)n\cdot\hat{r}_{21} \qquad \Rightarrow$$

$$E\left(P_0\right)=\frac{A}{i\lambda}\iint_{\Sigma}\frac{\exp\left(ik\left(r_{10}+r_{21}\right)\right)}{r_{21}r_{10}}\cos\left(n,\hat{r}_{21}\right)dS$$

Vemos que las expresiones varían sólo en el factor de oblicuidad y que para el caso de fuente y punto de observación lejanos, estos coinciden.

Se puede ver, además, que si cambiamos el punto fuente por el punto campo el resultado no varía. Esto se conoce como principio de reciprocidad

GENERALIZACIÓN DE LA FORMULACIÓN DE RAYLEIGH – SOMMERFELD PARA FUENTES NO MONOCROMÁTICAS

Con anterioridad supusimos que las fuentes eran monocromáticas, veamos cómo se modifica esta formulación cuando ello no sucede. (Lo hacemos para R – S pero para Kirchhoff es semejante).

Para ondas monocromáticas teníamos que

$$E_{Mon}\left(P_0,t\right)=E_0\left(P_0,\nu\right)\exp\left(-2\pi i\nu t\right) \quad ; \quad 2\pi\nu=\omega$$

Dado que las ecuaciones de Maxwell son lineales, podemos sintetizar una onda policromática como la suma de ondas monocromáticas, así

$E_{Pol}\left(P_0,t\right)=\int_{\nu} E_0\left(P_0,\nu\right)\exp\left(-2\pi i\nu t\right)d\nu$ será la expresión para una onda policromática, donde $E_0\left(P_0,\nu\right)$ son las soluciones de Rayleigh-Somerfeld o Kirchhoff. Si por ejemplo utilizamos la expresión (6) y, teniendo en cuenta que, $k=\frac{2\pi\nu}{c}=\frac{2\pi}{\lambda}$

$$E_{Pol}\left(P_0,t\right)=-\frac{i}{c}\int_{\nu}\iint_{\Sigma}\nu E_0\left(P_1,\nu\right)\frac{\exp\left(2\pi i\frac{\nu}{c}r_{10}\right)}{r_{10}}\exp\left(-2\pi i\nu t\right)\cos\left(n,\hat{r}_{10}\right)d\nu\, dS$$

Dividiendo y multiplicando por 2π tenemos que

$$E_{Pol}\left(P_0,t\right)=\frac{1}{2\pi c}\iint_{\Sigma}\int_{\nu}-2\pi i\,\nu E_0\left(P_1,\nu\right)\exp\left(-2\pi i\nu\left(t-\frac{r_{10}}{c}\right)\right)d\nu\,\frac{\cos\left(n,\hat{r}_{10}\right)}{r_{10}}dS$$

Ahora bien, la onda que llega a *P1* la podemos escribir como

$E_{Pol}\left(P_1,t\right)=\int_{\nu} E_0\left(P_1,\nu\right)\exp\left(-2\pi i\nu t\right)d\nu$ además

$\frac{\partial}{\partial t}E_{Pol}\left(P_1,t\right)=\int_{\nu}-2\pi i\nu\, E_0\left(P_1,\nu\right)\exp\left(-2\pi i\nu t\right)d\nu$ por lo tanto tendremos que:

$$\boxed{E_{Pol}\left(P_0,t\right)=\frac{1}{2\pi c}\iint_{\Sigma}\frac{\partial}{\partial t}E_{Pol}\left(P_1,t-\frac{r_{10}}{c}\right)\frac{\cos\left(n,\hat{r}_{10}\right)}{r_{10}}dS} \qquad (8)$$

Así vemos que la perturbación en P_0 es linealmente proporcional a la derivada temporal de la perturbación en cada punto P_1 de la abertura. Dado que la onda tarda un tiempo $\frac{r_{10}}{c}$ en llegar de P_1 a P_0 la onda observada depende de la derivada temporal de la incidente, evaluada en $t-\frac{r_{10}}{c}$

Ejercicio 2:

Probar que si a P_1 llega una onda monocromática se recupera la expresión $E_{Mon}\left(P_0,t\right)=E_0\left(P_0,\nu\right)e^{-2\pi i\nu t}$

PRINCIPIO DE BABINET

Supongamos que queremos calcular el campo producido por una pantalla con una abertura de determinada forma Σ en un punto P_0 . Según lo visto, tenemos que:

$$E\left(P_0\right)=\iint_{\Sigma} E\left(P_1\right)\underbrace{\frac{\exp\left(ikr_{10}\right)}{i\lambda r_{10}}\cos\left(n,\hat{r}_{10}\right)}_{f\left(r_{10}\right)}dS=\iint_{\Sigma} E\left(P_1\right)f\left(r_{10}\right)dS$$

El campo siempre lo podemos escribir como

$$E\left(P_1\right)=E\left(P_1\right)\left[1-C\left(P_1\right)\right]+E\left(P_1\right)C\left(P_1\right)$$

donde $C\left(P_1\right)$ es una función que se anula en algunos puntos y vale 1 en otros

$$E\left(P_0\right)=\iint_{\Sigma} E\left(P_1\right)\left[1-C\left(P_1\right)\right]f\left(r_{10}\right)dS+\iint_{\Sigma} E\left(P_1\right)C\left(P_1\right)f\left(r_{10}\right)dS$$

Dado que $C\left(P_1\right)$ es un factor de forma también lo podemos tener en cuenta, en vez de en el integrando, cambiando los límites de integración

$$E\left(P_0\right)=\iint_{\Sigma} E\left(P_1\right)\ f\left(r_{10}\right)dS=\iint_{\Sigma'} E\left(P_1\right)f\left(r_{10}\right)dS+\iint_{\Sigma''} E\left(P_1\right)f\left(r_{10}\right)dS$$

Es evidente, por su definición, que $\left[1-C\left(P_1\right)\right]$ y $C\left(P_1\right)$ son funciones complementarias, por lo tanto Σ' y Σ'' son aberturas complementarias. Vale decir que la distribución de campo que produce una pantalla Σ será igual a la suma de los campos producidos por Σ' y Σ'' . Esto es $E_{\Sigma}=E_{\Sigma'}+E_{\Sigma''}$.

Por ejemplo:

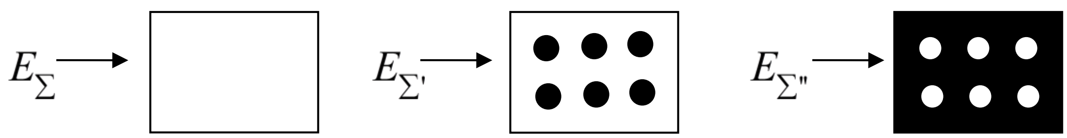


Figura 7: *Pantallas complementarias*

Comentarios:

- Es evidente que si una de las pantallas es opaca, por ejemplo Σ', entonces Σ'' será totalmente transparente y producirá igual campo que Σ
- El principio de Babinet vale punto a punto, por lo tanto si para un cierto P_0 logramos hacer que $E_{\Sigma}(P_0)=0$ entonces $E_{\Sigma'}(P_0)=-E_{\Sigma''}(P_0)$. Es decir, el campo producido por la pantalla Σ' será igual al producido por la pantalla Σ'' pero desfasado en π. Esto implica que la intensidad en P_0 será la misma en ambos casos dado que $I(P_0)=\left|E(P_0)\right|^2$

Esto último puede implementarse de la siguiente forma:

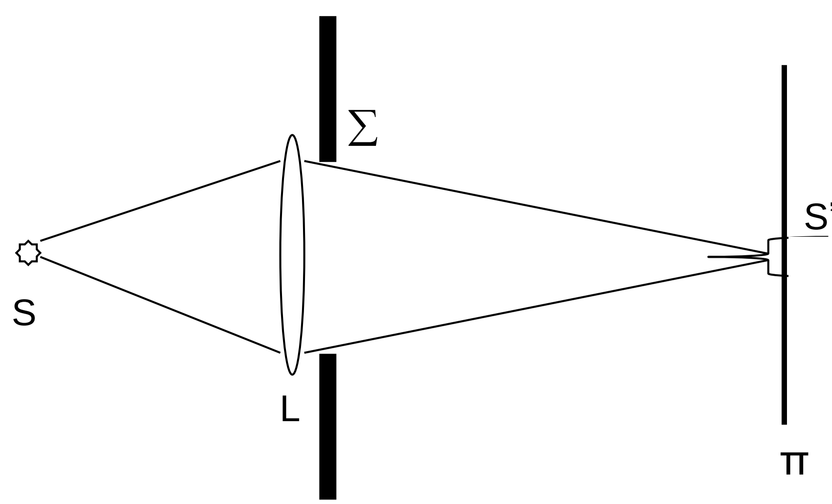


Figura 8: *Disposición para obtener figuras de difracción similares a partir de pantallas complementarias*

Una lente se ubica entre la fuente puntual y el plano de observación π. La imagen de S sobre π no será estrictamente un punto pero sí una función picuda, vale decir que $\forall P_0 \neq S' \Rightarrow E(P_0) = 0$ Luego $E_{\Sigma'}(P_0) = e^{i\pi} E_{\Sigma''}(P_0) \Rightarrow I_{\Sigma'}(P_0) = I_{\Sigma''}(P_0)$.

Pero esta configuración (como lo veremos próximamente) provee ni más ni menos que la difracción de Fraunhofer de ∑, esto es su transformada de Fourier. Si queremos analizar este fenómeno, por el momento, con una configuración más tradicional (pero que es un caso particular de la descripta) podemos iluminar a la lente con un haz plano (fuente puntual en el infinito) y sobre el plano focal de la misma obtendremos los diagramas de difracción de Fraunhofer correspondientes a ∑' y ∑'' que serán idénticos, salvo en el punto imagen de la fuente. Sintetizando, podemos decir que *pantallas complementarias producen el mismo diagrama de difracción de Fraunhofer*.

DIFRACCIÓN DE FRESNEL Y FRAUNHOFER

Aproximaciones iniciales

Hasta ahora el tratamiento, si bien escalar, fue totalmente general. Vamos a ver algunas aproximaciones adicionales que permiten calcular figuras de difracción con métodos matemáticos menos complejos.

Habíamos visto que para ondas monocromáticas, y según la teoría de Rayleigh – Sommerfeld, la expresión para el campo difractado en un punto P_0 era

$$E(P_0) = \frac{1}{i\lambda} \iint_{\Sigma} E(P_1) \boxed{\frac{\exp\left(i \frac{2\pi}{\lambda} r_{10}\right)}{r_{10}}} \cos\left(n, \hat{r}_{10}\right) dS \qquad (6)$$

Dado que $r_{10} \gg \lambda$ la función recuadrada es rápidamente variable en la fase y lentamente variable en amplitud. Por ejemplo si r_{10} cambia en una longitud de onda, la fase varía en 2π mientras que la amplitud casi no cambia. Así deberemos tener cuidado con qué clase de aproximaciones tomamos para una y otra magnitud. Veamos el siguiente dibujo

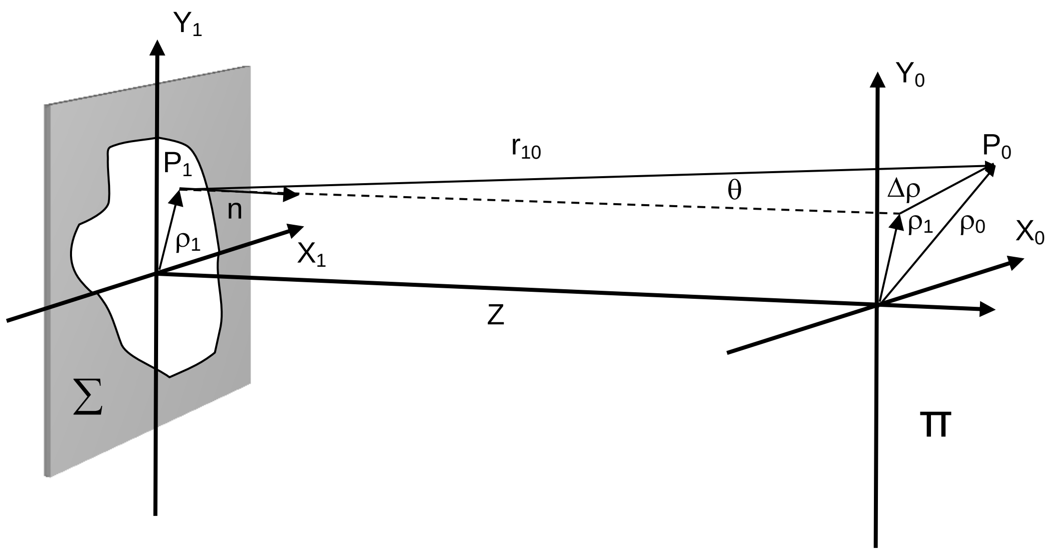


Figura 9: *Geometría empleada para analizar la difracción por una pantalla plana*

Asumiremos que
$$\rho_{Max\sum} = \sqrt{x_1^2 + y_1^2} \ll z$$
$$\rho_{Max\prod} = \sqrt{x_0^2 + y_0^2} \ll z$$

Esta es la aproximación paraxial ya que los ángulos que intervienen son pequeños. Si observamos el dibujo vemos que

$$\cos(\theta) = \frac{z}{\sqrt{z^2 + \Delta\rho^2}} \quad ; \quad \Delta\rho = \left|\vec{\rho_0} - \vec{\rho_1}\right|$$

$$\cos(\theta) = \frac{1}{\sqrt{1 + \left(\frac{\Delta\rho}{z}\right)^2}} \cong 1 - \frac{1}{2}\left(\frac{\Delta\rho}{z}\right)^2$$

Veamos que sucede si tomamos *cos(θ)=1*, esto es despreciamos el factor de oblicuidad.

En este caso cometemos un error $\frac{1}{2}\left(\frac{\Delta\rho}{z}\right)^2 \leq \varepsilon \;\Rightarrow\; \frac{\Delta\rho}{z} \leq 1.41\sqrt{\varepsilon}$ pero

$$\frac{\Delta\rho}{z} = \tan(\theta) \leq \theta \leq 1.41\sqrt{\varepsilon}$$

Así si deseamos trabajar con un error porcentual del 1% $\Rightarrow \theta \leq 1.41 \cdot 0.1 \cong 8º$, esto es podemos trabajar con un cono de 16º. Si en cambio admitimos un error del 5% podemos trabajar con un cono de 36º.

Analicemos ahora que sucede con el factor $\dfrac{1}{r_{10}}$

$$r_{10} = \left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2 + z^2\right]^{\frac{1}{2}} = z\sqrt{1+u} \qquad \text{con}$$

$$u = \left(\frac{x_0 - x_1}{z}\right)^2 + \left(\frac{y_0 - y_1}{z}\right)^2 \ll 1 \qquad \Rightarrow$$

$\dfrac{1}{r_{10}} = \dfrac{1}{z\sqrt{1+u}} \simeq \dfrac{1}{z}\left(1 - \dfrac{u}{2} + \dfrac{u^2}{8} \cdots\cdots\right)$, basta tomar para la amplitud $\dfrac{1}{r_{10}} \simeq \dfrac{1}{z}$

Veamos ahora que sucede con las fases. Como dijimos, la fase es una función de variación rápida por lo que tendremos que tomar órdenes superiores en los desarrollos en serie. Del tipo de aproximación que tomemos surgirán los tratamientos de Fresnel y Fraunhofer.

Aproximación de Fresnel

Primero desarrollemos en serie r_{10}

$r_{10} \simeq z\left(1 + \dfrac{u}{2} - \dfrac{u^2}{8} \cdots\cdots\right)$ Vamos a quedarnos, en principio, con el desarrollo hasta el término lineal, despreciando el cuadrático y superiores. Veamos en qué condiciones esto es válido.

En realidad estamos aproximando una onda esférica $\dfrac{e^{ikr_{10}}}{r_{10}}$ por superficies cuadráticas

$$\frac{\exp(ik\,z)}{z}\exp\left[\frac{ik}{2z}\left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2\right]\right]$$

El error que se comete al despreciar el orden inmediato superior, esto es $-z\frac{u^2}{8}$, significa que estamos considerando $e^{-\frac{ikzu^2}{8}} \cong 1$. En general se toma como razonable que $kz\frac{u^2}{8} \ll 1 radian \Rightarrow \quad z^3 \gg \frac{\pi}{4\lambda}\left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2\right]^2$.

Si le asignamos valores a esta desigualdad vemos que, por ejemplo, para una abertura y zona de observación del orden de 1mm y una longitud de onda de 500nm, obtendríamos mediciones precisas para z >> 20cm. Sin embargo se observa experimentalmente que esta aproximación es válida a distancias mucho menores.

Veamos entonces cuál es la explicación, recurramos para ello al método de la fase estacionaria. Tenemos una integral que es del tipo

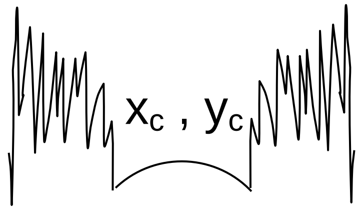
$$\iint E\left(x_1, y_1\right) \exp\left(ik\ f\left(x_1, y_1\right)\right) dx_1\ dy_1; \quad k = \frac{2\pi}{\lambda}; \quad f\left(x_1, y_1\right) = \frac{1}{2z}\left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2\right]$$

E y f son independientes de k y f >> λ. Quiere decir que la exponencial oscilará muchísimo y en promedio dará cero salvo en aquellas zonas donde la oscilación sea más lenta. Veamos si para la función $f\left(x_1, y_1\right)$ existen dos puntos críticos $\left(x_c, y_c\right)$ tales que sus derivadas parciales se anulen, esto es:

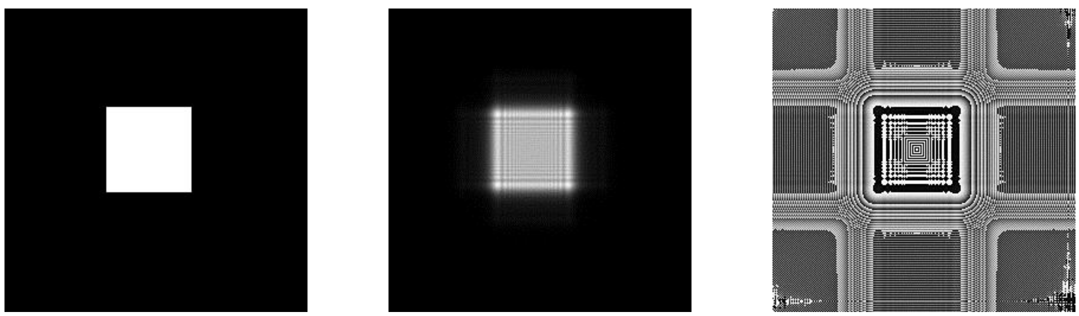
$$\left.\frac{\partial f\left(x_1, y_1\right)}{\partial x_1}\right)_{x_c, y_c} = -\frac{1}{z}\left.\left(x_0 - x_1\right)\right)_{x_c, y_c} = 0 \quad \Rightarrow \quad x_c = x_1 = x_0$$

$$\left.\frac{\partial f\left(x_1, y_1\right)}{\partial y_1}\right)_{x_c, y_c} = -\frac{1}{z}\left.\left(y_0 - y_1\right)\right)_{x_c, y_c} = 0 \quad \Rightarrow \quad y_c = y_1 = y_0$$

Vemos que para $f\left(x_1, y_1\right)$ existen tales puntos críticos y que en ellos su derivada es nula. Es decir que alrededor de $\left(x_0, y_0\right)$, $f\left(x_1, y_1\right)$ es constante o levemente variable y por lo tanto el término de fase oscilará menos y su contribución promedio será distinta de cero. Esto no ocurre para el resto de la función (ver JOSA **71**, 7, 1981)



Se podría argumentar que también en los extremos de $f\left(x_1, y_1\right)$ la derivada es nula, pero la función es tan fluctuante y apretada que en los extremos la derivada prácticamente no está definida. Entonces vemos que el término lineal en u, esto es cuadrático en $\left(x, y\right)$, contribuye sólo en las inmediaciones de $x_1 = x_0$ y $y_1 = y_0$ es decir en aquellos puntos que se hallan fuera de la sombra geométrica. Veamos esto en la figura siguiente



objeto módulo difracción fase difracción

En la zona dentro de la sombra geométrica la amplitud es casi nula y la fase varía rápidamente. Se puede demostrar que los términos de orden superior no contribuyen a la integral, ni aún en las proximidades de $x_1 = x_0$ y $y_1 = y_0$. Esto nos dice que no hace falta pedir $e^{-\frac{ikzu^2}{8}} \cong 1$ y de ahí la condición sobre Z ya que al integrar ese término no aporta.

Así pues, el campo $E\left(P_0\right)$ en la aproximación de Fresnel vendrá dado por

$$\boxed{E\left(P_0\right) = \frac{\exp\left(ik\,z\right)}{i\lambda z} \iint_{\Sigma} E\left(x_1, y_1\right) \exp\left(i\,\frac{k}{2z}\left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2\right]\right) dx_1\, dy_1} \qquad (9)$$

Aproximación de Fraunhofer

La ecuación anterior la podemos escribir como

$$E\left(x_0,y_0\right)=\overbrace{\frac{\exp\left(ik\,z\right)}{i\lambda z}\exp\left(i\,\frac{k}{2z}\left(x_0^2+y_0^2\right)\right)}^{\mathbf{1}}.$$

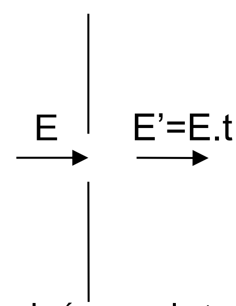
$$\underbrace{.\iint_{\Sigma} E\left(x_1,y_1\right)\exp\left(-i\,\frac{k}{z}\left[x_0x_1+y_0y_1\right]\right)\exp\left(i\,\frac{k}{2z}\left(x_1^2+y_1^2\right)\right)dx_1\,dy_1}_{\mathbf{2}}$$

El término **2** es la transformada de Fourier de $E\left(x_1,y_1\right)$ a menos del factor $e^{i\frac{k}{2z}\left(x_1^2+y_1^2\right)}$.

Veamos en qué condiciones es posible eliminarlo.

Si $e^{i\frac{k}{2z}\left(x_1^2+y_1^2\right)}\cong 1 \;\Rightarrow\; z\gg\frac{k}{2}\left(x_1^2+y_1^2\right)_{Max}$ Pero justamente esta es la condición de difracción de Fraunhofer. Esto es, la figura de difracción de Fraunhofer, a menos del factor **1** es la transformada de Fourier del campo en la abertura Σ . Ahora bien, el factor **1**, a menos de la atenuación 1/λz es básicamente un factor global de fase. Dado que lo que se registra es la intensidad (I=E.E*) esa fase se elimina y la figura que se obtiene coincidirá con el módulo de la transformada de Fourier.

Veremos que esta no es la única manera de obtener $\mathbb{F}\left(E\right)$. Por ahora recordemos que la clase pasada dijimos que si en lugar de tener una abertura, la misma tiene un factor de transmisión $t\left(x_1,y_1\right)$ y por otra parte el campo que llega a la misma es $E\left(x_1,y_1\right)$, entonces el campo inmediatamente emergente será $E'\left(x_1,y_1\right)=E\left(x_1,y_1\right)\cdot t\left(x_1,y_1\right)$ y ese será el que debe ir en la integral para calcular $E\left(P_0\right)$



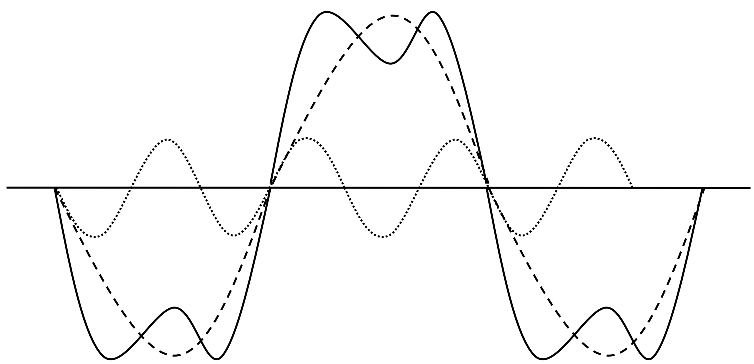
Ahora bien, $t\left(x_1,y_1\right)$ en principio es algo que puede ser sólo de fase, sólo de amplitud o combinar ambas magnitudes. Entonces si en dicha abertura ubicásemos un filtro con un factor de transmisión $t\left(x_1,y_1\right)=e^{-i\frac{k}{2z}\left(x_1^2+y_1^2\right)}$, dicho factor se cancelaría con el otro y no necesitaríamos imponer restricciones sobre z para obtener la transformada de Fourier. Veremos más adelante que una lente convergente de distancia focal f es justamente el filtro que introduce el desfasaje $e^{-i\frac{k}{2f}\left(x_1^2+y_1^2\right)}$.

II. SERIES E INTEGRALES DE FOURIER CLASE 3

Habiendo llegado a este punto es conveniente hacer un paréntesis, en lo que a óptica se refiere, para dedicarnos a realizar una revisión de las definiciones y principales propiedades matemáticas de las series e integrales de Fourier.

FUNCIONES PERIÓDICAS

Para comenzar veamos que sucede con una función periódica arbitraria. Supongamos que sumamos dos funciones armónicas (senos o cosenos) de distinta amplitud y período.



Vemos que tenemos una función periódica pero no senoidal. Es decir que si tenemos una función arbitraria periódica, la podríamos sintetizar mediante la suma de senos y cosenos con amplitudes, períodos y fases relativas adecuadas. Justamente esto es en lo que se basa la teoría desarrollada por el físico francés Jean Baptiste Joseph Fourier (1768-1830). Dicho de otra forma, Fourier elige una base de senos y cosenos para expandir una función periódica. También pueden elegirse otras bases, por ejemplo la constituida por funciones rectángulo, que conduce a las transformadas de Hadamard.

Volviendo al caso que nos ocupa, el teorema de Fourier establece que si $f(x)$ es una función de período espacial λ_0 con $k_0 = \frac{2\pi}{\lambda_0}$ (si fuese una función temporal en período sería T y $k_0 = \frac{2\pi}{T_0}$) entonces

$$\boxed{f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos(nk_0 x) + b_n \sin(nk_0 x)\right]} \qquad (10)$$

donde $\cos\left(k_0 x\right);\ sen\left(k_0 x\right)$ constituyen las componentes fundamentales. Luego, dado que $nk_0 = n\frac{2\pi}{\lambda_0} = \frac{2\pi}{\lambda_0 / n}$ $\quad n \in \mathbb{N}$, los armónicos superiores serán ondas cuyo período será una fracción del período de la función.

Ahora debemos encontrar las expresiones para los a_n y b_n. Para ello recurramos a la propiedad de ortogonalidad de las funciones armónicas

$$\int_0^{\lambda_0} \sin\left(nk_0 x\right)\cos\left(mk_0 x\right)dx = 0$$

$$\int_0^{\lambda_0} \cos\left(nk_0 x\right)\cos\left(mk_0 x\right)dx = \frac{\lambda_0}{2}\delta nm \qquad \delta nm = \begin{matrix} 1\ si\ n = m \\ 0\ si\ n \neq m \end{matrix}$$

$$\int_0^{\lambda_0} \sin\left(nk_0 x\right)\sin\left(mk_0 x\right)dx = \frac{\lambda_0}{2}\delta nm$$

δnm es la delta de Kronecker

Entonces multipliquemos (10) por $\cos\left(mk_0 x\right)$ e integremos

$$\int_0^{\lambda_0} f\left(x\right)\cos\left(mk_0 x\right)dx = \frac{a_0}{2}\int_0^{\lambda_0} \cos\left(mk_0 x\right)dx +$$
$$+\sum_{n=1}^{\infty}\left[a_n \int_0^{\lambda_0} \cos\left(mk_0 x\right)\cos\left(nk_0 x\right)dx + b_n \int_0^{\lambda_0} \cos\left(mk_0 x\right)\ sen\left(nk_0 x\right)dx\right]$$

$$\int_0^{\lambda_0} f\left(x\right)\cos\left(mk_0 x\right)dx = \frac{a_m\,\lambda_0}{2} \quad \Rightarrow \quad a_m = \frac{2}{\lambda_0}\int_0^{\lambda_0} f\left(x\right)\cos\left(mk_0 x\right)dx \qquad (11)$$

m=0,1,2…

La expresión (11) también es válida para *a₀* y el término $\frac{a_0}{2} = \frac{1}{\lambda_0}\int_0^{\lambda_0} f\left(x\right)dx$ equivale al promedio de la función en un período.

Multiplicando (10) por $sen\left(mk_0 x\right)$ e integrando, obtenemos

$$b_m = \frac{2}{\lambda_0}\int_0^{\lambda_0} f\left(x\right)\sin\left(mk_0 x\right)dx \qquad (12)$$

m=1,2…

Cabe destacar que la integración puede realizarse sobre cualquier intervalo espacial igual a λ_0. En general entre x' y $x'+\lambda_0$

Desarrollo de $f\left(x\right)$ como una serie finita

Vimos que una función periódica $f\left(x\right)$ se puede sintetizar como una sumatoria infinita de senos y cosenos. Podemos preguntarnos cuál es el error que se comete si no tomamos infinitos términos. Esto es

$$f\left(x\right)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left[a_n\cos\left(nk_0 x\right)+b_n\sin\left(nk_0 x\right)\right]=\frac{a_0}{2}+\sum_{n=1}^{N}+\xi_N\left(x\right)$$

Si definimos $\sigma_N^2\left(x\right)$ como la desviación cuadrática media

$$\sigma_N^2\left(x\right)=\frac{\int_0^{\lambda_0}\xi_N^2\left(x\right)}{\lambda_0}$$

Puede demostrarse que:

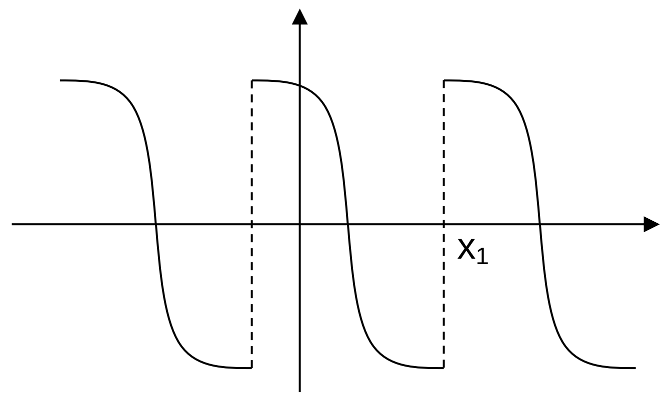
$$\sigma_N^2\left(x\right)=\frac{\int_0^{\lambda_0} f^2\left(x\right)dx}{\lambda_0}-\frac{a_0^2}{4}-\frac{1}{2}\sum_{n=1}^{N}\left(a_n^2+b_n^2\right)$$

Condiciones de existencia y propiedades

Con anterioridad dijimos que $f\left(x\right)$ era una función periódica arbitraria, sin embargo debe cumplir algunas propiedades para que pueda existir su síntesis de Fourier.

i) Si $\exists$ discontinuidades de $f\left(x\right)$ en un período, estas deben ser un número finito y con valores finitos.

Supongamos que en x_1 hay una discontinuidad. Entonces la serie de Fourier en x_1 toma el valor $f\left(x_1\right)=\dfrac{f\left(x_1^-\right)+f\left(x_1^+\right)}{2}$



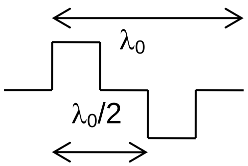
ii) $f\left(x\right)$ tiene que tener un número finito de máximos y mínimos en un período

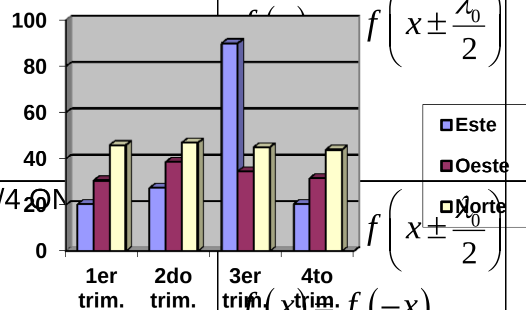
iii) $\int_0^{\lambda_0}\left|f\left(x\right)\right|dx<\infty$; $f\left(x\right)$ debe ser de módulo integrable ya que es proporcional a *a_0*

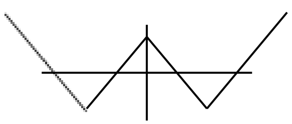
Sea $f\left(x\right)$ periódica y tal que $f\left(x\right)=\dfrac{a_0}{2}+\sum_{n=1}^{\infty}$. Si además $f'\left(x\right)\equiv\dfrac{\partial f\left(x\right)}{\partial x}$ es también continua por tramos entonces

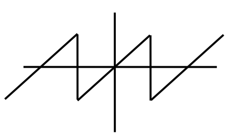
$$f'\left(x\right)=\sum_{n=1}^{\infty}nk_0\left[-a_n\sin\left(nk_0x\right)+b_n\cos\left(nk_0x\right)\right]$$

Existen algunas consideraciones de simetría que deben realizarse antes de iniciar el desarrollo de la función a fin de ahorrar esfuerzos en el cálculo de los coeficientes a_n y b_n. En el siguiente cuadro resumimos algunas de estas propiedades.









SIMETRÍA	CONDICIONES	FORMA DE LA SERIE	COEFICIENTES
PAR	$f(x) = f(-x)$	$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n k_0 x)$	$a_n = \frac{4}{\lambda_0} \int_0^{\lambda_0/2} f(x) \cos(n k_0 x) dx$ $b_n = 0$
IMPAR	$f(x) = -f(-x)$	$f(x) = \sum_{n=1}^{\infty} b_n \, sen(n k_0 x)$	$a_n = 0$ $b_n = \frac{4}{\lambda_0} \int_0^{\lambda_0/2} f(x) \, sen(n k_0 x) dx$
1/2 ONDA	$f(x) = -f\left(x \pm \frac{\lambda_0}{2}\right)$	$f(x) = \sum_{n=1}^{\infty} \left[a_{2n-1} \cos((2n-1)k_0 x) + b_{2n-1} \, sen((2n-1)k_0 x) \right]$	$a_{2n-1} = \frac{4}{\lambda_0} \int_0^{\lambda_0/2} f(x) \cos((2n-1)k_0 x) dx$ $b_{2n-1} = \frac{4}{\lambda_0} \int_0^{\lambda_0/2} f(x) \, sen((2n-1)k_0 x) dx$
1/4 ON	$f\left(x \pm \frac{\lambda_0}{2}\right)$ $f(x) = f(-x)$	$f(x) = \sum_{n=1}^{\infty} a_{2n-1} \cos((2n-1)k_0 x)$	$a_{2n-1} = \frac{8}{\lambda_0} \int_0^{\lambda_0/4} f(x) \cos((2n-1)k_0 x) dx$ $b_{2n-1} = 0$
1/4 ONDA IMPAR	$f(x) = -f\left(x \pm \frac{\lambda_0}{2}\right)$ $f(x) = -f(-x)$	$f(x) = \sum_{n=1}^{\infty} b_{2n-1} \, sen((2n-1)k_0 x)$	$a_{2n-1} = 0$ $b_{2n-1} = \frac{8}{\lambda_0} \int_0^{\lambda_0/4} f(x) \, sen((2n-1)k_0 x) dx$

Forma compleja

Otra forma de representar las series de Fourier es la forma compleja. Sabemos que:

$$\cos(k_0 nx) = \frac{e^{ik_0 nx} + e^{-ik_0 nx}}{2} \quad ; \quad \sin(k_0 nx) = \frac{e^{ik_0 nx} - e^{-ik_0 nx}}{2i}$$

con lo cual podemos expresar

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[\frac{(a_n - ib_n)}{2} e^{ik_0 nx} + \frac{(a_n + ib_n)}{2} e^{-ik_0 nx} \right] \qquad \text{ó}$$

$$\boxed{f(x) = \sum_{n=-\infty}^{\infty} C_n e^{ik_0 nx}} \qquad (13) \qquad \text{con} \quad C_n \equiv \frac{(a_n - ib_n)}{2}, \quad C_{-n} \equiv \frac{(a_n + ib_n)}{2}$$

Usando las propiedades de ortogonalidad también se puede llegar (usar las definiciones de a_n y b_n) a que

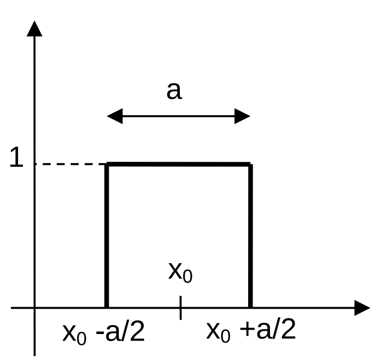
$$C_{\pm n} = \frac{1}{\lambda} \int_0^{\lambda_0} f(x) e^{\mp ink_0 x} dx \qquad (14)$$

Si $f(x) \in \mathbb{R} \quad \Rightarrow \quad C_{-n} = C_n^*$

Ejemplos de cálculo de series de Fourier

Antes de entrar en los ejemplos es conveniente definir la nomenclatura de algunas funciones muy usadas

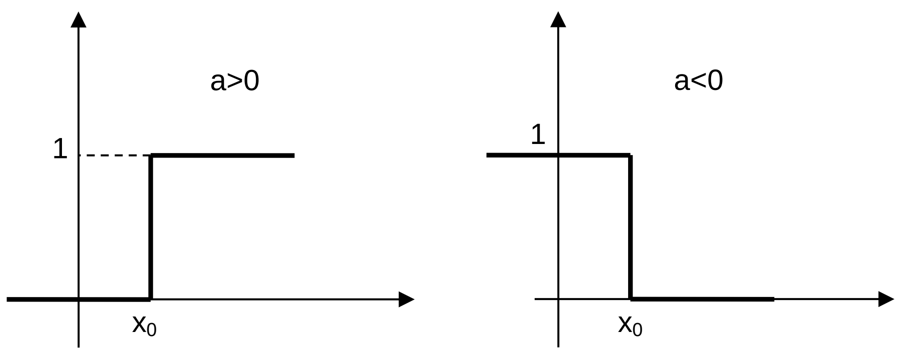
- <u>Función rectángulo</u>



$$\mathrm{rect}\left(\frac{x-x_0}{a}\right)=\begin{cases} 0 & \dfrac{|x-x_0|}{a}>\dfrac{1}{2} \\[2ex] \dfrac{1}{2} & \dfrac{|x-x_0|}{a}=\dfrac{1}{2} \\[2ex] 1 & \dfrac{|x-x_0|}{a}<\dfrac{1}{2} \end{cases} \quad *$$

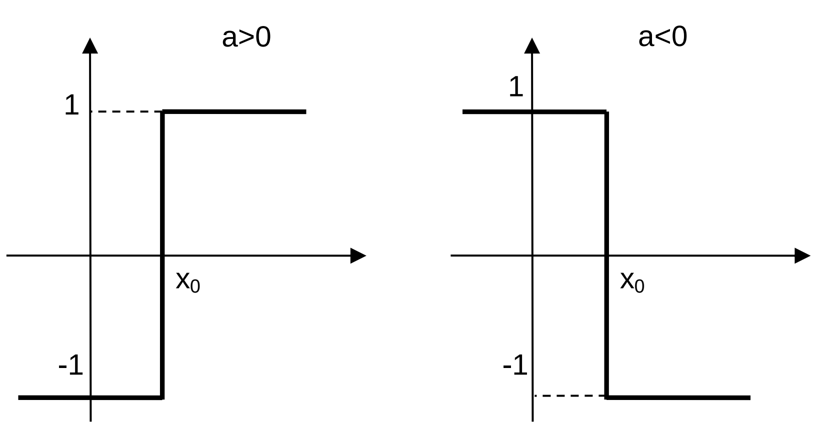
* recordemos que un desarrollo de Fourier da el valor medio de la función en un punto de discontinuidad

- <u>Función escalón</u>



$$\mathrm{step}\left(\frac{x-x_0}{a}\right)=\begin{cases} 0 & \dfrac{(x-x_0)}{a}<0 \\[2ex] \dfrac{1}{2} & x-x_0=0 \\[2ex] 1 & \dfrac{(x-x_0)}{a}>0 \end{cases} \qquad a=1 \, o -1$$

- <u>Función signo</u>

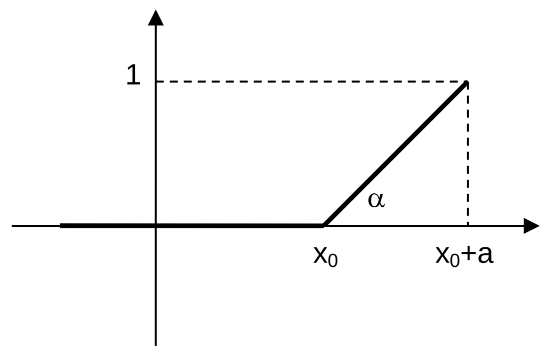


$$\operatorname{sign}\left(\frac{x-x_0}{a}\right)=\begin{cases}-1 & \dfrac{(x-x_0)}{a}<0\\ 0 & x-x_0=0\\ 1 & \dfrac{(x-x_0)}{a}>0\end{cases}\qquad a=1\,o\,-1$$

$$\operatorname{sign}\left(\frac{x-x_0}{a}\right)=2\operatorname{step}\left(\frac{x-x_0}{a}\right)-1$$

- <u>Función rampa</u>

$$\operatorname{ramp}\left(\frac{x-x_0}{a}\right)=\begin{cases}0 & x\le x_0\\ \dfrac{|x-x_0|}{a} & x>x_0\end{cases}$$

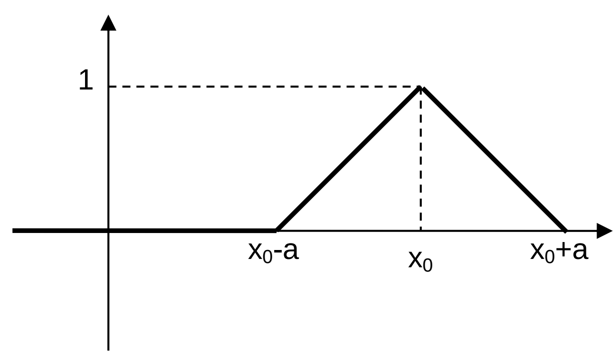


$$\tan(\alpha)=\frac{1}{a}$$

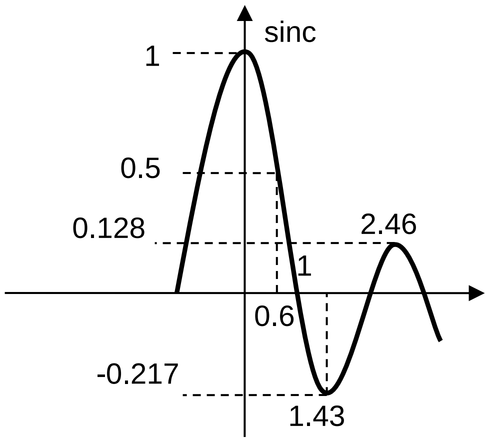
- <u>Función triángulo</u>

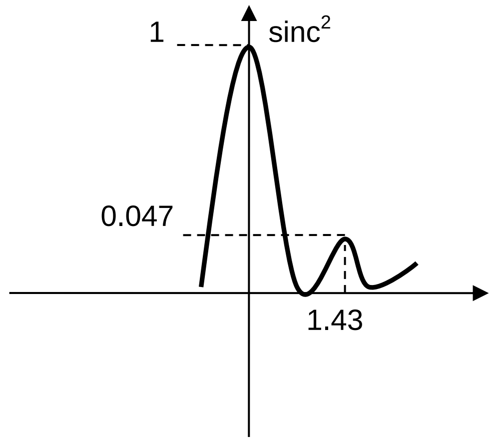
$$\operatorname{tri}\left(\frac{x-x_0}{a}\right)=\begin{cases}0 & \dfrac{|x-x_0|}{a}\ge 1\\ 1-\dfrac{|x-x_0|}{a} & \dfrac{|x-x_0|}{a}<1\end{cases}$$

$$\acute{a}rea=|a|$$



- Función sinc

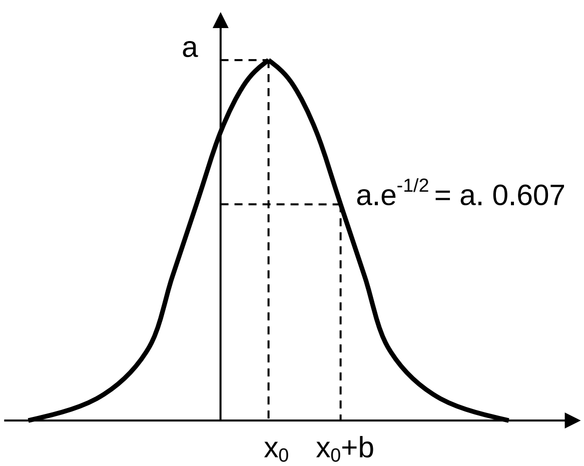




$$\mathrm{sinc}\left(\frac{x-x_0}{a}\right)=\frac{sen\left[\pi\left(\frac{x-x_0}{a}\right)\right]}{\pi\left(\frac{x-x_0}{a}\right)}$$

x_0 nos dice dónde está centrada y |a| es el área

- Función gaussiana

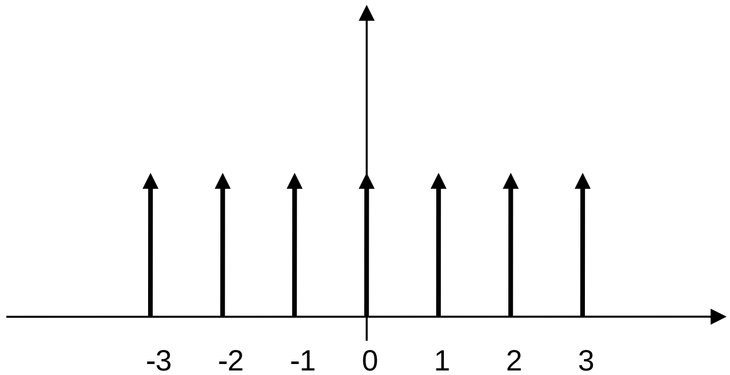


$$\mathrm{gaus}\left(\frac{x-x_0}{b}\right)=a\,e^{-\frac{(x-x_0)^2}{2b^2}}$$

$$\int_{-\infty}^{\infty} a\,e^{-\frac{(x-x_0)^2}{2b^2}}\,dx=ab\sqrt{2\pi}$$

a, x_0 y b pertenecen a los reales. El valor de la integral es 1 si $a=\frac{1}{b\sqrt{2\pi}}$, en cuyo caso la función gaussiana es la función densidad de una variable aleatoria con distribución normal de media x_0 y varianza *b*.

- <u>Función peine de Dirac</u>

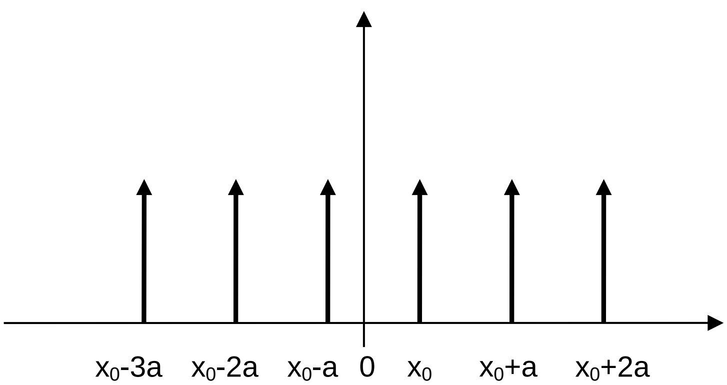


$$\mathrm{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n)$$

$$\delta(x-n) = \begin{cases} \infty & x = n \\ 0 & x \neq n \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)\delta(x-n)dx = f(n)$$

donde $\delta(x-n)$ es la delta de Dirac. Si queremos desplazarlo y cambiarle el paso

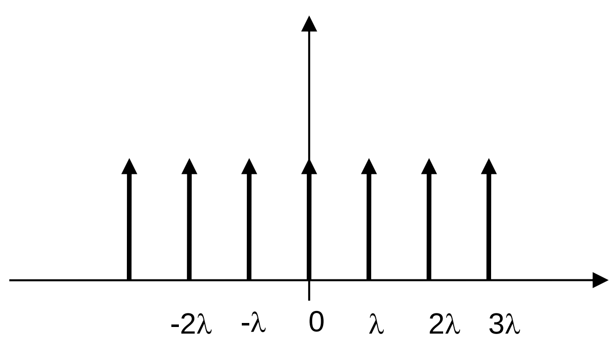


$$\mathrm{comb}\left(\frac{x-x_0}{a}\right) = |a| \sum_{n=-\infty}^{\infty} \delta(x-x_0-na)$$

Vayamos ahora a algunos ejemplos de cálculo de transformadas. Comencemos con uno simple

1. $f(x) = \sum_{n=-\infty}^{\infty} \delta(x-n\lambda)$

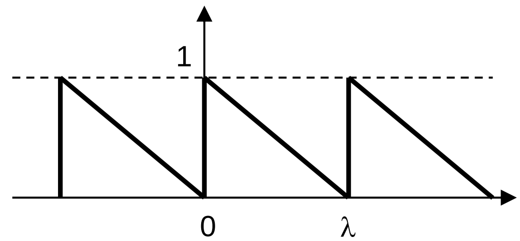
es una función par $\Rightarrow$ $b_n = 0$



$$a_n = \frac{2}{\lambda}\int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} \delta(x-n'\lambda)\cos(n k x)dx = \frac{2}{\lambda}\cos(nn'k\lambda) = \frac{2}{\lambda}$$

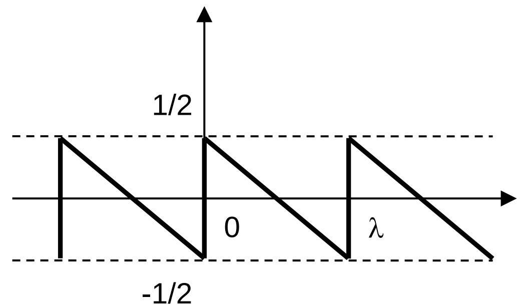
$$\Rightarrow \quad f(x) = \frac{1}{\lambda} + \frac{2}{\lambda}\sum_{n=1}^{\infty}\cos(n k x)$$

2. Veamos otro ejemplo $f(x) = 1 - \frac{x}{\lambda} \quad 0 \le x \le \lambda$



Conviene redefinir la función para buscar una simetría

$g(x) = \frac{1}{2} - \frac{x}{\lambda} \quad 0 \le x \le \lambda$



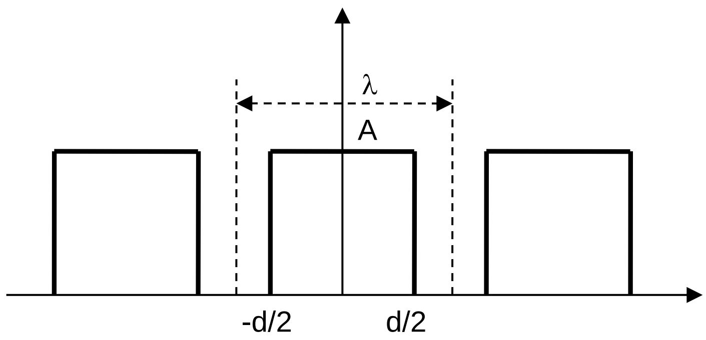
Ahora tenemos una función impar g(x) = -g(-x) ⇒ $a_n = 0$

$$b_n = \frac{4}{\lambda}\int_0^{\lambda/2}\left(\frac{1}{2}-\frac{x}{\lambda}\right)sen(n k x)dx = \frac{4}{\lambda}\left[-\frac{\cos(n k x)}{2nk}\Bigg|_0^{\lambda/2} - \int_0^{\lambda/2}\frac{x}{\lambda}sen(n k x)dx\right] =$$

$$= \frac{4}{\lambda}\left\{\frac{(-1)^{n+1}+1}{2nk} - \left[-\frac{x}{\lambda}\frac{\cos(nkx)}{nk} + \frac{1}{\lambda}\frac{sen(nkx)}{n^2k^2}\right]\Bigg|_0^{\lambda/2}\right\} = \frac{1}{n\pi}\left[-1^n+1\right] - \frac{1}{n\pi}(-1)^n = \frac{1}{n\pi}$$

$$g(x) = \sum_{n=1}^{\infty}\frac{1}{n\pi}sen(nkx) \quad \Rightarrow \quad f(x) = \frac{1}{2} + \sum_{n=1}^{\infty}\frac{1}{n\pi}sen(nkx)$$

3. Veamos ahora el siguiente ejemplo $f(x) = \begin{cases} A & -\frac{d}{2} \le x \le \frac{d}{2} \\ 0 & \end{cases}$



La función es par $\Rightarrow$ $b_n = 0$;

$$a_0 = \frac{2}{\lambda}\int_{-d/2}^{d/2} A dx = \frac{2Ad}{\lambda}$$

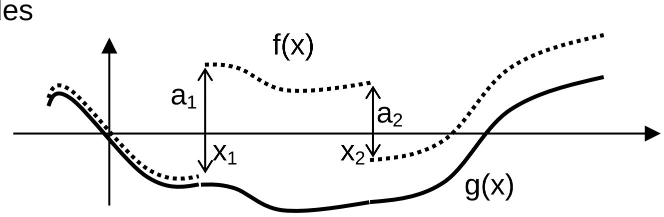
$$a_n = \frac{2}{\lambda}\int_{-d/2}^{d/2} A\cos(nkx)dx = \frac{2A}{\lambda nk}\sin(nkx)\Big|_{-d/2}^{d/2} = \frac{2A}{n\pi}\sin\left(nk\frac{d}{2}\right) = \frac{2dA}{\lambda}\frac{\sin\left(nk\frac{d}{2}\right)}{nk\frac{d}{2}}$$

$$f(x) = \frac{Ad}{\lambda} + \frac{2Ad}{\lambda}\sum_{n=1}^{\infty}\frac{\sin\left(\frac{n\pi d}{\lambda}\right)}{\frac{n\pi d}{\lambda}}\cos\left(\frac{2n\pi x}{\lambda}\right)$$

Vemos que los coeficientes a_n consisten en la clásica función *sinc* que, si recordamos de óptica, está relacionada con la difracción de Fraunhofer por una ranura. Habíamos visto con anterioridad que la difracción en campo lejano correspondía a la transformada de Fourier de la abertura. Luego volveremos sobre este tema.

Ejercicio 3: Llegar al mismo resultado que encontramos usando la siguiente propiedad

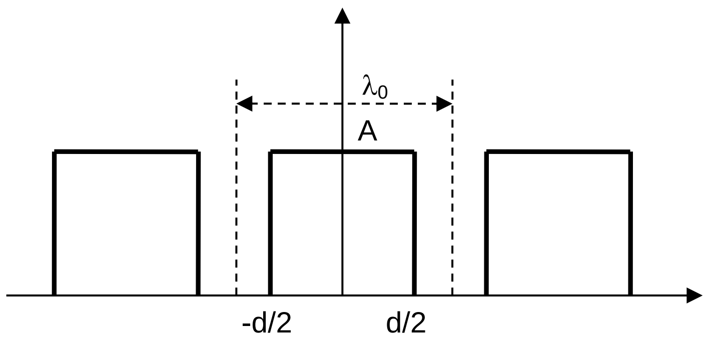
Sea $f(x)$ una función continua por tramos, con discontinuidades en los puntos x_j. Además su derivada $f'(x)$ está definida en todas partes excepto en dichas discontinuidades; entonces $f'(x) = g'(x) + \sum_j a_j \delta(x - x_j)$. Donde $g(x)$ es $f(x)$ sin las discontinuidades



CLASE 4

FUNCIONES NO PERIÓDICAS

Hasta ahora habíamos analizado como representar funciones periódicas, veamos cómo se puede hacer un desarrollo para aquellas que no lo son. Para ello comencemos con la siguiente función que vimos con anterioridad



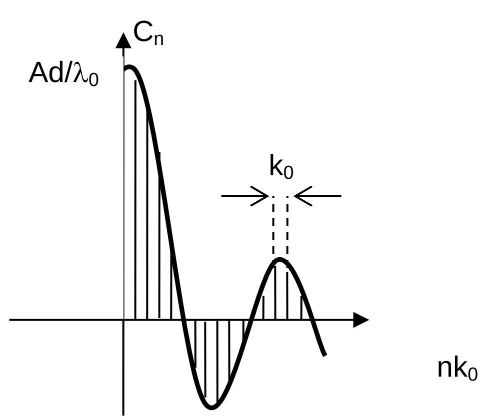
$$f(x)=\begin{cases} A & -\frac{d}{2}\le x\le \frac{d}{2} \\ 0 & \end{cases}$$

Vamos a escribir su desarrollo en serie en la forma compleja, de modo que

$$f(x)=\sum_{n=-\infty}^{\infty} C_n e^{ink_0 x} \qquad ; \qquad k_0 = 2\pi/\lambda_0$$

$$C_n = \frac{1}{\lambda_0}\int_{-\lambda_0/2}^{\lambda_0/2} f(x)\exp(-ink_0 x)\,dx = \frac{Ad}{\lambda_0}\frac{\sin\left(\frac{nk_0 d}{2}\right)}{\frac{nk_0 d}{2}}$$

Si hacemos el gráfico de los C_n tenemos



La separación entre armónicos es $\Delta k = \left|k_n - k_{n-1}\right| = \left|nk_0 - (n-1)k_0\right| = k_0$.

Ahora bien, si queremos una función no periódica podemos, por ejemplo, hacer tender el período espacial λ_0 a ∞. Esto es, en el gráfico sólo aparecerá la parte central, dejando de ser la función periódica anterior.

En este caso $\Delta k = k_0 = \dfrac{2\pi}{\lambda_0} \xrightarrow[\lambda_0 \to \infty]{} 0 \Rightarrow \Delta k \to dk$ y $k_n = nk_0 \to k$ pasa a ser una variable continua.

Pensemos en un análogo óptico en el que tenemos una red de difracción constituida por bandas oscuras y transparentes como la de la figura

Los órdenes difractados por esa red estarán separados de acuerdo a la frecuencia espacial de la misma, cuanto más alta más separados. La cantidad de luz que va a cada orden estará modulada por la campana de difracción, que corresponderá a la de una sola ranura. A medida que vamos disminuyendo esa frecuencia espacial, esto es aumentando el período, dichos órdenes se irán juntando. En el límite nos quedaríamos con la luz difractada por una sola ranura que coincide con la envolvente que antes modulaba los órdenes. Ahora los órdenes están tan juntos que constituyen un continuo dentro de esa campana.

Volvamos al caso que nos ocupa. Para ver que sucede en este límite escribamos nuevamente $f(x)$ con los coeficientes C_n expresados explícitamente.

$$f(x)=\sum_{n=-\infty}^{\infty} C_n \exp(i\,nk_0\,x)=\sum_{n=-\infty}^{\infty}\left[\frac{\Delta k}{2\pi}\int_{-\lambda_0/2}^{\lambda_0/2} f(x')\exp(-i\,nk_0\,x')\,dx'\right]\exp(i\,nk_0\,x)$$

recordemos que $\frac{\Delta k}{2\pi}=\frac{1}{\lambda_0}$. Cuando λ_0 tiende a ∞

$$\boxed{f(x)=\int_{-\infty}^{\infty}\frac{1}{2\pi}\underbrace{\left[\int_{-\infty}^{\infty} f(x')\exp(-i\,k\,x')\,dx'\right]}_{F(k)}\exp(i\,k\,x)\,dk} \tag{15}$$

Este es el *Teorema de la integral de Fourier*

$$\boxed{f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F(k)\exp(i\,k\,x)\,dk \quad ; \quad F(k)=\int_{-\infty}^{\infty} f(x)\exp(-i\,k\,x)\,dx} \tag{16}$$

$F(k)$ es la transformada de Fourier de $f(x)$. Ahora $\sum \;\rightarrow \int$ y los $C_n \rightarrow F(k)$

Al igual que con las funciones periódicas, una función deberá cumplir ciertas condiciones para poder expresarla de esta forma.

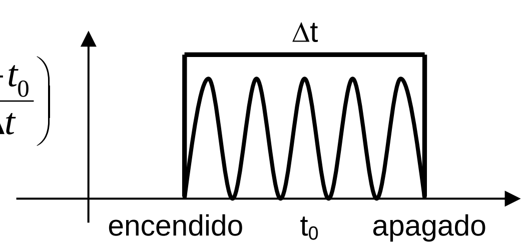
i) $f(x)$ debe ser tal que $\int_{-\infty}^{\infty}|f(x)|\,dx<\infty$ es decir, debe ser absolutamente integrable.

En realidad las funciones que no son integrables no tienen sentido físico.

Este requisito no lo cumple en general ninguna función periódica. Por ejemplo una onda plana, la usamos como herramienta pero no tiene sentido físico; lo que haremos en adelante para poder usar funciones tales como $\cos(kx); e^{ikx}\ldots$ es tomar

$f(x)=f(x)\cdot\mathrm{rect}\left(\frac{x}{a}\right)$ donde $f(x)$ es la función que no es integrable.

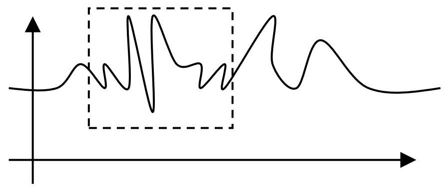
Por ejemplo, un voltaje $V(t)=V_0\cos(wt)\cdot\mathrm{rect}\left(\frac{t-t_0}{\Delta t}\right)$



Las funciones con sentido físico son aquellas que están acotadas espacial y temporalmente, es decir, cuando escribimos coseno, en realidad estamos usando una función recortada que dura en forma limitada, pero nosotros medimos un $\Delta t' < \Delta t$ de manera de despreciar los efectos de borde. Por simplicidad no se trabaja matemáticamente con funciones limitadas, de modo que muchas veces i) se viola.

ii) $f(x)$ debe tener un número finito de discontinuidades finitas

iii) $f(x)$ debe tener un número finito de máximos y mínimos en cualquier rectángulo finito (o cubo si es tridimensional)



Propiedades y teoremas básicos

Vimos que $F(k) = \mathbb{F}\left[f(x)\right] = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$. Esto también puede expresarse como

$$\mathbb{F}\left[f(x)\right] = \int_{-\infty}^{\infty} f(x)\cos(kx)dx - i\int_{-\infty}^{\infty} f(x)\sin(kx)dx = \mathbb{F}_C\left[f(x)\right] - i\mathbb{F}_S\left[f(x)\right]$$

donde $\mathbb{F}_C\left[f(x)\right]$ se denomina transformada coseno y $\mathbb{F}_S\left[f(x)\right]$ transformada seno.

Si la función es par $\mathbb{F}_S = 0$ si es impar $\mathbb{F}_C = 0$

Resulta útil también saber cómo será la transformada a partir de las características de la función $f(x)$. Podemos resumirlo en un cuadro.

$f(x)$	$F(k)$
REAL Y PAR	REAL Y PAR
IMAGINARIA Y PAR	IMAGINARIA Y PAR
REAL E IMPAR	IMAGINARIA E IMPAR
IMAGINARIA E IMPAR	REAL E IMPAR

Vamos a enunciar algunos otros teoremas básicos

- $\mathbb{F}\left[a\, f(x)+b\, g(x)\right]=a\, F(k)+b\, G(k)$
- $\mathbb{F}\left[\, f(x)\right]=F(k) \;\; \Rightarrow \;\; \mathbb{F}\left[\, f(ax)\right]=\frac{1}{|a|}F\left(\frac{k}{a}\right)$
- $\mathbb{F}\left[\, f(x)\right]=F(k) \;\; \Rightarrow \;\; \mathbb{F}\left[\, f(x\pm a)\right]=e^{\pm ika}\, F(k)$
- $\iint_{-\infty}^{\infty} \left| f(x,y)\right|^2 dx\, dy=\iint_{-\infty}^{\infty} \left|F\left(k_x,k_y\right)\right|^2 dk_x dk_y$

  Esta es la *igualdad de Parseval* que representa la conservación de la energía. Esto es el contenido energético de un campo coincide con el de su espectro.

- $\mathbb{F}\left[\, f(-x)\right]=F(-k)$
- $\mathbb{F}\left[\, f^*(x)\right]=F^*(-k)$
- Si $f(x) \in$ *Re* $\Rightarrow \;\; F^*(k)=F(-k)$
- Si la función es separable, esto es, si

$$t(x,y)= f(x)\cdot g(y)\Rightarrow \mathbb{F}\left[t(x,y)\right]=\mathbb{F}\left[\, f(x)\right]\cdot\mathbb{F}\left[g(y)\right]$$

- *Teorema integral de Fourier*

  En todo punto donde no exista discontinuidad

$$\mathbb{F}\left[\mathbb{F}^{-1}\left[\, f(x)\right]\right]=\mathbb{F}^{-1}\left[\mathbb{F}\left[\, f(x)\right]\right]= f(x)$$

$$F(k)=\mathbb{F}\left[\, f(x)\right]=\int_{-\infty}^{\infty} f(x)\exp(-i\,k\,x)\,dx \quad \Rightarrow$$

$$\Rightarrow \mathbb{F}^{-1}\left[F(k)\right]=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x)\exp(-i\,k\,x)\,dx\right]\exp(i\,k\,x')\,dk =$$

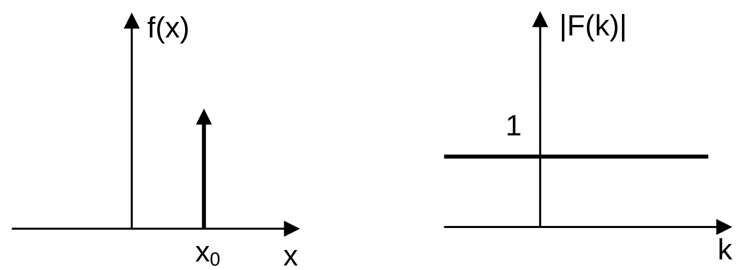
$$= \int_{-\infty}^{\infty} f(x)\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-ik\left(x-x'\right)\right)dk\, dx\equiv f(x) \quad \Rightarrow \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-ik\left(x-x'\right)\right)dk = \delta\left(x-x'\right)$$

Ejemplos de transformadas

Veamos ahora algunos ejemplos sencillos

1. $f(x)=\delta(x-x_0)$ es la delta de Dirac $\Rightarrow$

$$\mathbb{F}\left[f(x)\right]=\int_{-\infty}^{\infty} f(x)e^{-ikx}dx=\int_{-\infty}^{\infty} \delta(x-x_0)e^{-ikx}dx=e^{-ikx_0}$$



Vemos que la transformada de la delta es una constante. Desde la óptica este es un resultado conocido; una fuente puntual en el ∞ da lugar a una onda plana. Si bien este ejemplo constituye un extremo, vale para ver que el ancho de la transformada es inversamente proporcional al de la función. Esto en cuántica da lugar al principio de incerteza; pero es una propiedad general entre variables conjugadas de Fourier. Así, por ejemplo, si tomamos las variables t y ν (tiempo y frecuencia) llegamos a que el contenido de frecuencias de un tren de ondas es inversamente proporcional a su duración. Esto es, si tenemos una onda monocromática (ν es una delta) el tiempo de duración de la misma es ∞. En el otro extremo, un pulso de muy corta duración (por ejemplo un relámpago) tiene un ancho de banda muy grande. Por esta razón el relámpago o un chispazo en general, introduce ruido en todas las estaciones de radio.

2. $f(x)=\cos(k_0x-\varepsilon)=\dfrac{e^{i(k_0x-\varepsilon)}+e^{-i(k_0x-\varepsilon)}}{2}=\dfrac{e^{-i\varepsilon}}{2}e^{ik_0x}+\dfrac{e^{i\varepsilon}}{2}e^{-ik_0x}$

$$\mathbb{F}\left[f(x)\right]=\frac{e^{-i\varepsilon}}{2}\mathbb{F}\left[e^{ik_0x}\right]+\frac{e^{i\varepsilon}}{2}\mathbb{F}\left[e^{-ik_0x}\right]$$

pero de acuerdo a lo anteriormente visto

$$\mathbb{F}\left[e^{ik_0x}\right]=\int_{-\infty}^{\infty}e^{ik_0x}e^{-ikx}dx=\int_{-\infty}^{\infty}e^{-i(k-k_0)x}dx=2\pi\,\delta\left(k-k_0\right)$$

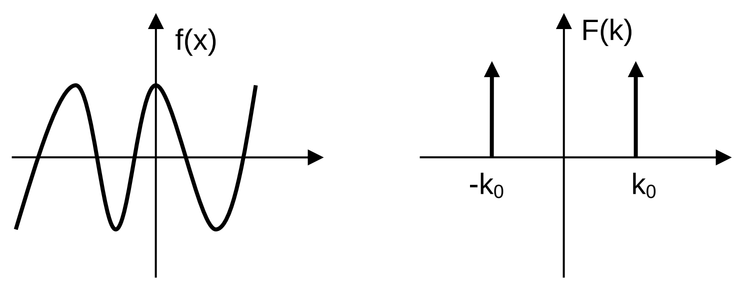
Análogamente

$$\mathbb{F}\left[e^{-ik_0x}\right]=\int_{-\infty}^{\infty}e^{-ik_0x}e^{-ikx}dx=\int_{-\infty}^{\infty}e^{-i(k+k_0)x}dx=2\pi\,\delta\left(k+k_0\right)$$

$$\Rightarrow\quad \mathbb{F}\left[f(x)\right]=\frac{e^{-i\varepsilon}}{2}2\pi\,\delta\left(k-k_0\right)+\frac{e^{i\varepsilon}}{2}2\pi\,\delta\left(k+k_0\right)$$

Si $\varepsilon=0$ es la transformada del coseno

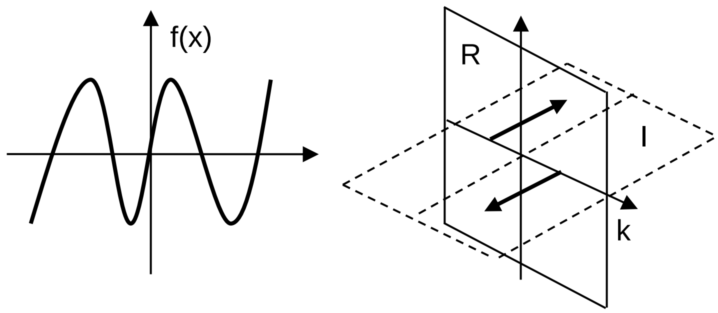
$$\mathbb{F}\left[\cos\left(k_0x\right)\right]=2\pi\left[\frac{\delta\left(k-k_0\right)}{2}+\frac{\delta\left(k+k_0\right)}{2}\right]$$



Si $\varepsilon=\pi/2$ es la transformada del seno

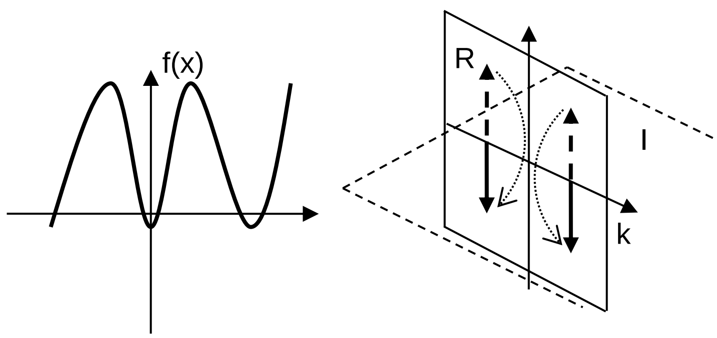
$$\mathbb{F}\left[\sin\left(k_0x\right)\right]=i2\pi\left[-\frac{\delta\left(k-k_0\right)}{2}+\frac{\delta\left(k+k_0\right)}{2}\right]$$

Aca se debe tener cuidado en cómo se grafican las deltas ya que son imaginarias. En muchas publicaciones figuran graficadas como $\uparrow\downarrow$ sin considerar en qué plano están



Si $\varepsilon = \pi$ $\Rightarrow \mathbb{F}\left[f\left(x\right)\right] = 2\pi\left[-\frac{\delta\left(k-k_0\right)}{2} - \frac{\delta\left(k+k_0\right)}{2}\right]$

Veamos como evolucionan las deltas a medida que ε crece



Ejercicio 4: Demostrar que la transformada de un peine de Dirac es otro peine de Dirac

Convolución y correlación

Vamos a definir ahora dos funciones que resulta importantísimo conocer cuando uno trabaja con transformadas de Fourier. Por ahora sólo veremos su interpretación matemática, para luego aplicarlas en problemas de difracción y sistemas lineales.

- <u>Correlación</u>

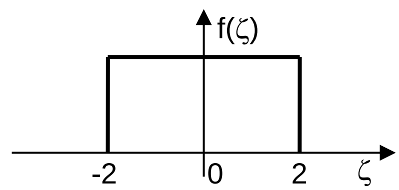
Se define la correlación entre dos funciones $f\left(x\right)$ y $g\left(x\right)$ como

$$\boxed{q_{fg}(x) = f(x) \odot g(x) = \int_{-\infty}^{\infty} f(\zeta)\, g^*(\zeta - x)\, d\zeta} \qquad (17)$$

Si $f(x) = g(x)$ se denomina autocorrelación.

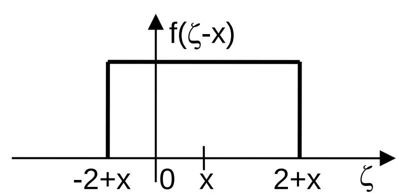
Examinemos el proceso que realiza esta operación en forma gráfica ya que no resulta obvio. Para ello estudiemos la autocorrelación $q_{ff}(x)$ de la función

$$f(\zeta) = \mathrm{rect}\left(\frac{\zeta}{4}\right)$$

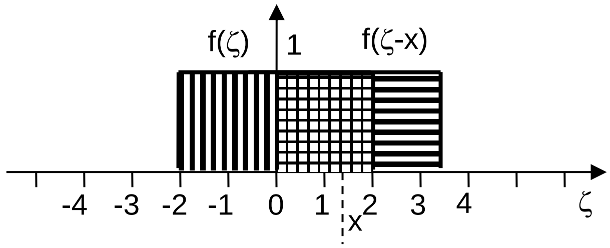


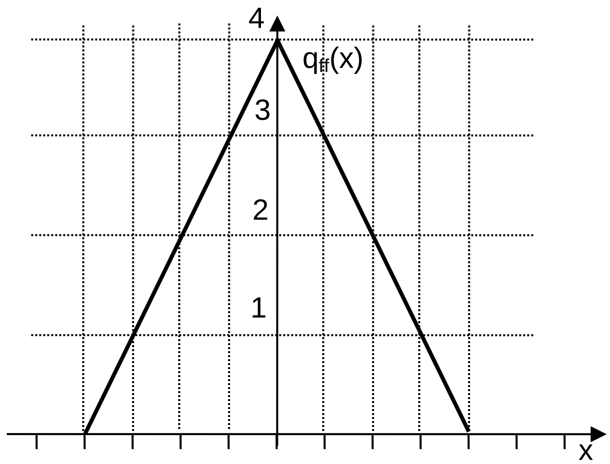
Ahora bien, la función corrida en x será

$$f(\zeta - x) = \mathrm{rect}\left(\frac{\zeta - x}{4}\right)$$



Vamos a resolver geométricamente la integral (17) para comprender que significa. Para cada valor de x la función $f(\zeta - x)$ se correrá y se superpondrá más o menos a $f(\zeta)$ dando como resultado un área mayor o menor de solapamiento. La integral será el valor de esa área.





El ancho de la correlación es siempre mayor que cualquiera de las dos funciones intervinientes y su forma es más suave.

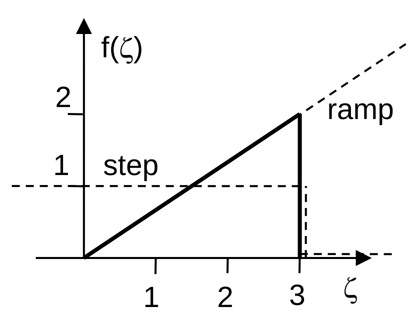
- <u>Convolución</u>

Se define la convolución entre las funciones $f(x)$ y $g(x)$ como

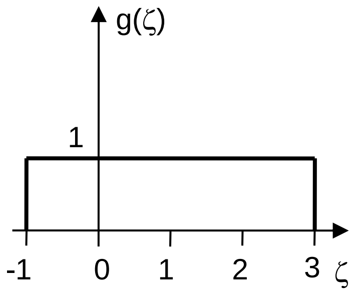
$$\boxed{C_{fg}(x) = f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(\zeta)\, g(x-\zeta)\, d\zeta} \tag{18}$$

Veamos ahora un ejemplo gráfico de convolución

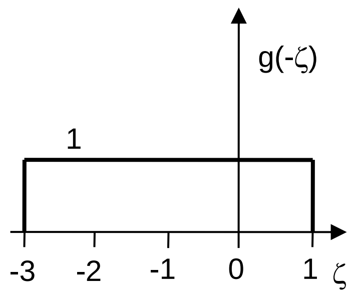
$$f(\zeta) = \mathrm{ramp}\left(\frac{\zeta}{3/2}\right) \cdot \mathrm{step}\left(\frac{\zeta - 3}{-1}\right)$$



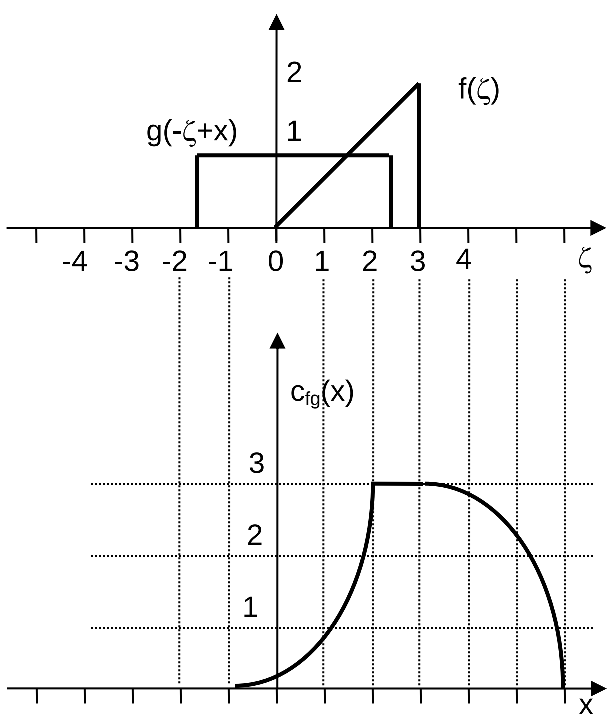
$$g(\zeta)=\mathrm{rect}\left(\frac{\zeta-1}{4}\right)$$



$$g(-\zeta)=\mathrm{rect}\left(\frac{-\zeta-1}{4}\right)$$



Nuevamente acá para obtener la integral (18) deberemos multiplicar $f(\zeta)$ por $g(-\zeta+x)$ y ver cuál es el área de superposición para cada valor de x. A medida que x cambie, el rectángulo que representa $g(-\zeta)$ se irá desplazando y dicha área de superposición variará.



De -1 a 2 crece en forma parabólica, se mantiene constante de 2 a 3 y luego decrece parabólicamente pero con la concavidad cambiada.

Debemos tener en cuenta que las condiciones para que $q_{fg}(x)$ y $c_{fg}(x)$ existan son básicamente que el área de superposición esté bien definida.
Veamos ahora algunos teoremas fundamentales:

- $f(x)\otimes g(x) = g(x)\otimes f(x)$ conmutatividad
- $f(x)\otimes\big(g(x)\otimes t(x)\big) = \big(f(x)\otimes g(x)\big)\otimes t(x)$ asociatividad
- $f(x)\otimes\big(g(x)+t(x)\big) = f(x)\otimes g(x) + f(x)\otimes t(x)$ propiedad distributiva
- $\left[f(x)\otimes g(x)\right]^* = f(x)^*\otimes g(x)^*$ conjugación
- $f(x-a)\otimes\delta(x-b) = f(x-a-b)$
- $f(-x)\otimes\delta(x) = f(-x)$
- $f(-x)\otimes\delta(x-a) = f(-x+a)$
- $f(x)\odot f(x) = f(-x)\odot f(-x)$
- $f(x)\odot g(x) \neq g(x)\odot f(x)$ en general no es conmutativa
- $f(x), g(x)\in\mathbb{R} \;\Rightarrow\; f(x)\odot g(x) = g(-x)\odot f(-x)$
- $f(x), g(x)\in\mathbb{R} \;;\; g(x) = g(-x) \;\Rightarrow\; f(x)\otimes g(x) = f(x)\odot g(x)$
- $f(x)\otimes g^*(x) = f(x)\odot g(-x)$

Veamos algunas propiedades importantes, relacionada con las transformadas de Fourier

- $\mathbb{F}\left[f(x)\bullet g(x)\right] = F(k)\otimes G(k)$
- $\mathbb{F}\left[f(x)\bullet g^*(x)\right] = F(k)\otimes G^*(-k) = F(k)\odot G(k)$
- $\mathbb{F}\left[f(x)\otimes g(x)\right] = F(k)\bullet G(k)$ Teorema de la convolución
- $\mathbb{F}\left[f(x)\odot g(x)\right] = F(k)\bullet G^*(-k)$

  si $G(k)\in\mathbb{R} \;\Rightarrow\; \mathbb{F}\left[f(x)\odot g(x)\right] = F(k)\bullet G(-k)$ y si además

$$G(k)=G(-k) \;\Rightarrow\; \mathbb{F}\left[f(x)\odot g(x)\right]=F(k)\bullet G(k)$$

- $\mathbb{F}\left[f(x)\otimes f(x)\right]=\left|F(k)\right|^2$ si $F(k)$ es par
- $\mathbb{F}\left[f(x)\odot f(x)\right]=\left|F(k)\right|^2$ si $F(k)\in\mathbb{R}$

III. DIFRACCIÓN DE FRESNEL Y FRAUNHOFER – PROPIEDAD TRANSFORMADORA DE LAS LENTES

CLASE 5

INTRODUCCIÓN

Volvamos ahora a los temas de óptica que nos ocupan, pero tratemos de comparar las expresiones anteriormente obtenidas con las funciones matemáticas que hemos estudiado.
La operación convolución en dos dimensiones puede expresarse como

$$C_{fg}(x,y)= f(x,y)\otimes g(x,y)=\int\int_{-\infty}^{\infty} f(\zeta,\eta)\,g(x-\zeta,y-\eta)\,d\zeta\, d\eta$$

Comparemos esta ecuación con la expresión para el campo en un punto P_0 según la integral de difracción en la aproximación de Fresnel

$$E(x_0,y_0)=\iint_{\Sigma} E(x_1,y_1,z=0)\underbrace{\frac{\exp(ikz)}{i\lambda z}\exp\left(i\frac{k}{2z}\left[(x_0-x_1)^2+(y_0-y_1)^2\right]\right)}_{h}dx_1\,dy_1$$

La integral puede extenderse a $\pm\infty$ ya que $E(x_1,y_1,z=0)=0$ fuera de Σ . Vemos que entonces es posible escribir

$$E(x_0,y_0)=E_{\Sigma}\otimes h$$

Volveremos sobre esta expresión cuando veamos sistemas lineales.

Analicemos ahora la ecuación obtenida en la aproximación de Fraunhofer

$$E\left(x_0, y_0\right) = \frac{\exp\left(ik\,z\right)}{i\lambda z} \exp\left(i\,\frac{k}{2z}\left(x_0^2 + y_0^2\right)\right) \iint_{\Sigma} E\left(x_1, y_1\right) \exp\left[-i2\pi\left(\frac{x_0}{\lambda z}\;x_1 + \frac{y_0}{\lambda z}\;y_1\right)\right] dx_1\,dy_1$$

Si la comparamos con la transformada de Fourier bidimensional

$$\mathbb{F}\left[\,f\left(x, y\right)\right] = \int\int_{-\infty}^{\infty} f\left(x, y\right) \exp\left[-i\left(k_x\,x + k_y\,y\right)\right] dx\,dy \quad \text{con} \quad k_x = \frac{2\pi}{\lambda_x}\,;\, k_y = \frac{2\pi}{\lambda_y}$$

Si definimos las frecuencias espaciales como $\begin{cases} f_x \equiv \dfrac{x_0}{\lambda z} = \dfrac{1}{\lambda_x} \\ f_y \equiv \dfrac{y_0}{\lambda z} = \dfrac{1}{\lambda_y} \end{cases}$ entonces

$$E\left(x_0, y_0\right) = \frac{\exp\left[i\varphi\left(x_0, y_0, z\right)\right]}{i\lambda z}\; \mathbb{F}_{f_x f_y}\left[E\left(x_1, y_1\right)\right]$$

EJEMPLOS DE DIAGRAMAS DE DIFRACCIÓN

A fin de aclarar un poco más estos conceptos veamos algunos diagramas de difracción básicos y de suma importancia.

Ejemplos en la aproximación de Fresnel

Comenzaremos analizando el fenómeno de difracción en la aproximación de Fresnel. Se debe resolver cada caso con un tratamiento particular ya que debido a su complejidad no existe un tratamiento general.
Habíamos encontrado que la expresión para el campo en un punto (x_0,y_0) era:

$$E\left(P_0\right) = \frac{\exp\left(ik\,z\right)}{i\lambda z} \iint_{\Sigma} E\left(x_1, y_1\right) \exp\left\{i\,\frac{k}{2z}\left[\left(x_0 - x_1\right)^2 + \left(y_0 - y_1\right)^2\right]\right\} dx_1\,dy_1$$

Abertura rectangular

Comencemos estudiando el caso de la abertura rectangular. Podemos escribir

$$E(x_1, y_1) = \mathrm{rect}\left(\frac{x_1}{a}\right) \cdot \mathrm{rect}\left(\frac{y_1}{b}\right)$$

$$E(P_0) = \frac{\exp(ikz)}{i\lambda z} \int_{-a/2}^{a/2} \exp\left[i\frac{k}{2z}(x_0 - x_1)^2\right] dx_1 \int_{-b/2}^{b/2} \exp\left[i\frac{k}{2z}(y_0 - y_1)^2\right] dy_1$$

Hagamos el cambio de variables

$$u = \sqrt{\frac{k}{\pi z}}(x_1 - x_0) \;\Rightarrow\; dx_1 = du\sqrt{\frac{\pi z}{k}} \quad ; \quad v = \sqrt{\frac{k}{\pi z}}(y_1 - y_0) \;\Rightarrow\; dy_1 = dv\sqrt{\frac{\pi z}{k}}$$

Los límites serán

$$-\frac{a}{2} \;\rightarrow\; u_1 = -\sqrt{\frac{k}{\pi z}}\left(\frac{a}{2} + x_0\right) \quad ; \quad -\frac{b}{2} \;\rightarrow\; v_1 = -\sqrt{\frac{k}{\pi z}}\left(\frac{b}{2} + y_0\right)$$

$$\frac{a}{2} \;\rightarrow\; u_2 = \sqrt{\frac{k}{\pi z}}\left(\frac{a}{2} - x_0\right) \quad ; \quad \frac{b}{2} \;\rightarrow\; v_2 = \sqrt{\frac{k}{\pi z}}\left(\frac{b}{2} - y_0\right)$$

Entonces

$$E(x_0, y_0) = \frac{\exp(ikz)}{i\lambda z}\frac{\pi z}{k}\int_{u_1}^{u_2} \exp\left(i\frac{\pi}{2}u^2\right) du \int_{v_1}^{v_2} \exp\left(i\frac{\pi}{2}v^2\right) dv$$

Es conveniente llevar estas integrales a la forma $\int_0^{u'}$, esto es

$$E(x_0, y_0) = \frac{e^{ikz}}{2i}\left\{\left[\int_{u_1}^{0} + \int_{0}^{u_2}\right] \cdot \left[\int_{v_1}^{0} + \int_{0}^{v_2}\right]\right\} = \frac{e^{ikz}}{2i}\left\{\left[-\int_{0}^{u_1} + \int_{0}^{u_2}\right] \cdot \left[-\int_{0}^{v_1} + \int_{0}^{v_2}\right]\right\}$$

Hemos hecho esto porque existen dos funciones, extensamente estudiadas y cuyos valores numéricos se hallan tabulados, que se pueden aplicar a la resolución de este problema. Dichas funciones son las llamadas integrales de Fresnel y actualmente forman parte de muchos programas de cálculo (ver por ejemplo http://mathworld.wolfram.com/FresnelIntegrals.html)

$$C(w)=\int_0^w \cos\left(\frac{\pi}{2}w'^2\right)dw' \quad ; \quad S(w)=\int_0^w \sin\left(\frac{\pi}{2}w'^2\right)dw'$$

Ahora bien, en nuestro caso

$$\int_0^{u_n} \exp\left(i\frac{\pi}{2}u^2\right)du=\int_0^{u_n}\left[\cos\left(\frac{\pi}{2}u^2\right)+i\sin\left(\frac{\pi}{2}u^2\right)\right]du=C(u_n)+i\,S(u_n) \qquad n=1,2$$

Entonces

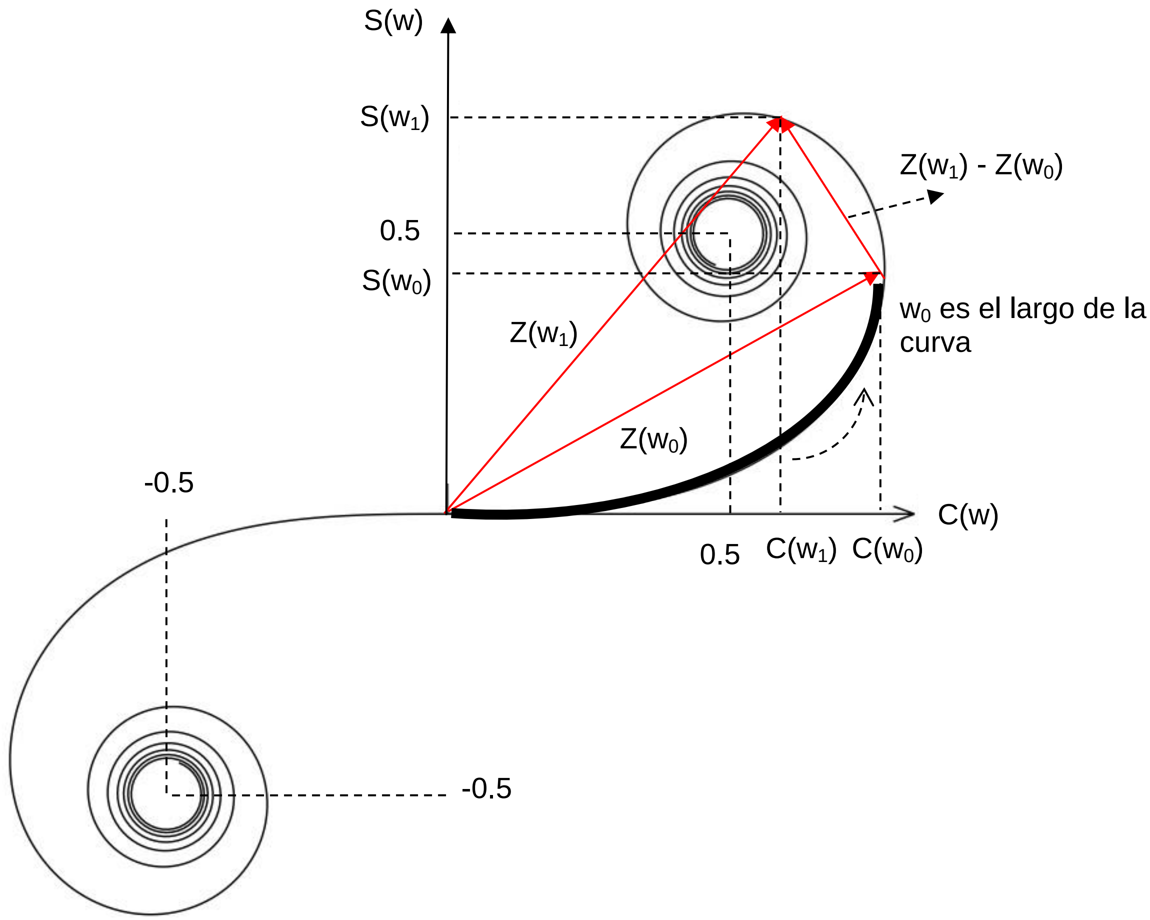
$$E(x_0,y_0)=\frac{e^{ikz}}{2i}\left\{\left[C(u_2)+i\,S(u_2)-C(u_1)-i\,S(u_1)\right]\cdot\left[C(v_2)+i\,S(v_2)-C(v_1)-i\,S(v_1)\right]\right\}$$

$$E(x_0,y_0)=\frac{e^{ikz}}{2i}\left\{\left[C(u_2)-C(u_1)\right]+i\left[S(u_2)-S(u_1)\right]\right\}\cdot\left\{\left[C(v_2)-C(v_1)\right]+i\left[S(v_2)-S(v_1)\right]\right\}$$

La intensidad viene dada por

$$I(x_0,y_0)=\frac{1}{4}\left\{\left[C(u_2)-C(u_1)\right]^2+\left[S(u_2)-S(u_1)\right]^2\right\}\cdot\left\{\left[C(v_2)-C(v_1)\right]^2+\left[S(v_2)-S(v_1)\right]^2\right\}$$

Sólo por completitud mencionaremos que históricamente se solía recurrir a la construcción gráfica denominada *espiral de Cornú* (Francia 1841-1902) para la resolución de este tipo de integrales.



Definamos un número complejo tal que

$$Z(w) = C(w) + i\,S(w) = \int_0^w \cos\left(\frac{\pi}{2} w'^2\right) dw' + i \int_0^w \sin\left(\frac{\pi}{2} w'^2\right) dw'$$

Sea *dl* un diferencial de arco a lo largo de la espiral, entonces

$$dl^2 = dC^2 + dS^2 = \left[\left(\frac{dC}{dw}\right)^2 + \left(\frac{dS}{dw}\right)^2\right] dw^2 = \left[\cos^2\left(\frac{\pi}{2} w^2\right) + \sin^2\left(\frac{\pi}{2} w^2\right)\right] dw^2 = dw^2$$

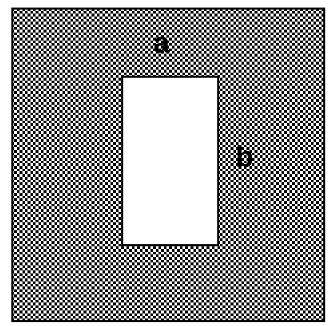
$$\Rightarrow \quad dl = dw$$

Las representaciones de w_0, siendo este un número arbitrario, y la correspondiente a $Z(w_0) = C(w_0) + iS(w_0)$ pueden observarse en la figura. Del mismo modo puede graficarse $Z(w_1) - Z(w_0) = \left[C(w_1) - C(w_0)\right] + i\left[S(w_1) - S(w_0)\right]$. Pero esta última expresión es justamente la que aparece dentro de cada llave de la ecuación correspondiente al campo $E(x_0, y_0)$. Por otra parte cada llave de la ecuación que describe la intensidad $I(x_0, y_0)$ no es más que $\left|Z(w_1) - Z(w_0)\right|^2$.

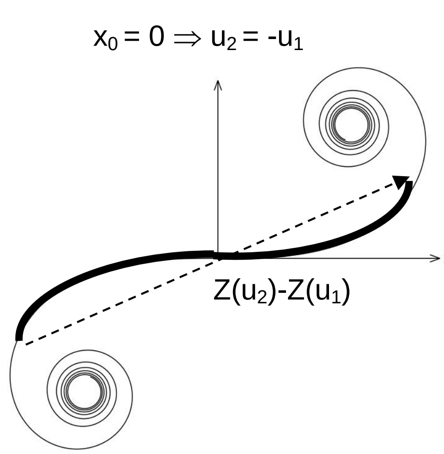
Veamos entonces que sucede en el caso de nuestra abertura rectangular. Recordemos que:

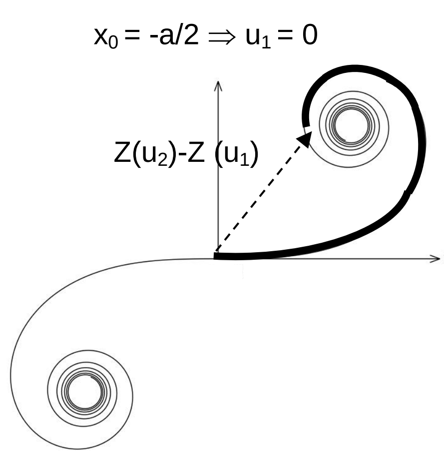
$$u_1 = -\sqrt{\frac{k}{\pi z}}\left(\frac{a}{2} + x_0\right) \quad ; \quad v_1 = -\sqrt{\frac{k}{\pi z}}\left(\frac{b}{2} + y_0\right)$$

$$u_2 = \sqrt{\frac{k}{\pi z}}\left(\frac{a}{2} - x_0\right) \quad ; \quad v_2 = \sqrt{\frac{k}{\pi z}}\left(\frac{b}{2} - y_0\right)$$



Si hacemos $u_2 - u_1 = \sqrt{\frac{k}{\pi z}}\, a \quad ; \quad v_2 - v_1 = \sqrt{\frac{k}{\pi z}}\, b$ encontramos expresiones que son independientes del punto de observación, sólo dependen de la geometría de la abertura y son las que determinan el largo de la curva sobre la espiral (no su posición). Por ejemplo





Veamos qué sucede en algunos casos límite.

Supongamos primero que z es tal que el punto de observación está muy cerca de la abertura; esto es, un z muy chico pero lo suficientemente grande como para que siga valiendo la aproximación de Fesnel. En este caso $\sqrt{\frac{k}{\pi z}}$ es muy grande y podemos tomar

$$u_1 \cong \begin{cases} -\infty & x_0 > -\frac{a}{2} \\ \infty & x_0 < -\frac{a}{2} \end{cases} \quad ; \quad u_2 \cong \begin{cases} -\infty & x_0 > \frac{a}{2} \\ \infty & x_0 < \frac{a}{2} \end{cases}$$

$$v_1 \cong \begin{cases} -\infty & y_0 > -\frac{b}{2} \\ \infty & y_0 < -\frac{b}{2} \end{cases} \quad ; \quad v_2 \cong \begin{cases} -\infty & y_0 > \frac{b}{2} \\ \infty & y_0 < \frac{b}{2} \end{cases}$$

Los valores en la espiral de Cornú serán:

$$C(u_1) = S(u_1) \cong \begin{cases} -\frac{1}{2} & x_0 > -\frac{a}{2} \\ \frac{1}{2} & x_0 < -\frac{a}{2} \end{cases} \quad ; \quad C(u_2) = S(u_2) \cong \begin{cases} -\frac{1}{2} & x_0 > \frac{a}{2} \\ \frac{1}{2} & x_0 < \frac{a}{2} \end{cases}$$

$$C(v_1) = S(v_1) \cong \begin{cases} -\frac{1}{2} & y_0 > -\frac{b}{2} \\ \frac{1}{2} & y_0 < -\frac{b}{2} \end{cases} \quad ; \quad C(v_2) = S(v_2) \cong \begin{cases} -\frac{1}{2} & y_0 > \frac{b}{2} \\ \frac{1}{2} & y_0 < \frac{b}{2} \end{cases}$$

Substituyendo estos valores en la expresión hallada para $E(x_0, y_0)$ tenemos

$$E(x_0, y_0)\Big|_{x_0 < \left|\frac{a}{2}\right|; y_0 < \left|\frac{b}{2}\right|} = \frac{\exp(i k z)}{2i} \left\{ \left[\frac{1}{2} - \left(-\frac{1}{2} \right) \right] + i \left[\frac{1}{2} - \left(-\frac{1}{2} \right) \right] \right\} \cdot \left\{ \left[\frac{1}{2} - \left(-\frac{1}{2} \right) \right] + i \left[\frac{1}{2} - \left(-\frac{1}{2} \right) \right] \right\} =$$

$$= \frac{\exp(i k z)}{2i} 2i = \exp(i k z) \operatorname{rect}\left(\frac{x_0}{a} \right) \operatorname{rect}\left(\frac{y_0}{b} \right)$$

$$E\left(x_0, y_0\right)\Big|_{x_0 > \left|\frac{a}{2}\right|; y_0 > \left|\frac{b}{2}\right|} = 0$$

Vale decir que es la sombra geométrica de $E\left(x_1, y_1\right)$, lo que es un resultado conocido

Ejercicio 5: Analizar el diagrama de difracción de Fresnel para una pantalla semi-infinita

Usualmente la difracción de Fresnel puede calcularse numéricamente en base a la Transformada Rápida de Fourier (FFT). Este es un algoritmo que permite computar muy velozmente la transformada de Fourier. Sin embargo si escribimos la ecuación que describe al campo en la aproximación de Fresnel como

$$E\left(x_0, y_0\right) = \frac{\exp\left(ik\,z\right)}{i\lambda z} \exp\left[i\frac{k}{2z}\left(x_0^2 + y_0^2\right)\right] \iint_{\Sigma} \underbrace{E\left(x_1, y_1\right) \exp\left[i\frac{k}{2z}\left(x_1^2 + y_1^2\right)\right]}_{*} .$$

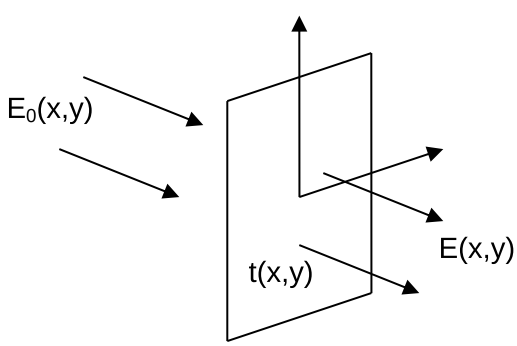
$$.\exp\left[-i\frac{k}{z}\left(x_0 x_1 + y_0 y_1\right)\right] dx_1\, dy_1$$

vemos que podemos interpretarla como la transformada de Fourier del término señalado con la llave, o sea del campo en la abertura por un factor de fase cuadrático. Para más detalles puede verse el trabajo [Appl. Opt. **38**, 7085 (1999)].

Ejemplos en la aproximación de Fraunhofer

En esta aproximación veremos los diagramas generados por una red cuya amplitud varía sinusoidalmente, una cuya fase varía sinusoidalmente y una abertura circular.

En general resultará útil introducir el concepto de *función transmisión* $t\left(x, y\right)$



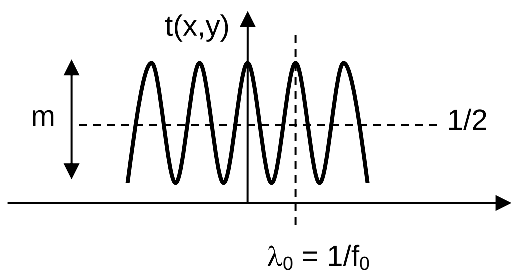
$$t\left(x, x\right) = \frac{E\left(x, y\right)}{E_0\left(x, y\right)} = \left|t\left(x, x\right)\right| e^{i\varphi(x,y)}$$

Cuando φ *=0* el objeto es sólo de amplitud, y cuando $|t(x,x)|$ =1 el objeto es sólo de fase

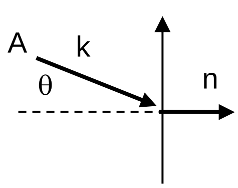
1. <u>Red de amplitud sinusoidal</u>

$$t(x,y)=\left[\frac{1}{2}+\frac{m}{2}\cos\left(2\pi\, f_0 x\right)\right]\mathrm{rect}\left(\frac{x}{a}\right)\mathrm{rect}\left(\frac{y}{b}\right)$$

El factor ½ se debe poner para evitar transmisiones negativas, f_0 es la frecuencia espacial de la red. Las funciones rectángulo dan cuenta del tamaño finito de la red y $0<m\leq 1$ es la modulación.



Supongamos que sobre la red incide normalmente una onda plana, monocromática de amplitud unitaria $\Rightarrow E_0(x,y)=1 \;\Rightarrow\; E(x,y)=1\cdot t(x,y)$. Si no fuese así deberíamos incluir una amplitud y una fase debida al ángulo de incidencia



$$Ae^{i\,k\cdot n}$$

Ahora bien, en la aproximación de Fraunhofer teníamos que

$$E(x_0,y_0)=\frac{\exp\left[i\varphi(x_0,y_0,z)\right]}{i\lambda z}\;\mathbb{F}\left[E(x_1,y_1)\right]$$

Calculemos entonces $\mathbb{F}\left[t\left(x,y\right)\right]$. Para ello vamos a desarrollar el coseno y emplearemos las propiedades anteriormente vistas

$$\mathbb{F}\left[t\left(x,y\right)\right]=\mathbb{F}\left[\frac{1}{2}+\frac{m}{4}\exp\left(i2\pi f_0 x\right)+\frac{m}{4}\exp\left(-i2\pi f_0 x\right)\right]\otimes\mathbb{F}\left[\mathrm{rect}\left(\frac{x}{a}\right)\cdot\mathrm{rect}\left(\frac{y}{b}\right)\right]$$

En el producto de los rectángulos no hace falta aplicar la convolución dado que son variables separables. Calculemos entonces las transformadas

$$\mathbb{F}\left[\frac{1}{2}\right]=\int\int_{-\infty}^{\infty}\frac{1}{2}\exp\left(-i2\pi\, f_x x\right)\exp\left(-i2\pi\, f_y y\right)dx\,dy=\frac{1}{2}\delta\left(f_x, f_y\right)$$

conviene recordar que $\int_{-\infty}^{\infty} e^{-ikx}\,dx=2\pi\,\delta\left(k\right)$ pero existe una propiedad de las deltas que establece que $\delta\left(cte\cdot x\right)=\frac{1}{|cte|}\delta\left(x\right) \;\Rightarrow\; 2\pi\,\delta\left(k\right)=2\pi\,\delta\left(2\pi\, f\right)=\delta\left(f\right)$

$$\mathbb{F}\left[\frac{m}{4}\exp\left(i2\pi\, f_0 x\right)\right]=\int\int_{-\infty}^{\infty}\frac{m}{4}\exp\left[-i2\pi\left(f_x-f_0\right)x\right]\exp\left(-i2\pi\, f_y y\right)dx\,dy=$$

$$=\frac{m}{4}\delta\left(f_x-f_0, f_y\right)$$

Análogamente

$$\mathbb{F}\left[\frac{m}{4}\exp\left(-i2\pi\, f_0 x\right)\right]=\frac{m}{4}\delta\left(f_x+f_0, f_y\right)$$

Por otra parte tenemos que

$$\mathbb{F}\left[\operatorname{rect}\left(\frac{x}{a}\right)\cdot\operatorname{rect}\left(\frac{y}{b}\right)\right]=\int_{-\infty}^{\infty}\operatorname{rect}\left(\frac{x}{a}\right)\exp\left(-i2\pi\, f_x x\right)dx\,\cdot\int_{-\infty}^{\infty}\operatorname{rect}\left(\frac{y}{b}\right)\exp\left(-i2\pi\, f_y y\right)dy=$$

$$=\int_{-a/2}^{a/2}\exp\left(-i2\pi\, f_x x\right)dx\,\cdot\int_{-b/2}^{b/2}\exp\left(-i2\pi\, f_y y\right)dy=$$

$$=\int_{-a/2}^{a/2}\left[\cos\left(2\pi\, f_x x\right)-i\sin\left(2\pi\, f_x x\right)\right]dx\,\cdot\int_{-b/2}^{b/2}\left[\cos\left(2\pi\, f_y y\right)-i\sin\left(2\pi\, f_y y\right)\right]dy=$$

$$=\left\{\left.\frac{\sin\left(2\pi\, f_x x\right)}{2\pi\, f_x}\right|_{-a/2}^{a/2}+i\left.\frac{\cos\left(2\pi\, f_x x\right)}{2\pi\, f_x}\right|_{-a/2}^{a/2}\right\}\cdot\left\{\left.\frac{\sin\left(2\pi\, f_y y\right)}{2\pi\, f_y}\right|_{-b/2}^{b/2}+i\left.\frac{\cos\left(2\pi\, f_y y\right)}{2\pi\, f_y}\right|_{-b/2}^{b/2}\right\}=$$

$$=\frac{\sin\left(\pi\, f_x a\right)}{\pi\, f_x}\cdot\frac{\sin\left(\pi\, f_y b\right)}{\pi\, f_y}=a\,b\operatorname{sinc}\left(f_x a\right)\operatorname{sinc}\left(f_y b\right)$$

Así pues

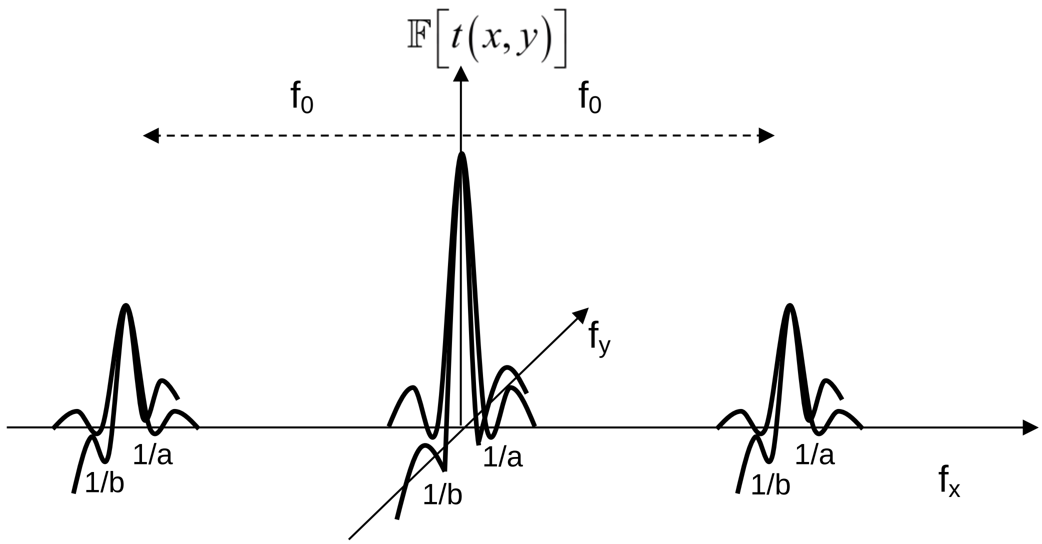
$$\mathbb{F}\left[t\left(x,y\right)\right]=\left[\frac{1}{2}\delta\left(f_x,f_y\right)+\frac{m}{4}\delta\left(f_x-f_0,f_y\right)+\frac{m}{4}\delta\left(f_x+f_0,f_y\right)\right]\otimes$$

$$\otimes\,a\,b\operatorname{sinc}\left(f_x a\right)\operatorname{sinc}\left(f_y b\right)=$$

$$=\frac{a\,b}{2}\operatorname{sinc}\left(f_y b\right)\left[\operatorname{sinc}\left(f_x a\right)+\frac{m}{2}\operatorname{sinc}\left(\left(f_x-f_0\right)a\right)+\frac{m}{2}\operatorname{sinc}\left(\left(f_x+f_0\right)a\right)\right]$$

Teniendo en cuenta que $\operatorname{sinc}\left(px\right)=\dfrac{\sin\left(\pi\,px\right)}{\pi\,px}$ tiene el máximo principal en $x=0$

y el primer cero cuando $\pi\,px=\pm\pi \quad\Rightarrow\quad x=\pm\dfrac{1}{p}$, grafiquemos $E\left(x_0,y_0\right)$



Si a y b tienden a infinito entonces las funciones sinc tienden a ser una delta. Por otra parte dado que $f_0 \gg \dfrac{2}{a}$ la intensidad será

$$I = E.E^* = \frac{a^2b^2}{4\lambda^2z^2}\,\mathrm{sinc}^2\left(b\,f_y\right)\left[\mathrm{sinc}^2\left(a\,f_x\right)+\frac{m^2}{4}\mathrm{sinc}^2\left(a\left(f_x-f_0\right)\right)+\frac{m^2}{4}\mathrm{sinc}^2\left(a\left(f_x+f_0\right)\right)\right]$$

ya que los términos cruzados se pueden despreciar por estar las tres funciones sinc centradas en puntos muy alejados. Vemos entonces que una red cuya transmisión en amplitud varía sinusoidalmente da lugar al orden 0 y sólo al ± 1. Obviamente siempre que $d_0 >> \lambda$ que es una de nuestras hipótesis de trabajo.

Examinemos este resultado que es muy importante. Habíamos visto que las frecuencias espaciales que aparecían en la expresión de la transformada de Fourier eran

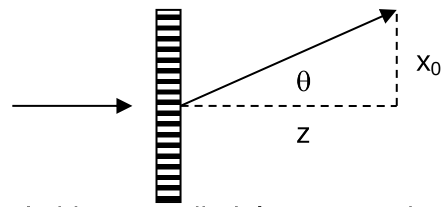
$$\begin{cases} f_x \equiv \dfrac{x_0}{\lambda z} & \Rightarrow \quad x_0 = f_x\lambda z \\ f_y \equiv \dfrac{y_0}{\lambda z} & \Rightarrow \quad y_0 = f_y\lambda z \end{cases}$$

Por otra parte acabamos de ver que un objeto con frecuencia espacial f_0 en el eje x produce máximos de difracción centrados en $f_x = \pm f_0$, esto es en $x_0 = f_0 \lambda z$.

Ahora bien, si tenemos una red de difracción de período d_0, esto es de frecuencia f_0, la ecuación de la red para incidencia normal predice la posición de los órdenes ±1 en

$\sin\theta = m\lambda f_0 \xrightarrow{m=1} = \lambda f_0$ pero

$$\sin\theta \cong \frac{x_0}{z} \quad \Rightarrow \quad \frac{x_0}{z} = \lambda f_0 \quad \Rightarrow \quad x_0 = \lambda f_0 z$$



Vale decir que cada frecuencia espacial presente en el objeto contribuirá con un orden en el plano transformado, cuya posición estará unívocamente relacionada con su frecuencia espacial y la orientación de la misma. A la inversa, si nosotros montamos la experiencia de Young de forma tal que la separación entre fuentes sea x_0, a una distancia z obtendremos una figura de interferencia cosenoidal con frecuencia $f_0 = \dfrac{x_0}{\lambda z}$.

Este ejemplo que acabamos de ver es muy importante porque, de acuerdo a lo que estudiamos de Fourier, una función arbitraria (por ejemplo una imagen en una diapositiva) la podemos descomponer en senos y cosenos de distintas frecuencias y orientaciones, por lo tanto podemos imaginarla como conformada por una superposición de redes sinusoidales de amplitudes, frecuencias y orientaciones distintas; ahora ya sabemos cómo se comporta cada una de ellas en el plano transformado.

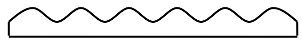
CLASE 6

Es muy común decir que una red sinusoidal produce sólo tres órdenes difractados, el 0, el 1 y el -1. Debe remarcarse que esto es cierto si la red es de amplitud. Para estudiar qué sucede si lo que varía sinusoidalmente es la fase, veamos este segundo ejemplo.

2. Red de fase sinusoidal

Sea un objeto con transmisión

$$t(x,y)=\exp\left\{i\,k\,m\left[\Delta+\frac{1}{2}\cos(2\pi f_0 x)\right]\right\}\mathrm{rect}\left(\frac{x}{a}\right)\mathrm{rect}\left(\frac{y}{b}\right)$$

Una forma de hacer esto es, por ejemplo, modelando un sustrato transparente con una forma de superficie  . Otra, es elaborar una placa de un material transparente pero cuyo índice de refracción varíe de forma sinusoidal (cuando estudiemos materiales de registro veremos esto en más detalle). El término $e^{i\Delta}$ puede eliminarse ya que es un factor global de fase.

Para resolver este problema es conveniente usar la siguiente ecuación (Table of integrals……Gradshteyn, pg 973):

$$\exp(i\,a\cos\theta)=\sum_{n=-\infty}^{\infty}J_n(a)\exp\left(i\,n\frac{\pi}{2}\right)\exp(i\,n\,\theta)$$

Donde $J_n(a)$ es la función J de Bessel de primera especie y de orden n.

Entonces podemos escribir

$$t(x,y)=\sum_{n=-\infty}^{\infty}J_n\left(\frac{\rho}{2}\right)\exp\left(i\frac{n\pi}{2}\right)\exp(i\,n\,2\pi f_0 x)\ \mathrm{rect}\left(\frac{x}{a}\right)\mathrm{rect}\left(\frac{y}{b}\right)$$

donde ρ =*km* es el factor de modulación de fase.

Como en el ejemplo anterior, consideraremos que incide normalmente una onda plana monocromática de amplitud unitaria.

$$E(x_0,y_0)\propto\mathbb{F}\left[t(x,y)\right]=\sum_{n=-\infty}^{\infty}J_n\left(\frac{\rho}{2}\right)\exp\left(i\frac{n\pi}{2}\right)\mathbb{F}\left[\exp(i\,n\,2\pi f_0 x)\right]\otimes$$

$$\otimes\,a.b.\mathrm{sinc}(a\,f_x).\mathrm{sinc}(b\,f_y)$$

pero $\mathbb{F}\left[\exp(i\,n\,2\pi f_0 x)\right]=\delta\left(f_x-n\,f_0,\,f_y\right)$ con lo cual tenemos que

$$\mathbb{F}\left[t(x,y)\right]=a.b.\mathrm{sinc}(b\,f_y)\sum_{n=-\infty}^{\infty}J_n\left(\frac{\rho}{2}\right)\exp\left(i\frac{n\pi}{2}\right)\mathrm{sinc}\left(a\left(f_x-n\,f_0\right)\right)$$

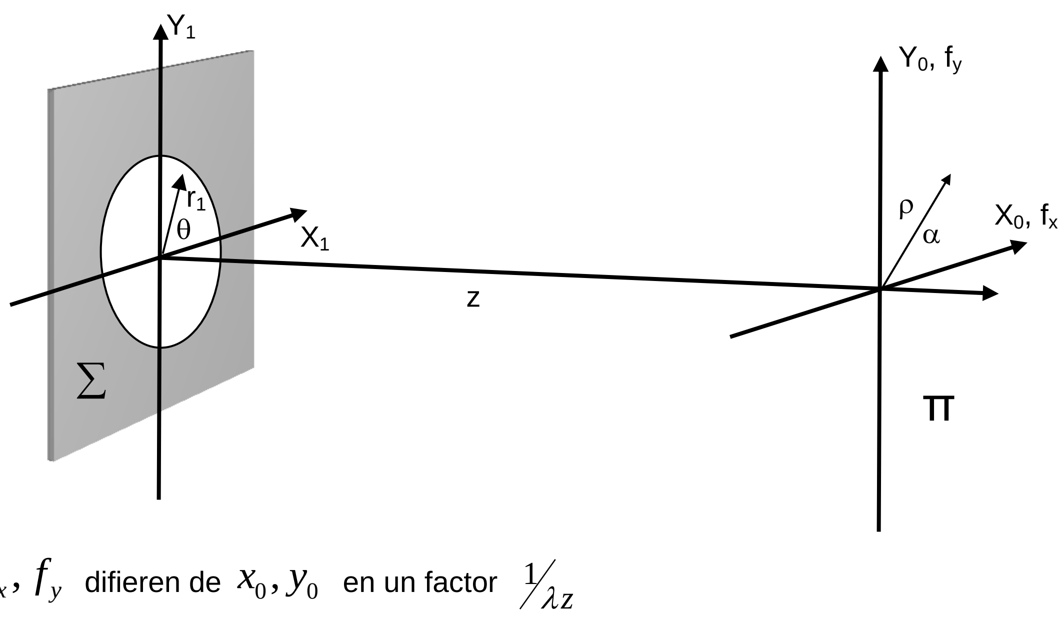
Nuevamente se obtienen órdenes en el eje f_x centrados en los puntos $n\,f_0$ y de ancho 2/a. Vemos que la separación entre órdenes así como su ancho es una cuestión geométrica ya que es igual en la red de amplitud, sin embargo acá aparecen órdenes superiores y la amplitud de cada uno de ellos está determinada por la función $J_n\left(\frac{\rho}{2}\right)e^{i\frac{n\pi}{2}}$. Suponiendo nuevamente que $f_0 \gg \frac{2}{a}$

$$I(x,y)=\left[\frac{a.b}{\lambda z}\operatorname{sinc}\left(b\,f_y\right)\right]^2.\sum_{n=-\infty}^{\infty}J_n^2\left(\frac{\rho}{2}\right).\operatorname{sinc}^2\left(\frac{a}{\lambda z}\left(x_0-n\,\lambda z f_0\right)\right)$$

Con este tipo de redes es posible distribuir más energía en los órdenes laterales que en el central, según el ρ elegido.

3. <u>Abertura circular</u>

Vayamos ahora al caso importante de las aberturas circulares (recordemos que la mayoría de las lentes lo son). Debido a la simetría del problema hagamos las siguientes consideraciones.



f_x, f_y difieren de x_0, y_0 en un factor $1/\lambda z$

$$r_1 = \sqrt{x_1^2 + y_1^2}$$
$$\theta = \arctan\left(\frac{y_1}{x_1}\right)$$

$$\rho = \sqrt{f_x^2 + f_y^2}$$
$$\alpha = \arctan\left(\frac{f_y}{f_x}\right)$$
$$r_0 = \sqrt{x_0^2 + y_0^2} = \lambda\, z\, \rho$$

$$t(x_1, y_1) = t(r_1) = \mathrm{circ}\left(\frac{r_1}{a}\right) = \begin{cases} 1 & r_1 < a \\ 0 & r_1 > a \end{cases}$$

Nuevamente sobre la abertura incide normalmente una onda plana, monocromática de amplitud unitaria

$$E(x_0, y_0) \propto \mathbb{F}\left[t(x_1, y_1)\right] = \int\int_{-\infty}^{\infty} t(x_1, y_1) \exp\left[-i\, 2\pi\left(f_x x_1 + f_y y_1\right)\right] dx_1\, dy_1$$

Tengamos en cuenta que

$$\begin{aligned} x_1 &= r_1 \cos\theta \quad ; \quad f_x = \rho\cos\alpha \\ y_1 &= r_1 \sin\theta \quad ; \quad f_y = \rho\sin\alpha \quad ; \quad dx_1\, dy_1 = r_1\, dr_1\, d\theta \end{aligned}$$

$$\mathbb{F}\left[t(r_1)\right] = \int_0^{2\pi}\int_0^{\infty} \mathrm{circ}\left(\frac{r_1}{a}\right) \exp\left[-i\, 2\pi\, \rho\, r_1\left(\cos\alpha\cos\theta + \sin\alpha\sin\theta\right)\right] r_1\, dr_1\, d\theta =$$

$$= \int_0^{\infty} \mathrm{circ}\left(\frac{r_1}{a}\right) r_1 \left[\int_0^{2\pi} \exp\left[-i\, 2\pi\, r_1\, \rho\, \cos(\theta - \alpha)\right] d\theta\right] dr_1$$

Expresemos el término entre corchetes en una forma alternativa, para ello recurramos a las funciones de Bessel

$$J_n(x) = \frac{i^{-n}}{2\pi} \int_0^{2\pi} \exp(i\,x\cos\phi)\exp(i\,n\phi)\,d\phi \quad para\; n=0 \quad J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(i\,x\cos\phi)\,d\phi$$

Luego si identificamos $\begin{cases} x & \rightarrow & 2\pi\,\rho\,r_1 \\ \phi & \rightarrow & \theta-\alpha \end{cases}$ tenemos que $d\phi = d\theta$ y los límites de integración que para θ eran de 0 a 2π pasan a ser para ϕ de $-\alpha$ a $2\pi-\alpha$.

Por otra parte, teniendo en cuenta que el signo de la exponencial no tiene importancia por ser J_0 una función par y que integrar entre 0 y 2π ó entre $-\alpha$ y $2\pi-\alpha$ es igual por ser una función de período 2π. Entonces podemos escribir

$$\mathbb{F}\left[t(r_1)\right] = 2\pi\int_0^{\infty} \mathrm{circ}\left(\frac{r_1}{a}\right) r_1\, J_0\left(2\pi\,\rho\,r_1\right) dr_1$$

Esta es la *transformada de Fourier-Bessel*. Luego

$$\mathbb{F}\left[t(r_1)\right] = 2\pi\int_0^{a} r_1\, J_0\left(2\pi\,\rho\,r_1\right) dr_1$$

Llamando $u = 2\pi\, r_1\,\rho \quad ; \quad du = 2\pi\,\rho\, dr_1$ tenemos que

$$\mathbb{F}\left[t(u)\right] = \frac{2\pi}{\left(2\pi\,\rho\right)^2} \int_0^{2\pi a\rho} u\, J_0(u)\, du$$

pero existe una propiedad de las funciones J de Bessel que establece que

$$\frac{d}{du}\left[u^{n+1}\, J_{n+1}(u)\right] = u^{n+1}\, J_n(u) \quad \Rightarrow \quad \mathbb{F}\left[t(u)\right] = \frac{2\pi}{\left(2\pi\,\rho\right)^2}\, u\, J_1(u)\Big|_0^{2\pi a\rho}$$

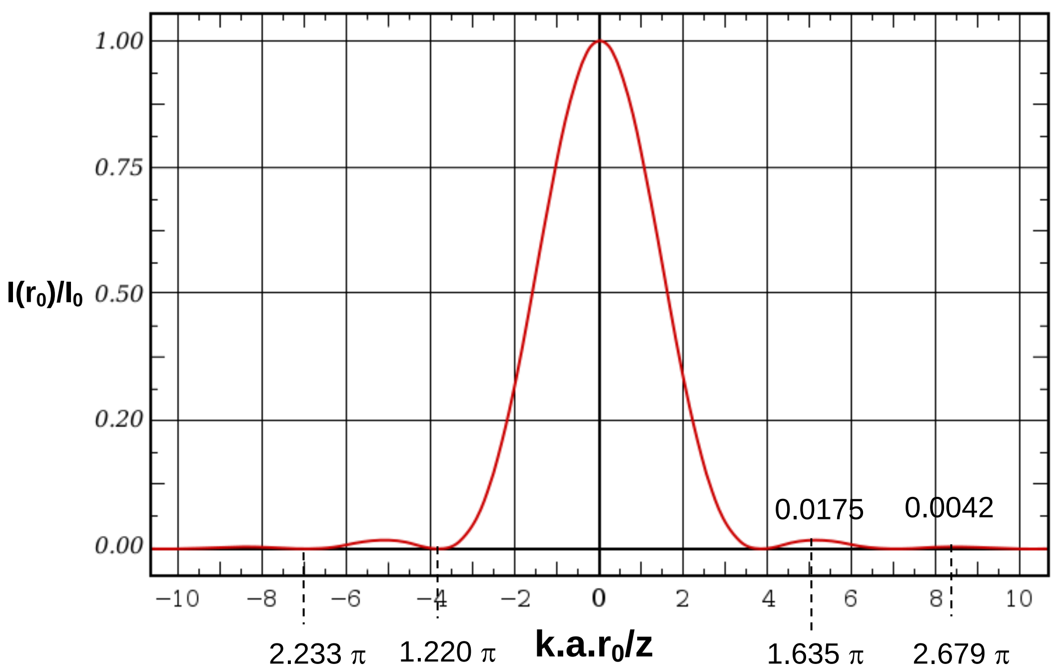
Entonces

$$E(x_0, y_0) = \frac{\exp(ikz)}{i\lambda z} \exp\left(i\,\frac{kr_0^2}{2z}\right) \frac{\lambda z a}{r_0}\, J_1\left(\frac{2\pi a\, r_0}{\lambda z}\right)$$

donde se tuvo en cuenta que $\rho = \frac{r_0}{\lambda z}$. La intensidad vendrá dada por

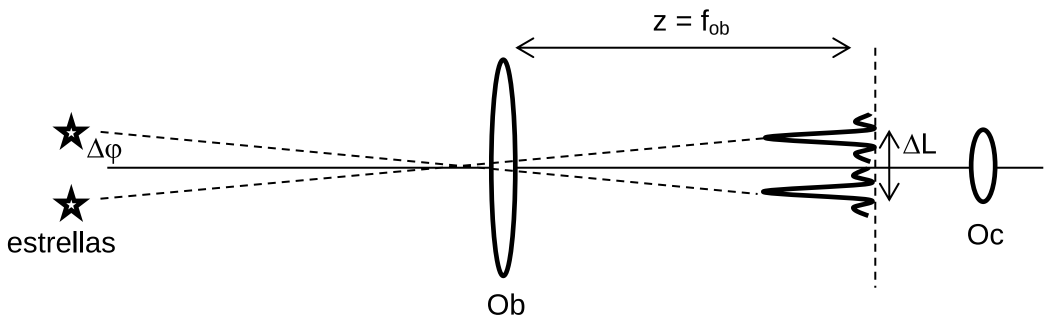
$$I(r_0) = I_0 \left[2 \frac{J_1\left(\frac{kar_0}{z}\right)}{\frac{kar_0}{z}} \right]^2 \quad ; \quad I_0 = \left(\frac{a^2 k}{2z} \right)^2$$

Se elige agruparlo de esta forma para que la expresión entre corchetes tienda a 1 cuando r_0 tiende a 0. Veamos algunas características de esta figura conocida como *diagrama de difracción de Airy*.



Vemos que la intensidad de los máximos decae muy rápidamente. El 84% de la energía total está concentrada en la campana central y el 91% de la misma si incluimos los primeros máximos secundarios.

Teniendo en cuenta lo visto analicemos el poder resolverte de una lente. Supongamos para ello que tenemos un telescopio con el que observamos dos estrellas que se hallan muy cerca una de la otra



En el telescopio *z* = *f0b* ; el ocular sólo se usa para ver adecuadamente la imagen que se forma en el plano focal del objetivo, pero lo que el objetivo no resuelva no lo podrá resolver el telescopio en su conjunto.

Ahora bien, las dos fuentes (estrellas) son incoherentes por lo tanto la intensidad total sobre el plano focal del objetivo será la suma de las intensidades que produce cada estrella. La luz proveniente de cada una de ellas sufre difracción en la abertura circular que constituye la lente Ob, en consecuencia producirá un diagrama de Airy. Cuando las estrellas estén muy próximas se superpondrán dichos diagramas y no podrá resolverse la imagen.

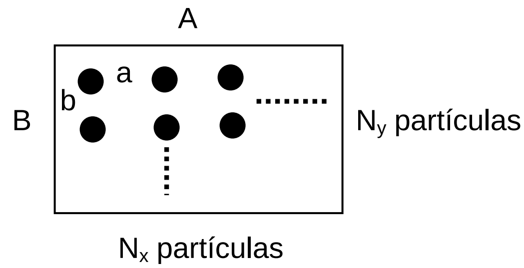
El criterio de Rayleigh establece que dos puntos están justamente resueltos cuando el centro del diagrama de Airy que produce uno coincide con el primer mínimo del otro. Lo que sucede con los otros máximos no tiene importancia debido a su baja intensidad.

Habíamos visto que el primer mínimo se producía en

$$\frac{kar_0}{z} = 1.22\,\pi \quad \Rightarrow \quad r_0^{1\min} = 1.22\,\frac{\lambda z}{2a}$$

En este caso *2.a* es el diámetro de la lente y *z* = *f0b* su distancia focal. Entonces si observamos el dibujo vemos que $\Delta L_{\min} = 1.22\,\dfrac{\lambda\, f_{ob}}{2a}$ ó escrito en términos angulares, y aproximando la tangente por el ángulo $\Delta\varphi_{\min} = 1.22\,\dfrac{\lambda}{2a}$

Ejercicio 6: Considere una distribución de partículas idénticas como en la figura, su espaciado es *a* y *b* periódico. Calcule el diagrama de difracción de Fraunhofer que produce teniendo en cuenta que el diámetro de las partículas es *d*



Ejercicio 7: Dibuje cualitativamente el diagrama de difracción de Fraunhofer de las siguientes aberturas. Discuta los mismos en base a las propiedades de las transformadas

PROPIEDADES DE TRANSFORMACIÓN DE UNA LENTE

Con anterioridad habíamos visto que la integral de difracción de Fresnel venía dada por

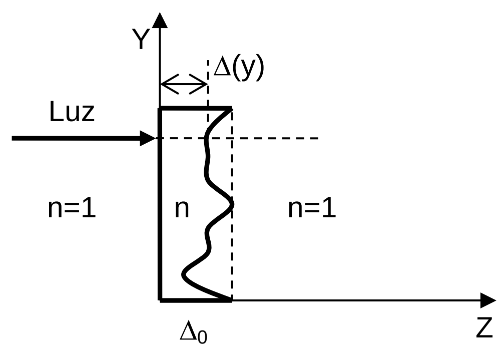
$$E\left(x_0, y_0\right) = \overbrace{\frac{\exp\left(ik\,z\right)}{i\lambda z} \exp\left[i\frac{k}{2z}\left(x_0^2 + y_0^2\right)\right]}^{\mathbf{1}}.$$

$$\underbrace{.\iint_{\Sigma} E\left(x_1, y_1\right) \exp\left[-i\frac{k}{z}\left(x_0 x_1 + y_0 y_1\right)\right] \exp\left[i\frac{k}{2z}\left(x_1^2 + y_1^2\right)\right] dx_1\, dy_1}_{\mathbf{2}}$$

y que el término **2** era la transformada de Fourier de $E\left(x_1, y_1\right)$ a menos del factor $e^{i\frac{k}{2z}\left(x_1^2+y_1^2\right)}$. También vimos que una de las posibilidades para eliminarlo era pedir $z \gg \frac{k}{2}\left(x_1^2 + y_1^2\right)_{Max}$ esto es, estar en las condiciones de difracción de Fraunhofer. La otra consistía en introducir un factor de fase $e^{-i\frac{k}{2z}\left(x_1^2+y_1^2\right)}$. Vamos a estudiar ahora que dicho factor de fase es justamente el introducido por una lente convergente.

Efecto de una lente sobre un frente de ondas

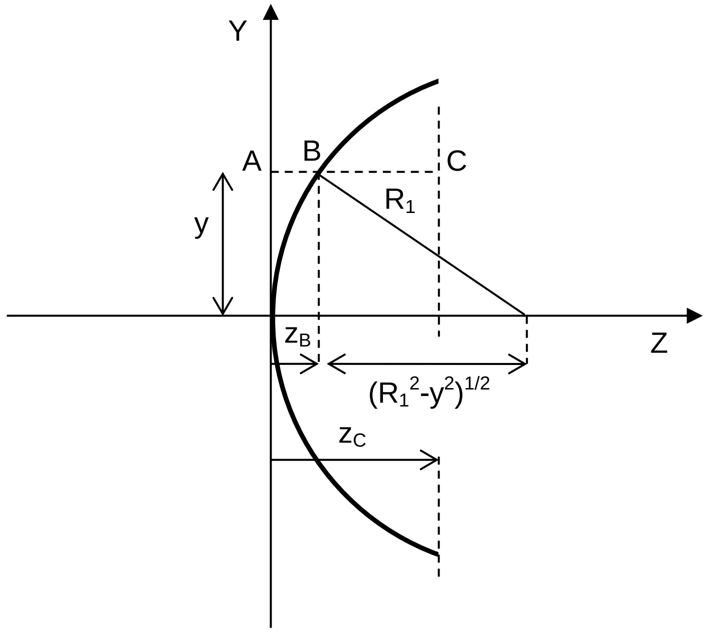
Cuando un haz de luz atraviesa un elemento constituido por un medio transparente (en el caso de las lentes generalmente es vidrio), su amplitud permanece casi inalterada pero su distribución de fase puede variar fuertemente, dependiendo de la forma de dicho elemento. Consideremos el siguiente dibujo



Si sobre el elemento incide un campo $E(x,y)$ a la salida tendremos un campo $E'(x,y)=E(x,y)\cdot t(x,y)$. En una dimensión la fase introducida por $t(x,y)$ será

$$\phi(y)=k\Big[n\Delta(y)+1\big(\Delta_0-\Delta(y)\big)\Big]=k\,\Delta_0+k(n-1)\Delta(y)$$

$k(n-1)\Delta(y)$ es el desfasaje que sufre el rayo con respecto a otro que hubiese viajado la distancia Δ_0 en aire. Obviamente dicho desfasaje depende de $\Delta(y)$, esto es, de la forma del elemento transparente. Veamos entonces la expresión de $\Delta(y)$ para una lente esférica



Vamos a introducir una convención de signos: la luz incide desde la izquierda y las superficies convexas que encuentra (→( ) tienen radio positivo, mientras que las cóncavas (→)) lo tienen negativo. Por ejemplo en una lente () la primer superficie tiene radio positivo y la segunda negativo. Vamos a separar entonces la lente en dos partes. Sobre la primer superficie tendremos:

$$\begin{cases} z_B = R_1 - \sqrt{R_1^2 - y^2} \\ z_C \equiv \Delta_{01} \end{cases} \quad ; \quad \Delta^{S1}(y) = z_c - z_b = \Delta_{01} - R_1\left(1 - \sqrt{1 - \frac{y^2}{R_1^2}}\right)$$

Análogamente para la segunda superficie obtenemos

$$\Delta^{S2}(y) = \Delta_{02} + R_2\left(1 - \sqrt{1 - \frac{y^2}{R_2^2}}\right)$$

Esto, para simplificar, lo hemos planteado en una dimensión pero suponemos que tiene simetría de revolución (en el caso de tratarse de una lente cilíndrica, convergente, obtendríamos la transformada de Fourier en una sola dimensión). Entonces

$$\Delta(x,y)=\Delta^{S1}(x,y)+\Delta^{S2}(x,y)=\Delta_{01}+\Delta_{02}-R_1\left(1-\sqrt{1-\frac{x^2+y^2}{R_1^2}}\right)+R_2\left(1-\sqrt{1-\frac{x^2+y^2}{R_2^2}}\right)$$

$\Delta_{01}+\Delta_{02}=\Delta_0$ es el espesor en el centro de la lente.

Esta expresión es muy compleja y suele utilizarse su aproximación paraxial, esto es, para *x,y* pequeños. En este caso $\sqrt{1-\frac{x^2+y^2}{R^2}}\cong 1-\frac{x^2+y^2}{2R^2}$. Así resulta

$$\Delta(x,y)=\Delta_0-\frac{x^2+y^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

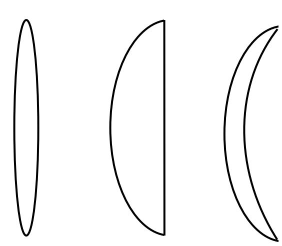
En consecuencia la fase introducida por la lente será:

$$\phi(x,y)=k\,\Delta_0+k(n-1)\left[\Delta_0-\frac{x^2+y^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)\right]=k\,n\,\Delta_0-k(n-1)\frac{x^2+y^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

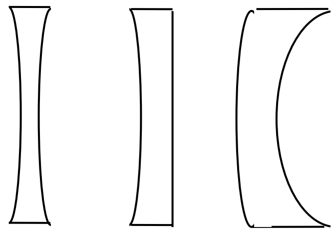
Recordemos la fórmula del constructor de lentes que relaciona las propiedades de la lente (n, R_1 y R_2) ; $\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$. Con lo cual la expresión para la fase queda

$$\phi(x,y)=k\,n\,\Delta_0-\frac{k\left(x^2+y^2\right)}{2f}$$

Cabe destacar que a pesar de la elección de la forma **()** de la lente adoptada para esta deducción, la expresión de la fase es válida para cualquier combinación de radios, siempre que se tenga en cuenta cuales son positivos y cuales negativos. Según su geometría tenemos que:



Tienen f >0



Tienen f <0

En general entonces el factor de transmisión de una lente será:

$$\boxed{t_L(x,y) = P(x,y)\cdot \exp\left[i\,k\,n\,\Delta_0\right]\cdot \exp\left[-i\frac{k}{2f}\left(x^2+y^2\right)\right]} \quad (19)$$

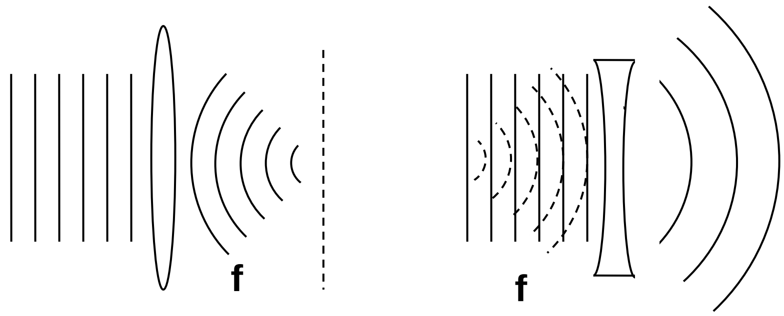
$P(x,y)$ es la función pupila y tiene en cuenta absorciones, el tamaño de la lente y desfasajes introducido por aberraciones. El factor $e^{i\,k\,n\,\Delta_0}$ es una fase constante y por ello suele no tomarse en cuenta.

Analicemos el tercer factor. Habíamos visto que en la aproximación de Fresnel una onda esférica $\frac{A}{r}\cdot e^{\,i\,k\,r}$ era reemplazada por su aproximación cuadrática $\frac{A}{z}e^{i\,k\,z}\cdot e^{\,i\frac{k}{2z}\left(x^2+y^2\right)}$

Consideremos que sobre una lente incide normalmente una onda plana de amplitud unitaria; por el momento no tendremos en cuenta el factor $P(x,y)\cdot e^{i\,k\,n\,\Delta_0}$. Entonces

$$E'(x,y) = 1\cdot t(x,y) = \exp\left[-i\frac{k}{2f}\left(x^2+y^2\right)\right]$$

Vemos que este término puede considerarse como una aproximación cuadrática a una onda esférica que, si *f >0* converge al plano focal detrás de la lente (justamente *f >0* corresponde a lentes convergentes) y si *f <0* la onda diverge de una fuente virtual situada a una distancia *f* delante de la lente (en este caso la lente es divergente)



Lente convergente Lente divergente

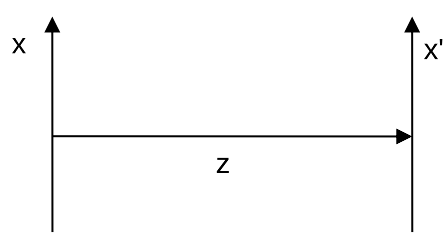
CLASE 7

Transformada de Fourier de una transparencia mediante una lente

Vamos a estudiar ahora cuál es la figura de difracción de una transparencia a través de una lente. Analizaremos dos casos: cuando el objeto está delante de la lente y cuando está detrás de la lente.

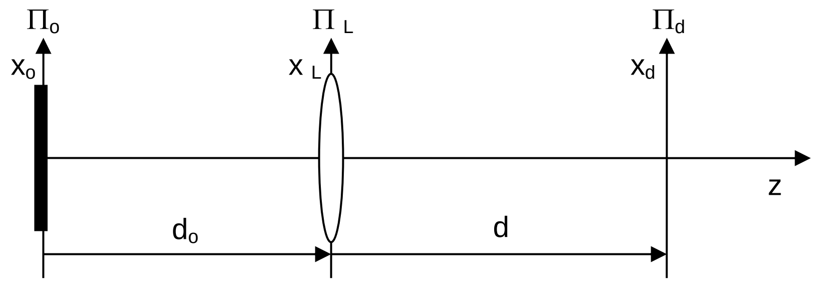
Para realizar esto vamos a utilizar la integral de difracción de Fresnel en una dimensión y luego generalizaremos los resultados obtenidos a dos dimensiones. En ella denominaremos, de forma genérica, con variables sin primar a los puntos desde donde se parte en el cálculo y con variables primadas a los puntos de llegada, esto es:

$$E(x') = \frac{\exp(ikz)}{\sqrt{i\lambda z}} \cdot \exp\left(i\frac{k\,x'^2}{2z}\right)\int_{-\infty}^{\infty} E(x)\cdot\exp\left(i\frac{k\,x^2}{2z}\right)\cdot\exp\left(-i\frac{k\,x'x}{z}\right)dx$$



- <u>Objeto delante de la lente (fuente en el infinito)</u>

Consideremos la siguiente disposición



Primero debemos efectuar una propagación libre desde el objeto situado en el plano Π_o hasta el plano de la lente Π_L . Entonces, según la convención adoptada deberemos hacer los siguientes reemplazos en la integral

$$\begin{cases} z = d_o \\ x = x_o \\ x' = x_L \end{cases} \Rightarrow \quad E\left(x_L^-\right) = \frac{\exp\left(i\,k\,d_o\right)}{\sqrt{i\lambda d_o}} \cdot \exp\left(i\,\frac{k\,x_L^2}{2d_o}\right) \int_{-\infty}^{\infty} E\left(x_o\right) \cdot \exp\left(i\,\frac{k\,x_o^2}{2d_o}\right) \cdot \exp\left(-i\,\frac{k\,x_L x_o}{d_o}\right) dx_o$$

Debemos considerar que $E\left(x_o\right) = E_0 \cdot t\left(x_o\right)$.

Al atravesar la lente tendremos

$$E\left(x_L^+\right) = E\left(x_L^-\right) \cdot t_L\left(x_L\right) = E\left(x_L^-\right) \cdot P\left(x_L\right) \exp\left(-i\,\frac{k\,x_L^2}{2\,f}\right)$$

Ahora haremos nuevamente una propagación libre desde el plano de la lente Π_L hasta un plano genérico Π_d situado a una distancia *d* de la lente.

$$E(x_d)=\frac{\exp(i\,k\,d)}{\sqrt{i\lambda d}}\cdot\exp\left(i\,\frac{k\,x_d^2}{2d}\right).$$

$$.\int_{-\infty}^{\infty}\left[\frac{\exp(i\,k\,d_o)}{\sqrt{i\lambda d_o}}\cdot\exp\left(i\,\frac{k\,x_L^2}{2d_o}\right)\int_{-\infty}^{\infty}E(x_o)\cdot\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_o}{d_o}\right)dx_o\right]\cdot$$

$$.P(x_L)\exp\left(-i\,\frac{k\,x_L^2}{2f}\right)\cdot\exp\left(i\,\frac{k\,x_L^2}{2d}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_d}{d}\right)dx_L$$

$$E(x_d)=\frac{\exp\left[i\,k\,(d+d_o)\right]}{i\lambda\sqrt{dd_o}}\cdot\exp\left(i\,\frac{k\,x_d^2}{2d}\right).$$

$$.\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E(x_o)P(x_L)\cdot\exp\left[i\,\frac{k}{2}\left(\frac{1}{d}+\frac{1}{d_o}-\frac{1}{f}\right)x_L^2\right]\cdot\exp\left[-i\,k\left(\frac{x_o}{d_o}+\frac{x_d}{d}\right)x_L\right].\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)dx_o dx_L$$

Ahora bien, utilicemos el siguiente resultado

$$\int_{-\infty}^{\infty}\exp\left[i\left(a\,x^2-b\,x\right)\right]dx=\sqrt{\frac{\pi}{a}}\cdot\exp\left(i\,\pi/4\right)\cdot\exp\left(-i\,\frac{b^2}{4a}\right)$$

Con lo cual si tomamos $a\equiv\frac{k}{2}\left(\frac{1}{d}+\frac{1}{d_o}-\frac{1}{f}\right)$; $b\equiv k\left(\frac{x_o}{d_o}+\frac{x_d}{d}\right)$; $P(x_L)=1$

tenemos que:

$$E(x_d)=\frac{\exp\left[i\,k\,(d+d_o)\right]}{i\lambda\sqrt{dd_o}}\cdot\exp\left(i\,\frac{k\,x_d^2}{2d}\right)\exp\left(i\,\frac{\pi}{4}\right)\sqrt{\frac{\pi}{\frac{k}{2}\left(\frac{1}{d}+\frac{1}{d_o}-\frac{1}{f}\right)}}.$$

$$.\int_{-\infty}^{\infty}E(x_o)\cdot\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left[-i\,\frac{k^2\left(\frac{x_o}{d_o}+\frac{x_d}{d}\right)^2}{4\,\frac{k}{2}\left(\frac{1}{d}+\frac{1}{d_o}-\frac{1}{f}\right)}\right]dx_o$$

Esta integral se simplifica bastante si elegimos la distancia a la que ubicamos el plano Π_d igual a la distancia focal de la lente. Entonces $d \equiv f \quad ; \quad x_d \equiv x_f \quad ; \quad \Pi_d \equiv \Pi_f$

En este caso

$$E\left(x_f\right)=\frac{\exp\left[i\,k\left(f+d_o\right)\right]}{i\sqrt{\lambda f}}\cdot\exp\left[i\,\frac{k\,x_f^2}{2f}\left(1-\frac{d_o}{f}\right)\right]\exp\left(i\,\frac{\pi}{4}\right)\int_{-\infty}^{\infty}E\left(x_o\right)\exp\left(-i\,\frac{k\,x_f\,x_o}{f}\right)dx_o$$

Recordando que $k=\frac{2\pi}{\lambda}$ y que la frecuencia espacial $f_{X_f}=\frac{x_f}{\lambda f}$ podemos escribir

$$\mathbb{F}_{f_{X\,f}}\left[E\left(x_o\right)\right]=\int_{-\infty}^{\infty}E\left(x_o\right)\exp\left(-i\,2\pi\,f_{X_f}\,x_o\right)dx_o$$

Con lo cual llegamos al resultado final, que en dos dimensiones toma la forma

$$\boxed{E\left(x_f,y_f\right)=\frac{\exp\left[i\,k\left(f+d_o\right)\right]}{i\lambda f}\cdot\exp\left[i\,\frac{k\left(x_f^2+y_f^2\right)}{2f}\left(1-\frac{d_o}{f}\right)\right]\cdot\mathbb{F}_{f_{X\,f},f_{Y\,f}}\left[E\left(x_o,y_o\right)\right]} \qquad (20)$$

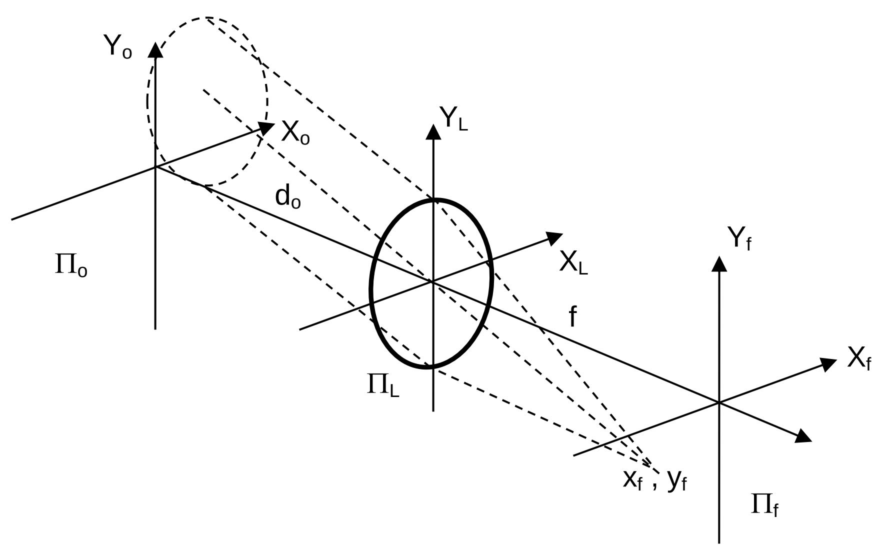
Vemos entonces que el campo en el plano focal de la lente es, a menos de un factor global de fase, proporcional a la transformada de Fourier del campo que emerge de la transparencia. Como lo que se mide, en general, es la intensidad $I=E\cdot E^*=\frac{1}{\lambda^2 f^2}\mathbb{F}\cdot\mathbb{F}^*$ el factor global de fase se elimina. De todos modos vemos que si ubicamos el objeto en $d_o=f$ entonces $E\left(x_f,y_f\right)$ es la transformada de Fourier de $E\left(x_o,y_o\right)$ por una constante ya que el término cuadrático de fase también desaparece, dando la transformada de Fourier exacta.

Otro caso particular a considerar es cuando el objeto está contra la lente. Basta tomar para ello $d_o = 0$.

Para llegar a la ecuación (20) hemos supuesto $P(x_L, y_L) = 1$ despreciando el tamaño finito de la lente. Si así no lo hiciéramos el campo en (x_f, y_f) sería

$$E(x_f, y_f) \propto \int\int_{-\infty}^{\infty} E(x_o, y_o) \exp\left[\frac{ik}{2d_o}\left(x_o^2 + y_o^2\right)\right] \cdot$$
$$\cdot \left\{\int\int_{-\infty}^{\infty} P(x_L, y_L) \cdot \exp\left[\frac{ik}{2d_o}\left(x_L^2 + y_L^2\right) - ik\left(\frac{x_o}{d_o} + \frac{x_f}{f}\right)x_L - ik\left(\frac{y_o}{d_o} + \frac{y_f}{f}\right)y_L\right] dx_L\, dy_L\right\} dx_o\, dy_o$$

Si queremos extraer $P(x_L, y_L)$ fuera de la integral encerrada entre llaves, lo que debemos hacer es proyectar la pupila sobre el objeto y expresarla en coordenadas de (x_o, y_o). Para ello analicemos el siguiente dibujo:



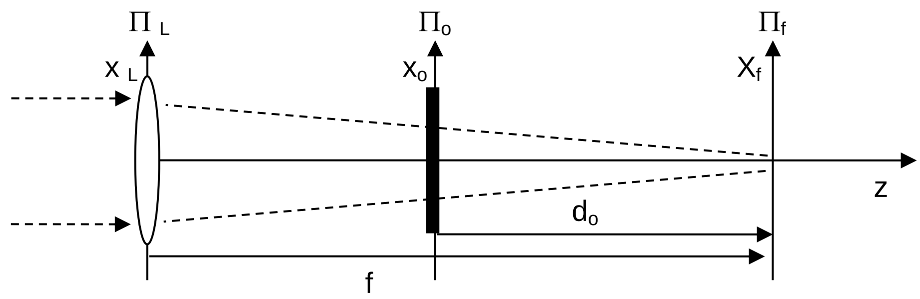
Dado que el punto $\left(x_f, y_f\right)$ se halla en el plano focal, la luz que arriba a dicho punto provendrá de un haz paralelo cuya dirección está especificada por los ángulos $\alpha = \frac{x_f}{f}\,;\, \beta = \frac{y_f}{f}$. Sin embargo de todos los rayos que viajen paralelos a dicha dirección sólo algunos entrarán en la lente, debido a su tamaño finito. Los que entren serán justamente los que están contenidos en la proyección hacia atrás, en la dirección $\alpha\,;\,\beta$, de la pupila. Esto nos delimita en el plano objeto Π_o cuáles son los puntos que contribuyen a la formación del punto luminoso $\left(x_f, y_f\right)$. Si suponemos que en dicha proyección no hay deformaciones, entonces sobre el objeto, la pupila estará centrada en el punto $x_o = -d_o\,\alpha = -d_o\,\frac{x_f}{f}\,;\, y_o = -d_o\,\beta = -d_o\,\frac{y_f}{f}$.

Ahora podemos extraer $P\left(x_L, y_L\right)$ fuera de la integral encerrada entre llaves, la cual se resuelve de igual modo que demostramos anteriormente. Así se obtiene

$$\boxed{\begin{aligned} E\left(x_f, y_f\right) = & \frac{\exp\left[i\,k\left(f + d_o\right)\right]}{i\lambda f} \cdot \exp\left[i\,\frac{k\left(x_f^2 + y_f^2\right)}{2f}\left(1 - \frac{d_o}{f}\right)\right] \cdot \\ & \mathbb{F}_{f_{X\,f}, f_{Y\,f}}\left[E\left(x_o, y_o\right) \cdot P\left(x_o + \frac{d_o x_f}{f}, y_o + \frac{d_o y_f}{f}\right)\right] \end{aligned}} \qquad (21)$$

- <u>Objeto detrás de la lente (fuente en el infinito)</u>

Analicemos ahora el caso del objeto ubicado a una distancia d_o del plano focal de la lente



Sobre la lente incide normalmente un frente de ondas plano, en consecuencia sobre el objeto incide una onda esférica que converge sobre el plano focal Π_f .

En este caso también debemos estudiar la propagación por tramos. Para ello analicemos, al igual que antes, el caso unidimensional para simplificar los cálculos y luego extendemos el resultado final a dos dimensiones. Inmediatamente a la salida de la lente el campo será

$$E\left(x_L^+\right) = A \cdot P\left(x_L\right) \exp\left(-i \frac{k\, x_L^2}{2f}\right)$$

En esta ecuación despreciamos el término de fase constante Δ_0

La zona del objeto que será iluminada corresponderá a la intersección del cono de luz emergente de la lente con el plano Π_0. Esta área se representa matemáticamente por la proyección de la pupila $P\left(x_L\right)$ a través del cono de rayos que cae sobre el objeto. Esto es, sobre el plano objeto tenemos una función pupila efectiva $P\left(\frac{x_o f}{d_0}\right)$ ya que $\frac{x_L}{f} = \frac{x_o}{d_o}$.

Por ejemplo si la lente es circular y de radio R entonces $P\left(x_L\right) = circ\left(\frac{x_L}{R}\right)$. Ahora bien, la proyección de este círculo sobre el plano Π_0 corresponderá a otro círculo de radio

$\frac{Rd_o}{f}$ con lo cual $P'(x_o) = \text{circ}\left(\frac{x_o}{\frac{Rd_o}{f}}\right) = \text{circ}\left(\frac{x_o f}{Rd_o}\right) = P\left(\frac{x_o f}{d_o}\right)$. Si el área iluminada es mayor que la extensión del objeto entonces se toma $P\left(\frac{x_o f}{d_0}\right) = 1$.

Luego de este análisis, hagamos la propagación libre desde la lente hasta el objeto. Entonces ahora, de acuerdo a la nomenclatura utilizada, tenemos que:

$$\begin{cases} z = f - d_o \\ x = x_L \\ x' = x_o \end{cases} \Rightarrow \quad E\left(x_o^-\right) = \frac{\exp\left[i\,k\left(f - d_o\right)\right]}{\sqrt{i\lambda\left(f - d_o\right)}} \cdot \exp\left[i\,\frac{k\,x_o^2}{2\left(f - d_o\right)}\right] \cdot$$

$$\cdot \int_{-\infty}^{\infty} AP\left(\frac{x_o f}{d_o}\right) \cdot \exp\left(-i\,\frac{k\,x_L^2}{2f}\right) \cdot \exp\left[i\,\frac{k\,x_L^2}{2\left(f - d_o\right)}\right] \cdot \exp\left(-i\,\frac{k\,x_L x_o}{f - d_o}\right) dx_L$$

$$E\left(x_o^-\right) = \frac{\exp\left[i\,k\left(f - d_o\right)\right]}{\sqrt{i\lambda\left(f - d_o\right)}} \cdot \exp\left[i\,\frac{k\,x_o^2}{2\left(f - d_o\right)}\right] \cdot AP\left(\frac{x_o f}{d_o}\right) \cdot$$

$$.\int_{-\infty}^{\infty} \exp\left[i\,\frac{k\,d_o\,x_L^2}{2f\left(f - d_o\right)}\right] \cdot \exp\left(-i\,\frac{k\,x_o x_L}{f - d_o}\right) dx_L$$

Recordando la resolución de la integral

$$\int_{-\infty}^{\infty} \exp\left[i\left(a\,x^2 - b\,x\right)\right] dx = \sqrt{\frac{\pi}{a}} \cdot \exp\left(i\,\pi/4\right) \cdot \exp\left(-i\,\frac{b^2}{4a}\right)$$

y tomando $a \equiv \frac{kd_o}{2f\left(f - d_o\right)}$; $b \equiv \frac{k\,x_o}{f - d_o}$ tenemos que

$$E\left(x_o^-\right)=\exp\left[i\,k\left(f-d_o\right)\right]\sqrt{\frac{f}{id_o}}\cdot\exp\left(-i\,\frac{k\,x_o^2}{2d_o}\right)\cdot AP\left(\frac{x_o f}{d_o}\right)\cdot\exp\left(i\,\frac{\pi}{4}\right)$$

El campo a la salida del objeto será $E\left(x_o^+\right)=E\left(x_o^-\right)\cdot t\left(x_o\right)$ donde $t\left(x_o\right)$ es el factor de transmisión del mismo.

Ahora debemos hacer nuevamente una propagación libre desde el objeto hasta el plano focal. Entonces

$$\begin{cases} z=d_o \\ x=x_o \\ x'=x_f \end{cases} \Rightarrow \quad E\left(x_f\right)=\frac{\exp\left(i\,k\,d_o\right)}{\sqrt{i\lambda d_o}}\cdot\exp\left(i\,\frac{k\,x_f^2}{2d_o}\right)\cdot\exp\left[i\,k\left(f-d_o\right)\right]\cdot\exp\left(i\,\pi/4\right)\sqrt{\frac{f}{id_o}}A.$$

$$\cdot\int_{-\infty}^{\infty}t\left(x_o\right)P\left(\frac{x_o f}{d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_f\,x_o}{d_o}\right)dx_o$$

$$E\left(x_f\right)=\exp\left(i\,k\,f\right)\cdot\exp\left(i\,\pi/4\right)\sqrt{\frac{f}{i^2\lambda d_o^2}}A\cdot\exp\left(i\,\frac{k\,x_f^2}{2d_o}\right)\cdot\mathbb{F}_{f_{X\,f}}\left[t\left(x_o\right)\cdot P\left(\frac{x_o f}{d_o}\right)\right]$$

Extendiendo este resultado a dos dimensiones tenemos que

$$\boxed{\begin{array}{l} E\left(x_f,y_f\right)=\exp\left(i\,k\,f\right)\cdot\dfrac{A\,f}{i\lambda d_o^2}\cdot\exp\left[i\,\dfrac{k\left(x_f^2+y_f^2\right)}{2d_o}\right]\cdot \\ \qquad .\mathbb{F}_{f_{X\,f,Y\,f}}\left[t\left(x_o,y_o\right)\cdot P\left(\dfrac{x_o f}{d_o},\dfrac{y_o f}{d_o}\right)\right] \end{array}} \qquad (22)$$

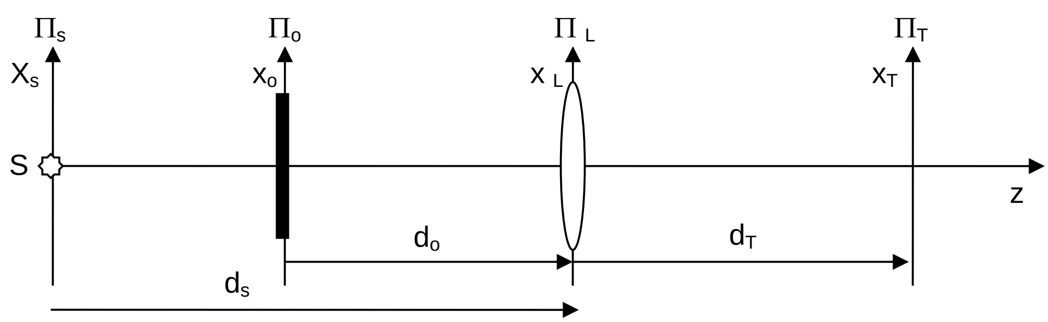
El resultado que hemos obtenido es prácticamente el mismo al encontrado para el objeto delante de la lente. De hecho si aquí tomamos $d_o=f$ y en la expresión (21) $d_o=0$ obtenemos el mismo resultado ya que ambos corresponden al caso del objeto contra la lente.

Sin embargo cuando trabajamos con el objeto delante de la lente obtuvimos frecuencias espaciales $f_{X_f} = \dfrac{x_f}{\lambda f}$; $f_{Y_f} = \dfrac{y_f}{\lambda f}$ para cualquier posición del objeto. Al contrario, para el objeto situado detrás de la lente, encontramos que $f_{X_f} = \dfrac{x_f}{\lambda d_o}$; $f_{Y_f} = \dfrac{y_f}{\lambda d_o}$ por lo tanto el tamaño de la transformada depende de la posición del objeto.

Esta propiedad puede resultar muy útil en las aplicaciones de filtrado espacial, tal como veremos más adelante.

- Objeto delante de la lente (fuente a distancia finita)

Es muy usual pensar que la transformada de Fourier de un objeto se halla en el plano focal de la lente utilizada para realizarla. Esta configuración constituye un caso especial en el que la fuente se halla en el infinito, esto es, se ilumina al sistema con ondas planas. Veremos ahora que sucede si la fuente luminosa se halla a distancia finita ya que esta es una configuración habitual en el laboratorio pero está muy poco tratada en los libros.



Consideremos que en el plano Π_s hay una fuente puntual ubicada en la coordenada x_S a la que representaremos como una delta de Dirac. Vale decir que sobre el objeto llegarán ondas esféricas. Comencemos entonces calculando la propagación desde el sitio donde se halla la fuente hasta el plano objeto, justo después de atravesar el mismo.

$$E\left(x_o^+\right) = \frac{\exp\left[i\,k\left(d_s - d_o\right)\right]}{\sqrt{i\lambda\left(d_s - d_o\right)}} \cdot \exp\left[i\,\frac{k\,x_o^2}{2\left(d_s - d_o\right)}\right] \cdot At\left(x_o\right) \cdot \exp\left[i\,\frac{k\,x_s^2}{2\left(d_s - d_o\right)}\right] \cdot \exp\left(-i\,\frac{k\,x_o x_s}{d_s - d_o}\right)$$

El campo detrás de la lente, considerando la propagación hasta la misma y teniendo en cuenta su factor de transmisión, será

$$E\left(x_L^+\right)=\frac{\exp\left(i\,k\,d_s\right)}{i\lambda\sqrt{\left(d_s-d_o\right)d_o}}\cdot A\cdot\exp\left[i\,\frac{k\,x_s^2}{2\left(d_s-d_o\right)}\right]\cdot\exp\left(i\,\frac{k\,x_L^2}{2d_o}\right)\cdot$$
$$\cdot\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left[i\,\frac{k\,x_o^2}{2\left(d_s-d_o\right)}\right]\cdot\exp\left(-i\,\frac{k\,x_o x_s}{d_s-d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_o}{d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_L^2}{2f}\right)dx_o$$

Luego la propagación libre hasta el plano Π_T dará la expresión

$$E\left(x_T\right)=\frac{\exp\left[i\,k\left(d_s+d_T\right)\right]}{\left(i\lambda\right)^{3/2}\sqrt{\left(d_s-d_o\right)d_o d_T}}\cdot A\cdot\exp\left[i\,\frac{k\,x_s^2}{2\left(d_s-d_o\right)}\right]\cdot\exp\left(i\,\frac{k\,x_T^2}{2d_T}\right)\cdot$$
$$\cdot\int\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left(i\,\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left(i\,\frac{k\,x_L^2}{2d_o}\right)\cdot\exp\left[i\,\frac{k\,x_o^2}{2\left(d_s-d_o\right)}\right].\exp\left(-i\,\frac{k\,x_o x_s}{d_s-d_o}\right)$$
$$\cdot\exp\left(-i\,\frac{k\,x_L x_o}{d_o}\right)\cdot\exp\left(-i\,\frac{k\,x_L^2}{2f}\right)\cdot\exp\left(i\,\frac{k\,x_L^2}{2d_T}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_T}{d_T}\right)dx_o\,dx_L$$

Vamos a resolver la integral doble agrupándola de la siguiente manera

$$\int\int_{-\infty}^{\infty}\cdots=\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left[i\,\frac{k\,x_o^2 d_s}{2d_o\left(d_s-d_o\right)}\right]\cdot\exp\left(-i\,\frac{k\,x_o x_s}{d_s-d_o}\right)\cdot$$
$$\cdot\left\{\int_{-\infty}^{\infty}\exp\left[i\,k\left(\frac{1}{2d_o}+\frac{1}{2d_T}-\frac{1}{2f}\right)x_L^2\right]\cdot\exp\left[-i\,k\left(\frac{x_o}{d_o}+\frac{x_T}{d_T}\right)x_L\right]dx_L\right\}\cdot dx_o$$

Primero resolveremos la integral entre llaves teniendo en cuenta que

$$\int_{-\infty}^{\infty}e^{i\left(a\,x^2-b\,x\right)}dx=\sqrt{\frac{\pi}{a}}\cdot e^{i\pi/4}\cdot e^{-i\frac{b^2}{4a}}\;;\;a\equiv k\left[\frac{1}{2d_o}+\frac{1}{2d_T}-\frac{1}{2f}\right];b\equiv k\left(\frac{x_o}{d_o}+\frac{x_T}{d_T}\right)$$

Resulta así que

$$\int\int_{-\infty}^{\infty}\cdots=\sqrt{\frac{\pi}{\frac{k}{2}\left[\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right]}}\cdot\exp\left(i\,\pi/4\right)\cdot\exp\left[-i\frac{k\,x_T^2}{2d_T^2\left(\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right)}\right]\cdot$$

$$.\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left\{i\,k\left[\frac{d_s}{2d_o\left(d_s-d_o\right)}-\frac{1}{2d_o^2\left(\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right)}\right]\cdot x_o^2\right\}\cdot$$

$$.\exp\left\{-i\,k\left[\frac{x_s}{\left(d_s-d_o\right)}+\frac{x_T}{d_T d_o\left(\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right)}\right]x_o\right\}\cdot dx_o$$

Ahora bien, nosotros tomamos el origen de coordenadas en el plano donde se halla la fuente, por lo tanto hacia la derecha las distancias son positivas. Sin embargo si queremos que sea válida la ecuación de formación de imágenes, el origen debe tomarse en la lente con lo cual las distancias a la izquierda de ella serán negativas. Es decir $d_o<0$ y $d_s<0$, en consecuencia debemos reemplazar d_o por $-d_o$ y d_s por $-d_s$ para establecer una coherencia entre la dirección hacia donde se difracta la onda y la ecuación de las lentes. En este caso tenemos

$$E\left(x_T\right)\propto\exp\left\{i\frac{k}{2d_T}\left[1-\frac{1}{d_T\left(-\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right)}\right]x_T^2\right\}\cdot$$

$$\cdot\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left\{i\,k\left[\frac{d_s}{2d_o\left(d_o-d_s\right)}-\frac{1}{2d_o^2\left(-\frac{1}{d_o}+\frac{1}{d_T}-\frac{1}{f}\right)}\right]\cdot x_o^2\right\}.$$

$$\cdot \exp\left\{ -i\,k\left[\frac{x_s}{\left(d_o - d_s\right)} - \frac{x_T}{d_T d_o \left(-\frac{1}{d_o} + \frac{1}{d_T} - \frac{1}{f} \right)} \right] x_o \right\} \cdot dx_o$$

Si elegimos que el plano Π_T sea el plano conjugado del de la fuente, entonces $\frac{1}{d_T} - \frac{1}{d_s} = \frac{1}{f}$ esto es $\frac{1}{d_T} = \frac{1}{d_s} + \frac{1}{f}$. Reemplacemos esta expresión en la de $E\left(x_T\right)$

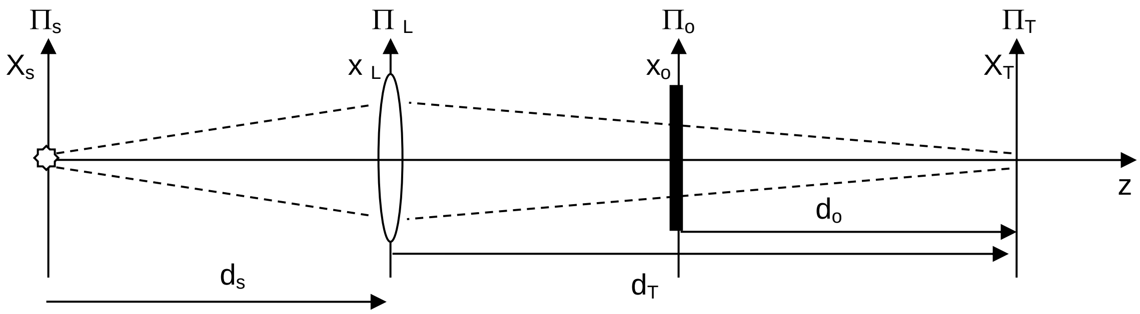
$$E\left(x_T\right) \propto \exp\left\{ i\,\frac{k\left(d_s + f\right)}{2d_s\, f} \left[\frac{f\,d_s + d_s\,d_o}{f\,d_s - f\,d_o} \right] x_T^2 \right\} \cdot$$

$$\cdot \int_{-\infty}^{\infty} t\left(x_o\right) \cdot \exp\left\{ -i\,k\left[\frac{x_s}{\left(d_o - d_s\right)} + \frac{x_T\left(1 + \frac{f}{d_s}\right)}{f\left(1 - \frac{d_o}{d_s}\right)} \right] x_o \right\} \cdot dx_o$$

Vemos que, aunque con una expresión complicada para las frecuencias espaciales, la integral corresponde a la transformada de Fourier de $t\left(x_o\right)$ que ahora se obtiene no en el plano focal sino en el plano conjugado de la fuente. Si dicha fuente está centrada, esto es si $x_s = 0$ y hacemos tender $d_s \to \infty$ entonces recuperamos la expresión que obtuvimos anteriormente para el objeto delante de la lente. Por otra parte si observamos la exponencial delante de la integral vemos que cuando $d_o = -f$ su valor es 1 al igual que antes.

- <u>Objeto detrás de la lente (fuente a distancia finita)</u>

Analicemos por último el caso en el que el objeto está detrás de la lente y la fuente se encuentra a una distancia finita d_s de ella.



Consideraremos, al igual que antes, que la fuente es puntual y tiene coordenadas x_S sobre el plano Π_S. Supondremos que la función pupila vale 1, para simplificar los cálculos. Vamos a hacer una propagación libre desde la fuente hasta la lente.

$$E\left(x_L^-\right)=\frac{\exp\left(i\,k\,d_s\right)}{\sqrt{i\lambda d_s}}\cdot A\cdot\exp\left(i\,\frac{k\,x_L^2}{2d_s}\right)\cdot\exp\left(i\,\frac{k\,x_s^2}{2d_s}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_s}{d_s}\right)$$

Luego de atravesar la lente el campo será

$$E\left(x_L^-\right)=\frac{\exp\left(i\,k\,d_s\right)}{\sqrt{i\lambda d_s}}\cdot A\cdot\exp\left[i\,\frac{k\,x_L^2}{2}\left(\frac{1}{d_s}-\frac{1}{f}\right)\right]\cdot\exp\left(i\,\frac{k\,x_s^2}{2d_s}\right)\cdot\exp\left(-i\,\frac{k\,x_L x_s}{d_s}\right)$$

A continuación realicemos una propagación libre justo hasta después del objeto

$$E\left(x_o^+\right)\propto\exp\left(i\,\frac{k\,x_s^2}{2d_s}\right)\cdot\exp\left[i\,\frac{k\,x_o^2}{2\left(d_T-d_o\right)}\right]\cdot$$
$$\cdot\int_{-\infty}^{\infty}t\left(x_o\right)\cdot\exp\left[i\,\frac{k}{2}\left(\frac{1}{d_s}-\frac{1}{f}\right)x_L^2\right]\exp\left[i\,\frac{k\,x_L^2}{2\left(d_T-d_o\right)}\right]\cdot\exp\left(-i\,\frac{k\,x_s x_L}{d_s}\right)\cdot\exp\left[-i\,\frac{k\,x_o x_L}{\left(d_T-d_o\right)}\right]dx_L$$

Luego si propagamos desde el objeto hasta el plano Π_T obtenemos

$$E\left(x_T\right) \propto \exp\left(i\frac{k\,x_T^2}{2d_o}\right)\cdot\int\int_{-\infty}^{\infty} t\left(x_o\right)\cdot\exp\left[i\frac{k\,x_o^2}{2\left(d_T-d_o\right)}\right]\cdot\exp\left(i\frac{k\,x_o^2}{2d_o}\right)\cdot\exp\left[i\frac{k}{2}\left(\frac{1}{d_s}-\frac{1}{f}\right)x_L^2\right]\cdot$$
$$\cdot\exp\left[i\frac{k\,x_L^2}{2\left(d_T-d_o\right)}\right].\exp\left(-i\frac{k\,x_s x_L}{d_s}\right)\cdot\exp\left(-i\frac{k\,x_o x_T}{d_o}\right)\cdot\exp\left[-i\frac{k\,x_o x_L}{\left(d_T-d_o\right)}\right]dx_L$$

Acá nuevamente debemos compatibilizar desde donde medimos las distancias con la convención de signos de la fórmula de las lentes, de modo tal que cambiaremos d_s por $-d_s$. Asimismo agrupemos las integrales de la siguiente manera:

$$E\left(x_T\right) \propto \exp\left(i\frac{k\,x_T^2}{2d_o}\right)\cdot\int_{-\infty}^{\infty} t\left(x_o\right)\cdot\exp\left[i\frac{k}{2}\left(\frac{1}{d_T-d_o}+\frac{1}{d_o}\right)x_o^2\right]\cdot\exp\left(-i\frac{k\,x_o x_T}{d_o}\right)\cdot$$
$$\cdot\left\{\int_{-\infty}^{\infty}\exp\left\{i\frac{k}{2}\left(-\frac{1}{d_s}-\frac{1}{f}+\frac{1}{\left(d_T-d_o\right)}\right)x_L^2\right\}\cdot\exp\left\{-i\,k\left(-\frac{x_s}{d_s}+\frac{x_o}{\left(d_T-d_o\right)}\right)x_L\right\}\cdot dx_L\right\}\cdot dx_o$$

Para resolver la integral entre llaves recurriremos nuevamente al resultado

$$\int_{-\infty}^{\infty}\exp\left[i\left(a\,x^2-b\,x\right)\right]dx=\sqrt{\frac{\pi}{a}}\cdot\exp\left(i\,\pi/4\right)\cdot\exp\left(-i\frac{b^2}{4a}\right)$$

y tomaremos $a\equiv k\left[-\frac{1}{2d_s}+\frac{1}{2\left(d_T-d_o\right)}-\frac{1}{2f}\right]$; $b\equiv k\left(-\frac{x_s}{d_s}+\frac{x_o}{\left(d_T-d_o\right)}\right)$

Al igual que en el caso anterior el plano de interés donde obtendremos la transformada corresponderá al plano conjugado de la fuente con lo cual $\frac{1}{d_s}=\frac{1}{d_T}-\frac{1}{f}$ y las expresiones anteriores resultan

$$a \equiv \frac{k}{2}\frac{d_o}{d_T\left(d_T - d_o\right)} \quad ; \quad b \equiv k\left(-\frac{x_s}{d_T}+\frac{x_s}{f}+\frac{x_o}{\left(d_T-d_o\right)}\right)$$

Resolviendo obtenemos que

$$E\left(x_T\right) \propto \exp\left(i\frac{k\,x_T^2}{2d_o}\right)\cdot \exp\left[i\frac{k}{2}\left(d_T-d_o\right)\left(-\frac{1}{d_T d_o}-\frac{d_T}{d_o f^2}+\frac{2}{d_o f}\right)x_s^2\right].$$
$$.\int_{-\infty}^{\infty} t\left(x_o\right)\cdot \exp\left[-i\,k\left(\frac{x_T}{d_o}-\frac{x_s}{d_o}+\frac{x_s d_T}{f\,d_o}\right)x_o\right]dx_o$$

Si la fuente está centrada y tomamos $d_T = f$ de modo que el plano conjugado de la fuente sea el plano focal (fuente en el infinito) obtenemos el mismo resultado que para el sistema iluminado con ondas planas.

IV. COHERENCIA PARCIAL

CLASE 8

Hasta ahora todo el análisis que hemos hecho ha sido para iluminación monocromática, esto es, coherente. Sin embargo sabemos que ninguna fuente es estrictamente monocromática (ni siquiera un láser) y por otra parte todas las fuentes presentan un cierto grado de coherencia (aún el sol).

Si bien el procesado óptico coherente es elegante desde el punto de vista matemático, experimentalmente presenta algunas desventajas. Por ejemplo el ruido coherente (speckle) afecta la calidad de la imagen final, es necesario trabajar con elementos perfectamente limpios y en un ambiente libre de polvo ya que cualquier partícula causará difracción, limita el tipo de fuentes luminosas que pueden emplearse, admite sólo ciertos objetos de entrada (quedan excluidos los objetos autoluminosos, y aquellos que introduzcan fases espurias), la luz empleada es monocromática por lo tanto se pierde la información relacionada con el color del objeto, etc. Sin embargo no todo tipo de procesado puede llevarse a cabo con iluminación incoherente.

Qué significa que una fuente sea coherente, incoherente o parcialmente coherente? Que significa coherencia temporal y coherencia espacial? Algunos de estos conceptos los estudiaremos en esta capítulo ya que son necesarios para analizar el fenómeno de formación de imágenes y todos los procesos relacionados desde un punto de vista más completo.

Nosotros veremos solamente los fundamentos de la *Teoría de la Coherencia Parcial* ya que es sumamente extensa, de hecho toda una rama de la óptica se dedica a realizar estos estudios. Un tratamiento más extenso puede encontrarse en *Statistical Optics* de J.Goodman o en *Principles of Optics* de M.Born y E.Wolf y aún más profundo en Optical Coherence and Quantum Optics e Introduction to the Theory of Coherence and Polarization of Light, ambos de E.Wolf.

INTRODUCCIÓN: COHERENCIA ESPACIAL Y TEMPORAL

Las primeras investigaciones sobre este tema fueron realizadas por Verdet en 1869, quien encontró que bajo ciertas condiciones fuentes comúnmente consideradas incoherentes (como por ejemplo el sol) producían interferencia al realizar una experiencia tipo Young.

Sin embargo fue Michelson en 1890 el primero en establecer la relación existente entre el contraste de las franjas de interferencia y la distribución de intensidad de una fuente extensa.

Algunos de los desarrollos más importantes sobre la teoría de coherencia parcial fueron realizados por Van Cittert en 1934 y Zernike en 1938. Ellos determinaron cuál es el grado de coherencia entre dos puntos cualesquiera de una pantalla iluminada por una fuente extensa. Otro avance importante lo realizó Wolf en 1957 con la introducción de la función coherencia mútua.

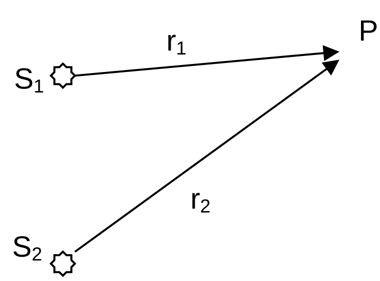
Antes de comenzar con el desarrollo de la teoría conviene aclarar algunos conceptos. A menudo resulta útil dividir los efectos de coherencia en dos clasificaciones: coherencia temporal y espacial. La primera se relaciona con el ancho de banda finito de la fuente y la segunda con el tamaño de la misma. Esta división resulta un tanto artificial ya que ambas características suelen estar, como veremos, íntimamente relacionadas.

Por ejemplo a veces se habla de una fuente “extensa monocromática” para describir a una fuente extensa que emite con una cierta longitud de onda media $\bar{\lambda}$ y ancho de banda $\Delta\lambda << \bar{\lambda}$. En realidad si la onda fuese estrictamente monocromática, esta debe estar asociada a una fuente puntual. Una fuente extensa es aquella en la que cada punto emite una onda cuya diferencia de fase con la emitida por otro punto varía al azar, esto es, tal diferencia no se mantiene constante. Por definición una onda monocromática existe desde $t=-\infty$ por lo que no tiene sentido hablar de diferencias de fase variables en el tiempo. Por ejemplo tomemos un láser como una aproximación a una onda monocromática (teniendo claro que no lo es) y para simular una fuente extensa hagámoslo incidir sobre un vidrio esmerilado. Si el difusor permanece quieto dicha fuente **no es extensa** ya que la diferencia de fase entre dos puntos cualesquiera permanece constante. Será en todo caso un frente de ondas con retardos distintos por zonas, pero fijos. Si en cambio tal difusor se hace rotar, tenemos entonces una buena aproximación a una fuente extensa con la característica $\Delta\lambda << \bar{\lambda}$.

En el otro extremo una fuente temporalmente incoherente, esto es con un gran ancho de banda, no puede ser puntual ya que ella debe estar constituida por varios emisores.

Para aclarar un poco más estos conceptos básicos hagamos un repaso de interferencia. Supongamos que tenemos dos fuentes S_1 y S_2 que emiten ondas esféricas de igual longitud de onda (nuevamente acá, tal como se hace generalmente, se supone que las fuentes emiten con una sola longitud de onda pero se les puede asignar una fase inicial. Esto es una inconsistencia pero resulta útil para simplificar la línea de razonamiento). Los campos producidos por dichas fuentes en un punto P ubicado a distancia r_1 se S_1 y a r_2 de S_2 serán

$$\vec{E}_1 = \vec{E}_{01}(r_1)\exp\left[i\left(k\,r_1 - \omega t + \varepsilon_1\right)\right]$$
$$\vec{E}_2 = \vec{E}_{02}(r_2)\exp\left[i\left(k\,r_2 - \omega t + \varepsilon_2\right)\right]$$



donde ε_1 y ε_2 son las fases iniciales.

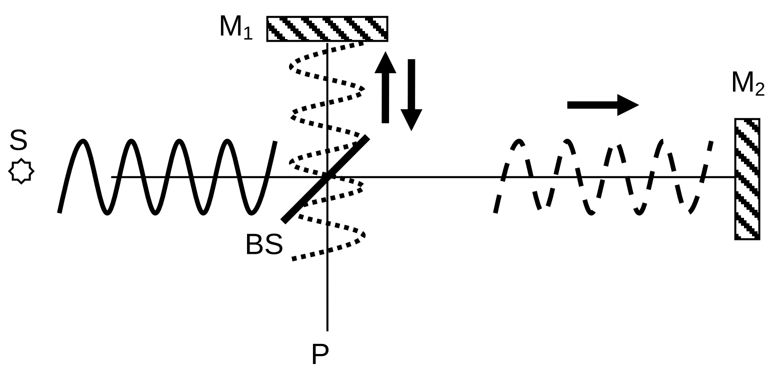
El campo total será $\vec{E} = \vec{E}_1 + \vec{E}_2$ y la intensidad vendrá dada por $I = \left\langle \vec{E}\cdot\vec{E}^* \right\rangle$ donde el símbolo $\langle\ \rangle$ significa promedio temporal. Dado que $\lambda_1 = \lambda_2 = \lambda \quad \Rightarrow \quad \omega_1 = \omega_2 = \omega$ entonces $I = \vec{E}\cdot\vec{E}^*$ ya que la expresión es independiente del tiempo.

$$\boxed{\begin{aligned} I &= E_{01}^2 + E_{02}^2 + \vec{E}_{01}\cdot\vec{E}_{02}\left\{\exp\left\{i\left[k\left(r_1-r_2\right)+\left(\varepsilon_1-\varepsilon_2\right)\right]\right\} + \exp\left\{-i\left[k\left(r_1-r_2\right)+\left(\varepsilon_1-\varepsilon_2\right)\right]\right\}\right\} = \\ &= E_{01}^2 + E_{02}^2 + 2\vec{E}_{01}\cdot\vec{E}_{02}\cos\left[k\left(r_1-r_2\right)+\left(\varepsilon_1-\varepsilon_2\right)\right] \qquad (23)\end{aligned}}$$

En el tratamiento escalar supondremos que $\vec{E}_{01} \parallel \vec{E}_{02}$. Claramente si $\vec{E}_{01} \perp \vec{E}_{02}$ no existe interferencia. En los casos intermedios podemos descomponer en una componente paralela que interfiere y otra perpendicular que contribuirá a un fondo uniforme.

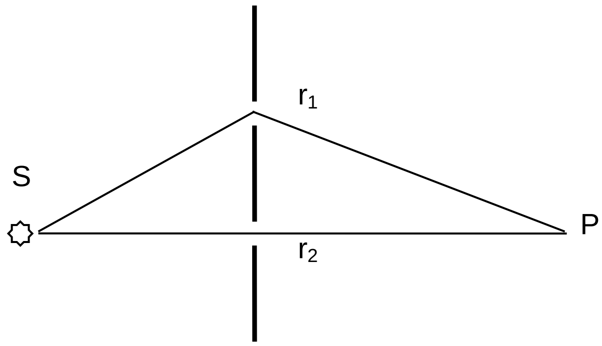
Volviendo a la expresión encontrada para la intensidad, vemos que si $(\varepsilon_1 - \varepsilon_2)$ no permanece constante en el tiempo el diagrama de interferencia no es estable. Esto se soluciona generalmente creando dos fuentes secundarias a partir de una primaria, sin embargo es conveniente analizar esta situación con un poco más de detalle.

Dijimos que a pesar de considerar que los campos son monocromáticos, los trenes de onda son finitos. Qué sucede entonces si la diferencia de caminos es mayor que la longitud de coherencia? Para analizar esta situación veamos el siguiente dibujo



La diferencia de caminos entre ambos brazos de este interferómetro de Michelson es tal que los trenes de ondas que se superponen en *P* no provienen del mismo tren de ondas inicial. Es decir, llega un tren de ondas al *beam splitter* y se separa en dos. El tren que sigue su recorrido hacia el espejo M_1 llega al mismo, se refleja y comienza su recorrido inverso hacia el punto *P*; mientras tanto el que viaja hacia el espejo M_2 aún no ha llegado a reflejarse. En consecuencia en el plano de observación se superpondrá el tren que se reflejó en el espejo M_1 con uno que se reflejo en el espejo M_2 pero que se generó en la fuente un cierto tiempo antes y por lo tanto con otra fase inicial. Acá no se mencionó en ningún momento el tamaño espacial de la fuente, este es claramente un problema de coherencia temporal.

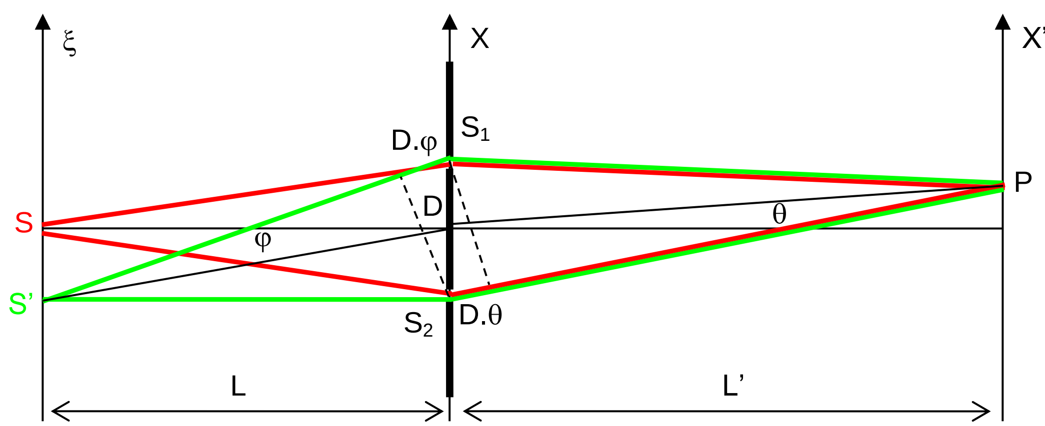
Cabe aclarar que pusimos el ejemplo del interferómetro por división de amplitud pero el mismo razonamiento es válido para uno por división de frente de ondas tal como el de Young. Si la diferencia de caminos es mayor que la longitud de coherencia, en el punto *P* no se superpondrán simultáneamente ondas provenientes del mismo tren emitido por la fuente *S*.



Claramente, en ambos casos, si las ondas fueran realmente monocromáticas tendrían un tren de ondas infinito e interferirían para cualquier diferencia de caminos.

Analicemos ahora el caso de coherencia espacial. Para ello vamos a asumir que el interferómetro está ajustado de forma tal que los trenes de onda emitidos por cada elemento de una fuente extensa interfieren en el punto de observación *P*.

Analicemos el siguiente dibujo donde se esquematiza una experiencia de Young. La diferencia $\left(\varepsilon_1 - \varepsilon_2\right)$ de las fases iniciales de las fuentes S_1 y S_2 estará relacionada con la diferencia de camino entre $\overline{SS_1}$ y $\overline{SS_2}$ por lo tanto podremos incorporarlo a la diferencia de caminos total Δr.



Para estudiar el efecto que tiene sobre la figura de interferencia el tamaño de la fuente comencemos analizando que sucede si tenemos dos fuentes puntuales primarias *S* y *S'*, de igual intensidad.

Supongamos también que ambas aberturas S_1 y S_2, tienen la misma intensidad

$I_1 = E_{01}^2 = I_2 = E_{02}^2 \equiv I_0$. Si reemplazamos en la ecuación (23) obtenemos que cada fuente primaria producirá sobre el plano X' un diagrama de franjas descripto por

$$I = 4 I_0 \cos^2\left(\frac{k}{2}\Delta r\right)$$

Para calcular Δr supongamos además que $L, L' \gg D \quad \Rightarrow \quad \sin\varphi \cong \varphi\,; \sin\theta \cong \theta$

Así obtenemos:

Para *S*

$$\Delta r = \left(\overline{SS_2} + \overline{S_2P}\right) - \left(\overline{SS_1} + \overline{S_1P}\right) = \overline{S_2P} - \overline{S_1P} \cong D\cdot\theta$$

Para *S'*

$$\Delta r' = \left(\overline{S'S_2} + \overline{S_2P}\right) - \left(\overline{S'S_1} + \overline{S_1P}\right) \cong D\cdot\theta - D\cdot\varphi$$

Vemos que la franja de orden cero, que corresponde a la igualdad de caminos $\Delta r = 0$ en un caso (para *S*) cae en $\theta = 0$ y en el otro (para *S'*) en $\theta = \varphi$. Así pues, tendremos dos sistemas de franjas uno para *S* y otro para *S'* que se suman en intensidad, porque *S* y *S'* son incoherentes, y que se hallan desplazados provocando una caída del contraste.

$$I = 4 I_0 \left[\cos^2\left(\frac{k}{2}D\cdot\theta\right) + \cos^2\left(\frac{k}{2}D\cdot\left(\theta - \varphi\right)\right)\right]$$

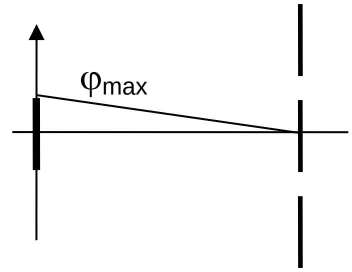
Este es un claro problema de coherencia espacial.

Avancemos un poco más y veamos qué sucede si en vez de *S* y *S'* tenemos una verdadera fuente extensa. En este caso cada elemento de la fuente extensa se comporta como una fuente puntual incoherente con los vecinos. Cada uno de ellos producirá sobre X' un sistema de franjas dado por

$$dI\left(\theta\right) = A\cos^2\left(\frac{kD}{2}\left(\theta - \varphi\right)\right)d\varphi$$

Así la intensidad total será

$$I(\theta)=\int_{-\varphi\max}^{\varphi\max} A\cos^2\left(\frac{kD}{2}(\theta-\varphi)\right)d\varphi=$$
$$=A\varphi_{\max}\left[1+\frac{\sin(k\,D\,\varphi_{\max})}{k\,D\,\varphi_{\max}}\cos(k\,D\,\theta)\right]$$



Michelson define empíricamente el contraste como

$$C=\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}}$$ que en nuestro caso es $$C=\left|\frac{\sin(k\,D\,\varphi_{\max})}{(k\,D\,\varphi_{\max})}\right|$$

Vemos que depende de la extensión de la fuente, a través de φ_{max}, de la longitud de onda y de la distancia entre ranuras *D*.

Después de estos ejemplos básicos de coherencia espacial y temporal pasemos a un tratamiento más riguroso.

REPRESENTACIÓN COMPLEJA DE CAMPOS POLICROMÁTICOS

Hasta ahora nosotros trabajamos con campos monocromáticos y adoptamos, por ser útil matemáticamente, la notación compleja.

Esto es, a un campo real $E^R(t)$ le asociábamos un campo complejo $E(t)$ tal que $E^R(t)=\Re\left[E(t)\right]$

Dado que ahora comenzaremos a trabajar con campos no monocromáticos deseamos encontrar una notación compleja para representarlos, es decir, buscamos una generalización de lo hecho para campos monocromáticos.

Si $E^R(t)$ es una perturbación policromática, podemos representarla en término de la integral de Fourier como $E^R(t)=\int_{-\infty}^{\infty}\varepsilon(\nu)\exp(-2\pi\,i\,\nu\,t)\,d\nu$

$$E^R(t) = \int_{-\infty}^{0} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu + \int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu =$$
$$= \int_{0}^{\infty} \varepsilon(-\nu)\exp(2\pi i \nu t)\,d\nu + \int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu$$

Pero $\varepsilon(\nu)$ es la antitransformada de $E^R(t)$, o sea $\varepsilon(\nu) = \int_{-\infty}^{\infty} E^R(t)\exp(2\pi i \nu t)\,dt$

Además $\left[\varepsilon(\nu)\right]^* = \int_{-\infty}^{\infty} E^R(t)\exp(-2\pi i \nu t)\,dt = \varepsilon(-\nu)$ ya que $E^R(t) = \left[E^R(t)\right]^* \in \Re$

Entonces

$$\int_{0}^{\infty} \varepsilon(-\nu)\exp(2\pi i \nu t)\,d\nu = \int_{0}^{\infty} \varepsilon^*(\nu)\exp(2\pi i \nu t)\,d\nu = \left[\int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu\right]^*$$

Con lo cual

$$E^R(t) = 2\Re\left[\int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu\right] = \Re\left[2\int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu\right]$$

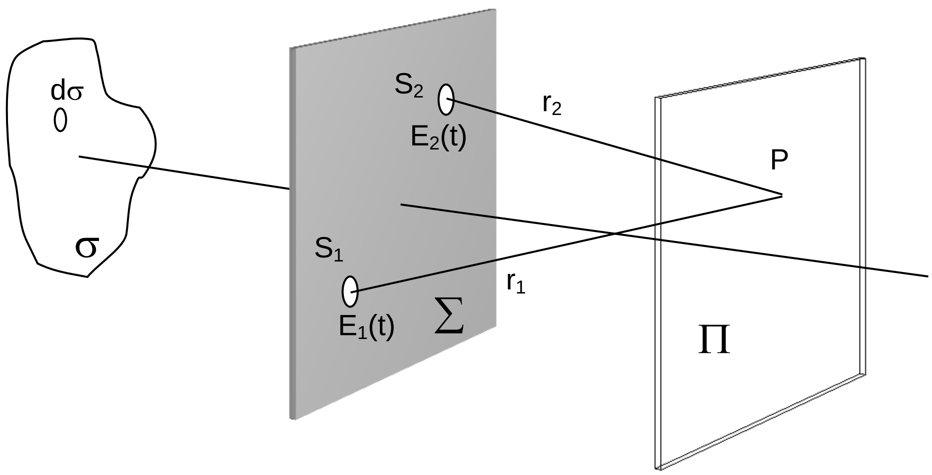
Si ahora llamamos $E^I(t) = \Im\left[2\int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu\right]$ podemos escribir un campo policromático complejo como

$$E(t) = E^R(t) + i\,E^I(t) = 2\int_{0}^{\infty} \varepsilon(\nu)\exp(-2\pi i \nu t)\,d\nu$$

Ahora que encontramos esta expresión generalizada, trabajaremos con $E(t)$

FUNCIÓN COHERENCIA MÚTUA Y GRADO DE COHERENCIA

Consideremos una fuente extensa σ con un ancho de banda finito y dos puntos arbitrarios en el espacio S_1 y S_2 . En dichos puntos la perturbación óptica a tiempo t es $E_1(t)$ y $E_2(t)$ respectivamente. Podemos imaginar que dichos puntos genéricos son fuentes secundarias que reemiten y estamos interesados en ver cómo interactúan los campos reemitidos por S_1 y S_2 . Para ello estos dos puntos se aíslan con dos *pinholes* circulares centrados en S_1 y S_2 sobre una pantalla opaca, esto es, estamos en una situación similar a la experiencia de Young.



Queremos saber cuál es la perturbación resultante en un punto P sobre el plano Π. Las distancias de S_1 y S_2 al punto de observación son r_1 y r_2 respectivamente. Así, si el ancho de banda de la fuente es angosto, el campo en P será

$$E_P(t) = K_1\, E_1(t-t_1) + K_2\, E_2(t-t_2) \qquad \text{donde} \qquad t_1 = \frac{nr_1}{c}\,;\, t_2 = \frac{nr_2}{c}$$ para simplificar

vamos a tomar el índice de refracción n=1.

Veamos como interpretar esta expresión. El campo resultante en P a tiempo t estará originado por la superposición del que a tiempo $(t-t_1)$ fue emitido por S_1 y el que a tiempo $(t-t_2)$ fue emitido por S_2. Los factores K_1 y K_2 están relacionados con las integrales de difracción de Sommerfeld que hemos deducido en el primer capítulo.

Vale la pena aclarar un poco más este punto. Habíamos visto que para un campo policromático la expresión de $E(P,t)$ era

$$E_{Pol}\left(P,t\right)=\frac{1}{2\pi c}\iint_{\Sigma}\frac{\partial}{\partial t}E_{Pol}\left(P_1,t-\frac{r_1}{c}\right)\frac{\cos\left(n,\hat{r}_1\right)}{r_1}\;dS$$

Si suponemos que los *pinholes* son muy pequeños y que el campo sobre su superficie se mantiene constante, entonces podemos sacarlo de la integral de modo que

$$E_{Pol}\left(P,t\right)=\left[\frac{1}{2\pi c}\iint_{\Sigma}\frac{\cos\left(n,\hat{r}_1\right)}{r_1}dS\right]\frac{\partial}{\partial t}E_{Pol}\left(P_1,t-\frac{r_1}{c}\right)$$

Tengamos en cuenta esta expresión para más adelante y recordemos que el factor entre corchetes es real; llamaremos a dicho factor $\tilde{K}$.

Si el campo policromático tiene un ancho de banda muy angosto, entonces la expresión

$$E_{Pol}\left(P_1,t-\frac{r_1}{c}\right)=\int_{\nu}\varepsilon\left(P_{1,}\nu\right)\exp\left[-2\pi i\nu\left(t-\frac{r_1}{c}\right)\right]d\nu$$ se reduce a

$$E\left(P_1,t-\frac{r_1}{c}\right)=\varepsilon\left(P_{1,}\bar{\nu}\right)\exp\left[-2\pi i\bar{\nu}\left(t-\frac{r_1}{c}\right)\right]$$ y por lo tanto

$$\frac{\partial}{\partial t}E\left(P_1,t-\frac{r_1}{c}\right)=-2\pi i\bar{\nu}\,\varepsilon\left(P_{1,}\bar{\nu}\right)\exp\left[-2\pi i\bar{\nu}\left(t-\frac{r_1}{c}\right)\right]=\frac{2\pi\bar{\nu}}{i}E\left(P_1,t-\frac{r_1}{c}\right)$$

con lo cual

$$E\left(P,t\right)=\left[\frac{\bar{\nu}}{i\,c}\iint_{\Sigma}\frac{\cos\left(n,\hat{r}_1\right)}{r_1}dS\right]E\left(P_1,t-\frac{r_1}{c}\right)$$

Vemos que justamente el factor entre corchetes es el propagador K_1 que en este caso es un número imaginario puro.

Volviendo al tema que nos ocupaba, la intensidad en el punto *P* producida por los dos *pinholes* será $I\left(P\right)=\left\langle E_P\left(t\right)\cdot E_P^*\left(t\right)\right\rangle$

$$\begin{aligned}I\left(P\right)=&K_1K_1^*\left\langle E_1\left(t-t_1\right)E_1^*\left(t-t_1\right)\right\rangle+K_2K_2^*\left\langle E_2\left(t-t_2\right)E_2^*\left(t-t_2\right)\right\rangle+\\&+K_1K_2^*\left\langle E_1\left(t-t_1\right)E_2^*\left(t-t_2\right)\right\rangle+K_1^*K_2\left\langle E_1^*\left(t-t_1\right)E_2\left(t-t_2\right)\right\rangle\end{aligned}$$

Supondremos ahora, como en general sucede, que los campos son estacionarios, es decir que no alteran su naturaleza estadística con el tiempo. Así pues, el promedio temporal es independiente de cualquier momento que se escoja.
Entonces en el primer término de la ecuación anterior haremos un corrimiento del origen en t_1 y en los tres restantes en t_2. Resulta así

$$I(P)=|K_1|^2\left\langle E_1(t)E_1^*(t)\right\rangle+|K_2|^2\left\langle E_2(t)E_2^*(t)\right\rangle+$$
$$+K_1K_2^*\left\langle E_1(t-t_1+t_2)E_2^*(t)\right\rangle+K_1^*K_2\left\langle E_1^*(t-t_1+t_2)E_2(t)\right\rangle$$

Si definimos $\mathrm{T}\equiv(t_2-t_1)$ y recordamos la propiedad de los números complejos que establece que dado $z_1;z_2 \quad\Rightarrow\quad z_1\,z_2^*+z_1^*\,z_2=2\Re\left(z_1\,z_2^*\right)$ entonces

$$I(P)=|K_1|^2\left\langle E_1(t)E_1^*(t)\right\rangle+|K_2|^2\left\langle E_2(t)E_2^*(t)\right\rangle+2\Re\left[K_1K_2^*\left\langle E_1(t+\mathrm{T})E_2^*(t)\right\rangle\right]$$

Ahora bien, vamos a definir la *función coherencia mutua* entre los campos luminosos en S_1 y S_2 como

$$\boxed{\Gamma_{12}(\mathrm{T})\equiv\left\langle E_1(t+\mathrm{T})E_2^*(t)\right\rangle=Lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{\tau}E_1(t+\mathrm{T})E_2^*(t)\,dt} \qquad (24)$$

Si recordamos lo expresado en el capítulo 2 vemos que esta ecuación corresponde a una correlación. El significado físico de la misma es que efectúa una comparación entre el campo existente en S_1 a tiempo $t+T$ con el que había en S_2 a tiempo t. Justamente T es la diferencia de tiempo que le demanda llegar al punto P a un campo y al otro. Así pues si a un cierto tiempo dos campos se suman en P, y uno tardó en llegar de S_1 a P un tiempo t_1 y el otro de S_2 a P un tiempo t_2 significa que debieron salir a tiempos distintos. Lo que estamos analizando es cuán correlacionados están esos campos, esto es, que coherencia habrá entre ellos. Planteémoslo de esta forma $E_1(t+\mathrm{T})$ y $E_2(t)$ son dos perturbaciones que se generan en diferentes puntos del espacio y en distintos instantes de tiempo. Las amplitudes y fases de dichas perturbaciones fluctúan en el tiempo. Si las fluctuaciones en S_1 y S_2 son completamente independientes entonces $\Gamma_{12}(\mathrm{T})$, que es un promedio temporal, tendrá valor cero. Si en cambio el campo $E_1(t+\mathrm{T})$ estuviese perfectamente correlacionado con $E_2(t)$ su fase relativa permanecerá inalterada a pesar de las fluctuaciones individuales.

Ahora bien, denotemos $I_n(P)=|K_n|^2 I_{Sn}$ a la irradiancia producida en P por S_n, siendo I_{Sn} la irradiancia en S_n.

Recordando que para fuentes de ancho de banda no muy grande K_1 y K_2 son números imaginarios puros, podemos escribir

$$I(P)=I_1(P)+I_2(P)+2|K_1||K_2|\Re\left[\Gamma_{12}(\mathrm{T})\right]$$

<u>Comentario</u>

Vale la pena aclarar que si la fuente tiene un ancho de banda extenso, entonces de acuerdo a lo visto

$$E_P(t)=\tilde{K}_1\frac{d}{dt}\left[E_1(t-t_1)\right]+\tilde{K}_2\frac{d}{dt}\left[E_2(t-t_2)\right] \qquad \tilde{K}_1\,;\tilde{K}_2\in\Re$$

Entonces

$$I(P)=\left\langle E_P(t)\cdot E_P^*(t)\right\rangle=\left|\tilde{K}_1\right|^2\left\langle\frac{d}{dt}\left[E_1(t-t_1)\right]\frac{d}{dt}\left[E_1^*(t-t_1)\right]\right\rangle+$$
$$+\left|\tilde{K}_2\right|^2\left\langle\frac{d}{dt}\left[E_2(t-t_2)\right]\frac{d}{dt}\left[E_2^*(t-t_2)\right]\right\rangle+2\Re\left[\tilde{K}_1\tilde{K}_2^*\left\langle\frac{d}{dt}\left[E_1(t-t_1)\right]\frac{d}{dt}\left[E_2^*(t-t_2)\right]\right\rangle\right]$$

Se llega a que

$$I(P)=I_1(P)+I_2(P)+2\left|\tilde{K}_1\right|\left|\tilde{K}_2\right|\Re\left[\frac{\partial^2}{\partial\mathrm{T}^2}\Gamma_{12}(\mathrm{T})\right]$$

Continuemos con la expresión que habíamos obtenido para $I(P)$

$$I(P)=I_1(P)+I_2(P)+2\underbrace{|K_1|\sqrt{I_{S1}}}_{\sqrt{I_1(P)}}\underbrace{|K_2|\sqrt{I_{S2}}}_{\sqrt{I_2(P)}}\Re\left[\frac{\Gamma_{12}(\mathrm{T})}{\sqrt{I_{S1}}\sqrt{I_{S2}}}\right]$$

Si recordamos la ecuación (24) tenemos que $\sqrt{I_{S1}}=\sqrt{\Gamma_{11}(0)}$; $\sqrt{I_{S2}}=\sqrt{\Gamma_{22}(0)}$. Vamos a definir entonces *la función grado de coherencia*

$$\boxed{\gamma_{12}(\mathrm{T})\equiv\frac{\Gamma_{12}(\mathrm{T})}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}}} \qquad (25)$$

Ahora bien $\gamma_{12}(\mathrm{T})$ es un número complejo que puede escribirse como $\gamma_{12}(\mathrm{T})=|\gamma_{12}(\mathrm{T})|\exp\left\{i\,ARG\left[\gamma_{12}(\mathrm{T})\right]\right\}$ Luego veremos que es conveniente expresar $ARG\left[\gamma_{12}(\mathrm{T})\right]=\alpha_{12}(\mathrm{T})-2\pi\bar{\nu}\,\mathrm{T}=\alpha_{12}(\mathrm{T})-\delta$ con $\delta=2\pi\bar{\nu}\,\mathrm{T}=\frac{2\pi}{\bar{\lambda}}(r_2-r_1)$

Resulta así

$$\boxed{I(P)=I_1(P)+I_2(P)+2\sqrt{I_1(P)}\sqrt{I_2(P)}\,|\gamma_{12}(\mathrm{T})|\cos\left[\alpha_{12}(\mathrm{T})-\delta\right]} \qquad (26)$$

Esta es la ley general de interferencia para campos estacionarios.

CLASE 9

Veamos entre qué valores está acotado $|\gamma_{12}(\mathrm{T})|$

$$|\Gamma_{12}(\mathrm{T})|^2=\left|Lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{\tau}E_1(t+\mathrm{T})E_2^*(t)\,dt\right|^2\leq$$

$$\leq Lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{\tau}E_1(t+\mathrm{T})E_1^*(t+\mathrm{T})\,dt\cdot Lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{\tau}E_2(t)E_2^*(t)\,dt$$

donde aplicamos la desigualdad de Schwartz. Entonces

$$\left|\Gamma_{12}(\mathrm{T})\right|^2 \leq \left\langle E_1(t)E_1^*(t)\right\rangle \cdot \left\langle E_2(t)E_2^*(t)\right\rangle = \Gamma_{11}(0)\cdot\Gamma_{22}(0)$$

con lo cual

$$\frac{\left|\Gamma_{12}(\mathrm{T})\right|^2}{\Gamma_{11}(0)\cdot\Gamma_{22}(0)} \leq 1 \quad \Rightarrow \quad \left|\gamma_{12}(\mathrm{T})\right| = \frac{\left|\Gamma_{12}(\mathrm{T})\right|}{\sqrt{\Gamma_{11}(0)}\cdot\sqrt{\Gamma_{22}(0)}} \leq 1$$

Vemos que si $\left|\gamma_{12}(\mathrm{T})\right| = 0$ el término de interferencia se anula por lo que este valor corresponde a una fuente incoherente. El máximo peso de dicho término ocurre cuando $\left|\gamma_{12}(\mathrm{T})\right| = 1$ esto es, para fuentes coherentes. Fuera de esos extremos tenemos coherencia parcial $0 < \left|\gamma_{12}(\mathrm{T})\right| < 1$. Justamente $\gamma_{12}(\mathrm{T})$ representa el grado de coherencia.

Comparemos ahora la ecuación (26)

$$I(P) = I_1(P) + I_2(P) + 2\sqrt{I_1(P)}\sqrt{I_2(P)}\,\left|\gamma_{12}(\mathrm{T})\right|\cos\left[\alpha_{12}(\mathrm{T}) - \delta\right]$$

con la (23), obtenida para fuentes puntuales monocromáticas

$$I = E_{01}^2 + E_{02}^2 + 2\vec{E}_{01}\cdot\vec{E}_{02}\cos\left[k(r_1 - r_2) + (\varepsilon_1 - \varepsilon_2)\right]$$

Recordando que $\delta = 2\pi\bar{\nu}\,\mathrm{T} = \frac{2\pi}{\bar{\lambda}}(r_2 - r_1)$, vemos que este factor tiene en cuenta las diferencias de camino entre las fuentes S_1 y S_2 con el punto P, es decir cumple el papel de $k(r_1 - r_2)$ evaluado en un $\bar{\lambda}$. En tanto $\alpha_{12}(\mathrm{T})$ tiene en cuenta las fases iniciales y cumple el papel de $(\varepsilon_1 - \varepsilon_2)$.

Por otra parte vimos que el contraste de las franjas se definía como $C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$

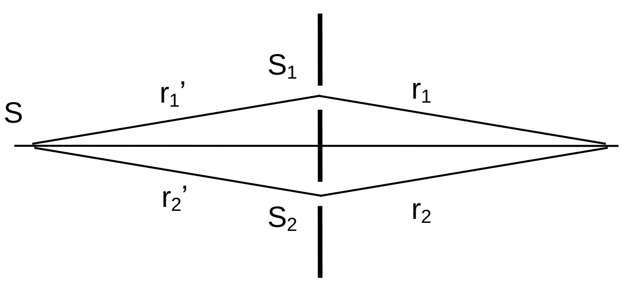
$I_{\max}$ e $I_{\min}$ se dan cuando $\cos[\ \ ]=\pm 1$ respectivamente. Con lo cual tenemos que

$$C = 2\frac{\sqrt{I_1(P)}\sqrt{I_2(P)}}{I_1(P)+I_2(P)}\left|\gamma_{12}(\mathrm{T})\right|$$

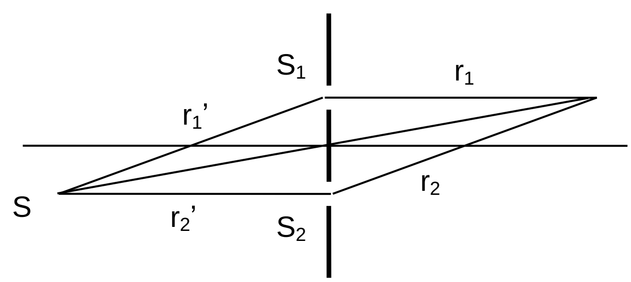
Si $I_1(P)=I_2(P) \quad \Rightarrow \quad C=\left|\gamma_{12}(\mathrm{T})\right|$. En consecuencia aunque no conozcamos las características geométricas de la figura de interferencia podemos decir que existe un término de *bias* dado por $I_1(P)+I_2(P)$ sobre el que se halla un patrón de franjas con un período determinado por $\overline{\lambda}$ modulado por una variación mucho mas lenta dada por $\left|\gamma_{12}(\mathrm{T})\right|$ y por $\alpha_{12}(\mathrm{T})$.

Experimentalmente $\left|\gamma_{12}(\mathrm{T})\right|$ puede medirse a partir del contraste de las franjas, en tanto el apartamiento de la franja central del punto $(r_1-r_2)=0$ es una medida de $\alpha_{12}(\mathrm{T})$.

Veamos esto en el siguiente esquema:



Franja central en $r_1=r_2 \Rightarrow r_1'=r_2' \Rightarrow \varepsilon_1=\varepsilon_2$

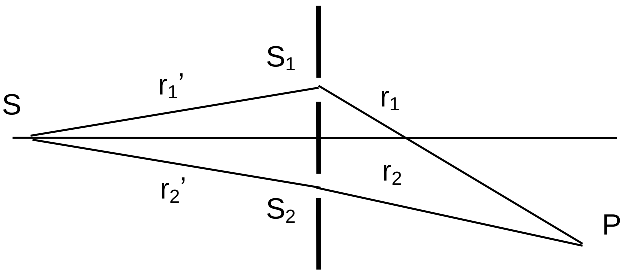


Franja central desplazada $\Rightarrow r_1 \neq r_2 \Rightarrow \varepsilon_1 \neq \varepsilon_2$

La franja central se produce cuando $r_1 + r_1' = r_2 + r_2'$.

Trataremos ahora de integrar el formalismo desarrollado con los conceptos de coherencia espacial y temporal. Como dijimos con anterioridad, aunque ambos efectos están relacionados, vamos a analizarlos separadamente.

Consideremos una fuente primaria, puntual, ubicada sobre el eje de simetría y con un ancho de banda finito



Dado que $r_1' = r_2'$ las perturbaciones ópticas en S_1 y S_2 serán idénticas. Entonces el estudio del diagrama de franjas estará relacionado con los efectos de coherencia temporal y la función coherencia mutua será igual a la autocoherencia, esto es

$$\Gamma_{12}(\mathrm{T}) = \Gamma_{11}(\mathrm{T}) = \Gamma_{22}(\mathrm{T}) = \left\langle E_1(t+\mathrm{T})E_1^*(t)\right\rangle$$

Esta función, así expresada, lo que hace es comparar el campo en un instante (t) con el mismo a tiempo $(t+\mathrm{T})$. Aclaremos un poco este punto, a S_1 y S_2 llega una perturbación idéntica, pero al punto P la perturbación que llega de S_1 emplea en arribar un tiempo $(t+\mathrm{T})$ mientras que la proveniente de S_2 lo hace en un tiempo (t). En consecuencia, para ver si se superponen en P los trenes de onda secundarios generados a partir del mismo tren de ondas primario, debemos comparar como eran los campos en S_1 y S_2 con esa diferencia de tiempos.

Una situación análoga ocurre en un interferómetro por división de amplitud como el Michelson. (T) es la diferencia de longitud de brazos dividida por la velocidad de la luz C.

Así pues, la expresión (26) para $I(P)$ contendrá a $\left|\gamma_{11}(\mathrm{T})\right|$ en vez de a $\left|\gamma_{12}(\mathrm{T})\right|$.

Veamos algunos ejemplos:

Supongamos que existen dos perturbaciones idénticas de la forma $E(t) = E_0 \exp\left[i\Phi(t)\right]$. Dichas perturbaciones son recombinadas por el interferómetro de modo tal que podemos escribir

$$\gamma_{11}(\mathrm{T}) = \frac{\left\langle E_1(t+\mathrm{T})E_1^*(t)\right\rangle}{\left|E_1\right|^2} = \left\langle \exp\left[i\,\Phi(t+\mathrm{T})\right]\cdot\exp\left[-i\,\Phi(t)\right]\right\rangle =$$

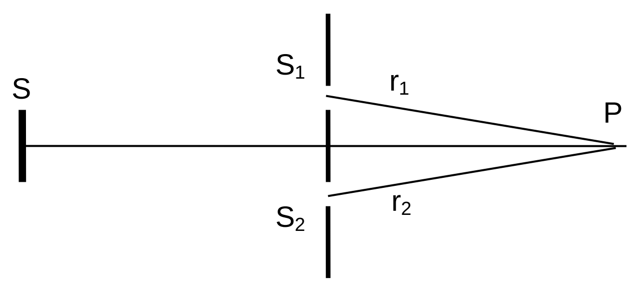
$$= Lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau}\exp\left[i\,\Phi(t+\mathrm{T})\right]\cdot\exp\left[-i\,\Phi(t)\right]dt = Lim_{\tau\to\infty}\frac{1}{\tau}\int_0^{\tau}\exp\left(i\,\Delta\Phi\right)dt$$

Para una onda monocromática $\Phi = \vec{k}\cdot\vec{r} - \omega t \quad \Rightarrow \Delta\Phi = -\omega\mathrm{T}$ por lo tanto

$\gamma_{11}(\mathrm{T}) = \exp(-i\,\omega\mathrm{T}) \quad \Rightarrow \left|\gamma_{11}(\mathrm{T})\right| = 1$ y tendremos coherencia completa.

En el caso opuesto, si el tren de ondas es finito y T es mayor que el tiempo de coherencia, esto es, $\mathrm{T} = \dfrac{r_1 - r_2}{c} > t_C = \dfrac{l_C}{C}$ donde l_C es la longitud de coherencia, entonces $\Delta\Phi$ varía al azar ya que estamos comparando campos provenientes de distintos trenes de onda. En consecuencia, al ser $\gamma_{11}(\mathrm{T})$ un promedio, este es nulo. Vemos entonces que $\gamma_{11}(\mathrm{T})$ es una medida del grado de coherencia temporal de la fuente.

Analicemos que sucede ahora con la coherencia espacial. Para ello retornemos al ejemplo de la experiencia de Young donde la fuente luminosa empleada es extensa y de ancho de banda muy angosto.



Si estudiamos que sucede en un punto *P* ubicado sobre el eje de simetría tendremos, en ese caso, $r_1 = r_2$ y las perturbaciones llegarán simultáneamente. Por lo tanto $\mathrm{T} = 0$ y la evaluación de $\Gamma_{12}(0)$ y $\gamma_{12}(0)$, a partir del contraste de las franjas, nos permitirá tener idea de la coherencia espacial de la fuente. Estas funciones nos permiten evaluar cuán parecidos son los campos en S_1 y S_2 en el mismo instante de tiempo.

Cálculo de $\Gamma_{12}(\mathrm{T})$ para campos policromáticos

Habíamos visto que se podía escribir $E(t)=2\int_0^\infty \varepsilon(\nu)\exp(-2\pi i\nu t)\,d\nu$ luego

$$\Gamma_{12}(\mathrm{T})\equiv\left\langle E_1(t+\mathrm{T})E_2^*(t)\right\rangle = Lim_{\tau\to\infty}\frac{1}{2\tau}\left\{\int_{-\tau}^{\tau}\left[2\int_0^\infty \varepsilon_1(\nu)\exp\left[-2\pi i\nu(t+\mathrm{T})\right]d\nu\right]\cdot\right.$$

$$\left.\cdot\left[2\int_0^\infty \varepsilon_2^*(\nu')\exp(2\pi i\nu' t)\,d\nu'\right]\right\}dt$$

$$\Gamma_{12}(\mathrm{T})=Lim_{\tau\to\infty}\frac{1}{2\tau}\int\int_0^\infty 4\varepsilon_1(\nu)\varepsilon_2^*(\nu')\exp(-2\pi i\nu\mathrm{T})\left[\int_{-\tau}^{\tau}\exp\left[-2\pi i(\nu-\nu')t\right]dt\right]d\nu d\nu'$$

$$\Gamma_{12}(\mathrm{T})=4\int\int_0^\infty Lim_{\tau\to\infty}\frac{\varepsilon_1(\nu)\varepsilon_2^*(\nu')}{2\tau}\exp(-2\pi i\nu\mathrm{T})\,\delta(\nu-\nu')\,d\nu d\nu'$$

$$\Gamma_{12}(\mathrm{T})=4\int_0^\infty \underbrace{Lim_{\tau\to\infty}\frac{\varepsilon_1(\nu)\varepsilon_2^*(\nu)}{2\tau}}_{G_{12}(\nu)}\exp(-2\pi i\nu\mathrm{T})\,d\nu$$

$G_{12}(\nu)$ es la *función densidad espectral mutua*. Entonces podemos expresar

$$\Gamma_{12}(\mathrm{T})=4\int_0^\infty G_{12}(\nu)\exp(-2\pi i\nu\mathrm{T})\,d\nu$$

INTERFERENCIA CON LUZ CUASI-MONOCROMÁTICA

Habíamos visto que $\gamma_{12}(\mathrm{T})=\left|\gamma_{12}(\mathrm{T})\right|\exp\left[i\left(\alpha_{12}(\mathrm{T})-2\pi\bar{\nu}\,\mathrm{T}\right)\right]=\dfrac{\Gamma_{12}(\mathrm{T})}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}}$

Dado que $\Gamma_{11}(0)$ y $\Gamma_{22}(0)$ son reales, la fase de $\gamma_{12}(\mathrm{T})$ será igual a la fase de $\Gamma_{12}(\mathrm{T})$. Entonces

$$\Gamma_{12}(\mathrm{T})=\left|\Gamma_{12}(\mathrm{T})\right|\exp\left[i\left(\alpha_{12}(\mathrm{T})-2\pi\bar{\nu}\,\mathrm{T}\right)\right]=4\int_0^{\infty}G_{12}(\nu)\exp\left(-2\pi i\nu\mathrm{T}\right)d\nu$$

Luego

$$\left|\Gamma_{12}(\mathrm{T})\right|\exp\left[i\,\alpha_{12}(\mathrm{T})\right]=4\int_0^{\infty}G_{12}(\nu)\exp\left[-2\pi i\left(\nu-\bar{\nu}\right)\mathrm{T}\right]d\nu$$

Ahora bien, la longitud de coherencia, esto es el largo del tren de ondas, viene dado por $l_C=C\,t_C$ o sea es la velocidad de la luz por el tiempo de coherencia, pero a su vez el tiempo de coherencia viene dado por $t_C=\dfrac{1}{\Delta\nu}=\dfrac{1}{\nu-\bar{\nu}}$. Si trabajamos con interferómetros en donde la diferencia de caminos son mucho menores que la longitud de coherencia entonces $\dfrac{\left|r_1-r_2\right|}{C}=\mathrm{T}\ll\dfrac{1}{\nu-\bar{\nu}} \quad\Rightarrow\quad \left(\nu-\bar{\nu}\right)\mathrm{T}\ll 1$. En consecuencia

$$\left|\Gamma_{12}(\mathrm{T})\right|\exp\left[i\,\alpha_{12}(\mathrm{T})\right]\cong 4\int_0^{\infty}G_{12}(\nu)\,d\nu=\Gamma_{12}(0)=\left|\Gamma_{12}(0)\right|\exp\left[i\,\alpha_{12}(0)\right]$$

Así, en el caso de fuentes cuasi-monocromáticas $\Gamma_{12}(\mathrm{T})$, $\gamma_{12}(\mathrm{T})$ y $\alpha_{12}(\mathrm{T})$ prácticamente no difieren de $\Gamma_{12}(0)$, $\gamma_{12}(0)$ y $\alpha_{12}(0)$. Se suelen definir

$J_{12}=\Gamma_{12}(0)=\left\langle E_1(t)\,E_2^*(t)\right\rangle$ *función intensidad mutua*

$$\mu_{12} = \gamma_{12}(0) = \frac{\Gamma_{12}(0)}{\left[\Gamma_{11}(0)\Gamma_{22}(0)\right]^{1/2}} = \frac{J_{12}}{\left[J_{11}J_{22}\right]^{1/2}}$$ *factor complejo de coherencia*

En este caso la expresión para la intensidad toma la forma

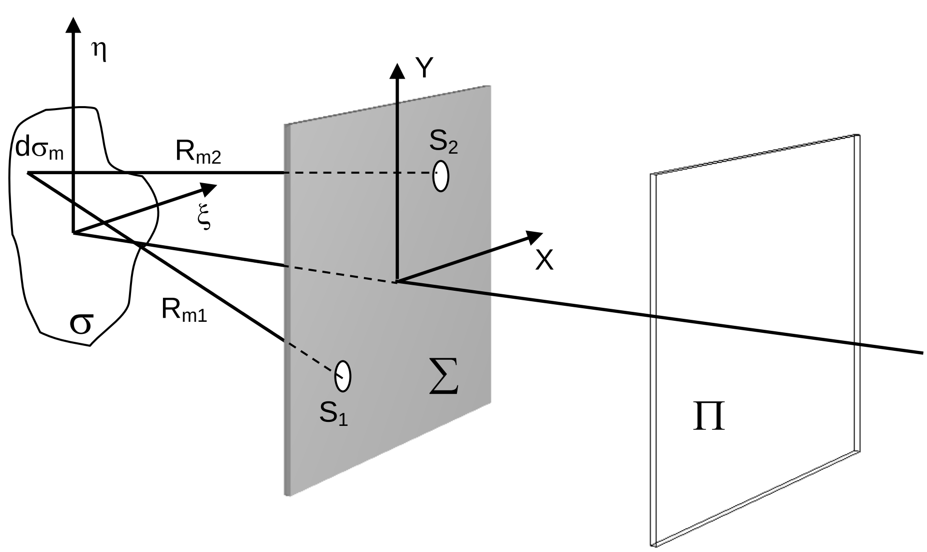
$$I(P) = I_1(P) + I_2(P) + 2\sqrt{I_1(P)}\sqrt{I_2(P)}\,|\mu_{12}|\cos\left[\alpha_{12}(0) - \delta\right]$$

TEOREMA DE VAN CITTERT – ZERNIKE

Este teorema es muy importante y permite establecer una relación entre el tamaño de la fuente utilizada para iluminar un interferómetro y el contraste de las franjas de interferencia obtenidas.

Supongamos que tenemos una fuente extensa y cuasi-monocromática σ que ilumina una plantalla Σ sobre la que se ubican los *pinholes* que actúan como fuentes secundarias S_1 y S_2 en una experiencia de Young. Lo que queremos establecer es qué área sobre dicha pantalla sería iluminada coherentemente de modo que si se ubicasen los *pinholes* dentro de ella producirían sobre un plano de observación Π franjas de interferencia con buen contraste.

Para analizar esta situación veamos el siguiente dibujo y consideremos que la fuente σ está conformada por emisores elementales de área $d\sigma_m$ cuyas dimensiones son tales que $\sqrt{d\sigma_m} \ll \bar{\lambda}$



Vamos a denotar $E_{m1}(t)$ al campo en S_1 producido por el elemento $d\sigma_m$, así resulta que el campo total en esa fuente secundaria será $E_1(t)=\sum_m E_{m1}(t)$.

Dado que estamos interesados en estudiar cuál es el efecto que produce la extensión espacial de la fuente sobre el contraste de las franjas, analizaremos la función intensidad mutua J_{12}

$$J_{12}=\left\langle E_1(t)E_2^*(t)\right\rangle=\sum_m\left\langle E_{m1}(t)E_{m2}^*(t)\right\rangle+\sum_m\sum_{l\neq m}\left\langle E_{m1}(t)E_{l2}^*(t)\right\rangle$$

El último término de esta ecuación es nulo ya que están promediadas en el tiempo contribuciones que provienen de distintos emisores elementales que son incoherentes entre sí (si las emisiones de estos elementos estuviesen parcialmente correlacionadas, debe considerarse dicho término y esta da origen al teorema de Van Cittert – Zernike generalizado).

Vamos a considerar que cada elemento $d\sigma_m$ se comporta como una fuente puntual, por lo tanto el campo $E_m(t)$ irradiado por cada uno de ellos corresponderá al de una onda esférica. Recordemos que una onda de este tipo la podemos describir matemáticamente como $\frac{A}{r}\exp\left[i\left(kr-\omega t+\varepsilon\right)\right]$, o teniendo en cuenta que $k=\frac{2\pi}{\lambda}=\frac{2\pi\nu}{c}\,;\omega=2\pi\nu$ y que podemos agrupar la amplitud y la fase inicial en un cierto A', tomará la forma $\frac{A'}{r}\exp\left[i2\pi\nu\left(\frac{r}{c}-t\right)\right]$. Con lo cual podemos escribir

$$E_{m1}(t)=A_m\left(t-\frac{R_{m1}n}{c}\right)\cdot\frac{\exp\left[-i2\pi\bar{\nu}\left(t-\frac{R_{m1}n}{c}\right)\right]}{R_{m1}}$$

Analicemos esta expresión; $E_{m1}(t)$ es el campo en S_1 producido por el elemento $d\sigma_m$. El tiempo que le demanda en llegar a la radiación desde ese elemento a S_1 es $\frac{R_{m1}n}{c}$ por lo tanto la amplitud y fase inicial están contempladas en $A_m\left(t-\frac{R_{m1}n}{c}\right)$, luego a S_1 llega esa amplitud dividida por R_{m1} y con la fase adicional $\exp\left[i\left(\bar{k}\,R_{m1}-\bar{\omega}t\right)\right]$.

Reemplacemos estas expresiones en el promedio anterior

$$J_{12}=\sum_m\left\langle A_m\left(t-\frac{R_{m1}n}{c}\right)A_m^*\left(t-\frac{R_{m2}n}{c}\right)\right\rangle\cdot\frac{\exp\left[i2\pi\bar{\nu}\,\frac{n}{c}\left(R_{m1}-R_{m2}\right)\right]}{R_{m1}R_{m2}}$$

Dado que los campos son estacionarios podemos efectuar un corrimiento en el tiempo sin afectar la estadística, por lo que podemos escribir

$$J_{12}=\sum_m\left\langle A_m\left(t\right)A_m^*\left(t-\left(R_{m2}-R_{m1}\right)\frac{n}{c}\right)\right\rangle\cdot\frac{\exp\left[i2\pi\bar{\nu}\,\frac{n}{c}\left(R_{m1}-R_{m2}\right)\right]}{R_{m1}R_{m2}}$$

Como estamos bajo la hipótesis de campos cuasi-monocromáticos ($\frac{n}{c}\left|R_{m2}-R_{m1}\right|\ll t_c$), cada elemento de la fuente emite trenes de onda lo suficientemente largos como para que luz proveniente del mismo tren llegue a las fuentes secundarias S_1 y S_2. Quiere decir entonces que la luz proveniente de cada elemento y que llega a S_1 y S_2 tiene la misma fase inicial. En consecuencia podemos escribir

$$J_{12}=\sum_m\left\langle A_m\left(t\right)A_m^*\left(t\right)\right\rangle\cdot\frac{\exp\left[i2\pi\bar{\nu}\,\frac{n}{c}\left(R_{m1}-R_{m2}\right)\right]}{R_{m1}R_{m2}}$$

Ahora debemos pasar al continuo y para ello vamos a definir la densidad de irradiancia de la fuente como $I\left(\xi_m,\eta_m\right)=\frac{\left\langle A_m\left(t\right)A_m^*\left(t\right)\right\rangle}{d\sigma_m}$. Con lo cual tenemos que

$$J_{12} = \sum_m I\left(\xi_m, \eta_m\right) \cdot \frac{\exp\left[i2\pi \bar{\nu} \frac{n}{c} \left(R_{m1} - R_{m2} \right) \right]}{R_{m1} R_{m2}} d\sigma_m$$

Ahora podemos hacer tender el área del elemento a cero y reemplazar la sumatoria por una integral, de modo que

$$J_{12} = \int\int_{\sigma} I\left(\xi, \eta\right) \cdot \frac{\exp\left[i2\pi \bar{\nu} \frac{n}{c} \left(R_1 - R_2 \right) \right]}{R_1 R_2} d\xi \, d\eta \qquad \text{con} \qquad \begin{array}{l} R_1 = R_1\left(\xi, \eta\right) \\ R_2 = R_2\left(\xi, \eta\right) \end{array}$$

Veamos ahora que aproximaciones geométricas podemos hacer en esta integral. Acá, tal como hicimos en la integral de difracción, se deberán tomar distintas aproximaciones para la amplitud y para la fase.

$R_{1,2} = \sqrt{\left(x_{1,2} - \xi\right)^2 + \left(y_{1,2} - \eta\right)^2 + R^2}$ como siempre trabajamos en la aproximación paraxial $R \gg \begin{cases} \sqrt{\xi^2 + \eta^2} \\ \sqrt{x_{1,2}^2 + y_{1,2}^2} \end{cases}$ en consecuencia podemos escribir

$$R_{1,2} \cong R + \frac{\left(x_{1,2} - \xi\right)^2 + \left(y_{1,2} - \eta\right)^2}{2R}$$

En la amplitud aproximaremos $\frac{1}{R_1 R_2} \cong \frac{1}{R^2}$

En la fase tomaremos

$$R_1 - R_2 \cong R + \frac{\left(x_1 - \xi\right)^2 + \left(y_1 - \eta\right)^2}{2R} - R - \left(\frac{\left(x_2 - \xi\right)^2 + \left(y_2 - \eta\right)^2}{2R} \right) =$$

$$= \frac{\left(x_1^2 + y_1^2\right) - \left(x_2^2 + y_2^2\right) - 2\left(x_1 - x_2\right)\xi - 2\left(y_1 - y_2\right)\eta}{2R}$$

Así tenemos que podemos escribir el factor complejo de coherencia (cuyo módulo gobernaba el contraste de las franjas de interferencia para el caso cuasi-monocromático)

$$\mu_{12} = \frac{\exp\left\{ i\pi \bar{\nu}\, \frac{n}{Rc}\left[\left(x_1^2 + y_1^2 \right) - \left(x_2^2 + y_2^2 \right) \right] \right\}}{R^2 \sqrt{I_1}\sqrt{I_2}} \cdot$$

$$\cdot \iint_{\sigma} I(\xi,\eta) \cdot \exp\left\{ -2i\pi \bar{\nu}\, \frac{n}{Rc}\left[(x_1 - x_2)\xi + (y_1 - y_2)\eta \right] \right\} d\xi\, d\eta$$

Teniendo en cuenta que

$$I_1 = J_{11} = \iint_{\sigma} \frac{I(\xi,\eta)}{R_1^2} d\xi\, d\eta \cong \iint_{\sigma} \frac{I(\xi,\eta)}{R^2} d\xi\, d\eta$$

$$I_2 = J_{22} = \iint_{\sigma} \frac{I(\xi,\eta)}{R_2^2} d\xi\, d\eta \cong \iint_{\sigma} \frac{I(\xi,\eta)}{R^2} d\xi\, d\eta$$

Obtenemos una expresión final para el factor complejo de coherencia dada por

$$\boxed{\mu_{12} = \frac{\exp\left\{ i\, \frac{\bar{k}}{2R}\left[\left(x_1^2 + y_1^2 \right) - \left(x_2^2 + y_2^2 \right) \right] \right\} \mathbb{F}_{f_{X_{1,2}}, f_{Y_{1,2}}}\left[I(\xi,\eta) \right]}{\iint_{\sigma} I(\xi,\eta)\, d\xi\, d\eta}} \qquad (27)$$

donde $f_{X_{1,2}} = \frac{(x_1 - x_2)}{\bar{\lambda}\, R}$; $f_{Y_{1,2}} = \frac{(y_1 - y_2)}{\bar{\lambda}\, R}$

Luego de esta deducción encontramos un resultado muy importante: para fuentes cuasi-monocromáticas el contraste de las franjas de interferencia es igual a la transformada de Fourier normalizada de la distribución de intensidad de la fuente.

Un caso particular en donde vimos esta propiedad con anterioridad fue cuando analizamos la experiencia de Young con una fuente extensa lineal, al comienzo de este capítulo. Allí habíamos encontrado que el contraste era igual a la función *sinc*

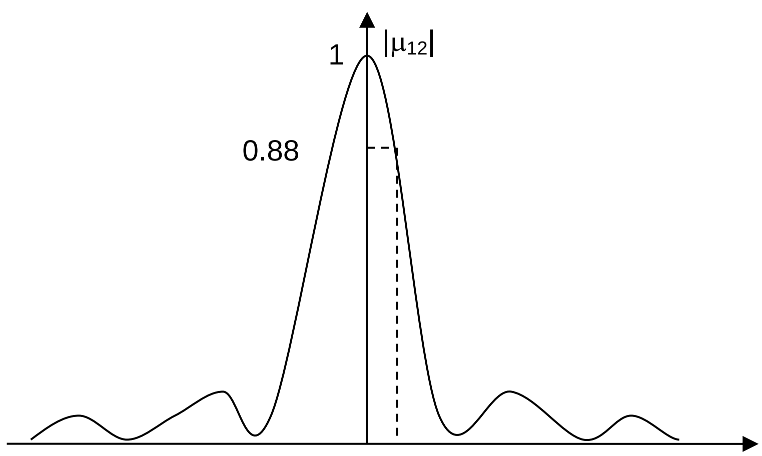
Veamos un ejemplo. Tomemos una fuente circular, uniforme de radio *h*.

$$I(\xi,\eta)=\begin{cases} I_0 & \sqrt{\xi^2+\eta^2}<h \\ 0 & \sqrt{\xi^2+\eta^2}>h \end{cases}$$

Entonces

$$|\mu_{12}|=\frac{\left|2\,J_1\left(\frac{h\bar{k}}{R}\rho\right)\right|}{\left|\frac{h\bar{k}}{R}\rho\right|} \qquad \text{con} \qquad \rho=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

Vemos que ρ es el radio de un círculo tomando como origen uno de los *pinholes*. Vamos a graficar la función que representa a $|\mu_{12}|$

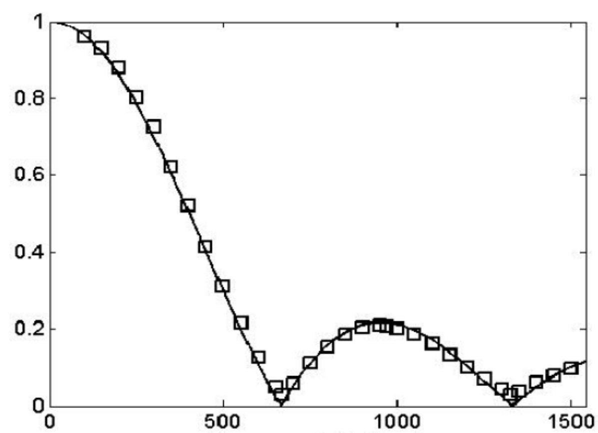


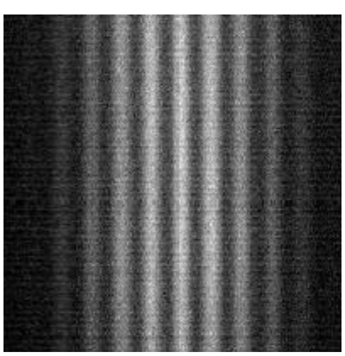
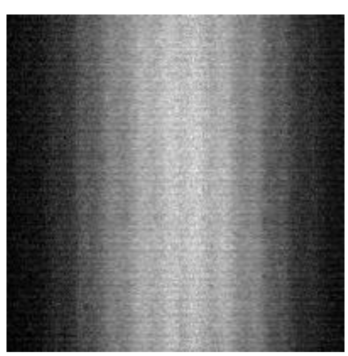
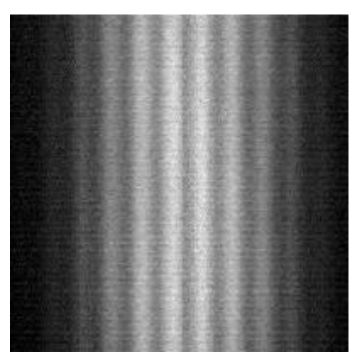
Si suponemos, como se hace habitualmente, que una buena coherencia corresponde a un contraste de 0.88 o mejor, entonces

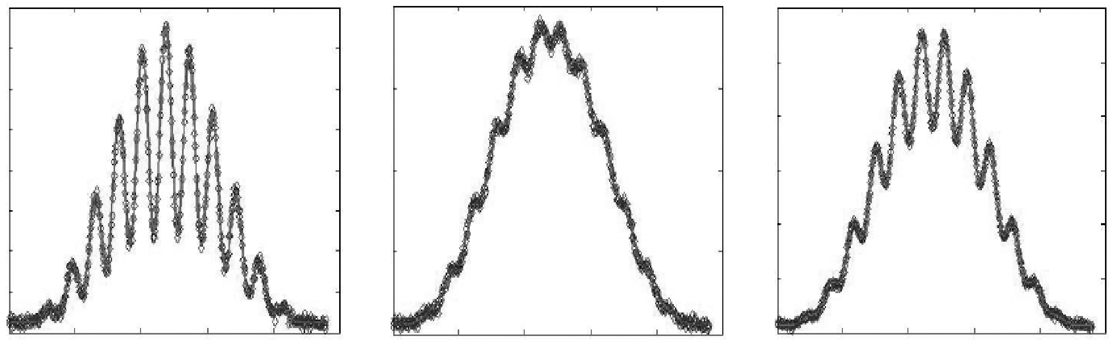
$$\frac{\bar{k}\,\rho h}{R}=1 \quad \Rightarrow \rho=\frac{R}{\bar{k}h}=\frac{R\bar{\lambda}}{2\pi h}=0.16\frac{R\bar{\lambda}}{h}$$

Ahora bien $\qquad \frac{h}{R}\cong\sin\alpha \quad \Rightarrow \rho=0.16\frac{\bar{\lambda}}{\sin\alpha}$

Así el área circular máxima que puede ser iluminada coherentemente por una fuente circular tiene un radio $\rho_{max} = 0.16 \dfrac{\bar{\lambda}}{\sin \alpha}$ O sea, existe un área para la cual la iluminación es coherente por más grande que sea la fuente. Es notable como en este caso el contraste luego de pasar por un mínimo vuelve a crecer a pesar de aumentar el tamaño de la fuente. Esta experiencia fue propuesta por Thompson y Wolf (JOSA **47**, 895, 1957) y luego repetida numerosas veces. A continuación se muestran los resultados obtenidos para una fuente extensa con forma de ranura cuyo ancho varía de 0 a 1500 μm. Esta fuente ilumina un par de ranuras de 40 μm de ancho separadas 250 μm entre sí y ubicadas a 28 cm de dicha fuente. En el grafico se ilustra el contraste teórico y los puntos medidos experimentalmente. Las fotografías corresponden a los casos de 400, 700 y 900 μm respectivamente. Se puede distinguir la oscilación en el módulo del contraste pero además es notorio el cambio de fase pasando la franja central de máximo, en el caso correspondiente a 400 μm, a mínimo en el caso correspondiente a 900 μm, lo que concuerda con la fase de la función sinc que describe la transformada de Fourier de una ranura.





400 μm 700 μm 900 μm

Como ejercicio interesante (a pesar de que no es un caso cuasi-monocromático) puede realizarse la siguiente experiencia: en una pantalla opaca haga dos orificios (*pinholes*) colóquela delante de su pupila y use como fuente primaria al sol. Calcule previamente la distancia máxima a la que pueden estar separados los *pinholes*.

CLASE 10

FÓRMULA DE HOPKINS

En la deducción de la fórmula de Van Cittert – Zernike asumimos que el medio entre la fuente σ y S_1 y S_2 es homogéneo. Esta fórmula se puede generalizar a medios heterogéneos tales como aquellas con diferentes medios de refracción, esto es, podemos incluir lentes.

Con anterioridad habíamos encontrado que el campo producido por cada elemento $d\sigma_m$ era

$$E_{m1,2}(t) = A_m\left(t - \frac{R_{m1,2}\,n}{c}\right) \cdot \frac{\exp\left[-i2\pi\bar{\nu}\left(t - \frac{R_{m1,2}\,n}{c}\right)\right]}{R_{m1,2}}$$

Ahora deberemos reemplazar el factor $\dfrac{\exp\left(i\bar{k}R_{m1,2}\right)}{R_{m1,2}}$ por una función más general $h\left(\xi,\eta,x,y,\bar{\nu}\right)$ que tendrá en cuenta la transmisión del medio. Esto es, el campo en *x,y* será para una perturbación monocromática

$$E(x,y)=\int\int_{-\infty}^{\infty} E(\xi,\eta)h(\xi,\eta,x,y,\bar{v})\,d\xi\,d\eta$$

Para un medio homogéneo tenemos, por el principio de Huygens – Fresnel, que $h(\xi,\eta,x,y,\bar{v})=-i\dfrac{\exp(i\bar{k}R)}{\bar{\lambda}R}$ donde R es la distancia del punto (ξ,η) al (x,y) y donde hemos despreciado el factor de oblicuidad.

Así, generalizando, puede intuirse que el factor $\dfrac{\exp(i\bar{k}R_{m1,2})}{R_{m1,2}}$ debe ser reemplazado por $i\,\bar{\lambda}\,h(\xi,\eta,x,y,\bar{v})$. Tomando esto en cuenta obtendremos ahora

$$J_{12}=\bar{\lambda}^2\int\int_{\sigma} I(\xi,\eta)\cdot h(\xi,\eta,x_1,y_1,\bar{v})h^*(\xi,\eta,x_2,y_2,\bar{v})\,d\xi\,d\eta$$

En consecuencia

$$\mu_{12}=\frac{\bar{\lambda}^2}{\sqrt{I(x_1,y_1)}\sqrt{I(x_2,y_2)}}\int\int_{\sigma} I(\xi,\eta)\cdot h(\xi,\eta,x_1,y_1,\bar{v})h^*(\xi,\eta,x_2,y_2,\bar{v})\,d\xi\,d\eta$$

Algunas veces resulta más útil hacer el siguiente reemplazo

$$U(\xi,\eta,x_1,y_1,\bar{v})=i\bar{\lambda}\sqrt{I(\xi,\eta)}\,h(\xi,\eta,x_1,y_1,\bar{v})$$
$$U(\xi,\eta,x_2,y_2,\bar{v})=i\bar{\lambda}\sqrt{I(\xi,\eta)}\,h(\xi,\eta,x_2,y_2,\bar{v})$$

Con lo cual

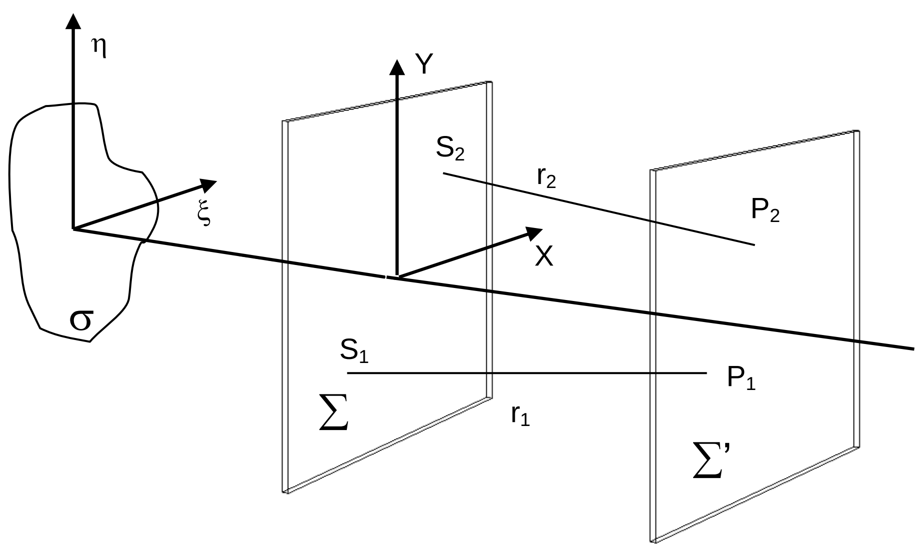
$$\boxed{\begin{aligned}\mu_{12}&=\frac{1}{\sqrt{I(x_1,y_1)}\sqrt{I(x_2,y_2)}}\int\int_{\sigma} U(\xi,\eta,x_1,y_1,\bar{v})U^*(\xi,\eta,x_2,y_2,\bar{v})\,d\xi\,d\eta\\ J_{12}&=\int\int_{\sigma} U(\xi,\eta,x_1,y_1,\bar{v})U^*(\xi,\eta,x_2,y_2,\bar{v})\,d\xi\,d\eta\end{aligned}} \tag{28}$$

Estas son las fórmulas de Hopkins y podrían interpretarse como el factor complejo de coherencia y la intensidad mutua entre dos puntos sobre el plano Σ sobre el que arriban las perturbaciones ópticas $U\left(\xi,\eta,x,y,\bar{\nu}\right)$ producidas por diferentes fuentes puntuales equivalentes ubicadas en los distintos puntos $\left(\xi,\eta\right)$ de σ que emiten con una frecuencia $\bar{\nu}$.

PROPAGACIÓN DE LA INTENSIDAD MUTUA

Supongamos, como lo hicimos con anterioridad, que una fuente cuasi-monocromática σ ilumina una pantalla virtual Σ sobre la que conocemos, para cualquier par de puntos sobre dicha pantalla, cuál es la intensidad mutua.

Deseamos ahora averiguar cómo se propaga la intensidad mutua de Σ a otra pantalla Σ', es decir, estamos interesados en saber cuál es la intensidad mutua entre dos puntos cualesquiera sobre Σ'.



De acuerdo a lo que hemos visto con las fórmulas de Hopkins, si $U\left(\xi,\eta,P_1\right)$ y $U\left(\xi,\eta,P_2\right)$ son las perturbaciones en los puntos P_1 y P_2, ubicados sobre Σ', debidas a una fuente puntual ($d\sigma$) monocromática ubicada en $\left(\xi,\eta\right)$ sobre σ, entonces

$$J_{P_1P_2} = \iint\limits_{\sigma} U\left(\xi,\eta,P_1\right) U^*\left(\xi,\eta,P_2\right) d\xi\, d\eta$$

Ahora bien, dado que *U* es una perturbación monocromática podemos aplicar para su propagación de Σ a Σ' la ecuación de Huygens – Fresnel, de modo que

$$U\left(\xi,\eta,P_1\right) = \iint\limits_{\Sigma} U\left(\xi,\eta,x_1,y_1\right) \cdot \frac{\exp\left(i\bar{k} r_1\right)}{r_1} q_1\, dx_1\, dy_1$$

donde $\left(x_1, y_1\right)$ son variables que pueden tomar cualquier valor sobre Σ y $q_1 = \frac{1}{i\bar{\lambda}} \cos\left(\hat{n},\hat{r}_1\right)'$. Análogamente

$$U\left(\xi,\eta,P_2\right) = \iint\limits_{\Sigma} U\left(\xi,\eta,x_2,y_2\right) \cdot \frac{\exp\left(i\bar{k} r_2\right)}{r_2} q_2\, dx_2\, dy_2$$

Con lo cual podemos expresar

$$J_{P_1P_2} = \iint\limits_{\Sigma} \iint\limits_{\Sigma} \frac{\exp\left[i\bar{k}\left(r_1 - r_2\right)\right]}{r_1\, r_2} q_1\, q_2^* \left\{ \iint\limits_{\sigma} U\left(\xi,\eta,x_1,y_1\right) U^*\left(\xi,\eta,x_2,y_2\right) d\xi\, d\eta \right\} \cdot dx_1\, dy_1 \cdot dx_2\, dy_2$$

Pero $$J_{S_1S_2} = \iint\limits_{\sigma} U\left(\xi,\eta,x_1,y_1\right) U^*\left(\xi,\eta,x_2,y_2\right) d\xi\, d\eta$$ con lo cual

$$\boxed{J_{P_1P_2} = \iint\limits_{\Sigma} \iint\limits_{\Sigma} \frac{\exp\left[i\bar{k}\left(r_1 - r_2\right)\right]}{r_1\, r_2} q_1\, q_2^* \cdot J_{S_1S_2} \cdot dx_1\, dy_1 \cdot dx_2\, dy_2} \qquad (29)$$

Esta es la fórmula deducida por Zernike para la propagación de la intensidad mutua. En el caso especial en que $P_1 = P_2 \equiv P$ entonces $J_{P_1P_2} = I\left(P\right)$ y resulta

$$I\left(P\right)=\iint\limits_{\Sigma}\iint\limits_{\Sigma}\sqrt{I\left(x_1,y_1\right)}\sqrt{I\left(x_2,y_2\right)}\cdot\mu_{S_1S_2}\,\frac{\exp\left[i\,\bar{k}\left(r_1-r_2\right)\right]}{r_1\,r_2}\,q_1\,q_2^{*}\cdot dx_1\,dy_1\cdot dx_2\,dy_2$$

Esta integral expresa a la intensidad en un punto P como la suma de las contribuciones de cada par de puntos S_1 y S_2 de una superficie Σ. La contribución de cada par de elementos depende de las intensidades sobre S_1 y S_2 y es pesada por el factor grado de coherencia $\mu_{S_1S_2}$. Podría pensarse que esta fórmula es el equivalente a Huygens – Fresnel para la propagación de la intensidad de campos parcialmente coherentes.

Aquí nuevamente supusimos que existe un medio homogéneo entre Σ y Σ', si así no fuera debemos reemplazar, tal como lo hicimos en la deducción de la fórmula de Hopkins, el factor $q\frac{\exp\left(i\,k\,r\right)}{r}$ por la apropiada función de transmisión $h\left(x,y,P\right)$ con lo cual

$$J_{P_1P_2}=\iint\limits_{\Sigma}\iint\limits_{\Sigma}h\left(x_1,y_1,P_1\right)h^{*}\left(x_2,y_2,P_2\right)q_1\,q_2^{*}\cdot J_{S_1S_2}\cdot dx_1\,dy_1\cdot dx_2\,dy_2$$

V. SISTEMAS FORMADORES DE IMÁGENES

El conocimiento de las bases de la teoría de la coherencia parcial nos permitirá avanzar otro paso en el análisis del fenómeno de formación y procesado de imágenes. Sin embargo antes de avocarnos a esta tarea resultará útil la introducción al concepto de sistemas lineales

SISTEMAS LINEALES

Existen muchos sistemas físicos cuya respuesta a un estímulo complejo es exactamente igual a la que se obtendría sumando las respuestas del mismo a cada estímulo individual que conforma el estímulo complejo. Tales sistemas se denominan lineales.

Matemáticamente esto lo podemos representar de la siguiente forma. Sea $L\{\ \}$ el operador que representa al sistema y que aplicado a una función de entrada $g_E(x,y)$ produce una función de salida $g_S(x',y')$, o sea $L\{g_E(x,y)\} = g_S(x',y')$.

El sistema se denomina lineal si

$$\begin{aligned} L\left\{a\cdot g_{E1}(x,y)+b\cdot g_{E2}(x,y)\right\} &= a\cdot L\left\{g_{E1}(x,y)\right\}+b\cdot L\left\{g_{E2}(x,y)\right\} = \\ &= a\cdot g_{S1}(x',y')+b\cdot g_{S2}(x',y') \end{aligned}$$

con a y b constantes complejas.

Como mencionamos anteriormente la ventaja de trabajar con sistemas lineales es que nos permite expresar la respuesta del sistema a una entrada arbitraria en término de las respuestas a ciertas entradas elementales. Así resulta de gran importancia encontrar una descomposición conveniente de la entrada. Una de las posibles descomposiciones la podemos realizar basándonos en la propiedad de la función δ que establece que

$$g_E(x,y) = \int\int_{-\infty}^{\infty} g_E(x_o,y_o)\,\delta(x-x_o,y-y_o)\,dx_o\,dy_o$$

Esto es, expresamos $g_E(x,y)$ como una combinación lineal de δ convenientemente pesadas y desplazadas. La respuesta del sistema a esta entrada será:

$$g_S(x',y') = L\left\{\int\int_{-\infty}^{\infty} g_E(x_o,y_o)\,\delta(x-x_o,y-y_o)\,dx_o\,dy_o\right\}$$

Si observamos la ecuación en donde se definía matemáticamente un sistema lineal, vemos que $g_E(x_o,y_o)$ cumple el papel de los coeficientes a, b. Entonces

$$g_S(x',y') = \int\int_{-\infty}^{\infty} g_E(x_o,y_o)\,L\left\{\delta(x-x_o,y-y_o)\right\}dx_o\,dy_o$$

Así $L\{\ \}$ opera sobre las funciones elementales. Se denota $h(x',y',x_o,y_o)$ a la respuesta del sistema en el punto (x',y') del espacio de salida para la función δ ubicada en (x_o,y_o) del espacio de entrada

$$h(x',y',x_o,y_o) = L\{\delta(x-x_o,y-y_o)\}$$

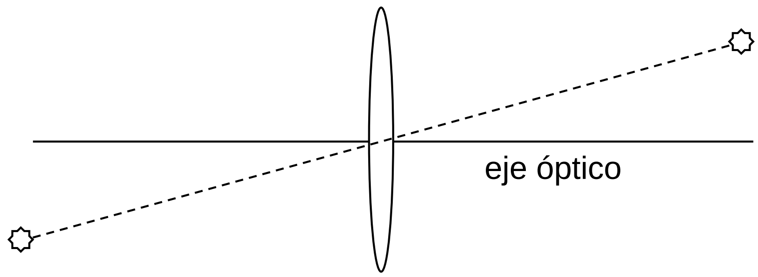
La función *h* es llamada la *respuesta al impulso del sistema*. Así la relación entre la señal de entrada y la de salida será

$$g_S(x',y') = \int\int_{-\infty}^{\infty} g_E(x_o,y_o)\,h(x',y',x_o,y_o)\,dx_o\,dy_o$$

Una subclase de sistemas lineales son los denominados invariantes. En óptica un sistema espacialmente invariante o *isoplanático* es aquel en el que la función $h(x',y',x_o,y_o)$ sólo depende de las distancias $(x'-x_o)\,;\,(y'-y_o)$. Para tales sistemas tenemos que

$$h(x',y',x_o,y_o) = h(x'-x_o,y'-y_o)$$

Así un sistema formador de imágenes será espacialmente invariante si la imagen de una fuente puntual sólo cambia de posición y no de forma funcional a medida que desplazamos la fuente. Obviamente esto no es válido si hay aberraciones, en consecuencia la mayoría de los sistemas ópticos no son isoplanáticos sobre todo el campo. Sin embargo, dado que la formación de imágenes es un fenómeno local, es decir los haces que forma determinado punto imagen provienen del punto objeto correspondiente, se puede hablar de isoplanatismo local alrededor del eje que pasa por el centro del sistema y une los dos puntos



En estos casos podemos expresar la ecuación anterior como:

$$\boxed{g_S\left(x',y'\right)=\int\int_{-\infty}^{\infty} g_E\left(x_o,y_o\right)h\left(x'-x_o,y'-y_o\right)dx_o\,dy_o} \qquad (30)$$

En esta ecuación se reconoce claramente una convolución bidimensional de la función objeto con la función respuesta al impulso del sistema $g_S = g_E \otimes h$. De acuerdo a las propiedades de las transformadas de Fourier podemos escribir

$$\mathbb{F}\left[g_S\right]=\mathbb{F}\left[g_E\right]\cdot\mathbb{F}\left[h\right] \text{ esto es } G_S\left(f_x,f_y\right)=G_E\left(f_x,f_y\right)\cdot H\left(f_x,f_y\right) \qquad (31)$$

$$\text{donde}\quad H\left(f_x,f_y\right)=\int\int_{-\infty}^{\infty} h\left(x_o,y_o\right)\exp\left[-2\pi i\left(f_x x_o+f_y y_o\right)\right]dx_o dy_o$$

La función *H* es llamada *función transferencia del sistema* y da cuenta de los efectos que este causa en el dominio de las frecuencias.

En esta deducción aparecen ligadas dos visiones distintas, desarrolladas casi paralelamente, del proceso de formación de imágenes.

Por un lado Lord Rayleigh (1848 – 1919) en el año 1896 propuso un modelo en el que visualizaba cada punto del objeto como una fuente puntual (una δ) cuyas ondas emitidas eran difractadas por la lente dando como imagen de cada una de ellas un diagrama de

Airy centrado en el punto imagen ideal correspondiente. Así el plano imagen estaba cubierto por una distribución de diagramas de Airy más o menos superpuestos.

Claramente esta línea de razonamiento es la que se sigue cuando escribimos la función de entrada como una suma de funciones δ y encontrábamos como actuaba el sistema sobre cada una de ellas.

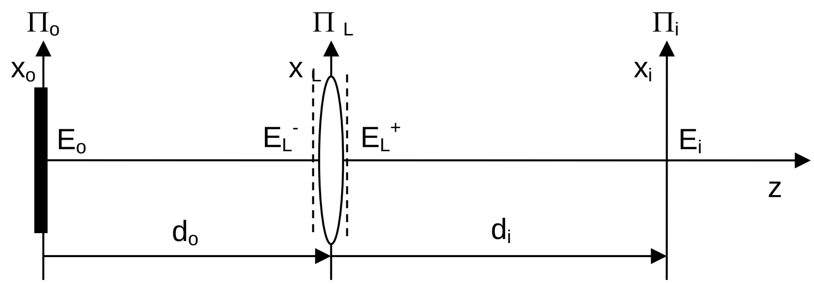
Alternativamente Ernst Abbe (1840 – 1905) en 1873 propuso que un objeto actuaba como una superposición de redes de difracción de distintas frecuencias y orientaciones, emitiendo luz en las direcciones correspondientes a sus órdenes. Si el sistema formador de imágenes no tiene un tamaño infinito entonces no podrá captar toda la luz que emerge del objeto. Actúa así como un filtro pasa-bajos que rechaza frecuencias espaciales mayores que un valor dado. En este caso la imagen no corresponde exactamente al objeto sino que más bien se relacionará con un objeto ficticio cuya transformada de Fourier coincida con la brindada por la lente. Los órdenes difractados presentes en el plano transformado se pueden pensar como fuentes secundarias con amplitudes, fases y posiciones necesarias como para sintetizar la imagen.

Esta visión coincide con la que hemos estudiado anteriormente y con lo expresado en la ecuación (31). Al transformar $g_E(x, y)$ se está descomponiendo esta función en sus frecuencias constitutivas $\left(f_x, f_y\right)$. Luego, multiplicando su espectro $G_E\left(f_x, f_y\right)$ por $H\left(f_x, f_y\right)$se están teniendo en cuenta los efectos del sistema sobre las frecuencias espaciales, por último la transformación inversa de $G_S\left(f_x, f_y\right)$ simplemente nos permite obtener la señal de salida $g_S(x', y')$.

Como vemos ambas teorías son esencialmente la misma y según el caso una es más o menos conveniente de usar que la otra.

FORMACIÓN DE IMÁGENES CON ILUMINACIÓN COHERENTE MONOCROMÁTICA

Analicemos primero el proceso de formación de imágenes a través de una lente simple. Para ello utilizaremos la misma geometría empleada cuando estudiamos la figura de difracción producida por un objeto a una distancia d_o delante de la lente.



El objeto está iluminado por una onda plana, monocromática y la distribución de campo inmediatamente después del mismo es $E_o\left(x_o, y_o\right)$. A una distancia *di* detrás de la lente se forma una imagen del mismo con una distribución de campo $E_i\left(x_i, y_i\right)$. Dado que el fenómeno de propagación de ondas es lineal podemos escribir:

$$E_i\left(x_i, y_i\right) = \int\int_{-\infty}^{\infty} h\left(x_i, y_i, x_o, y_o\right) E_o\left(x_o, y_o\right) dx_o\, dy_o$$

donde $h\left(x_i, y_i, x_o, y_o\right)$es el campo producido en $\left(x_i, y_i\right)$ por un emisor puntual ubicado en $\left(x_o, y_o\right)$. Claramente este enfoque sigue la línea de razonamiento de Rayleigh. Las propiedades del sistema óptico estarían completamente descriptas si se encontrase una expresión para $h\left(x_i, y_i, x_o, y_o\right)$.

Para hallar esta función tomemos como objeto al emisor puntual ubicado en $\left(x_o, y_o\right)$. De ese punto emergerán ondas esféricas que van a incidir sobre la lente luego, en la aproximación paraxial, el campo justo delante de la misma será

$$E_L^-\left(x_L, y_L\right) = \frac{1}{i\lambda\, d_o} \exp\left[\frac{i\,k}{2d_o}\left[\left(x_L - x_o\right)^2 + \left(y_L - y_o\right)^2\right]\right]$$

Al pasar la lente tendremos

$$E_L^+\left(x_L, y_L\right) = E_L^-\left(x_L, y_L\right) P\left(x_L, y_L\right) \exp\left[-\frac{i\,k}{2\,f}\left(x_L^2 + y_L^2\right)\right]$$

donde $P\left(x_L, y_L\right)$ es la función pupila de la lente y f es su distancia focal

Finalmente aplicando la integral de propagación de Fresnel hasta el plano imagen Π_i tenemos que

$$h\left(x_i, y_i, x_o, y_o\right) = \frac{1}{i\lambda\, d_i} \int\int_{-\infty}^{\infty} E_L^{+}\left(x_L, y_L\right) \exp\left[\frac{i\,k}{2d_i}\left[\left(x_i - x_L\right)^2 + \left(y_i - y_L\right)^2\right]\right] dx_L\, dy_L$$

Reemplazando la expresión obtenida para $E_L^{+}\left(x_L, y_L\right)$ y agrupando llegamos a que

$$\begin{aligned} h\left(x_i, y_i, x_o, y_o\right) = &\frac{1}{\lambda^2\, d_i d_o} \exp\left[\frac{i\,k}{2d_i}\left(x_i^2 + y_i^2\right)\right] \exp\left[\frac{i\,k}{2d_o}\left(x_o^2 + y_o^2\right)\right] \cdot \\ &\cdot \int\int_{-\infty}^{\infty} P\left(x_L, y_L\right) \exp\left[\frac{i\,k}{2}\left(\frac{1}{d_o} + \frac{1}{d_i} - \frac{1}{f}\right)\left(x_L^2 + y_L^2\right)\right] \cdot \\ &\cdot \exp\left[-i\,k\left[\left(\frac{x_o}{d_o} + \frac{x_i}{d_i}\right) x_L + \left(\frac{y_o}{d_o} + \frac{y_i}{d_i}\right) y_L\right]\right] dx_L\, dy_L \end{aligned}$$

Veamos cómo se puede simplificar esta expresión. El factor $\exp\left[\frac{i\,k}{2d_i}\left(x_i^2 + y_i^2\right)\right]$ sobre el plano imagen en general podrá eliminarse dado que todos los detectores (ojo, films, CCDs, etc) registran intensidad y no fase y al calcular $I = \left\langle E \cdot E^{*} \right\rangle$ ese término desaparece.

Por el contrario el factor $\exp\left[\frac{i\,k}{2d_o}\left(x_o^2 + y_o^2\right)\right]$ depende de las coordenadas del punto objeto de modo que cuando debamos resolver la integral $E_i = \int\int_{-\infty}^{\infty} h\, E_o\, dx_o\, dy_o$ deberemos integrar sobre esas variables. Sin embargo, tal como hemos dicho con anterioridad, la formación de imágenes es un fenómeno local y por lo tanto en la aproximación geométrica el punto imagen $\left(x_i, y_i\right)$ estará relacionado con el punto objeto $\left(x_o, y_o\right)$ a través del aumento. Esto es, podemos escribir $x_i = M \cdot x_o; \quad y_i = M \cdot$.

Luego si en un entorno de esa región la fase no cambia más que una fracción de radián se puede aproximar $\exp\left[\dfrac{ik}{2d_o}\left(x_o^2+y_o^2\right)\right]\cong\exp\left[\dfrac{ik}{2d_o}\left(\dfrac{x_i^2}{M^2}+\dfrac{y_i^2}{M^2}\right)\right]$. Ahora que dicho factor está escrito en términos de las coordenadas de la imagen podemos aplicar el mismo argumento que hemos empleado anteriormente para despreciar el factor previamente analizado.

Por último debemos tener en cuenta que si estamos en el plano imagen las distancias d_o y d_i cumplen con la ecuación gaussiana de las lentes $\dfrac{1}{d_o}+\dfrac{1}{d_i}=\dfrac{1}{f}$. Así la expresión para $h\left(x_i,y_i,x_o,y_o\right)$ queda

$$h\left(x_i,y_i,x_o,y_o\right)\cong\frac{1}{\lambda^2 d_i d_o}\int\int_{-\infty}^{\infty}P\left(x_L,y_L\right)\cdot\exp\left[-ik\left[\left(\frac{x_o}{d_o}+\frac{x_i}{d_i}\right)x_L+\left(\frac{y_o}{d_o}+\frac{y_i}{d_i}\right)y_L\right]\right]dx_L\,dy_L$$

Debemos considerar que el aumento M también puede escribirse como $M=\dfrac{d_i}{d_o}$ con lo cual obtenemos

$$h\left(x_i,y_i,x_o,y_o\right)\cong\frac{1}{\lambda^2 d_i d_o}\int\int_{-\infty}^{\infty}P\left(x_L,y_L\right)\cdot\exp\left[-\frac{i2\pi}{\lambda d_i}\left[\left(x_i+Mx_o\right)x_L+\left(y_i+My_o\right)y_L\right]\right]dx_L\,dy_L$$

Claramente $h\left(x_i,y_i,x_o,y_o\right)$ resulta igual a la transformada de Fourier de la función pupila $P\left(x_L,y_L\right)$ centrada en $x_i=-M\cdot x_o\,;\,y_i=-M\cdot y_o$. Este es un resultado que era previsible y que por otra parte coincide con lo expuesto por Rayleigh.

Para avanzar un poco más en el proceso de formación de imágenes hagamos el siguiente cambio de variables: $\tilde{x}_L\equiv\dfrac{x_L}{\lambda d_i}\,;\,\tilde{y}_L\equiv\dfrac{y_L}{\lambda d_i}$. Luego reemplazando en la ecuación anterior:

$$h\left(x_i,y_i,x_o,y_o\right)\cong M\int\int_{-\infty}^{\infty}P\left(\lambda d_i\tilde{x}_L,\lambda d_i\tilde{y}_L\right)\cdot\exp\left[-i2\pi\left[\left(x_i+Mx_o\right)\tilde{x}_L+\left(y_i+My_o\right)\tilde{y}_L\right]\right]d\tilde{x}_L\,d\tilde{y}_L$$

Como paso intermedio, y para comprobar si estamos obteniendo resultados razonables, hagamos el siguiente cálculo. Veamos qué obtenemos en el proceso de formación de imágenes cuando hacemos tender a cero la longitud de onda. En este caso deberíamos obtener la imagen ideal que se produce según la óptica geométrica, donde los efectos de difracción no son tenidos en cuenta. Ahora bien, esto es equivalente a que la función pupila $P(x_L, y_L)$ tienda a 1.

Antes de calcular la integral de superposición $E_i = \int\int_{-\infty}^{\infty} h\, E_o\, dx_o\, dy_o$ veamos que sucede con la expresión de $h(x_i, y_i, x_o, y_o)$ en este caso:

$$h(x_i, y_i, x_o, y_o) \longrightarrow M \int\int_{-\infty}^{\infty} \exp\left[-i\,2\pi\left[(x_i + Mx_o)\,\tilde{x}_L + (y_i + My_o)\,\tilde{y}_L\right]\right] d\tilde{x}_L\, d\tilde{y}_L =$$

$$= M\,\delta(x_i + Mx_o, y_i + My_o) = \frac{1}{M}\,\delta\left(\frac{x_i}{M} + x_o, \frac{y_i}{M} + y_o\right)$$

Aplicando esta expresión a la integral de superposición tenemos que

$$E_i(x_i, y_i) = \int\int_{-\infty}^{\infty} \frac{1}{M}\,\delta\left(\frac{x_i}{M} + x_o, \frac{y_i}{M} + y_o\right) E_o(x_o, y_o)\, dx_o\, dy_o = \frac{1}{M} E_o\left(-\frac{x_i}{M}, -\frac{y_i}{M}\right)$$

Luego en la aproximación en la que la longitud de onda tiende a cero se obtiene la imagen predicha por la óptica geométrica, esto es, una réplica exacta del objeto aumentado e invertido $E_{Gi}(x_i, y_i) = \frac{1}{M} E_o\left(-\frac{x_i}{M}, -\frac{y_i}{M}\right)$.

Habiendo hecho este paso intermedio, continuemos con nuestro análisis. Para ello, volvamos a la expresión hallada para $h(x_i, y_i, x_o, y_o)$ antes de hacer tender a cero la longitud de onda y hagamos un nuevo cambio de variables $\tilde{x}_o \equiv -M \cdot x_o$; $\tilde{y}_o \equiv -M \cdot y_o$
En cuyo caso queda

$$h\left(x_i, y_i, \tilde{x}_o, \tilde{y}_o\right) \cong M \int\int_{-\infty}^{\infty} P\left(\lambda d_i \tilde{x}_L, \lambda d_i \tilde{y}_L\right) \cdot \exp\left[-i\,2\pi\left[\left(x_i - \tilde{x}_o\right)\tilde{x}_L + \left(y_i - \tilde{y}_o\right)\tilde{y}_L\right]\right] d\tilde{x}_L\, d\tilde{y}_L$$

De esta forma logramos expresar a la función respuesta al impulso de forma tal que sea espacialmente invariante ya que sólo depende de las diferencias $\left(x_i - \tilde{x}_o\right)$; $\left(y_i - \tilde{y}_o\right)$.

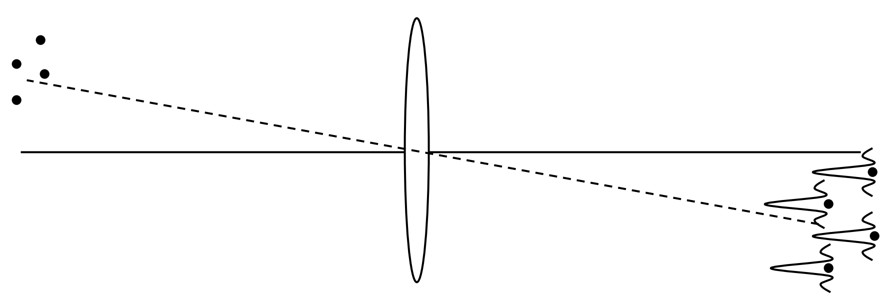
Vamos a definir por último $\tilde{h} \equiv \frac{1}{M} h$, de esta forma la integral de superposición será

$$E_i\left(x_i, y_i\right) = \int\int_{-\infty}^{\infty} \tilde{h}\left(x_i - \tilde{x}_o, y_i - \tilde{y}_o\right)\left[\frac{1}{M} E_o\left(-\frac{\tilde{x}_o}{M}, -\frac{\tilde{y}_o}{M}\right)\right] d\tilde{x}_o\, d\tilde{y}_o$$

Ahora bien, esta expresión corresponde a la convolución de $\tilde{h}$ con el campo que predecía la óptica geométrica. Luego

$$\boxed{E_i\left(x_i, y_i\right) = \tilde{h}\left(x_i, y_i\right) \otimes E_{Gi}\left(x_i, y_i\right)} \qquad (32)$$

Para la óptica geométrica cada punto objeto brinda un punto imagen perfecto. Al tener una lente de tamaño finito, existe difracción que ocasiona que ese punto imagen deba ser reemplazado por un patrón más o menos extenso que corresponderá a la figura de difracción de una fuente puntual a través de la lente. Esto es, tomamos cada punto objeto por separado y vemos cómo lo afecta la lente, luego reemplazamos el punto imagen ideal por esa respuesta individual. Claramente esta línea de razonamiento es la planteada por Rayleigh.



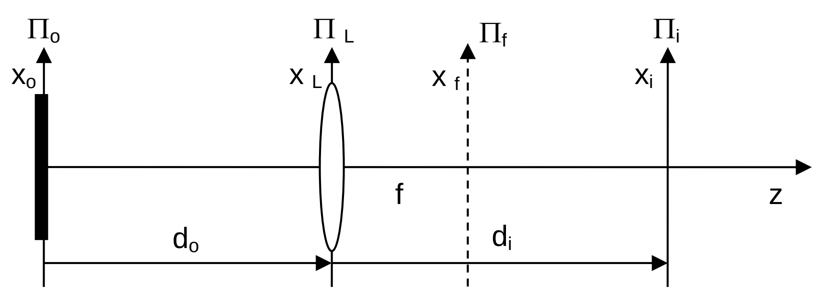
CLASE 11

Luego veremos en detalle el tratamiento de Abbe cuando estudiemos la función transmisión pero ahora analicemos el simple caso en el que la función pupila de la lente es tal que $P(x_L, y_L) = 1$. Bajo esta hipótesis deberíamos obtener, al igual que anteriormente, una imagen correspondiente al caso ideal que plantea la óptica geométrica.

Habíamos visto que, cuando teníamos un objeto a una distancia d_o delante de una lente y este era iluminado por una onda plana monocromática, la distribución del campo sobre el plano focal de la misma era

$$E(x_f, y_f) = \frac{\exp\left[i k (f + d_o)\right]}{i\lambda f} \cdot \exp\left[i \frac{k\left(x_f^2 + y_f^2\right)}{2f}\left(1 - \frac{d_o}{f}\right)\right] \cdot$$

$$\underbrace{\cdot \int\int_{-\infty}^{\infty} E_o(x_o, y_o) \exp\left[-i \frac{2\pi}{\lambda f}\left(x_f x_o + y_f y_o\right)\right] dx_o dy_o}_{\mathbb{F}_{f_{Xf}, f_{Yf}}\left[E_o(x_o, y_o)\right]}$$



Para calcular el campo sobre el plano imagen debemos hacer una propagación libre desde el plano focal

$$E_i\left(x_i,y_i\right)=\frac{\exp\left[i\,k\left(d_i-f\right)\right]}{i\lambda\left(d_i-f\right)}\exp\left[i\frac{k\left(x_i^2+y_i^2\right)}{2\left(d_i-f\right)}\right]\int\int_{-\infty}^{\infty}\exp\left[i\frac{k\left(x_f^2+y_f^2\right)}{2\left(d_i-f\right)}\right]\cdot$$
$$.\exp\left[-i\frac{k}{\left(d_i-f\right)}\left(x_ix_f+y_iy_f\right)\right]\cdot\left\{\frac{\exp\left[i\,k\left(f+d_o\right)\right]}{i\lambda f}\exp\left[i\frac{k\left(x_f^2+y_f^2\right)}{2f}\left(1-\frac{d_o}{f}\right)\right]\cdot\right.$$
$$\left.\cdot\int\int_{-\infty}^{\infty}E_o\left(x_o,y_o\right)\exp\left[-i\frac{2\pi}{\lambda f}\left(x_fx_o+y_fy_o\right)\right]dx_ody_o\right\}dx_fdy_f$$

Teniendo en cuenta que en formación de imágenes se cumple que $\frac{1}{d_o}+\frac{1}{d_i}=\frac{1}{f}$, resulta

$$E_i\left(x_i,y_i\right)=\frac{\exp\left[i\,k\left(d_o+d_i\right)\right]}{\left[i\lambda\left(d_i-f\right)\right]\left[i\lambda f\right]}\exp\left[i\frac{k\left(x_i^2+y_i^2\right)}{2\left(d_i-f\right)}\right]\int\int_{-\infty}^{\infty}\exp\left[-i\frac{k}{\left(d_i-f\right)}\left(x_ix_f+y_iy_f\right)\right]\cdot$$
$$\cdot\left\{\int\int_{-\infty}^{\infty}E_o\left(x_o,y_o\right)\exp\left[-i\frac{2\pi}{\lambda f}\left(x_fx_o+y_fy_o\right)\right]dx_ody_o\right\}dx_fdy_f$$

Pero $\frac{x_f}{\lambda f}=f_{X_f}$; $\frac{y_f}{\lambda f}=f_{Y_f}$ con lo cual $dx_f=\lambda f\cdot df_{X_f}$; $dy_f=\lambda f\cdot df_{Y_f}$.

Por otra parte si en el factor $\exp\left[-i\frac{k}{\left(d_i-f\right)}\left(x_ix_f+y_iy_f\right)\right]$ multiplicamos y dividimos el exponente por f tenemos que $\exp\left[-i\frac{2\pi f}{\left(d_i-f\right)}\left(x_i f_{X_f}+y_i f_{Y_f}\right)\right]$. Recordando nuevamente que $\frac{1}{d_o}+\frac{1}{d_i}=\frac{1}{f}$ entonces $d_o=\frac{f\,d_i}{\left(d_i-f\right)}$ $\Rightarrow\frac{1}{M}=\frac{d_o}{d_i}=\frac{f}{\left(d_i-f\right)}$.

Luego podemos escribir la exponencial como $\exp\left[-i\frac{2\pi}{M}\left(x_i f_{X_f}+y_i f_{Y_f}\right)\right]$ que al reemplazarla en la integral da

$$E_i\left(x_i,y_i\right)=-\frac{\exp\left[i\,k\left(d_o+d_i\right)\right]f}{\left(d_i-f\right)}\exp\left[i\frac{k\left(x_i^2+y_i^2\right)}{2\left(d_i-f\right)}\right]\cdot$$

$$\cdot\int\int_{-\infty}^{\infty}\mathbb{F}_{f_{X\,f},f_{Y\,f}}\left[E_o\left(x_o,y_o\right)\right]\exp\left[i\,2\pi\left(-\frac{x_i}{M}f_{X\,f}-\frac{y_i}{M}f_{Y\,f}\right)\right]df_{X\,f}df_{Y\,f}$$

Pero esta expresión la podemos escribir como

$$E_i\left(x_i,y_i\right)=-\frac{\exp\left[i\,k\left(d_o+d_i\right)\right]}{M}\exp\left[i\frac{k\left(x_i^2+y_i^2\right)}{2\left(d_i-f\right)}\right]\cdot\mathbb{F}^{-1}_{-\frac{x_i}{M},-\frac{y_i}{M}}\left[\mathbb{F}_{f_{X\,f},f_{Y\,f}}\left[E_o\left(x_o,y_o\right)\right]\right]$$

Vemos que si del plano objeto al plano focal de la lente obtenemos la transformada de Fourier de la entrada, del plano focal al plano imagen se origina la antitransformada y obtenemos la imagen

$$E_i\left(x_i,y_i\right)=-\exp\left[i\,k\left(d_o+d_i\right)\right]\exp\left[i\frac{k\left(x_i^2+y_i^2\right)}{2\left(d_i-f\right)}\right]\cdot\left[\frac{1}{M}E_o\left(-\frac{x_i}{M},-\frac{y_i}{M}\right)\right]$$

Esta ecuación nos dice que en el caso en el que $P\left(x_L,y_L\right)=1$ el campo en el plano imagen es igual a una fase global por el campo geométrico. Así se obtuvo igual resultado que aplicando la función respuesta al impulso.

FORMACIÓN DE IMÁGENES CON FUENTES PARCIALMENTE COHERENTES

Hasta el momento tratamos casos en los que el sistema óptico era iluminado por fuentes puntuales, monocromáticas. Analicemos ahora qué sucede si utilizamos fuentes que no lo son.

Cuando estudiamos la propagación de la intensidad mutua entre dos puntos *(S_1,S_2)* y *(P_1,P_2)* a través de sistemas ópticos, habíamos obtenido que:

$$J_{P_1P_2}=\iint_{\Sigma}\iint_{\Sigma} h\left(x_1,y_1,P_1\right)h^*\left(x_2,y_2,P_2\right)q_1\,q_2^*\cdot J_{S_1S_2}\cdot dx_1\,dy_1\cdot dx_2\,dy_2$$

Tomemos ahora $P_1=P_2=P=\left(x_i,y_i\right)$ como un punto sobre el plano imagen y llamemos $\left(x_1,y_1\right)=\left(x_o,y_o\right)$ a un punto sobre el plano objeto y $\left(x_2,y_2\right)=\left(x_o',y_o'\right)$ a otro punto sobre dicho plano. Además aproximaremos $q_1\cong q_2\cong 1$. Así la ecuación anterior toma la forma

$$I\left(P\right)=\int\int_{-\infty}^{\infty}\int\int h\left(x_o,y_o,x_i,y_i\right)h^*\left(x_o',y_o',x_i,y_i\right)\cdot J\left(x_o,y_o,x_o',y_o'\right)\cdot dx_o\,dy_o\cdot dx_o'\,dy_o'$$

Cabe aclarar que los límites de las integrales son $\pm\infty$ ya que estamos realizando el proceso de integración sobre todo un plano y eventualmente los límites del objeto estarán tenidos en cuenta en su factor de transmisión. Vamos a suponer que el sistema es isoplanático, luego tenemos que

$$I\left(P\right)=\int\int_{-\infty}^{\infty}\int\int h\left(x_i-x_o,y_i-y_o\right)h^*\left(x_i-x_o',y_i-y_o'\right)\cdot J\left(x_o,y_o,x_o',y_o'\right)\cdot dx_o\,dy_o\cdot dx_o'\,dy_o'$$

Ahora bien, sabemos que $J\left(x_o,y_o,x_o',y_o'\right)=\left\langle E_o\left(x_o,y_o\right)E_o^*\left(x_o',y_o'\right)\right\rangle$ donde E_o es el campo a la salida del objeto. A este último podemos escribirlo como el campo que llega al objeto por el factor de transmisión del mismo, esto es:

$E_o\left(x_o,y_o\right)=E^-\left(x_o,y_o\right)\cdot t\left(x_o,y_o\right)$; $E_o\left(x_o',y_o'\right)=E^-\left(x_o',y_o'\right)\cdot t\left(x_o',y_o'\right)$ con lo cual

$$J\left(x_o,y_o,x_o',y_o'\right)=t\left(x_o,y_o\right)t^*\left(x_o',y_o'\right)\left\langle E^-\left(x_o,y_o\right)E^{-*}\left(x_o',y_o'\right)\right\rangle$$

Vamos a escribir esta expresión de otra forma

$$\begin{aligned}J\left(x_o,y_o,x_o',y_o'\right)&=t\left(x_o,y_o\right)t^*\left(x_o',y_o'\right)J^-\left(x_o,y_o,x_o',y_o'\right)=\\&=t\left(x_o,y_o\right)t^*\left(x_o',y_o'\right)\sqrt{J^-\left(x_o,y_o\right)}\sqrt{J^-\left(x_o',y_o'\right)}\mu^-\left(x_o,y_o,x_o',y_o'\right)\end{aligned}$$

Ahora que tenemos esto expresado en términos del factor complejo de coherencia vamos a analizar la forma que toma la ecuación de la intensidad en un punto imagen para los casos límites.

Si la fuente de iluminación es espacialmente coherente, esto es, tenemos una fuente puntual, entonces de acuerdo a lo que habíamos visto cuando estudiamos la teoría de la coherencia parcial $\mu^{-}\left(x_o, y_o, x_o', y_o'\right)=1$. Resulta así

$$
\begin{aligned}
I(P) = & \int\int_{-\infty}^{\infty}\int\int h\left(x_i - x_o, y_i - y_o\right) h^*\left(x_i - x_o', y_i - y_o'\right)\cdot \\
& \cdot t\left(x_o, y_o\right) t^*\left(x_o', y_o'\right)\sqrt{J^{-}\left(x_o, y_o\right)}\sqrt{J^{-}\left(x_o', y_o'\right)}\cdot dx_o\, dy_o \cdot dx_o'\, dy_o' = \\
& = \int\int_{-\infty}^{\infty} h\left(x_i - x_o, y_i - y_o\right) t\left(x_o, y_o\right)\sqrt{J^{-}\left(x_o, y_o\right)}\, dx_o\, dy_o \cdot \\
& \cdot \int\int_{-\infty}^{\infty} h^*\left(x_i - x_o', y_i - y_o'\right) t^*\left(x_o', y_o'\right)\sqrt{J^{-}\left(x_o', y_o'\right)}\, dx_o'\, dy_o' = \\
& = \left|\int\int_{-\infty}^{\infty} h\left(x_i - x_o, y_i - y_o\right) t\left(x_o, y_o\right)\sqrt{J^{-}\left(x_o, y_o\right)}\, dx_o\, dy_o\right|^2
\end{aligned}
$$

Recordemos que $E_o\left(x_o, y_o\right)=t\left(x_o, y_o\right)\sqrt{J^{-}\left(x_o, y_o\right)}$ luego

$$I(P)=\left|\int\int_{-\infty}^{\infty} h\left(x_i - x_o, y_i - y_o\right) E\left(x_o, y_o\right) dx_o\, dy_o\right|^2 = \left|h\left(x_i, y_i\right)\otimes E\left(x_i, y_i\right)\right|^2$$

Anteriormente, cuando vimos formación de imágenes con iluminación monocromática, habíamos llegado a la ecuación (32) que establecía que

$$E_i\left(x_i, y_i\right)=\tilde{h}\left(x_i, y_i\right)\otimes E_{Gi}\left(x_i, y_i\right)$$

Luego la intensidad viene dada por

$$I(P)=\left\langle E_i\left(x_i, y_i\right) E_i^*\left(x_i, y_i\right)\right\rangle = \left|E_i\left(x_i, y_i\right)\right|^2 = \left|\tilde{h}\left(x_i, y_i\right)\otimes E_{Gi}\left(x_i, y_i\right)\right|^2$$

Vemos que llegamos a idéntico resultado.

En el otro extremo, es decir el de fuente extensa, cada punto es totalmente incoherente con el otro con lo cual $\mu^{-}\left(x_o, y_o, x_o', y_o'\right) = \delta\left(x_o - x_o', y_o - y_o'\right)$. Luego la intensidad en un punto sobre el plano imagen vendrá dada por

$$I(P) = \int \int_{-\infty}^{\infty} \left|h\left(x_i - x_o, y_i - y_o\right)\right|^2 \left|t\left(x_o, y_o\right)\right|^2 \left|J^{-}\left(x_o, y_o\right)\right| dx_o\, dy_o =$$

$$= \int \int_{-\infty}^{\infty} \left|h\left(x_i - x_o, y_i - y_o\right)\right|^2 \left|E\left(x_o, y_o\right)\right|^2 dx_o\, dy_o = \left|h\left(x_i, y_i\right)\right|^2 \otimes I\left(x_i, y_i\right)$$

De acuerdo a la identificación realizada anteriormente ahora podemos escribir

$$I(P) = \left|\tilde{h}\left(x_i, y_i\right)\right|^2 \otimes I_{Gi}\left(x_i, y_i\right)$$

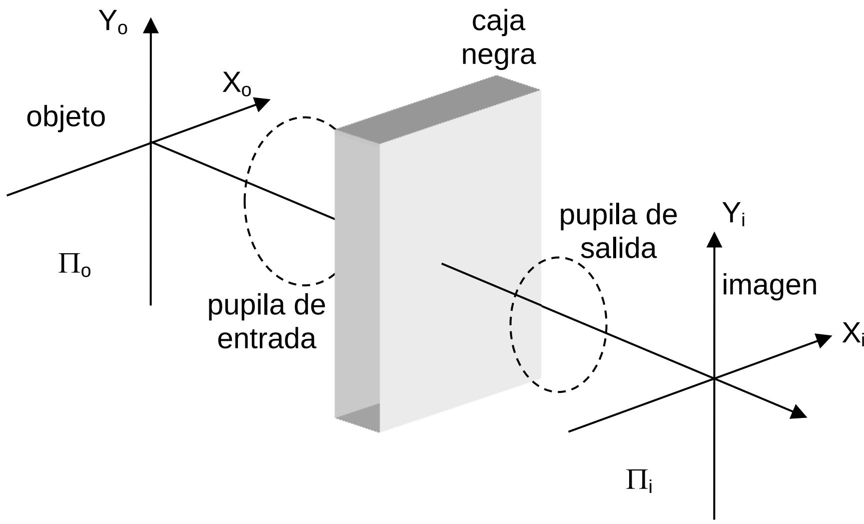
En general para coherencia parcial tendremos

$$I(P) = \int \int_{-\infty}^{\infty} \int \int \tilde{h}\left(x_i - \tilde{x}_o, y_i - \tilde{y}_o\right) \tilde{h}^*\left(x_i - \tilde{x}_o', y_i - \tilde{y}_o'\right) \cdot$$

$$\cdot \left\langle E_G\left(\tilde{x}_o, \tilde{y}_o\right) E_G^*\left(\tilde{x}_o', \tilde{y}_o'\right)\right\rangle \cdot d\tilde{x}_o\, d\tilde{y}_o \cdot d\tilde{x}_o'\, d\tilde{y}_o'$$

SISTEMAS FORMADORES DE IMÁGENES – TRATAMIENTO GENERALIZADO

Supongamos que el sistema óptico con el que trabajamos no está formado por una sola lente delgada sino por varias, incluso algunas pueden ser gruesas, otras divergentes, etc. Vamos a pensar dicho sistema como una caja negra cuyas propiedades pueden especificarse completamente a partir de las propiedades terminales del conjunto, esto es, conociendo la pupila de entrada (apertura real o efectiva que limita el cono de luz entrante al sistema) y la pupila de salida (apertura real o efectiva que limita el cono de salida).



Se asume generalmente que el pasaje de la luz entre la pupila de entrada y la de salida puede describirse mediante óptica geométrica. Esto es, la pupila de entrada se encuentra, como se hace usualmente, proyectando el diafragma de apertura a través de las lentes que lo preceden y la pupila de salida proyectándolo a través de las lentes que se hayan detrás del mismo; así ambas pupilas resultan proyecciones geométricas una de la otra. Entonces, dado que la óptica geométrica describe bien el pasaje de la luz de la pupila de entrada a la de salida, puede pensarse que los efectos de difracción están asociados a la propagación de la luz desde el objeto hasta la pupila de entrada y desde la pupila de salida hasta la imagen, siendo la resolución del sistema afectada por el tamaño finito de ambas aperturas.

Un sistema formador de imágenes se dice limitado por difracción si una onda esférica que parte de un punto objeto es convertido en una onda esférica que converge a un punto imagen. Si se aparta mucho de esta situación ideal se dice que está afectado por aberraciones.

En las deducciones anteriores vimos que cuando un sistema óptico era iluminado en forma coherente la respuesta era lineal en campo, esto es

$$E_i\left(x_i, y_i\right) = \tilde{h}\left(x_i, y_i\right) \otimes E_{Gi}\left(x_i, y_i\right)$$

En tanto que cuando era iluminado incoherentemente su respuesta era lineal en intensidad

$$I\left(x_i, y_i\right) = \left|\tilde{h}\left(x_i, y_i\right)\right|^2 \otimes I_{Gi}\left(x_i, y_i\right)$$

En estas ecuaciones la función respuesta al impulso venía dada por

$$\tilde{h}\left(x_i, y_i\right) = \int \int_{-\infty}^{\infty} P\left(\lambda d_i \tilde{x}_P, \lambda d_i \tilde{y}_P\right) \cdot \exp\left[-i\, 2\pi \left[x_i \tilde{x}_P + y_i \tilde{y}_P\right]\right] d\tilde{x}_P\, d\tilde{y}_P$$

donde ahora hemos reemplazado las variables $\left(\tilde{x}_L, \tilde{y}_L\right)$ de la lente por $\left(\tilde{x}_P, \tilde{y}_P\right)$ de la pupila de salida con $\tilde{x}_P = \dfrac{x_P}{\lambda d_i}$; $\tilde{y}_P = \dfrac{y_P}{\lambda d_i}$, siendo d_i la distancia del plano de la pupila de salida al plano imagen.

Veamos ahora cómo aplicar las propiedades de sistemas lineales a estos sistemas ópticos.

Función transferencia coherente

Dijimos que cuando el objeto está iluminado en forma coherente el campo a la salida es

$$E\left(x_i, y_i\right) = \int \int_{-\infty}^{\infty} \tilde{h}\left(x_i - \tilde{x}_o, y_i - \tilde{y}_o\right) E_G\left(\tilde{x}_o, \tilde{y}_o\right) d\tilde{x}_o\, d\tilde{y}_o$$

Dado el tamaño finito de las lentes, las frecuencias espaciales en el objeto no estarán todas presentes en la imagen. Claramente la imagen ideal que se obtendría si no hubiese difracción sí contendría todas las frecuencias espaciales y su espectro vendría dado por

$$\mathcal{E}_G\left(f_x, f_y\right) = \int \int_{-\infty}^{\infty} E_G\left(\tilde{x}_o, \tilde{y}_o\right) \cdot \exp\left[-i\, 2\pi \left(f_x \tilde{x}_o + f_y \tilde{y}_o\right)\right] d\tilde{x}_o\, d\tilde{y}_o$$

En tanto el que pertenece a la imagen real será

$$\mathcal{E}_i\left(f_x, f_y\right) = \int \int_{-\infty}^{\infty} E_i\left(x_i, y_i\right) \cdot \exp\left[-i\, 2\pi \left(f_x x_i + f_y y_i\right)\right] dx_i\, dy_i$$

Ahora bien, de acuerdo a las propiedades que vimos para la convolución tendremos que

$$E_i\left(x_i, y_i\right) = \tilde{h}\left(x_i, y_i\right) \otimes E_{Gi}\left(x_i, y_i\right) \rightarrow \mathcal{E}_i\left(f_x, f_y\right) = H\left(f_x, f_y\right) \cdot \mathcal{E}_G\left(f_x, f_y\right)$$

Vemos que la función $H\left(f_x, f_y\right)$ tiene en cuenta cómo se deterioran las frecuencias, la misma se denomina *función transferencia coherente* y es la transformada de Fourier de $\tilde{h}\left(x_i, y_i\right)$

$$H\left(f_x, f_y\right) = \int\int_{-\infty}^{\infty} \tilde{h}\left(x_i, y_i\right) \cdot \exp\left[-i\,2\pi\left(f_x x_i + f_y y_i\right)\right] dx_i\, dy_i$$

Recordemos que $\tilde{h}\left(x_i, y_i\right) = \mathbb{F}_{x_i, y_i}\left[P\left(\lambda d_i \tilde{x}_P, \lambda d_i \tilde{y}_P\right)\right]$ luego la función transferencia coherente es

$$H\left(f_x, f_y\right) = \mathbb{F}_{f_x, f_y}\left[\mathbb{F}_{x_i, y_i}\left[P\left(\lambda d_i \tilde{x}_P, \lambda d_i \tilde{y}_P\right)\right]\right] = P\left(\lambda d_i f_x, \lambda d_i f_y\right)$$

Si el sistema está limitado por difracción, esto es no existen aberraciones, la función pupila es real y sus valores son 0 ó 1. Luego en estos tipos de sistemas y bajo iluminación coherente las frecuencias espaciales que pasan no sufren atenuación o distorsión de fase, sólo existe una frecuencia de corte que impone un límite a la resolución del sistema óptico. Frecuencias superiores directamente no pasan.
Veamos el ejemplo de la pupila circular:

- Pupila circular de radio R

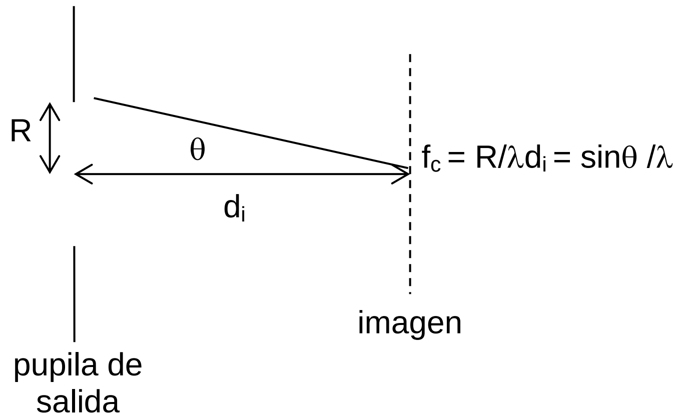
$P\left(x_P, y_P\right) = \mathrm{circ}\left(\dfrac{\sqrt{x_P^2 + y_P^2}}{R}\right)$ donde *R* es el radio de la pupila de salida

de acuerdo al cambio de variables hecho anteriormente tenemos que $\begin{array}{l} x_P = \lambda d_i\, \tilde{x}_P \\ y_P = \lambda d_i\, \tilde{y}_P \end{array}$

Luego la función pupila toma la forma $P\left(\lambda d_i \tilde{x}_P, \lambda d_i \tilde{y}_P\right) = \mathrm{circ}\left(\dfrac{\lambda d_i \sqrt{\tilde{x}_P^2 + \tilde{y}_P^2}}{R}\right)$ con lo cual la función transferencia coherente es

$$H\left(f_x, f_y\right) = P\left(\lambda d_i f_x, \lambda d_i f_y\right) = \mathrm{circ}\left(\frac{\lambda d_i \sqrt{f_x^2 + f_y^2}}{R}\right)$$

Existe una frecuencia de corte $f_c = \dfrac{R}{\lambda d_i}$ que establece cual es el valor de la máxima frecuencia espacial que podrá ser encontrada en la imagen



Función transferencia incoherente

Para iluminación incoherente habíamos obtenido que la intensidad en un punto de la imagen venía dado por

$$I\left(x_i, y_i\right) = \int\int_{-\infty}^{\infty} \left|\tilde{h}\left(x_i - \tilde{x}_o, y_i - \tilde{y}_o\right)\right|^2 I_G\left(\tilde{x}_o, \tilde{y}_o\right) d\tilde{x}_o\, d\tilde{y}_o$$

Al igual que con iluminación coherente, debido al tamaño finito de las lentes, habrá una degradación de las frecuencias espaciales.

El espectro normalizado ideal, es decir aquel correspondiente a $I_G\left(\tilde{x}_o, \tilde{y}_o\right)$ está dado por

$$\mathcal{I}_G\left(f_x, f_y\right) = \frac{\int\int_{-\infty}^{\infty} I_G\left(\tilde{x}_o, \tilde{y}_o\right) \cdot \exp\left[-i\,2\pi\left(f_x \tilde{x}_o + f_y \tilde{y}_o\right)\right] d\tilde{x}_o\, d\tilde{y}_o}{\int\int_{-\infty}^{\infty} I_G\left(\tilde{x}_o, \tilde{y}_o\right) d\tilde{x}_o\, d\tilde{y}_o}$$

en tanto que el espectro normalizado de la imagen real será

$$\mathcal{I}_i\left(f_x,f_y\right)=\frac{\displaystyle\iint_{-\infty}^{\infty} I_i\left(\tilde{x}_o,\tilde{y}_o\right)\cdot\exp\left[-i\,2\pi\left(f_x\tilde{x}_o+f_y\tilde{y}_o\right)\right]d\tilde{x}_o\,d\tilde{y}_o}{\displaystyle\iint_{-\infty}^{\infty} I_i\left(\tilde{x}_o,\tilde{y}_o\right)d\tilde{x}_o\,d\tilde{y}_o}$$

A partir de la ecuación $I\left(x_i,y_i\right)=\left|\tilde{h}\left(x_i,y_i\right)\right|^2\otimes I_{Gi}\left(x_i,y_i\right)$ podemos escribir que

$$\mathcal{I}_i\left(f_x,f_y\right)=\mathcal{H}\left(f_x,f_y\right)\cdot\mathcal{I}_G\left(f_x,f_y\right)$$

donde $\mathcal{H}\left(f_x,f_y\right)$ es la *función transferencia incoherente*, comúnmente llamada OTF (optical transfer function). Su expresión viene dada por:

$$\mathcal{H}\left(f_x,f_y\right)=\frac{\displaystyle\iint_{-\infty}^{\infty}\left|\tilde{h}\left(x_i-\tilde{x}_o,y_i-\tilde{y}_o\right)\right|^2\cdot\exp\left[-i\,2\pi\left(f_x\tilde{x}_o+f_y\tilde{y}_o\right)\right]d\tilde{x}_o\,d\tilde{y}_o}{\displaystyle\iint_{-\infty}^{\infty}\left|\tilde{h}\left(x_i-\tilde{x}_o,y_i-\tilde{y}_o\right)\right|^2 d\tilde{x}_o\,d\tilde{y}_o}$$

Hay dos razones por las que se toman los espectros normalizados, una menos importante es porque resultará más simple matemáticamente para futuros cálculos, la otra es que la intensidad nunca toma valores negativos y en general siempre está presente un fondo constante. Dado que la calidad visual de una imagen depende fuertemente del contraste, esto es de las intensidades relativas, conviene analizar el espectro normalizado por ese fondo.

Vimos que tanto la función transferencia coherente como la incoherente dependen de $\tilde{h}\left(x_i,y_i\right)$, por lo tanto podemos esperar que exista una relación entre ambas. Analicemos cuál es. Por un lado tenemos que

$H\left(f_x,f_y\right)=\mathbb{F}_{f_x,f_y}\left[\tilde{h}\left(x_i,y_i\right)\right]$; por el otro hemos encontrado la expresión:

$$\mathcal{H}\left(f_x,f_y\right)=\frac{\mathbb{F}_{f_x,f_y}\left[\left|\tilde{h}\left(x_i,y_i\right)\right|^2\right]}{\mathbb{F}_{f_x=0,f_y=0}\left[\left|\tilde{h}\left(x_i,y_i\right)\right|^2\right]}$$

pero esto lo podemos escribir como

$$\mathbb{F}_{f_x,f_y}\left[\left|\tilde{h}(x_i,y_i)\right|^2\right]=\mathbb{F}_{f_x,f_y}\left[\tilde{h}(x_i,y_i)\tilde{h}^*(x_i,y_i)\right]=\mathbb{F}_{f_x,f_y}\left[\tilde{h}(x_i,y_i)\right]\otimes\mathbb{F}_{f_x,f_y}\left[\tilde{h}^*(x_i,y_i)\right]$$

Si recordamos las propiedades de convolución y correlación, tenemos que

$$\mathbb{F}_{f_x,f_y}\left[\tilde{h}(x_i,y_i)\right]\otimes\mathbb{F}_{f_x,f_y}\left[\tilde{h}^*(x_i,y_i)\right]=\mathbb{F}_{f_x,f_y}\left[\tilde{h}(x_i,y_i)\right]\otimes\mathbb{F}^*_{-f_x,-f_y}\left[\tilde{h}(x_i,y_i)\right]=$$
$$=H\left(f_x,f_y\right)\odot H\left(f_x,f_y\right)$$

O sea que obtuvimos la autocorrelación de $H\left(f_x,f_y\right)$. Entonces expresaremos

$$\mathcal{H}\left(f_x,f_y\right)=\frac{\int\int_{-\infty}^{\infty}H\left(\mathrm{f}_x',\mathrm{f}_y'\right)H^*\left(\mathrm{f}_x'-f_x,\mathrm{f}_y'-f_y\right)d\mathrm{f}_x'd\mathrm{f}_y'}{\int\int_{-\infty}^{\infty}\left|H\left(\mathrm{f}_x',\mathrm{f}_y'\right)\right|^2d\mathrm{f}_x'd\mathrm{f}_y'}$$

donde $\left(\mathrm{f}_x',\mathrm{f}_y'\right)$ son variables mudas de integración que denominamos de esta forma sólo para señalar que tienen unidades de frecuencia espacial.

Vamos a simetrizar esta expresión, para ello hagamos el siguiente cambio de variables

$\mathrm{f}_x=\mathrm{f}_x'-\frac{f_x}{2}\,;\quad \mathrm{f}_y=\mathrm{f}_y'-\frac{f_y}{2}$. Tenemos entonces que

$$\mathcal{H}\left(f_x,f_y\right)=\frac{\int\int_{-\infty}^{\infty}H\left(\mathrm{f}_x+\frac{f_x}{2},\mathrm{f}_y+\frac{f_y}{2}\right)H^*\left(\mathrm{f}_x-\frac{f_x}{2},\mathrm{f}_y-\frac{f_y}{2}\right)d\mathrm{f}_xd\mathrm{f}_y}{\int\int_{-\infty}^{\infty}\left|H\left(\mathrm{f}_x,\mathrm{f}_y\right)\right|^2d\mathrm{f}_xd\mathrm{f}_y}$$

Usualmente esta función se escribe como $\mathcal{H}\left(f_x,f_y\right)=\left|\mathcal{H}\left(f_x,f_y\right)\right|\exp\left[i\Phi_{\mathcal{H}}\left(f_x,f_y\right)\right]$ donde $\left|\mathcal{H}\left(f_x,f_y\right)\right|$ se conoce como *función transferencia de modulación* (MTF) y a $\Phi_{\mathcal{H}}\left(f_x,f_y\right)$ como *función transferencia de fase*. Las principales propiedades son:

- $0\leq\left|\mathcal{H}\left(f_x,f_y\right)\right|\leq 1$

- $\mathcal{H}(0,0)=1$
- $\mathcal{H}(-f_x,-f_y)=\mathcal{H}^*(f_x,f_y)$
- Si la función pupila es par entonces $\mathcal{H}(f_x,f_y)\in\Re \quad \Rightarrow \quad \Phi_{\mathcal{H}}(f_x,f_y)=0$

CLASE 12

Veamos cómo actúa la función transferencia incoherente sobre la imagen de un objeto. Para ello analicemos un caso unidimensional simple.

Efecto de la OTF sobre un objeto de transmisión cosenoidal

Supongamos que tenemos un objeto cuya intensidad viene descripta por

$I_o=\dfrac{1+m\cos(2\pi\, f_o\, x_o)}{2}$ y recordemos que $I_i(x_i,y_i)=\left|\tilde{h}(x_i,y_i)\right|^2\otimes I_{Gi}(x_i,y_i)$.

En este caso tenemos que $I_{Gi}=\dfrac{1}{M^2}\left[\dfrac{1+m\cos\left(2\pi\, f_o\dfrac{\tilde{x}_o}{M}\right)}{2}\right]$ entonces

$$I_i(x_i)=\int_{-\infty}^{\infty}\left|\tilde{h}(x_i-\tilde{x}_o)\right|^2\cdot I_{Gi}(\tilde{x}_o)\,d\tilde{x}_o=\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2\cdot I_{Gi}(x_i-\tilde{x}_o)\,d\tilde{x}_o$$

En la igualdad anterior hemos empleado la propiedad de conmutatividad de la convolución Reemplacemos en la integral la distribución de intensidad geométrica

$$I_i(x_i)=\frac{1}{M^2}\left\{\frac{1}{2}\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2 d\tilde{x}_o+\frac{m}{2}\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2\cos\left[2\pi\frac{f_o}{M}(x_i-\tilde{x}_o)\right]d\tilde{x}_o\right\}=$$

$$=\frac{1}{M^2}\left\{\begin{array}{l}\displaystyle\frac{1}{2}\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2 d\tilde{x}_o+\frac{m}{2}\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2\cos\left[2\pi\frac{f_o}{M}x_i\right]\cos\left[2\pi\frac{f_o}{M}\tilde{x}_o\right]d\tilde{x}_o+\\ \displaystyle+\frac{m}{2}\int_{-\infty}^{\infty}\left|\tilde{h}(\tilde{x}_o)\right|^2\sin\left[2\pi\frac{f_o}{M}x_i\right]\sin\left[2\pi\frac{f_o}{M}\tilde{x}_o\right]d\tilde{x}_o\end{array}\right\}$$

Recordando las expresiones obtenidas al comienzo para la transformada seno y coseno, podemos escribir la expresión anterior como

$$I_i\left(x_i\right)=\frac{1}{M^2}\int_{-\infty}^{\infty}\left|\tilde{h}\left(\tilde{x}_o\right)\right|^2 d\tilde{x}_o\left\{\begin{array}{l}\frac{1}{2}+\frac{m}{2}\cos\left[2\pi\frac{f_o}{M}x_i\right]\cdot\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|\cdot\Re\left[\exp[\Phi_{\mathcal{H}}\left(\frac{f_o}{M}\right)]\right]+\\+\frac{m}{2}\sin\left[2\pi\frac{f_o}{M}x_i\right]\cdot\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|\cdot\Im\left[\exp[\Phi_{\mathcal{H}}\left(\frac{f_o}{M}\right)]\right]\end{array}\right\}$$

Entonces podemos reagruparlo como

$$I_i\left(x_i\right)=\frac{1}{M^2}\int_{-\infty}^{\infty}\left|\tilde{h}\left(\tilde{x}_o\right)\right|^2 d\tilde{x}_o\left\{\frac{1}{2}+\frac{m}{2}\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|\cos\left[2\pi\frac{f_o}{M}x_i-\Phi_{\mathcal{H}}\left(\frac{f_o}{M}\right)\right]\right\}$$

Vamos a calcular ahora el contraste del objeto.

$$C_o=\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}}=\frac{\frac{1+m}{2}-\frac{1-m}{2}}{\frac{1+m}{2}+\frac{1-m}{2}}=m$$

El contraste de la imagen ideal también tendrá este valor. Sin embargo veamos que sucede con la imagen real

$$C_i=\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}}=\frac{\frac{1+m\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|}{2}-\frac{1-m\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|}{2}}{\frac{1+m\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|}{2}+\frac{1-m\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|}{2}}=m\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|$$

Luego dado que $\left|\mathcal{H}\left(\frac{f_o}{M}\right)\right|\leq 1$ el contraste disminuye. Por otra parte $\Phi_{\mathcal{H}}\left(\frac{f_o}{M}\right)$ da cuenta de un corrimiento o distorsión en la fase.

Este punto está tratado en detalle e ilustrado con applets en
http://www.microscopyu.com/articles/optics/mtfintro.html

OTF de un sistema libre de aberraciones

Analicemos ahora la función transferencia incoherente de un sistema libre de aberraciones.
Habíamos visto que para sistemas ópticos iluminados coherentemente la función transferencia era $H\left(f_x, f_y\right)=P\left(\lambda d_i f_x, \lambda d_i f_y\right)$. Teniendo en cuenta la expresión que hallamos para $\mathcal{H}\left(f_x, f_y\right)$ en términos de $H\left(f_x, f_y\right)$, ahora podemos escribir

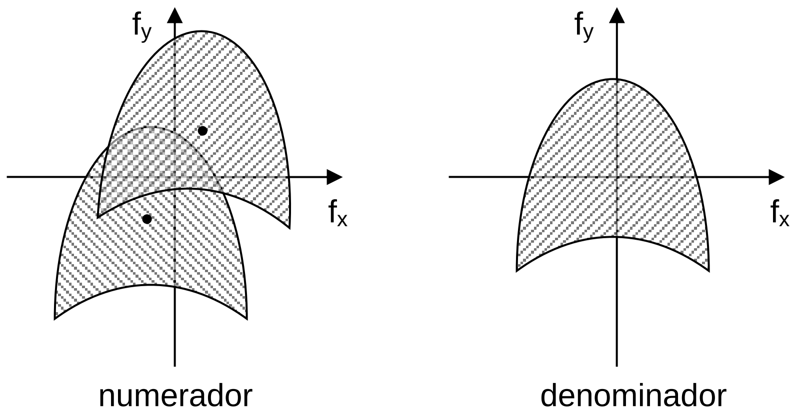
$$\mathcal{H}\left(f_x, f_y\right)=\frac{\int\int_{-\infty}^{\infty} P\left(\mathrm{f}_x+\frac{\lambda d_i f_x}{2}, \mathrm{f}_y+\frac{\lambda d_i f_y}{2}\right) P^*\left(\mathrm{f}_x-\frac{\lambda d_i f_x}{2}, \mathrm{f}_y-\frac{\lambda d_i f_y}{2}\right) d\mathrm{f}_x d\mathrm{f}_y}{\int\int_{-\infty}^{\infty}\left|P\left(\mathrm{f}_x, \mathrm{f}_y\right)\right|^2 d\mathrm{f}_x d\mathrm{f}_y}$$

Si el sistema está libre de aberraciones la función pupila es real y el hecho de que valga 0 ó 1 permite reemplazar $\left|P\left(\mathrm{f}_x, \mathrm{f}_y\right)\right|^2$ por $P\left(\mathrm{f}_x, \mathrm{f}_y\right)$. Luego la expresión anterior queda

$$\mathcal{H}\left(f_x, f_y\right)=\frac{\int\int_{-\infty}^{\infty} P\left(\mathrm{f}_x+\frac{\lambda d_i f_x}{2}, \mathrm{f}_y+\frac{\lambda d_i f_y}{2}\right) P\left(\mathrm{f}_x-\frac{\lambda d_i f_x}{2}, \mathrm{f}_y-\frac{\lambda d_i f_y}{2}\right) d\mathrm{f}_x d\mathrm{f}_y}{\int\int_{-\infty}^{\infty} P\left(\mathrm{f}_x, \mathrm{f}_y\right) d\mathrm{f}_x d\mathrm{f}_y}$$

Esta ecuación representa el cociente entre dos áreas. El numerador corresponde al valor del área de superposición de dos funciones pupila desplazadas; una de ellas está centrada en $\left(\frac{\lambda d_i f_x}{2}, \frac{\lambda d_i f_y}{2}\right)$ y la otra se halla centrada en el punto simétrico opuesto

$\left(-\frac{\lambda d_i f_x}{2}, -\frac{\lambda d_i f_y}{2}\right)$. El denominador es simplemente el área total de la pupila. Hagamos un gráfico para ilustrar esto

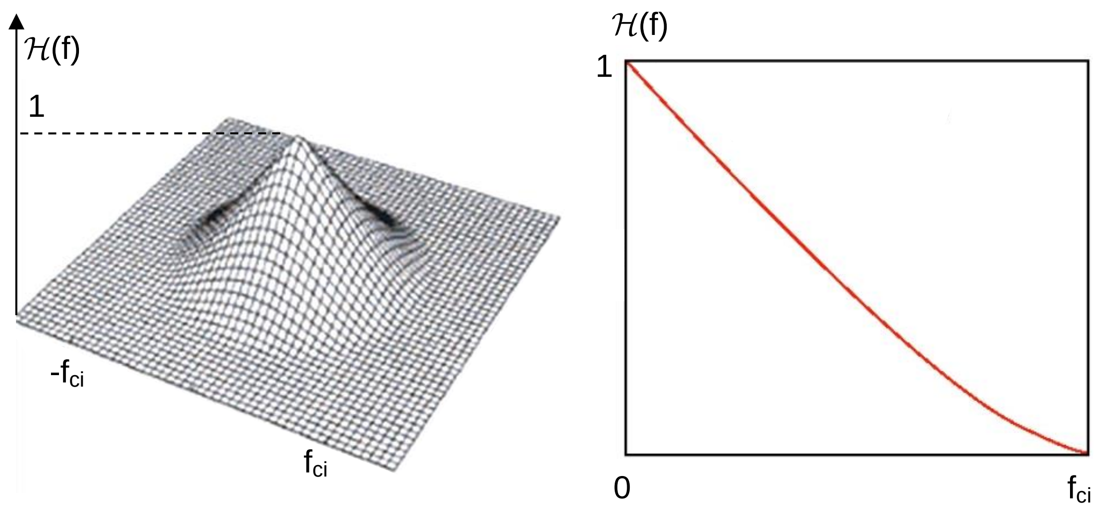


Cabe señalar además que para un sistema libre de aberraciones $\mathcal{H}\left(f_x, f_y\right) \geq 0$ y además $\mathcal{H}\left(f_x, f_y\right) \in \Re$.

Es interesante poner el resultado que se obtiene para el caso de la pupila circular de radio R, dado que en general las lentes son de esa forma. Puede demostrarse que

$$\mathcal{H}(f) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}\left(\frac{\lambda d_i f}{2R}\right) - \frac{\lambda d_i f}{2R}\sqrt{1-\left(\frac{\lambda d_i f}{2R}\right)^2}\right] & f \leq \frac{2R}{\lambda d_i} \\ 0 & afuera \end{cases}$$

Vemos que la frecuencia de corte incoherente $f_{ci} = \frac{2R}{\lambda d_i} = 2 f_{cc}$ es dos veces la frecuencia de corte coherente que habíamos obtenido con anterioridad. Sin embargo debe destacarse que en iluminación coherente las frecuencias pasaban o no lo hacían, pero había un corte brusco. En el caso incoherente, si bien el rango se extendió al doble, existe una atenuación gradual de las frecuencias. Un gráfico 3D y un corte transversal de esta OTF puede verse en la siguiente figura



Ejercicio 8: Encuentre $\mathcal{H}\left(f_x, f_y\right)$ para una pupila cuadrada de lado *a.* Demuestre que la frecuencia de corte incoherente es $f_{ci} = 2 f_{cc}$

Aberraciones y su efecto sobre la función transferencia

Cuando trabajábamos con sistemas limitados por difracción suponíamos que a la pupila de entrada llegaba una onda esférica y de la pupila de salida emergía una onda esférica. Si ahora tomamos en cuenta las aberraciones del sistema óptico, podemos pensar que la pupila de salida es iluminada por una onda esférica pero que en dicha apertura hay ubicada una placa transparente con una distribución de espesores tal que introduce corrimientos locales de fase. Así el apartamiento de la fase ideal en un punto $\left(x_p, y_p\right)$ de la pupila de salida será $kW\left(x_p, y_p\right)$ donde $W\left(x_p, y_p\right)$es el error efectivo en la longitud de camino. Luego la función pupila que tiene en cuenta las aberraciones la podemos escribir como $\mathcal{P}\left(x_p, y_p\right) = P\left(x_p, y_p\right)\exp\left[ikW\left(x_p, y_p\right)\right]$ y la denominaremos *función pupila generalizada*.

Veamos ahora cómo influyen las aberraciones en las funciones transferencia.

<u>Efectos sobre la función transferencia coherente</u>

La respuesta al impulso *h* de un sistema aberrado, iluminado coherentemente será simplemente la figura de difracción de Fraunhofer de una abertura con transmisión $\mathcal{P}\left(x_p, y_p\right)$, esto es, será simplemente su transformada de Fourier. Luego la función transferencia coherente vendrá dada por

$$H\left(f_x, f_y\right) = P\left(\lambda d_i f_x, \lambda d_i f_y\right)\exp\left[ikW\left(\lambda d_i f_x, \lambda d_i f_y\right)\right]$$

Evidentemente la frecuencia de corte debida al tamaño finito de la pupila será la misma que para el sistema sin aberraciones pero las distorsiones de fase pueden dañar severamente la imagen.

<u>Efectos sobre la función transferencia incoherente</u>

Habíamos visto para el caso sin aberraciones que

$$\mathcal{H}\left(f_x, f_y\right) = \frac{\displaystyle\iint_{\text{área de superposición}\left(f_x, f_y\right)} d\mathrm{f}_x d\mathrm{f}_y}{\displaystyle\iint_{\text{área de total}(0,0)} d\mathrm{f}_x d\mathrm{f}_y}$$

Cuando existen aberraciones esta expresión toma la forma

$$\mathcal{H}\left(f_x, f_y\right) = \frac{\displaystyle\iint_{\substack{\text{área de }\left(f_x, f_y\right)\\ \text{superposición}}} \exp\left[ik\left[W\left(\mathrm{f}_x + \frac{\lambda d_i f_x}{2}, \mathrm{f}_y + \frac{\lambda d_i f_y}{2}\right) - W\left(\mathrm{f}_x - \frac{\lambda d_i f_x}{2}, \mathrm{f}_y - \frac{\lambda d_i f_y}{2}\right)\right]\right] d\mathrm{f}_x d\mathrm{f}_y}{\displaystyle\iint_{\text{área de total}(0,0)} d\mathrm{f}_x d\mathrm{f}_y}$$

Puede mostrarse a través de la desigualdad de Schwarz que la MTF de un sistema aberrado es menor o igual que la de un sistema sin aberraciones.
Así en general las aberraciones producen una perdida de contraste para cada frecuencia espacial presente en la imagen. Si bien la frecuencia de corte no cambia, la frecuencia de corte efectiva es mucho menor que en el sistema sin aberraciones. Además vimos que para un sistema sin aberraciones $\mathcal{H}\left(f_x, f_y\right) \geq 0$. Esto no es cierto en este caso lo que quiere decir que existirán zonas de la imagen con el contraste invertido.
En general el estudio de las OTFs y de las aberraciones es muy complejo y existe toda una rama de la óptica que se dedica a ello.

Ejemplo del desenfoque como aberración

Un ejemplo relativamente sencillo y que ilustra muy bien lo que veníamos discutiendo es el caso de la aberración originada por desenfoque en un sistema con pupila cuadrada.
Supongamos entonces que tenemos una pupila cuadrada de lado *a* y que el error de desenfoque viene dado por $\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}=\varepsilon$.
Habíamos visto, cuando calculamos la función respuesta al impulso, que

$$h\left(x_i, y_i, x_o, y_o\right) \cong \frac{1}{\lambda^2 d_i d_o} \int\int_{-\infty}^{\infty} P\left(x_P, y_P\right) \cdot \exp\left[\frac{i k}{2}\left(\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}\right)\left(x_P^2+y_P^2\right)\right] \cdot$$
$$\cdot \exp\left[-\frac{i 2\pi}{\lambda d_i}\left[\left(x_i+M x_o\right) x_P+\left(y_i+M y_o\right) y_P\right]\right] dx_P\, dy_P$$

Luego en el plano imagen tomábamos $\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}=0$. Con lo cual ahora podemos escribir la pupila generalizada que describe este sistema desenfocado como

$$\mathcal{P}\left(x_p, y_p\right)=P\left(x_p, y_p\right) \exp\left[i k W\left(x_p, y_p\right)\right]=P\left(x_p, y_p\right) \exp\left[i \frac{k}{2} \varepsilon\left(x_P^2+y_P^2\right)\right]$$

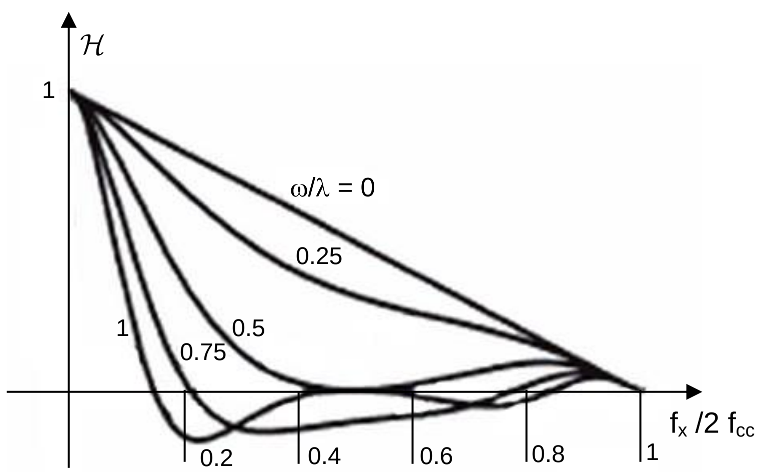
Así para una apertura cuadrada el máximo error de fase sobre un borde de la apertura será $k\varepsilon\dfrac{a^2}{8}$ siendo $\omega=\dfrac{\varepsilon a^2}{8}$ el máximo error de camino óptico. Este último número da una idea de cuan desenfocado está el sistema.

Reemplazando *W* en la expresión que hallamos para $\mathcal{H}\left(f_x, f_y\right)$ con aberraciones tenemos que

$$\mathcal{H}\left(f_x, f_y\right)=tri\left(\frac{f_x}{2f_{cc}}\right)tri\left(\frac{f_y}{2f_{cc}}\right)\mathrm{sinc}\left[\frac{8\omega}{\lambda}\left(\frac{f_x}{2f_{cc}}\right)\left(1-\frac{|f_x|}{2f_{cc}}\right)\right]\mathrm{sinc}\left[\frac{8\omega}{\lambda}\left(\frac{f_y}{2f_{cc}}\right)\left(1-\frac{|f_y|}{2f_{cc}}\right)\right]$$

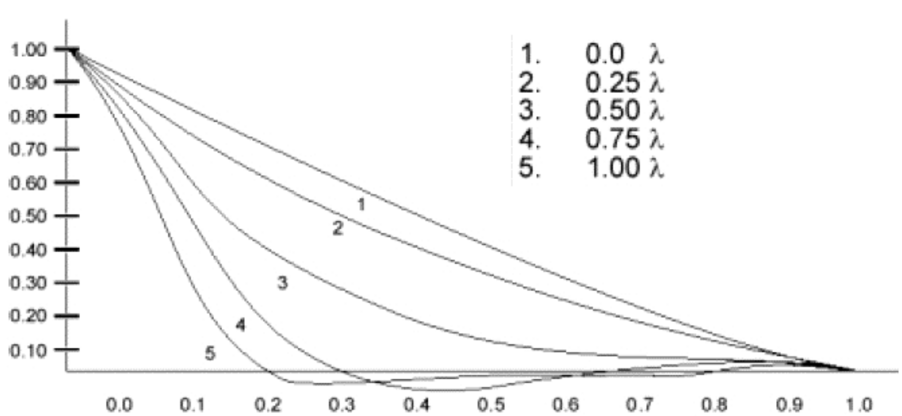
donde $f_{cc}=\dfrac{a}{2\lambda d_i}$.

Una sección transversal de $\mathcal{H}\left(f_x, f_y\right)$ se puede ver en el gráfico siguiente

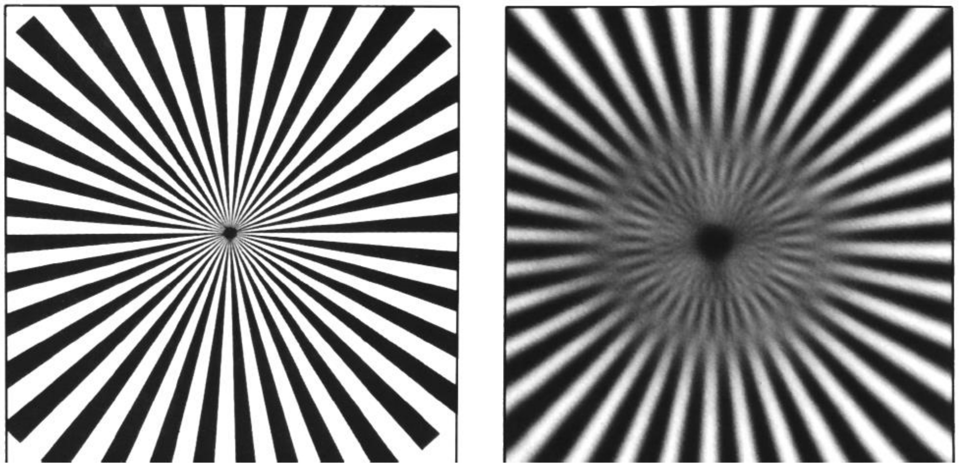


Se ve en esta figura que la función toma valores negativos, lo que indica una inversión del contraste.

Para el caso de las pupilas circulares, que es bastante más complicado desde el punto de vista matemático, vamos a ver un corte transversal de $\mathcal{H}\left(f_x, f_y\right)$



Acá también se ve que hay una inversión del contraste, fenómeno que se ilustra en las siguientes fotografías



Este es un clásico test, que se usa para evaluar sistemas ópticos, en el cual las frecuencias espaciales aumentan a medida que nos acercamos al centro. En este ejemplo se puede ver cómo cuando el sistema está desenfocado aparece la mencionada inversión de contraste.

Aquellos interesados en leer un poco más acerca de cómo se evalúa la MTF de un sistema óptico puede ir a

http://www.optikos.com/resource_files/pdfs/how_to_measure_mtf.pdf

Comparación entre iluminación coherente e incoherente

Hemos visto que la función transferencia incoherente poseía una frecuencia de corte igual al doble de la coherente, lo que nos tentaría presuponer que este tipo de iluminación da invariablemente imágenes de mejor calidad que la coherente. Esta conclusión no siempre es válida ya que estamos comparando frecuencias de corte que no son estrictamente comparables. La f_{cc} determina la componente de máxima frecuencia en la amplitud de la imagen, mientras que la f_{ci} se refiere a la máxima frecuencia en la intensidad de la misma.

Dado que la cantidad mensurable es la intensidad, deberíamos hacer la comparación de los dos tipos de iluminación en términos de esta magnitud.

Igualmente resultará difícil generalizar y decir qué tipo de iluminación es mejor dado que dicho concepto puede variar de un caso al otro, es decir de acuerdo a los detalles que queramos que estén presentes en la imagen. Aún así es posible establecer algunas pautas.

Análisis del espectro de frecuencias en la imagen

Comparemos los espectros de frecuencias en la distribución de intensidad de la imagen para ambos tipos de iluminación.

Para el caso coherente tenemos que $I_i = \left|\tilde{h} \otimes E_G\right|^2$ luego

$$\mathbb{F}\left[I_i\right] = \mathbb{F}\left[\tilde{h} \otimes E_G\right] \otimes \mathbb{F}\left[\tilde{h}^* \otimes E_G^*\right] = \mathbb{F}\left[\tilde{h}\right] \cdot \mathbb{F}\left[E_G\right] \otimes \mathbb{F}\left[\tilde{h}^*\right] \cdot \mathbb{F}\left[E_G^*\right] =$$
$$= H(f) \cdot \mathbb{F}_f\left[E_G\right] \otimes H^*(-f) \cdot \mathbb{F}^*_{-f}\left[E_G\right] = H(f) \cdot \mathcal{E}_G(f) \odot H(f) \cdot \mathcal{E}_G(f)$$

Esto es, obtenemos la autocorrelación del espectro del campo geométrico que está multiplicado por la función transferencia coherente.

Por otra parte, para el caso incoherente habíamos visto que $I_i = \left|\tilde{h}\right|^2 \otimes I_G$ con lo cual

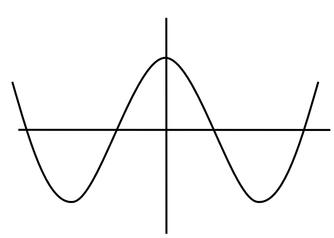
$$\mathbb{F}\left[I_i\right] = \left[\mathbb{F}\left[\tilde{h}\right] \otimes \mathbb{F}\left[\tilde{h}^*\right]\right] \cdot \left[\mathbb{F}\left[E_G\right] \otimes \mathbb{F}\left[E_G^*\right]\right] = \left[H(f) \otimes H^*(-f)\right] \cdot \left[\varepsilon_G(f) \otimes \varepsilon_G^*(-f)\right] =$$
$$= \left[H(f) \odot H(f)\right] \cdot \left[\varepsilon_G(f) \odot \varepsilon_G(f)\right]$$

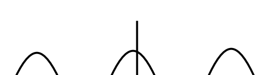
Evidentemente estos dos resultados no son iguales y por lo tanto el espectro de frecuencias es distinto. Además vemos que dependerá fuertemente de la distribución de intensidad y fase del objeto.

Para ilustrar esto veamos un ejemplo. Consideremos un objeto con la misma transmitancia en intensidad pero con distintas distribuciones de fase.

Tomemos la transmitancia en intensidad $T(x_o, y_o) = \cos^2(2\pi\, f_o x_o)$ con $\frac{f_{cc}}{2} < f_o < f_{cc}$.

Las transmitancias de ambos objetos en amplitud vienen dadas por:



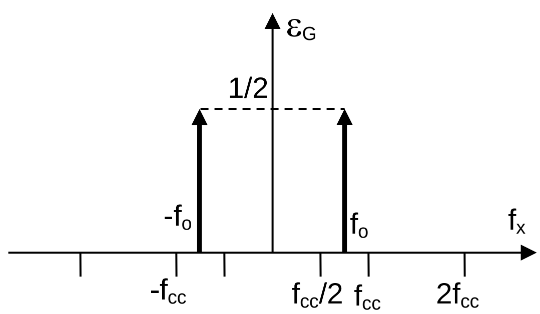
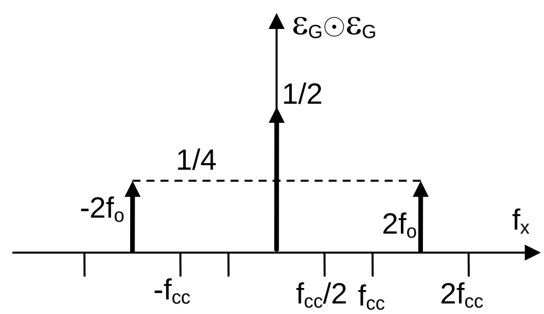
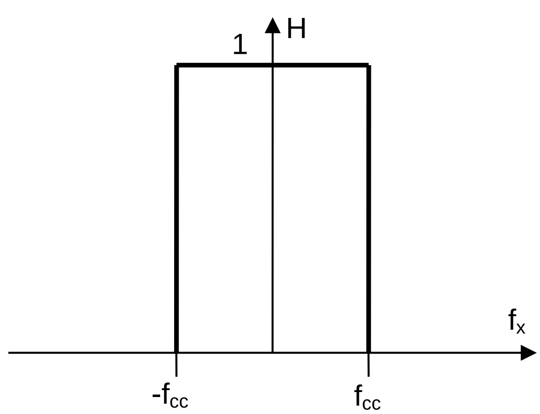
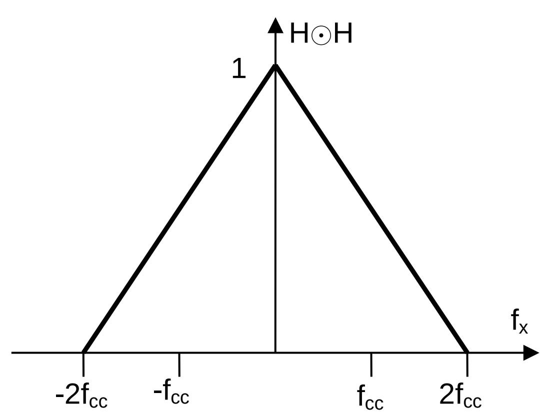


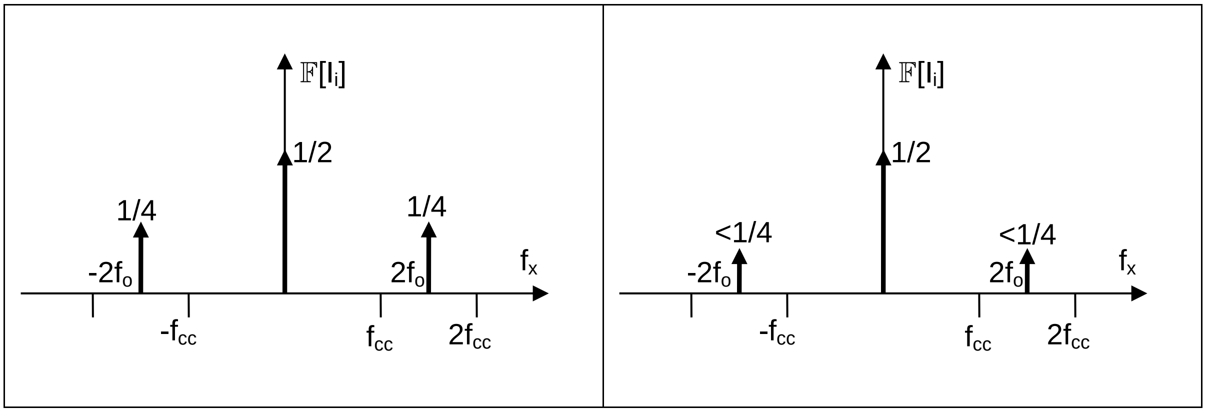
Objeto A: $t_A(x_o, y_o) = \cos(2\pi\, f_o x_o)$

Objeto B: $t_B(x_o, y_o) = \left|\cos(2\pi\, f_o x_o)\right|$

Supongamos ahora que nuestro sistema óptico está limitado por difracción y tiene una pupila cuadrada de lado *a* $P(x_p) = \mathrm{rect}\left(\dfrac{x_p}{a}\right)$.

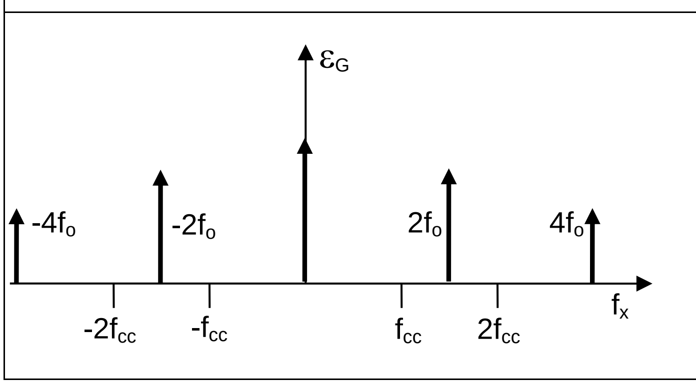
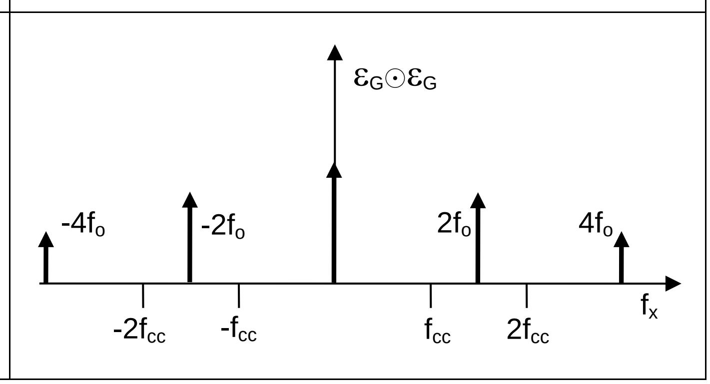
Comencemos comparando, en una tabla, los dos tipos de iluminación para el objeto A

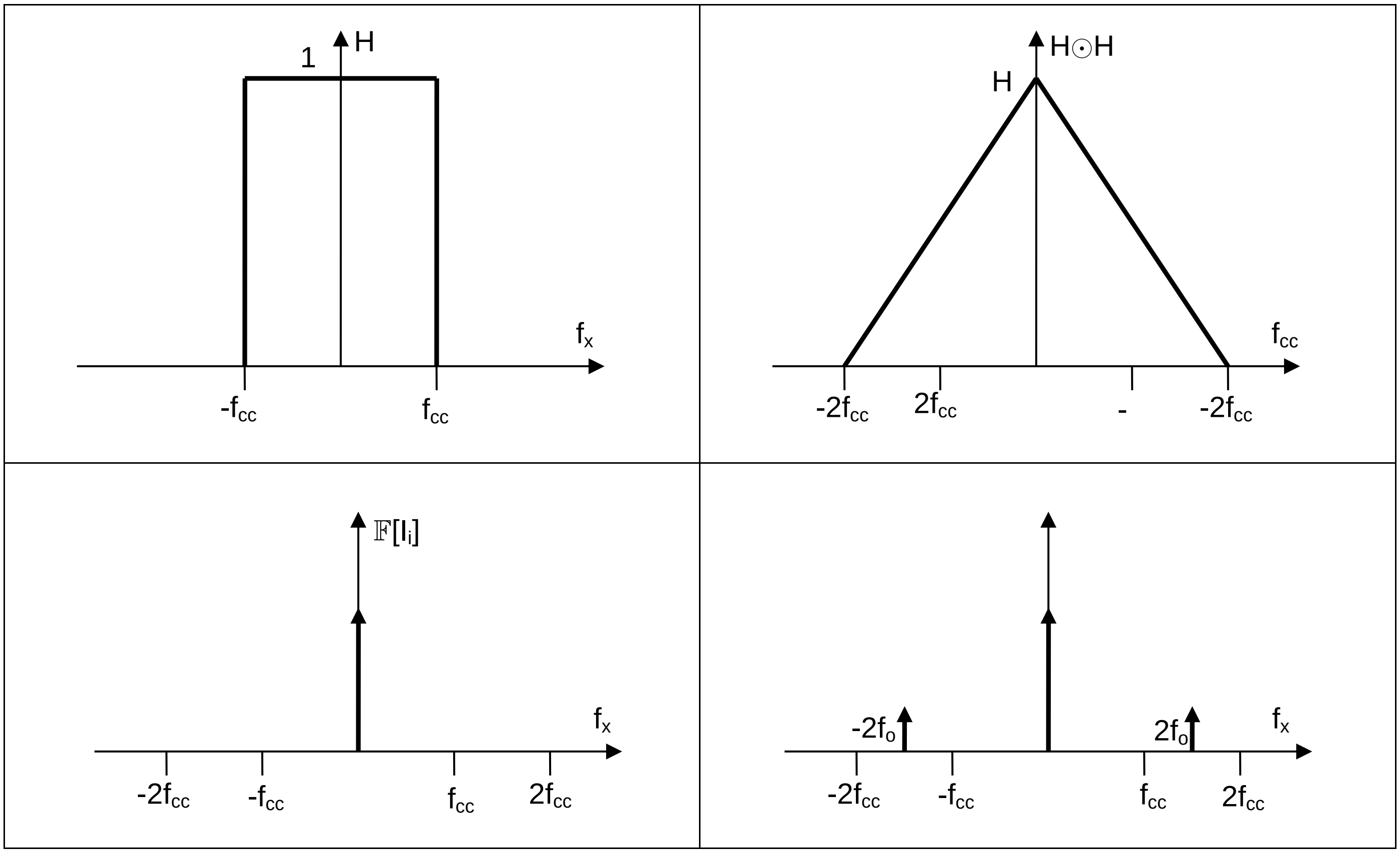
COHERENTE	INCOHERENTE
	
	



Vemos que para el objeto A resulta una imagen con mejor contraste aquella brindada por la iluminación coherente.

Estudiemos ahora que sucede con el objeto B. Para ello realicemos nuevamente la tabla comparativa con los dos tipos de iluminación teniendo en cuenta que la transformada de $\left|\cos\left(2\pi f_o x_o\right)\right|$ corresponde a un conjunto de deltas, cada vez de menor intensidad a medida que se alejan del centro y todas ellas separadas a una distancia $2f_o$. Acá nos limitaremos a dibujar la central y las dos más próximas.

COHERENTE	**INCOHERENTE**
	

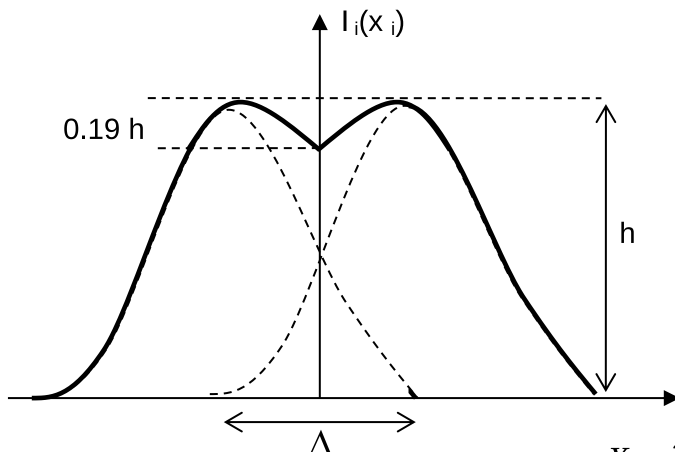


En este caso vemos que la iluminación incoherente resulta mejor que la coherente ya que en esta última se pierde toda la información. Este ejemplo muestra claramente que lo que denominaríamos como la iluminación más adecuada depende de cada caso particular.
Como vimos este es un criterio basado en el análisis de los espectros de frecuencia, estudiemos ahora otro criterio posible.

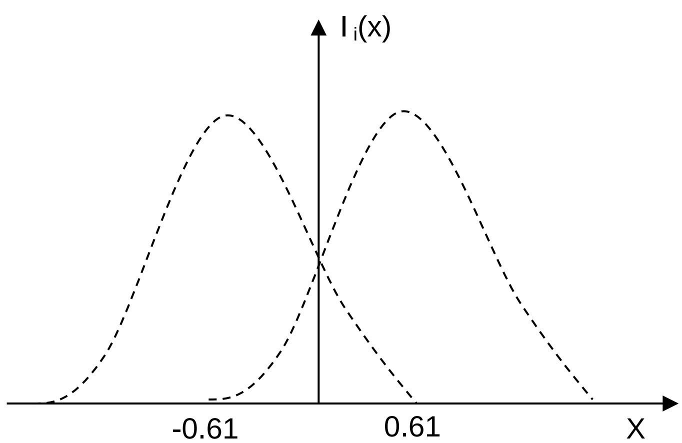
Resolución de dos puntos

Un segundo posible criterio de comparación es el relacionado con la capacidad del sistema óptico para resolver dos fuentes puntuales, según el tipo de iluminación empleada. Este criterio es muy usado para determinar calidad de lentes, particularmente en aplicaciones astronómicas.
De acuerdo con el criterio de Rayleigh dos fuentes puntuales incoherentes están justamente resueltas, por un sistema limitado por difracción, cuando en el plano imagen el máximo central del disco de Airy producido por una de ellas coincide con el primer mínimo del producido por la otra.

En el caso de iluminación incoherente habíamos visto que $I_i = \left|\tilde{h}\right|^2 \otimes I_G$. En este caso la imagen geométrica corresponde a una delta mientras que $\left|\tilde{h}\right|^2$ es justamente el diagrama de Airy (suponemos obviamente una pupila circular). La convolución da como resultado el diagrama centrado en las dos deltas, luego las dos fuentes puntuales están justamente resueltas cuando en el plano imagen la separación entre dichas funciones delta sea $\Delta = \dfrac{1.22\lambda\, d_i}{D}$ asumiendo que *D* es el diámetro de la lente y d_i la distancia al plano imagen



Vamos a hacer un cambio de v que llamamos $X = \dfrac{x_i D}{\lambda d_i}$ de esta forma el gráfico queda



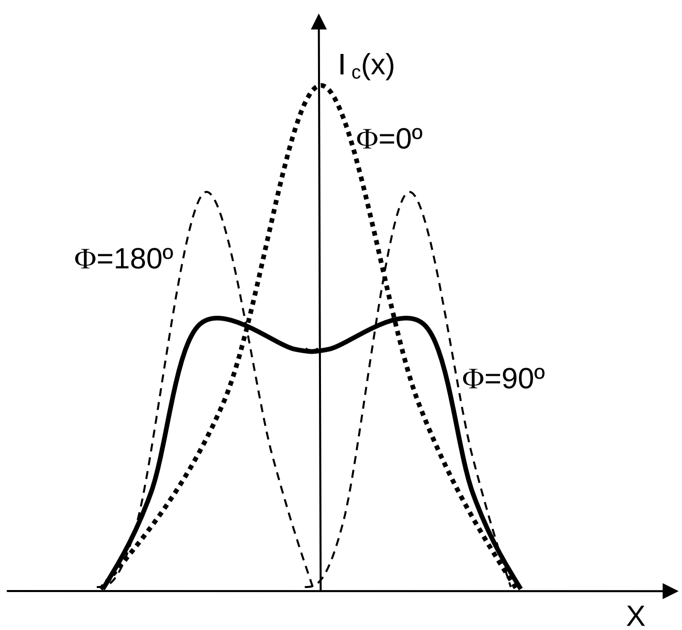
Claramente para λ, D, d_i fijos la magnitud Δ estará determinada por la separación entre los puntos objeto. Podríamos ahora preguntarnos que pasa si manteniendo fija dicha

separación iluminamos el objeto coherentemente. Esta cuestión por supuesto que no interesa en astronomía, donde siempre cada punto (estrella) es incoherente con el otro, pero si es de interés en microscopía donde el tipo de iluminación puede elegirse.
Así como discutimos cuando analizamos el criterio anterior, la iluminación más adecuada dependerá de la distribución de fase del objeto.

Para iluminación coherente teníamos que $I_c = \left|\tilde{h} \otimes E_G\right|^2$. Asumiendo el cambio de variables anteriormente mencionado, la intensidad vendrá dada por:

$$I_c\left(X\right) = \left| 2\frac{J_1\left[\pi\left(x-0.61\right)\right]}{\pi\left(x-0.61\right)} + e^{i\phi}\, 2\frac{J_1\left[\pi\left(x+0.61\right)\right]}{\pi\left(x+0.61\right)} \right|^2$$

donde Φ es fase relativa entre los dos puntos. En el siguiente gráfico puede verse que cuando los puntos están en cuadratura (Φ = 90º) la resolución es equivalente al caso incoherente, en cambio para Φ = 0º la resolución empeora y para Φ = 180º mejora notablemente.

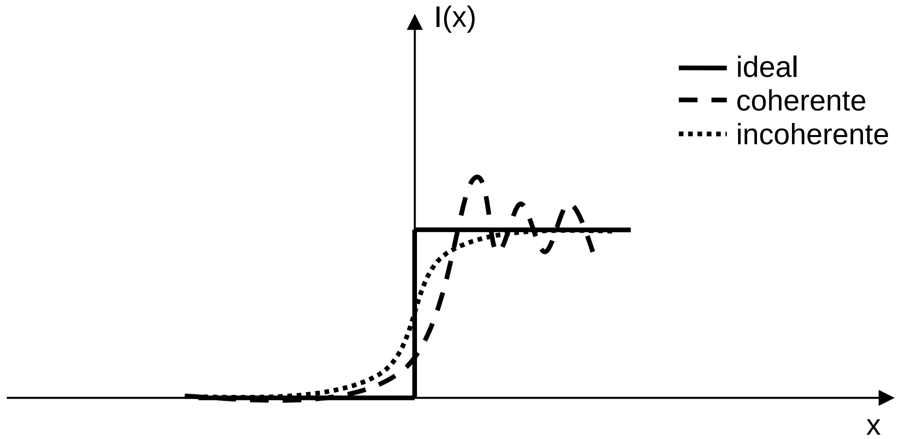


Otros efectos

Existen otras diferencias significativas entre un tipo de iluminación y otro

- Bordes abruptos

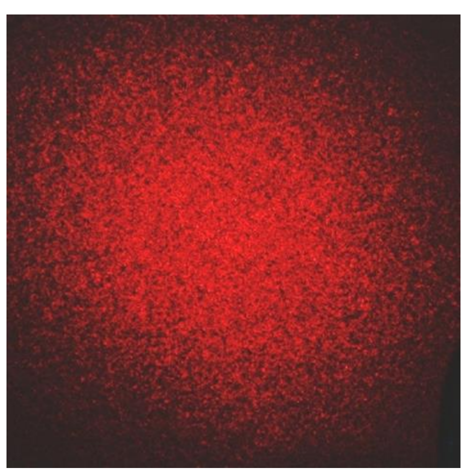
Supongamos que parte del objeto viene descripta por una función escalón, en el gráfico contiguo se ilustra el comportamiento distinto entre un tipo de iluminación y la otra. Con iluminación coherente aparece un efecto conocido como *ringing* debido a la difracción en el borde, además se produce un corrimiento de la sombra geométrica



- Speckle

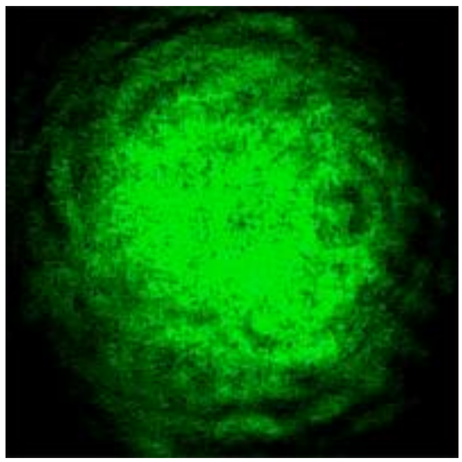
Es la apariencia granulosa de una imagen formada con iluminación altamente coherente. Su presencia se hace más evidente cuando la imagen corresponde a un objeto difusor, en este caso podemos pensar que cada punto objeto se comporta como una fuente puntual que emite ondas esféricas que al ser coherentes con las producidas por los vecinos interfieren. Luego volveremos sobre este tema.

Salvo que se utilice esta estructura para un fin muy especial, por ejemplo interferometría, en general resulta un efecto no deseado que degrada la calidad de la imagen.



- Imperfecciones en el sistema formador de imágenes

La iluminación altamente coherente es muy sensible a imperfecciones de los elementos ópticos utilizados en el montaje experimental. Rayas, picaduras, polvo, etc son todas imperfecciones que originan diagramas de difracción que se superponen a la imagen.



VI. FILTRADO ESPACIAL

CLASE 13

INTRODUCCIÓN

Con anterioridad hemos utilizado el concepto de sistemas lineales para describir el proceso de formación de imágenes y analizar la influencia de distintos parámetros, tales como la coherencia de la fuente, geometría del sistema óptico, etc., sobre dicho fenómeno. En particular vimos que era posible estudiar este proceso basándonos en un análisis de las frecuencias espaciales presentes en el objeto y en la imagen y cómo el sistema óptico modificaba este contenido.

Una de las ramas más importantes de la óptica moderna surgió a partir de la idea que el espectro de frecuencias podía además modificarse intencionalmente de tal manera que a la salida del sistema se obtuviera determinada información de la señal de entrada. Tal información puede o no estar relacionada con el concepto tradicional de imagen. Justamente esta amplia rama se llama tradicionalmente procesado óptico de la información, aunque con el correr de los años tiende a ser un procesado híbrido de la información, donde el término híbrido se refiere a los sistemas con una parte electrónica y una óptica. Las operaciones que pueden realizarse así como las áreas de aplicación son muy numerosas y variadas: tratamiento de imágenes biológicas, visión artificial, robótica, teledetección, control de calidad, computación óptica, encriptación de imágenes, etc. son sólo algunos ejemplos. En general estas operaciones requieren del manejo de un número muy grande de datos y es por eso que el paralelismo inherente a la óptica constituye un factor importante para disminuir el tiempo de procesado. Basta pensar que una lente permite obtener la transformada de Fourier de un objeto sin importar el tamaño del mismo (obviamente dentro de ciertos límites) a la velocidad de la luz.

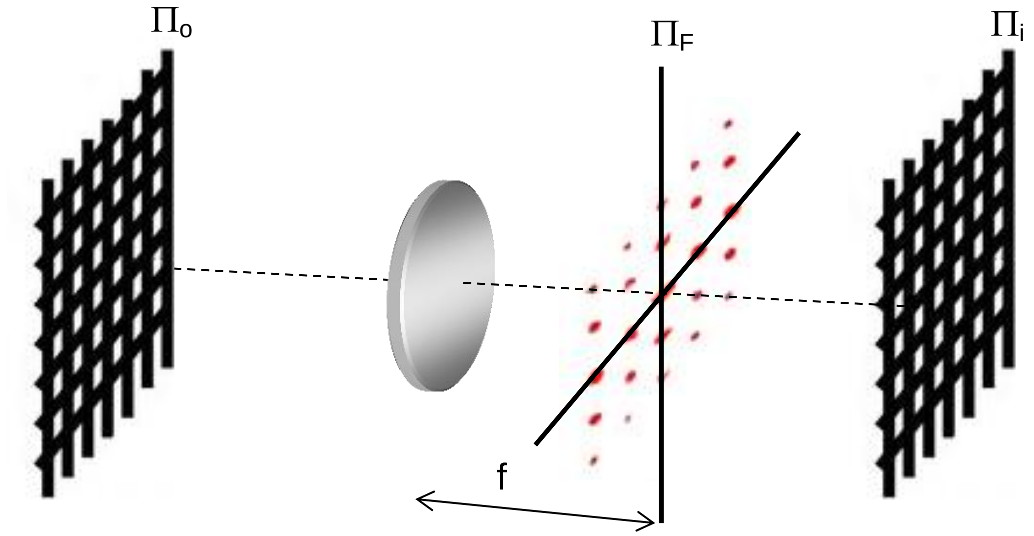
Para comenzar tal vez lo más sencillo sea analizar el proceso de filtrado espacial.

Un hecho cotidiano es el siguiente (es un ejemplo electrónico pero es útil para entender los conceptos): Tenemos una pieza de audio contaminada con ruido o que no nos gusta como suena; nos gustaría modificar esta señal. Tenemos un equipo que posee un ecualizador, este dispositivo no es más que un analizador de espectro que toma la señal de entrada, la transforma Fourier y luego envía las distintas frecuencias a determinadas bandas. El peso de estas bandas, esto es el volumen individual se puede controlar mediante potenciómetros. Así subiendo unos y bajando otros somos capaces de modificar el espectro de forma tal que una vez antitransformado sea emitido por los parlantes con el sonido deseado. Obviamente en un filtrado se eliminan o atenúan frecuencias de toda una banda de la señal de entrada, así si el ruido tiene una frecuencia coincidente, por ejemplo, con parte del espectro del cantante, cuando este es eliminado también se cambiará el timbre de la voz.

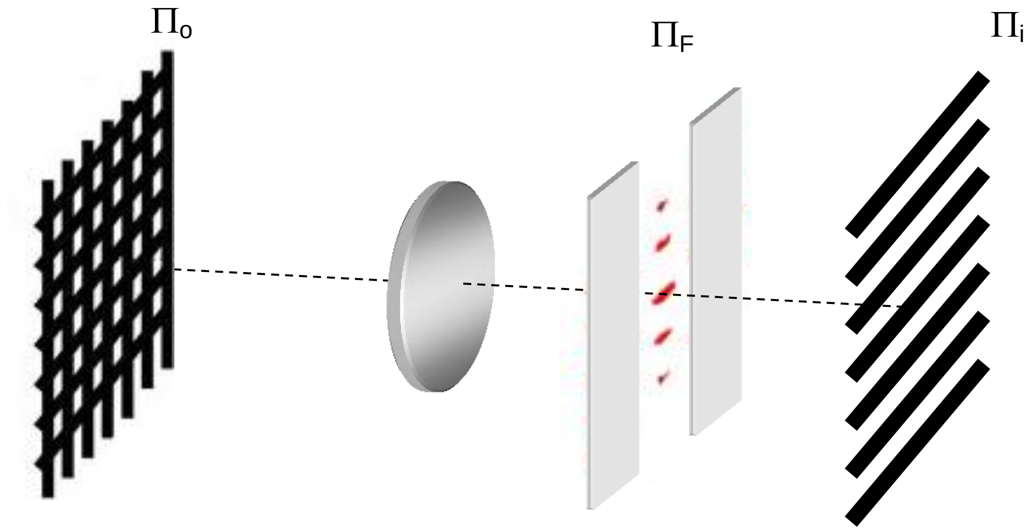
En óptica el primero en modificar intencionalmente las frecuencias espaciales de una imagen fue Abbe en 1893 y luego Porter en 1906. A continuación describiremos estas sencillas experiencias que muestran claramente el proceso de formación de imágenes y su relación con el análisis espectral.

Experiencia de Abbe – Porter

Abbe y Porter utilizaron como objeto una red cuadriculada de alambre y la iluminaron coherentemente en un dispositivo como el esquematizado en la figura



La imagen final en el plano Π_i era modificada obstruyendo con distintos tipos de pantallas (diafragmas, ranuras, etc.) frecuencias espaciales en el plano Π_F. En particular para una ranura se obtiene



Claramente los filtros usados en estas experiencias son los más elementales en donde la luz pasa o no pasa. Veamos ahora que sucede con un filtro que modifique la fase en vez de la amplitud.

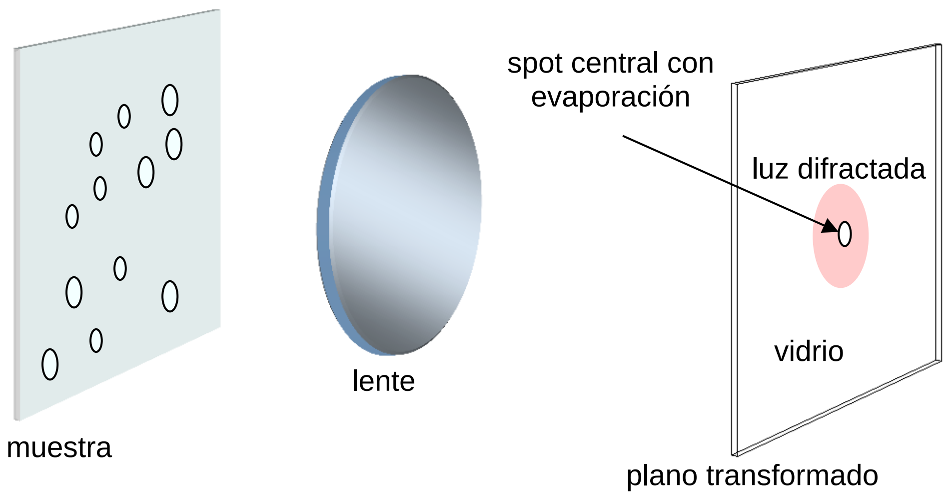
El microscopio de contraste de fase de Zernike

Muchos objetos de interés en microscopía son transparentes, por ejemplo las bacterias. Una solución que adoptan muchas veces los biólogos es teñirlas con lo cual generalmente observan bacterias muertas.
Sabemos que cuando la luz encuentra en su camino un obstáculo, esta se difracta ya sea el objeto de fase o de amplitud. Así la luz que atraviese la muestra transparente de líquido y bacterias se difractará por tener estas distinto índice de refracción que el medio que las rodea. Obviamente este solo hecho no basta para hacerlas visibles en un microscopio común. Sin embargo si esta muestra es iluminada coherentemente, en el plano transformado se tendrá un orden central, esto es correspondiente a bajas frecuencias, originado básicamente por la luz no difractada y una distribución de luz en frecuencias más altas que provendrá de la difracción en las bacterias. Una técnica posible para hacerlas visibles sería obstruir el orden central pero en este caso las variaciones de intensidad en la imagen no estarían linealmente relacionadas con las variaciones de fase.
En 1935 el físico holandés Fritz Zernike propuso una técnica de contraste de fase por la que le otorgaron el premio Nobel en 1953.
El estableció que bajo ciertas condiciones la intensidad observada podía relacionarse linealmente con el desfasaje introducido por el objeto.
Supongamos que dicho objeto al ser transparente sólo introduce cambios de fase, o sea que su transmisión en campo viene dada por $t(x,y)=\exp\left[i\,\phi(x,y)\right]$.
Si dichos desfasajes son menores que un radián, vale que $t(x,y)\cong 1+i\,\phi(x,y)$.
Esta condición es necesaria para que exista una relación lineal entre la intensidad en el plano final y la fase.
Supongamos que en un microscopio común observamos este objeto, iluminándolo con una fuente coherente. Consideremos además, por simplicidad, que la amplitud del campo que ilumina al objeto es 1 y que la pupila del sistema es muy grande con lo cual $\tilde{h}=\delta$. Entonces

$$I_{ic}=\left|\tilde{h}\otimes E_G\right|^2=\left|1+i\,\phi(x,y)\right|^2=1+\phi^2$$

Si, como dijimos las variaciones de fase son pequeñas el término ϕ^2 es despreciable y en consecuencia obtenemos una distribución de intensidad uniforme, como era de esperar.

Si observamos la ecuación anterior vemos que existen dos términos: uno que corresponde a la luz no difractada, y otro que contiene la información de la distribución de fase, y por lo tanto describe a la luz difractada, $\left[i\phi(x,y)\right]$. Este último está desfasado en 90º con respecto al fondo, ya que uno es real puro y el otro imaginario puro, físicamente esto significa que ambas ondas no interfieren entre sí. Zernike se dio cuenta de este hecho y propuso agregar un desfasaje adicional de 90º o 270º de manera que sí pudiesen hacerlo y lograr de esta forma un cambio de intensidad. Así pensó en introducir, en el plano transformado, un filtro de fase que retardara al orden cero con respecto a la luz difractada. Una manera de hacerlo es por ejemplo evaporando una sustancia transparente, de espesor adecuado, sobre un sustrato de vidrio. Un esquema del dispositivo puede verse en la figura siguiente



Luego, supongamos que el desfasaje introducido en el orden cero sea de π/2, entonces

$$I_{ic} = \left| e^{i\pi/2} + i\phi(x,y) \right|^2 = \left| i\left(1+\phi(x,y)\right) \right|^2 \cong 1 + 2\phi$$

Recordemos que en esta aproximación despreciamos ϕ^2 y órdenes superiores.

Vemos que en esta expresión existe una relación lineal entre la intensidad registrada y la fase del objeto.

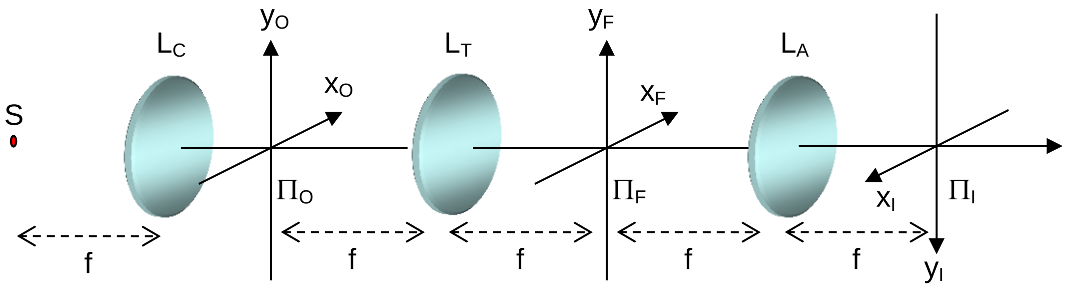
Análogamente si el desfasaje introducido en el orden cero es de 3π/2 obtenemos

$$I_{ic} = \left| e^{i\,3\pi/2} + i\,\phi(x,y) \right|^2 = \left| -i\left(1-\phi(x,y)\right) \right|^2 \cong 1-2\phi$$

que corresponde también a una relación lineal pero con el contraste cambiado.
Cabe destacar que en los microscopios de contraste de fase comerciales la fuente tiene forma de anillo (en vez de ser puntual) y el filtro correspondiente también ajusta a esa forma. Debemos recordar que el plano transformado es el conjugado de la fuente y por lo tanto allí se obtiene una imagen de la misma.
Quienes estén interesados en este tema pueden encontrar información en el sitio http://www.microscopyu.com/articles/phasecontrast/phasemicroscopy.html

PROCESADORES ÓPTICOS COHERENTES

Un procesador óptico es un sistema en el cual una señal de entrada es transformada, obteniéndose su espectro de frecuencias. Mediante el uso de filtros es posible modificar dicho espectro en el plano de Fourier y luego mediante una antitransformación se obtiene a la salida la señal deseada.
Una de las configuraciones posibles para implementar este tipo de operaciones es la denominada 4*f* y se esquematiza en la siguiente figura.



La luz proveniente de una fuente puntual monocromática *S*, ubicada en el foco de una lente colimadora L_C, incide sobre el plano objeto Π_O. Dicho plano coincide con el plano focal anterior de la lente L_T. El campo a la salida del objeto viene dado por $E_o(x_o, y_o)$ y constituye la señal de entrada. Luego, sobre el plano focal posterior Π_F de la lente transformadora, se obtiene su transformada de Fourier dada por $C_1\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$, donde C_1 es una constante compleja y $\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) = \mathbb{F}\left[E_o(x_o, y_o)\right]$. En dicho plano se ubica el filtro que actuará modificando la amplitud y la fase del espectro. La transmisión compleja de este filtro viene dada por $t_F\left(x_F, y_F\right) = C_2 H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$. Así la distribución de campo detrás del filtro será proporcional a $\mathcal{E} \cdot H$. Una segunda lente L_A, cuyo plano focal anterior coincide con el plano focal posterior de L_T, brinda la transformada de esta distribución de campo sobre su plano focal posterior Π_I. Esto es, se obtiene $\mathbb{F}\left[\mathcal{E} \cdot H\right] = E_o \otimes h$. Si observamos el dibujo correspondiente al procesador vemos que el sistema de coordenadas en el plano de salida está invertido. La razón de esto es que el sistema de lentes realiza una inversión en la imagen, lo cual es coincidente con el hecho de que en realidad la señal de entrada sufre dos transformadas. Ahora bien, dado que es más elegante matemáticamente decir que una lente transforma y la otra antitransforma, se opta por adoptar un cambio de signos en (x_I, y_I). Luego la intensidad a la salida viene dada por

$$I\left(x_i, y_i\right) = K\left|\int \int_{-\infty}^{\infty} E\left(\xi, \eta\right) h\left(x_i - \xi, y_i - \eta\right) d\xi \, d\eta\right|^2$$

Vemos la analogía con el resultado obtenido para formación de imágenes, sólo que antes h representaba la función respuesta al impulso del sistema óptico y ahora representa la respuesta al impulso del filtro. Este resultado es razonable ya que cuando estudiamos formación de imágenes el factor limitante para la luz de entrada era la pupila y justamente h era la transformada de Fourier de dicha función. Ahora el objeto que limitará la entrada de luz al sistema será el filtro y en consecuencia h es su transformada de Fourier. En este razonamiento estamos asumiendo que el sistema óptico no es el responsable del recorte de frecuencias espaciales, aunque en realidad la resolución del mismo debe tenerse en cuenta a la hora de diseñar un procesador en función del ancho de banda con el que se va a trabajar.

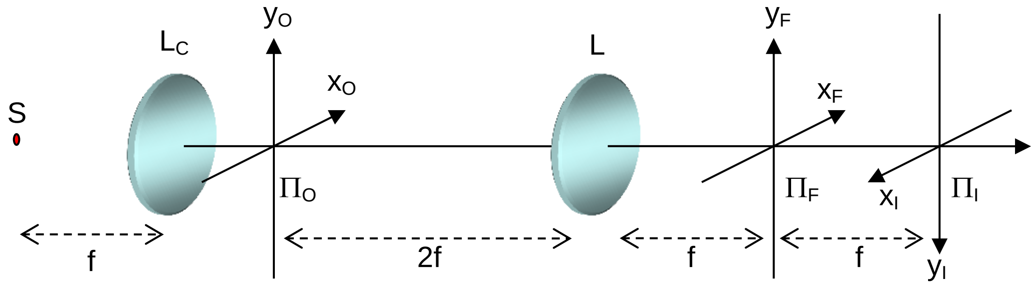
Tal como habíamos visto con anterioridad, el objeto algunas veces conviene ubicarlo detrás de la lente L_T de forma tal que el aumento de su transformada pueda regularse y no esté determinada por la distancia focal de esta lente. De esta forma resulta menos crítico el ajuste entre $\mathcal{E}$ y la escala con la que se realizó el filtro.

La configuración 4*f* es clásica porque, como vimos con anterioridad, no introduce fases espurias y permite obtener la transformada de Fourier exacta.

También existen otras configuraciones más sencillas en las que se emplean menos elementos ópticos. Veamos por ejemplo la esquematizada en la siguiente figura

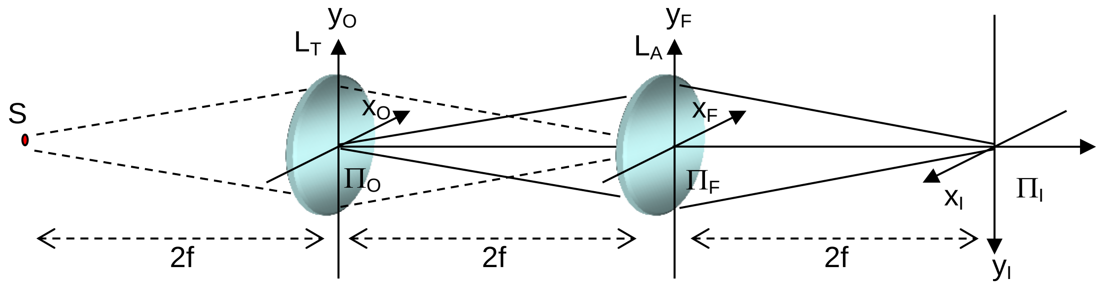
Acá la lente *L* provee la transformada de la entrada $E_o\left(x_o, y_o\right)$ sobre su plano focal Π_F (plano conjugado de la fuente) pero a la vez forma imagen del plano objeto Π_O

sobre el plano conjugado del mismo Π_I. La distancia 2*f* a la que se encuentra el objeto de la lente no es esencial sólo se toma así si se desea tener aumento unitario en el plano final. Lo único importante es que Π_O y Π_I sean planos conjugados.



Un inconveniente menor de este montaje es que al no estar el objeto sobre el plano focal anterior de la lente se introduce una fase espuria. El mayor problema es que al estar el objeto a una distancia mayor que *f* de la lente esta debe ser de mayor diámetro que la utilizada en la primera disposición para captar las mismas frecuencias espaciales.

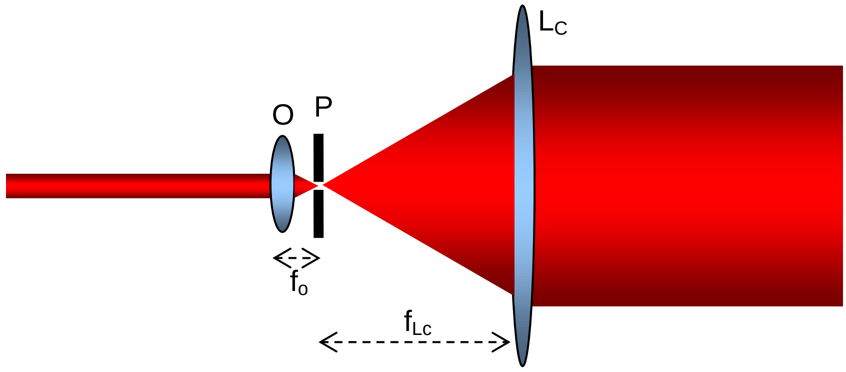
Variantes de las configuraciones analizadas, en las cuales la fuente no está en el infinito, también son muy utilizadas. Veamos el siguiente ejemplo, usualmente llamado procesador convergente.



En esta disposición se utiliza una primera lente para colectar la luz de la fuente puntual y para realizar la transformada de Fourier de la transparencia que se halla después de ella sobre el plano Π_O. El plano de frecuencias Π_F se ubica justo después de la segunda lente y es, tal como explicamos anteriormente, el plano conjugado de la fuente. A la vez esta segunda lente antitransforma el espectro sobre el plano de salida que es el plano conjugado del plano objeto. Este montaje en particular presenta la ventaja de que hay menos recorte de frecuencias espaciales que en el *4f* y que al igual que en este no se agrega ninguna fase adicional ya que el factor de fase cuadrática que se origina por propagación desde el plano objeto al transformado es cancelado por el factor de transmisión de la lente. La desventaja es que es un montaje más largo. Al igual que en los otros casos hay montajes similares en donde el objeto no esta justo detrás de la primera lente ni la disposición es *2f-2f*.

En todos estos procesadores siempre hablamos de una fuente puntual monocromática, veamos experimentalmente como se logra esto. Para ello analicemos la siguiente figura.

Un haz láser es expandido con un objetivo de microscopio *O* de modo que el haz plano se convierte en un frente de ondas esférico. Con el fin de eliminar el ruido coherente introducido por este objetivo, justo sobre el foco (que constituye la fuente puntual de donde parten ondas esféricas) se realiza un filtrado espacial ubicando un *pinhole P* que permite pasar sólo las frecuencias más bajas y eliminando en consecuencia la información del ruido. Luego una segunda lente L_C, colima el haz volviéndolo a hacer plano pero con un diámetro mucho mayor al original. Si observamos la disposición de lentes vemos que es un sistema telescópico invertido con focos coincidentes en la posición del *pinhole*.



Ejercicio 9: Vamos a suponer que se utiliza un láser de He-Ne (λ = 632.8 nm) que emite en un modo transversal TEM_{00} de perfil gaussiano cuya distribución de intensidad viene dada por

$$I(r) = \frac{2P_T}{\pi R^2}\exp\left(\frac{-2r^2}{R^2}\right) \quad \text{donde} \quad P_T = \int_0^\infty I(r)\, 2\pi\, r\, dr$$ es la potencia total y *R* el radio donde la intensidad cae a e^{-2}.

i) Calcule la distribución de intensidad en el plano focal del expansor

ii) Suponiendo que la distancia focal del objetivo es f_o = 20 mm y que *R*, antes de entrar al mismo es 1 mm, calcule cuanto vale *R* en el plano focal del expansor. Este es el diámetro que se elige para el pinhole

Ejercicio 10: Supongamos que iluminamos un objeto (circunferencia de radio A) con un haz plano de perfil gaussiano

i) Demostrar que

$$\frac{P_{incidente\ sobre\ el\ objeto}}{P_{Total}} + \frac{I(borde\ del\ objeto)}{I(centro\ del\ objeto)} = 1$$

ii) Si se desea que la iluminación entre el borde y el centro no difiera en más de un 70%, demostrar que el objeto sólo aprovecha el 30% de la potencia total emitida por el láser

Ejemplos con filtros compuestos

En esta sección vamos a ver algunos ejemplos que son “históricos” pero ilustran muy bien el proceso de filtrado.

En la introducción a este tema hemos visto dos tipos de filtros, uno de amplitud (Abbe-Porter) y otro de fase (Zernike). En general el procesado de una señal requerirá de filtros compuestos en los que tanto la amplitud como la fase sean alteradas simultáneamente. En un principio esto se lograba por la combinación de un filtro de amplitud y un filtro transparente que, mediante adecuados cambios de espesor, modificaba la fase. Esta técnica a menudo es bastante limitada ya que es muy difícil de implementar salvo algunos casos sencillos.

Para ubicarnos históricamente en el tema, cabe destacar que recién a comienzos de los ´50 el procesado óptico comienza a tener auge. Esto se debe fundamentalmente a que en esos años se hizo evidente la conexión entre óptica y la teoría de la información mediante publicaciones tales como “Fourier treatment of optical processes” [JOSA **42**, 127 (1952)] y “Optics and Communication Theory” [JOSA **43**, 229 (1953)] de P.Elias et. al y “Spatial filtering in optics” de E.O’Neil [IRE Trans. Inform. Theory **IT-2**, 56 (1956)]. También en la parte experimental A.Maréchal hizo grandes aportes principalmente en el mejoramiento de imágenes fotográficas con filtros compuestos. Algunos años antes, en 1949, D. Gabor [Proc. Roy. Soc. **A197**, 454, (1949)] había propuesto un nuevo proceso a partir del cual era posible reconstruir una imagen sin lentes. El lo denominó reconstrucción del frente de ondas ya que permitía modificar la amplitud y fase de un haz de luz; posteriormente fue conocido como holografía. Este proceso no fue tomado en un principio con mucho entusiasmo debido a la dificultad de su implementación, pero el advenimiento del láser en 1960 junto con una modificación de la técnica original sugerida por E.Leith y J.Upatnieks [JOSA **52**, 1123 (1962)] provocaron una verdadera revolución en este campo. La síntesis de filtros complejos avanzó rápidamente a partir de que estos pudieron realizarse de forma holográfica.

Por ahora analizaremos unos ejemplos sencillos y después de estudiar holografía volveremos sobre el tema para tratar filtros más complicados.

Como vimos anteriormente, el filtro tiene en general una transmisión compleja en campo dada por:

$$t_F\left(x_F, y_F\right) = C \cdot H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) \propto \left|H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right| \cdot \exp\left[i\,\phi\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right]$$

que debe cumplir las condiciones $\left| H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f} \right) \right| \leq 1$ ya que el filtro no puede amplificar, y $0 \leq \phi\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f} \right) \leq 2\pi$.

Implementación óptica de la operación derivada

Veamos ahora un ejemplo sencillo de computación óptica, la operación derivada. Supongamos que se desea obtener la derivada según la variable x de la función de entrada $E(x,y)$, esto es $\frac{\partial E(x,y)}{\partial x}$. Así en el plano de salida del procesador óptico buscamos $\frac{\partial E(x,y)}{\partial x} = \mathbb{F}^{-1}\left[\mathcal{E}\left(f_x, f_y \right) \cdot H\left(f_x, f_y \right) \right]$ donde $\mathcal{E}\left(f_x, f_y \right)$ es la transformada de Fourier de la entrada y $H\left(f_x, f_y \right)$ es la transmisión compleja del filtro. O sea $\mathcal{E}\left(f_x, f_y \right) \cdot H\left(f_x, f_y \right)$ es la distribución de campo a la salida del plano transformado.

Ahora bien, tenemos que $E(x,y) = \int \int_{-\infty}^{\infty} \mathcal{E}\left(f_x, f_y \right) \cdot e^{2\pi i\left(f_x \cdot x + f_y \cdot y \right)} df_x \, df_y$

Luego

$$\frac{\partial E(x,y)}{\partial x} = \int \int_{-\infty}^{\infty} \mathcal{E}\left(f_x, f_y \right) \cdot 2\pi i \, f_x \cdot e^{2\pi i\left(f_x \cdot x + f_y \cdot y \right)} df_x \, df_y = \mathbb{F}^{-1}\left[\mathcal{E}\left(f_x, f_y \right) \cdot 2\pi i \, f_x \right]$$

En consecuencia $H\left(f_x, f_y \right) = 2\pi i \, f_x$. Esta transmisión nos dice que la amplitud debe crecer linealmente con la frecuencia mientras que la fase debe ser $e^{i\frac{\pi}{2}}$ para frecuencias positivas y $e^{i\frac{3\pi}{2}}$ para negativas. Por lo tanto debemos hacer una transparencia fotográfica con máxima transmisión para las altas frecuencias y que disminuya linealmente a medida que estas bajan. El filtro de fase se puede realizar mediante dos evaporaciones de fluoruro de magnesio de espesor conveniente sobre un sustrato de vidrio plano.

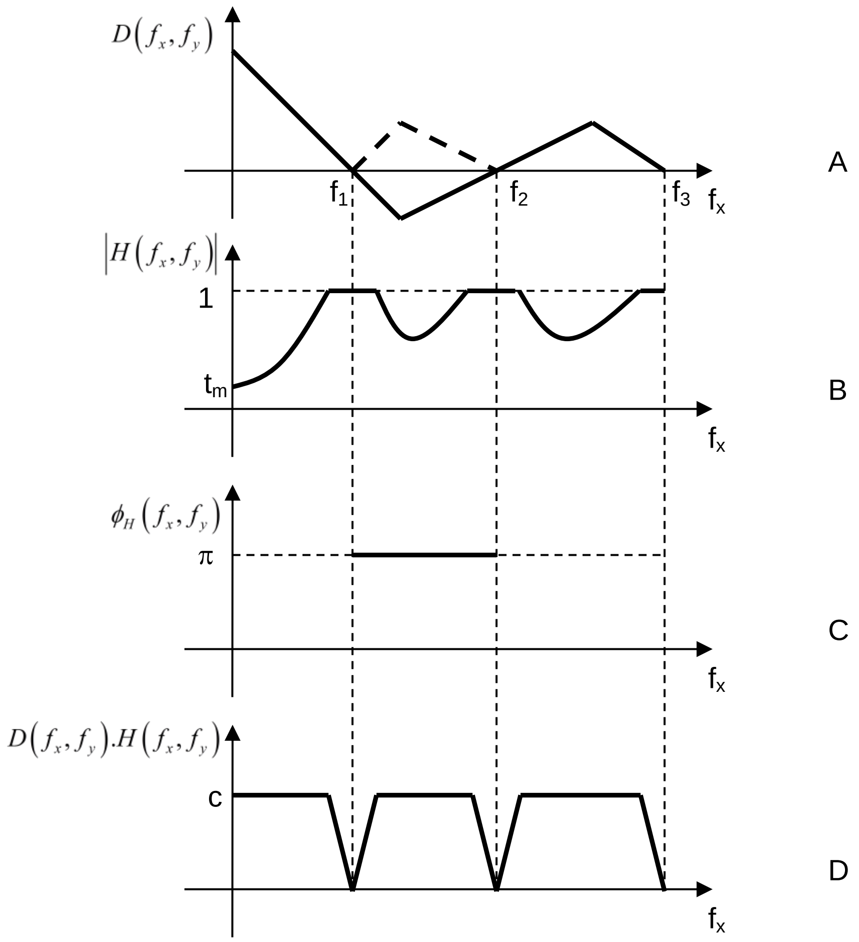
Más adelante veremos que esta operación se puede llevar a cabo en forma más sencilla con un filtro holográfico.

Filtrado inverso, aplicación a una fotografía movida

El filtrado inverso fue unas de las técnicas introducidas por Maréchal para mejorar imágenes empobrecidas por defectos en el registro de las mismas. Esto es imágenes movidas, fuera de foco, etc.

Llamemos $\mathcal{E}_D\left(f_x, f_y\right)$ a la transformada de Fourier de la imagen distorsionada, que podemos escribir como $\mathcal{E}_D\left(f_x, f_y\right)=\mathcal{E}\left(f_x, f_y\right)\cdot D\left(f_x, f_y\right)$, donde $\mathcal{E}\left(f_x, f_y\right)$ es la transformada de la imagen sin distorsionar y $D\left(f_x, f_y\right)$ es la función distorsión.

La función transferencia del filtro inverso que restaure la imagen debe ser $H\left(f_x, f_y\right)=\dfrac{C}{D\left(f_x, f_y\right)}$. Notemos que en general esta función será físicamente no realizable, por ejemplo para las frecuencias espaciales en las que $D\left(f_x, f_y\right)=0$ no tendremos solución. Para comprender un poco mejor este proceso analicemos los siguientes gráficos



Supongamos, por simplicidad, que $D\left(f_x, f_y\right)$ está compuesta por tramos rectos (ver Fig. A), esto es, cada tramo será de la forma $a\ f_x + b$. Luego $\left|H\left(f_x, f_y\right)\right|$ deben ser hipérbolas de la forma $\left|H\left(f_x, f_y\right)\right| = \dfrac{C}{\left|a\ f_x + b\right|}$ (ver Fig. B). Los tramos negativos deben compensarse con cambios de fase en π (ver Fig. C).

Vemos que la función $D\left(f_x, f_y\right)$ tiene tres ceros en $\left(f_1, f_2, f_3\right)$, luego habrá tres singularidades en $H\left(f_x, f_y\right)$. Habíamos visto que $\left|H\left(f_x, f_y\right)\right| \leq 1$ ya que el filtro no puede amplificar, entonces en esos puntos la función transferencia a lo sumo valdrá 1 y no infinito.

En el otro extremo, el filtro tendrá una transmisión mínima t_m (propia del filtro) que corresponderá al máximo valor de $D\left(f_x, f_y\right)$

$$t_m = \left|H_{\min}\right| = \frac{C}{\left|D_{\max}\right|} \quad \Rightarrow \quad C = t_m \cdot \left|D_{\max}\right|$$

Luego

$$H\left(f_x, f_y\right) = \frac{t_m \cdot \left|D_{\max}\right|}{D\left(f_x, f_y\right)}$$

Ahora bien, analicemos la Fig. D. Dado que $\left|H_{\max}\left(f_x, f_y\right)\right| = 1$ entonces de la ecuación $1 = \dfrac{t_m \cdot \left|D_{\max}\right|}{\left|a\ f_x + b\right|}$ podremos despejar las frecuencias que establecen los intervalos, alrededor de $\left(f_1, f_2, f_3\right)$, para los cuales no puede eliminarse la distorsión ya que

$$\mathcal{E}\left(f_x, f_y\right) \cdot D\left(f_x, f_y\right) \cdot H\left(f_x, f_y\right) = \mathcal{E}\left(f_x, f_y\right) \cdot D\left(f_x, f_y\right) \cdot 1$$

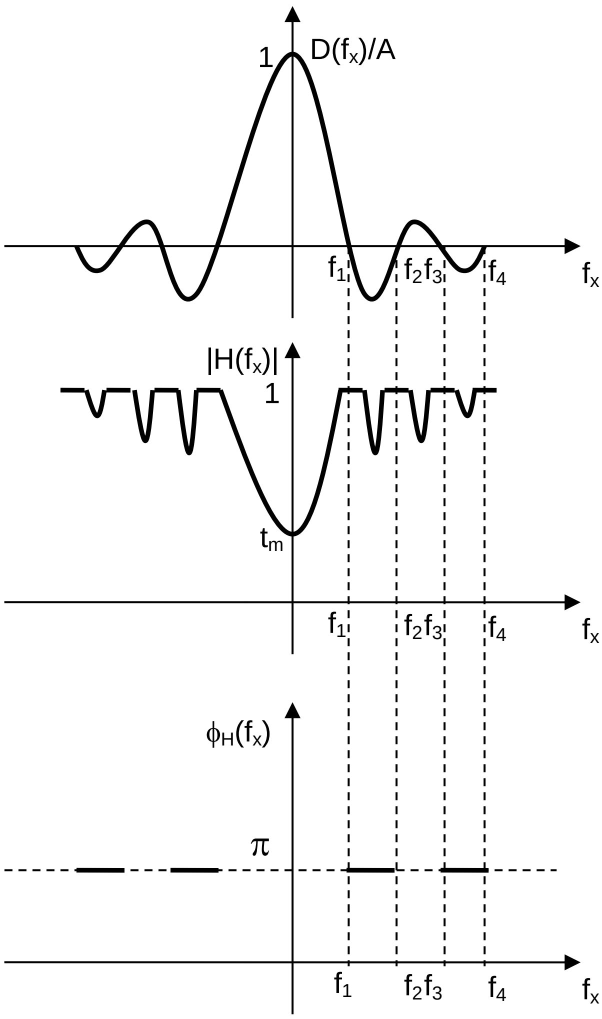
En las otras zonas, aquellas mesetas en las que se cumple que $D\left(f_x, f_y\right) \cdot H\left(f_x, f_y\right) = C$ con este tipo de filtros podremos restaurar la imagen.

Veamos el ejemplo de una fotografía movida. En este caso cada punto ideal de la transparencia, en vez de serlo, es un segmento de recta. Luego, si suponemos que el movimiento se produjo en la dirección X la función de entrada que lo describa individualmente será

$$d(x) = \begin{cases} 1 & -\frac{1}{2}\Delta x \leq x \leq \frac{1}{2}\Delta x \\ 0 & afuera \end{cases}$$

Donde Δx es la longitud del segmento debida al movimiento de la cámara o del objeto en el momento del registro.

Ahora bien para restaurar la fotografía debemos lograr transformar este segmento en un punto. Entonces si *d(x)* es la señal de entrada en un procesador coherente, sobre el plano transformado tendremos la función $D(f_x) = \mathbb{F}\left[d(x)\right] = \frac{A \sin\left(f_x \frac{\Delta x}{2}\right)}{f_x \frac{\Delta x}{2}}$ que es la que debemos filtrar. En las figuras siguientes se esquematizan el espectro del defecto y el filtro $H(f_x, f_y)$ (módulo y fase) que permite compensarlo.



CLASE 14

PROCESADORES ÓPTICOS INCOHERENTES

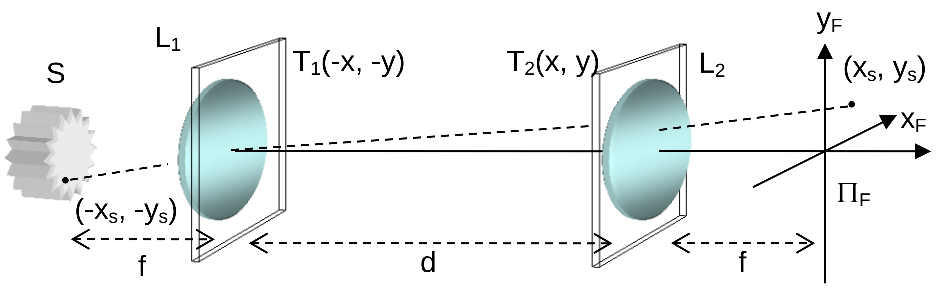
El procesado óptico incoherente presenta ciertas ventajas y desventajas con respecto al caso coherente, anteriormente analizado. Dentro de las ventajas podemos mencionar que debido a la redundancia de información presente en este tipo de iluminación no existe el ruido producido por la difracción de la luz en motas de polvo o defectos de los elementos ópticos, ni el fenómeno de speckle. Por otra parte en muchos casos puede utilizarse iluminación policromática y objetos autoluminosos. Una desventaja importante es que, en principio, estos sistemas no son lineales en campo y por lo tanto perdemos la información de la fase, además de trabajar siempre con cantidades positivas (por ejemplo no pueden restarse dos imágenes).

Procesado totalmente incoherente

Vamos a analizar dos tipos de procesadores, uno con fuente extensa y otro con fuente "puntual" (obviamente el término puntual en este caso no tiene las mismas implicancias que en el caso coherente).

<u>Implementación de la función convolución mediante un sistema basado en óptica geométrica</u>

Los sistemas procesadores basados e óptica geométrica utilizan de una u otra forma la proyección de imágenes (en general se denomina *shadow casting* o *image casting*). Veamos uno en particular que resulta interesante ya que permite implementar la operación de convolución entre dos funciones. Para ello analicemos la siguiente figura



Una fuente extensa *S* está ubicada en el plano focal anterior de la lente L_1. Inmediatamente después de dicha lente se ubica una transparencia cuya transmitancia en intensidad es $T_1(-x, -y)$. A distancia *d* de la misma e inmediatamente delante de la lente L_2 se ubica una segunda transparencia con transmitancia $T_2(x, y)$. La intensidad final se registra sobre el plano focal posterior de esta segunda lente.

Para entender como funciona el sistema consideremos un punto particular de la fuente $(-x_s, -y_s)$. Los rayos provenientes de ese punto emergerán de L_1, y en consecuencia de T_1, en forma paralela ya que la fuente se encuentra en el plano focal de esta primera lente. La distribución de intensidad que llega a la segunda transparencia es proporcional a $T_1\left(-x+\frac{d}{f}x_s,-y+\frac{d}{f}y_s\right)$. Después de pasar por T_2 los rayos se focalizan en el punto (x_s, y_s) del plano focal de la segunda lente. Sobre este plano final la distribución de intensidad será

$$I\left(x_F = x_s, y_F = y_s\right) = K\int\int_{-\infty}^{\infty} T_1\left(-x+\frac{d}{f}x_s,-y+\frac{d}{f}y_s\right)\cdot T_2\left(x,y\right)dx\,dy$$

que es justamente la buscada convolución.

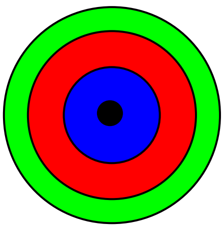
Un inconveniente que presentan los sistemas basados en óptica geométrica es que el ancho de banda espacial con el que se puede trabajar es bastante limitado. Aclaremos este punto. Para que valgan las aproximaciones de la óptica geométrica se deben despreciar todos los efectos de difracción. A medida que aumentan los detalles de la señal a procesar (aumenta el contenido de frecuencias espaciales) la luz comenzará a difractarse. Este hecho determina el límite de la cantidad de información que podremos procesar simultáneamente, es decir podremos aumentar el número de puntos en la señal de entrada hasta que estén tan apretados que comiencen a difractar la luz.

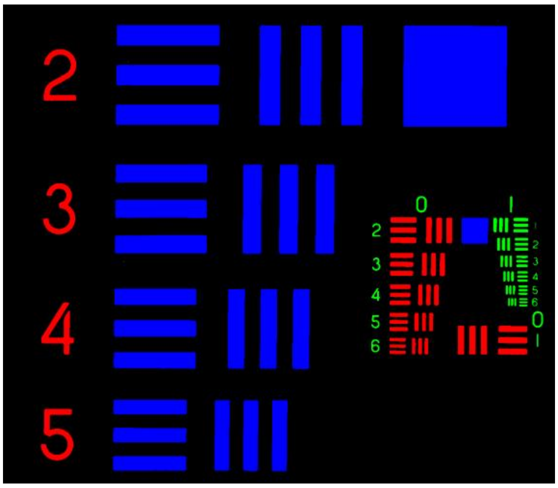
Procesador con fuente puntual: pseudocoloreado de bandas de frecuencia

Existen algunas aplicaciones de procesado óptico en las cuales se usa un montaje similar al empleado cuando se trabaja con iluminación coherente, pero se utiliza en estos casos una fuente “puntual” con ancho de banda finito. Si bien el dispositivo es similar, cuenta con algunas diferencias. Una de ellas es que las lentes utilizadas deben estar corregidas para ser acromáticas, es decir que su distancia focal debe ser aproximadamente igual para todas las longitudes de onda. Por otra parte la fuente de luz extensa debe concentrarse, mediante una lente, sobre un diafragma con el fin de obtener una fuente lo más estrecha y brillante posible. En estos casos el diámetro del diafragma es mucho mayor que el tamaño de los pinholes utilizados en iluminación coherente, además su función es diferente ya que se utilizan con el fin de limitar la extensión de la fuente y no para limpiar el haz.

Veamos un caso donde se aprovecha la policromaticidad de la iluminación empleada.

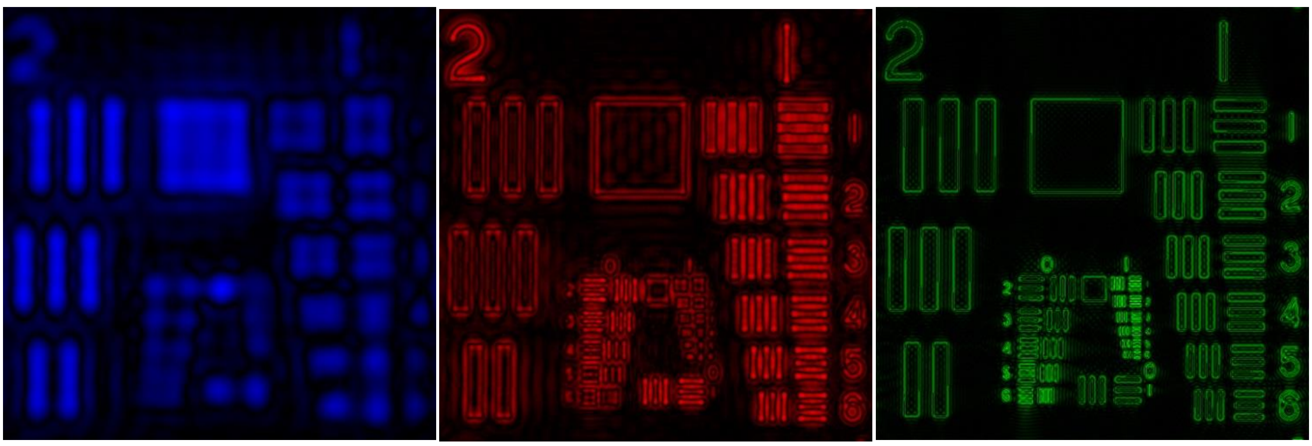
El pseudocoloreado es una técnica que se emplea para resaltar determinada información del objeto mediante una codificación de la misma en colores falsos. Por ejemplo supongamos que queremos resaltar la distribución de frecuencias espaciales en un objeto. En general el espectro de frecuencias de una imagen posee una distribución bastante compleja. Las bajas frecuencias corresponden a los detalles burdos y las altas a los detalles más finos. Si en el plano transformado ubicamos un filtro tal como el esquematizado en la figura, podremos obtener a la salida una imagen coloreada según el contenido de frecuencias de cada zona.

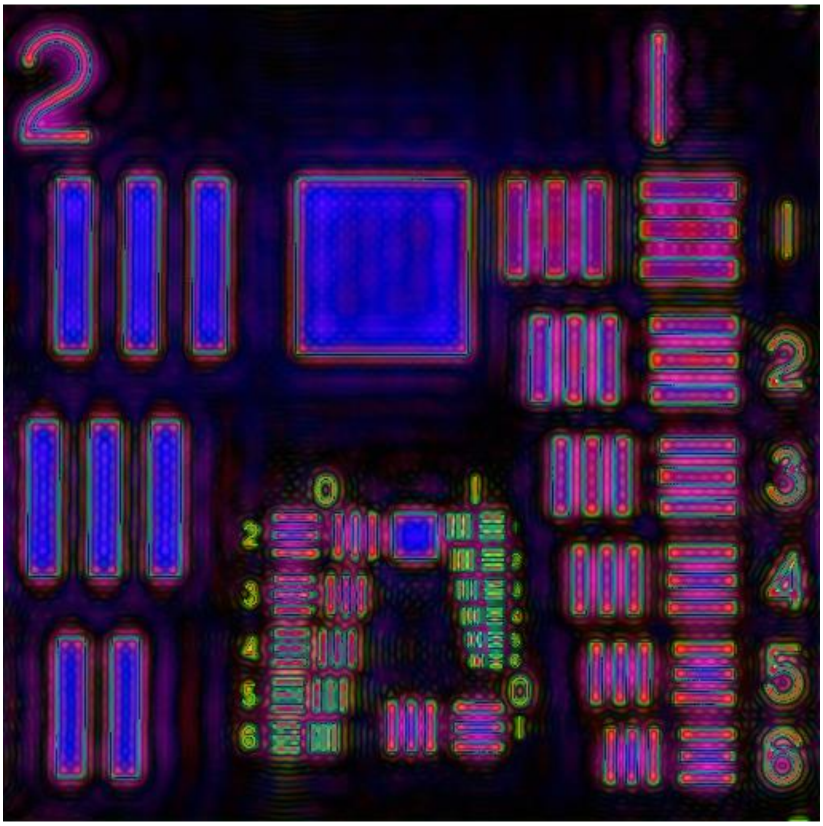




Obviamente la imagen anterior es sólo ilustrativa ya que los bordes bien definidos de los objetos más grandes también poseen cierto contenido de altas frecuencias y en consecuencia los colores no serán tan puros sino que habrá distinto tipo de mezclas.

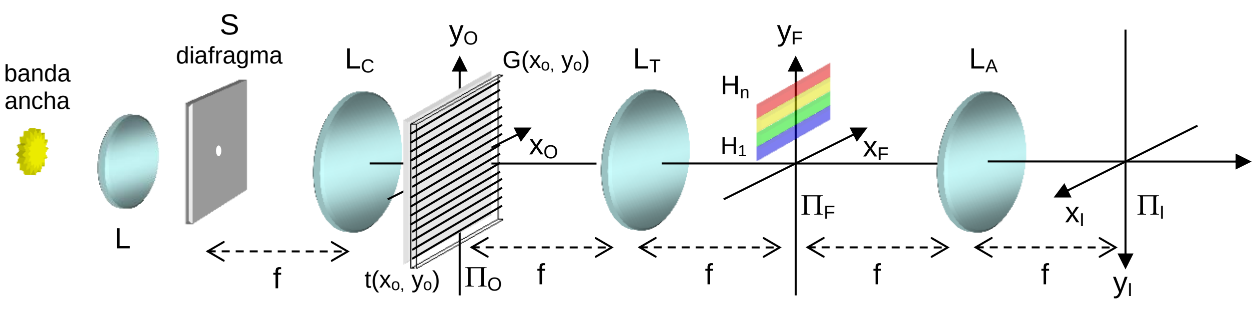
Vamos a ver como serían los canales reales y su combinación





Procesado parcialmente coherente

Veamos una disposición experimental que permite realizar ciertas operaciones de procesado de la información asociados generalmente a iluminación coherente pero utilizando ahora luz policromática. Esto es utilizaremos una fuente con un ancho de banda extenso pero el diseño del procesador nos permitirá trabajar con amplitudes complejas en lugar de con intensidades, como habitualmente sucede en las técnicas de procesado incoherente.



El procesador ilustrado en la figura es prácticamente un procesador convencional coherente, en disposición *4f*, con lentes acromáticas. La fuente de iluminación es de banda ancha y, tal como explicamos anteriormente, está focalizada sobre un diafragma que está ubicado en el plano focal anterior de una lente colimadora L_C. De esta forma sobre el objeto de transmisión $t(x_o, y_o)$ incide un haz de luz paralelo. Dicho objeto a su vez se encuentra en contacto con una red de difracción cuya transmisión podemos suponer de la forma $G(x_o, y_o) = 1 + \cos(2\pi\, p_o\, y_o)$. Así para cada longitud de onda el campo a la salida del plano objeto será

$$E(x_o, y_o, \lambda) = C(\lambda)\, t(x_o, y_o)\left[1 + \cos(2\pi\, p_o\, y_o)\right]$$

Sobre el plano transformado Π_F tendremos una distribución de campo que para cada longitud de onda será

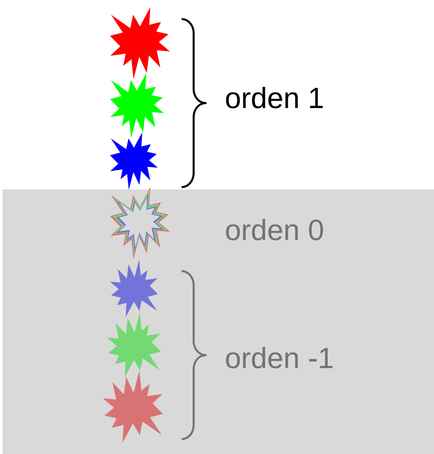
$$\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}, \lambda\right) = C \int\int_{-\infty}^{\infty} t(x_o, y_o)\left[1 + \cos(2\pi\, p_o\, y_o)\right] \cdot \exp\left[-2\pi i\left(\frac{x_F}{\lambda f} x_o + \frac{y_F}{\lambda f} y_o\right)\right] dx_o\, dy_o$$

donde C es una constante compleja. Resolviendo esta integral tenemos que

$$\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}, \lambda\right) = C_1 T\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) + C_2 T\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f} - p_o\right) + C_3 T\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f} + p_o\right)$$

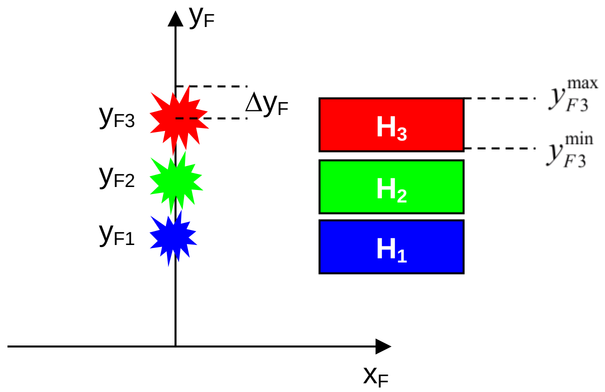
Vemos que el primer término es la transformada de Fourier tradicional, centrada en $(x_F, y_F) = 0$. La misma contendrá una pequeña dispersión cromática ya que depende de la longitud de onda, sin embargo debe destacarse que esta dispersión está originada por difracción en el objeto y no en la red ya que este término corresponde al orden 0 de la misma.

El segundo y tercer término provienen de los órdenes 1 y -1 de la red y sí presentarán una fuerte dispersión cromática producida por la misma. Estos espectros serán dispersados a lo largo del eje y_F y estarán centrados en $y_F = \pm\lambda f\, p_o$. Por ejemplo si tuviésemos sólo tres longitudes de onda (azul, verde y rojo) se vería algo como lo esquematizado en la siguiente figura



Basta quedarse con uno de los órdenes difractados y bloquear los otros.

Ahora bien, para cada longitud de onda estamos en condiciones de realizar un filtrado coherente, esto es, para λ_n deberemos construir un filtro $H\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)$ que debemos centrar en $y_{F\,n}=\lambda_n f\ p_o$

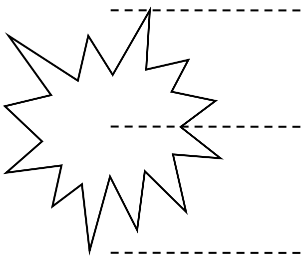


La extensión Δy_F claramente está relacionada con la distribución de frecuencias espaciales Δp propias del objeto $t(x_o, y_o)$. Cada filtro actúa sobre una extensión tal que

$$H\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)=\begin{cases} H\left(x_{Fn},y_{Fn}\right) & y_{Fn}^{\min}<y_{Fn}<y_{Fn}^{\max} \\ 0 & afuera \end{cases}$$

donde $y_{Fn}^{\max}=\lambda_n f\ p_o+\Delta y_F \quad ; \quad y_{Fn}^{\min}=\lambda_n f\ p_o-\Delta y_F$

Obviamente las longitudes de onda no estarán así separadas, en general será un continuo. Veamos cuales son las longitudes de onda extremas en los bordes del filtro. Para ello analicemos el siguiente dibujo



$$y_{Fn}^{\max}=\lambda_n f\left(\underset{\text{red}}{p_o}+\underset{\text{objeto}}{\Delta p}\right)$$

$$y_{Fn}$$

$$y_{Fn}^{\min}=\lambda_n f\left(p_o-\Delta p\right)$$

Cada espectro de frecuencias está centrado en $y_{Fn} = \lambda_n f\ p_o$. La máxima longitud de onda que llegue al borde superior será aquella dispersada por la mínima frecuencia espacial, esto es por $p_o - \Delta p$. Luego $\lambda_n^{\max} f\left(p_o - \Delta p\right) = \lambda_n f\left(p_o + \Delta p\right)$, con lo cual $\lambda_n^{\max} = \lambda_n \dfrac{p_o + \Delta p}{p_o - \Delta p}$

Análogamente la longitud de onda mínima admitida por el extremo inferior será la dispersada por la máxima frecuencia espacial, esto es por $p_o + \Delta p$.

Luego $\lambda_n^{\min} f\left(p_o + \Delta p\right) = \lambda_n f\left(p_o - \Delta p\right) \quad \Rightarrow \lambda_n^{\min} = \lambda_n \dfrac{p_o - \Delta p}{p_o + \Delta p}$

Vale decir que el ancho de banda que cae dentro del filtro n-ésimo es

$$\Delta\lambda_n = \lambda_n^{\max} - \lambda_n^{\min} = \lambda_n \frac{4 p_o \Delta p}{p_o^2 - \left(\Delta p\right)^2}$$

Como en general $p_o >> \Delta p$ entonces $\Delta\lambda_n \cong \lambda_n \dfrac{4\Delta p}{p_o}$

Cada operación de filtrado en particular requerirá una coherencia temporal distinta y se deberá estudiar específicamente. Una forma de regular dicho ancho de banda es eligiendo en forma adecuada la frecuencia espacial de la red empleada.

Si en vez de un espectro continuo tenemos uno discreto como el del dibujo, puede calcularse cuál debe ser la frecuencia de la red para tener en cada filtro un ancho de banda igual al ancho espectral de la línea.

Ahora bien, dado que este filtrado espacial es realizado por separado, en distintas bandas espectrales, las señales filtradas serán mutuamente incoherentes y se sumarán en intensidad. Esto es, el procesador parcialmente coherente es capaz de procesar la señal en campo (como si fuera coherente) para determinadas bandas espectrales, luego, al ser cada una de ellas incoherente con las demás, las respuestas individuales se suman en intensidad y es posible eliminar el molesto ruido coherente. Entonces para cada canal tenemos

$$E_{in}\left(x_i, y_i, \lambda_n\right) = t\left(x_i, y_i\right) \otimes h\left(x_i, y_i, \lambda_n\right)$$

Luego

$$I\left(x_i, y_i\right) \simeq \sum_{n=1}^{N} \Delta\lambda_n \left| t\left(x_i, y_i\right) \otimes h\left(x_i, y_i, \lambda_n\right) \right|^2$$

donde $h\left(x_i, y_i, \lambda_n\right)$ es la función respuesta al impulso de $H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$

VII. MODULACIÓN DEL FRENTE DE ONDAS

Hasta el momento describimos los filtros y las señales de entrada mediante su función transmisión pero no mencionamos cómo implementarlos experimentalmente. Esto se puede llevar a cabo por distintos métodos y utilizando diferentes materiales de registro o medios de representación. A la hora de elegir los mismos deben analizarse sus características principales tales como: resolución, eficiencia de difracción, sensibilidad espectral, energía requerida para el grabado de la información, si constituye un medio que permite modular sólo la fase o la amplitud, o ambas, si es reutilizable o no, si permite trabajar a tiempo real, su precio, etc.

De la gran variedad de elementos que permiten modular el frente de ondas acá sólo analizaremos algunos pocos de ellos y comenzaremos haciéndolo por los materiales fotográficos que, si bien están siendo cada vez menos empleados, constituyen un punto de partida para comprender el proceso de registro y representación de la información.

MATERIALES FOTOGRÁFICOS

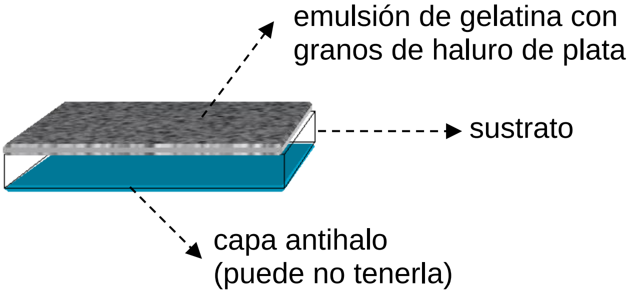
Fueron tradicionalmente los más utilizados por su bajo precio, fácil procesado, amplio rango espectral y sensibilidad. Se pueden elegir dentro de un gran rango de resoluciones por ejemplo:

Película o placa	Resolución (l/mm)	Sensibilidad (Asas)
Kodak Tri-X (film)	aprox. 50	400
Agfa Copex-pan (film)	aprox. 300	25
Holotest 8E-75 (placa y film)	> 3000	0.015

Para que tengamos una idea comparativa podemos mencionar que una cámara CCD tiene aproximadamente una resolución de 100 l/mm. Ahora bien, una cosa es registrar la imagen y otra poder representarla con igual resolución espacial.

El material fotográfico puede utilizarse tanto para representar la señal de entrada, como para sintetizar el filtro o como medio de registro de la señal de salida. Algunas de las desventajas que presentan es que no son reutilizables, no pueden usarse como medios dinámicos y no son lineales con la exposición.

La película fotográfica virgen está constituida por una gran cantidad de granos de haluro de plata suspendidos en un soporte de gelatina, el cual a la vez se halla adherido a un sustrato de acetato (film) o a una de vidrio (placa).



La capa antihalo es de un material absorbente que impide que la luz incidente, después de atravesar la emulsión sensible, se refleje parcialmente en la interfase sustrato-aire y vuelva a impresionar la película. Esta capa evita además que entre luz por la otra cara del film. Comúnmente se elimina en el proceso de fijado. En muchos casos (películas y placas holográficas) es posible elegir si se desea o no que vengan con esta capa de antihalo. La razón es que a veces es necesario incidir con luz por ambos lados del material de registro (luego veremos esta situación).

Analicemos ahora el proceso físico-químico de la exposición. Cuando el material fotográfico es expuesto a la luz, los granos de haluro de plata absorben energía y sufren un cambio complejo. En aquellos que absorbieron una cantidad suficiente se detectan pequeños centros de plata metálica, tales zonas se denominan centros de revelado. Cuando el material expuesto se somete al proceso químico subsiguiente, el revelador propaga este pequeño centro a todo el grano, convirtiéndolo en un grano de plata metálica altamente opaco a la luz. Los granos no expuestos no sufren cambios y son eliminados en el proceso de fijado.

Una película es más sensible a la luz (más Asas) cuanto mayor es el tamaño del grano de haluro de plata y, en consecuencia, es menor su resolución.

La energía por unidad de área que llega a una película es $W = I_o \mathcal{T}$, donde I_o es la intensidad y $\mathcal{T}$ es el tiempo de exposición.

La transmitancia del film, una vez revelado, será:

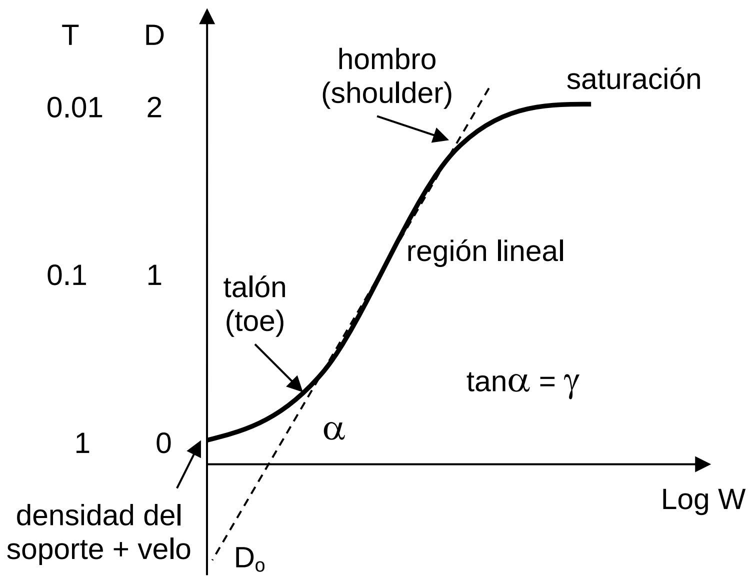
$$T\left(x,y\right)=\left\langle \frac{I\left(transmitida\ en\ x,y\right)}{I\left(incidente\ en\ x,y\right)}\right\rangle$$

donde < > significa un promedio local, esto es, un promedio sobre un área grande con respecto al tamaño de un grano del film pero pequeña comparada con un cambio significativo de la transmitancia *T*.

En 1890 F.Hurter y V.Driffield publicaron un trabajo (J.Soc.Chem.Indust. **9**, 445, 1890) en donde demuestran que $\log\left(\frac{1}{T}\right)=-\log\left(T\right)$ es proporcional a la masa de granos de plata por unidad de área de la transparencia revelada. Definieron a esa cantidad como la densidad fotográfica *D*.

$$D=\log\left(\frac{1}{T}\right)=-\log\left(T\right)\propto\frac{masa\ de\ granos\ de\ plata}{unidad\ de\ área\ de\ película\ revelada}$$

Una de las descripciones más usadas para caracterizar las propiedades de una película es la curva de Hurter-Driffield ó como se llama más habitualmente curva H-D. Ella es un gráfco de la densidad fotográfica *D* en función del logaritmo de la energía de exposición *W*



Energías por debajo del talón no impresionan el negativo. Aquellas por arriba del hombro tampoco producen un cambio de densidad ya que la película está saturada. En fotografía convencional se trabaja generalmente en la región lineal, la pendiente de esta región se conoce como el γ de la película. Aquellas con un γ grande (2,3) son las de alto contraste, mientras que las que tienen un γ chico (<1) son de bajo contraste. Esta región lineal corresponde al rango dinámico de la película. Un determinado valor de γ está influenciado por tres factores: tipo de emulsión, tipo de revelado y tiempo de revelado. Cuanto más tiempo es revelado el film, más aumenta el γ.

Qué tipo de este parámetro se desea elegir depende, en parte, de en qué tipo de procesador se utilizará la película.

Consideremos a continuación las propiedades de una película bajo distintos tipos de iluminación.

Película en un sistema óptico incoherente

Queremos analizar la relación entre la intensidad I_o que impresiona la película virgen y la intensidad I transmitida, luego de revelada. Asumamos que la película es usada en la región lineal de la curva H-D. Entonces

$$D = \gamma_n \log W - D_o = \gamma_n \log\left(I_o.\mathcal{T}\right) - D_o$$

donde γ_n es la pendiente de la curva y *n* denota que estamos trabajando con un negativo. $-D_o$ es el punto del eje D donde cortaría la prolongación de la zona lineal si no existiese la desviación del talón.

La intensidad durante la exposición I_o se relaciona con la transmitida, luego del revelado, a través de la definición de densidad $D = \log\left(\frac{1}{T_n}\right)$. Luego

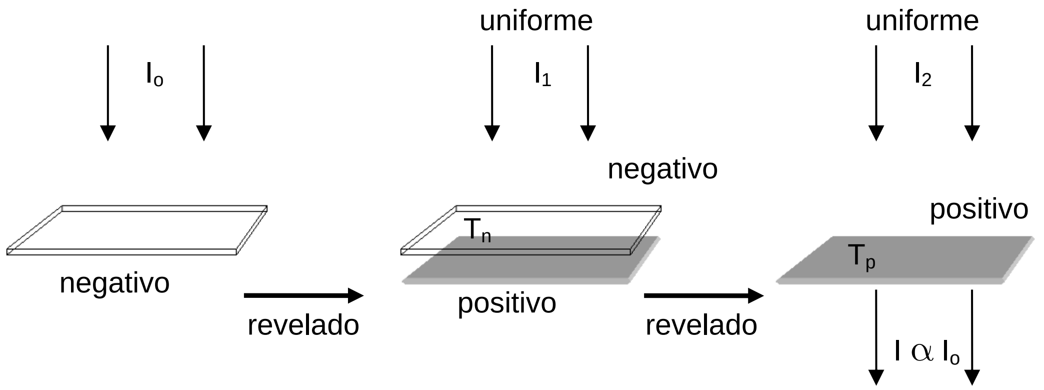
$$\log T_n = -\gamma_n \log\left(I_o.\mathcal{T}\right) + D_o \quad \Rightarrow \quad T_n = 10^{D_o}\left[I_o.\mathcal{T}\right]^{-\gamma_n} = K_n I_o^{-\gamma_n}$$

donde K_n es una constante positiva.

Vemos que no existe una relación lineal entre la intensidad de exposición y el factor de transmisión. Si buscamos que exista una relación lineal, entonces debemos proceder en dos etapas. Primero realizamos un negativo tal como hemos visto y, a continuación, realizamos una copia de contacto sobre otra película de forma tal que sobre este nuevo film arribe una intensidad $I_1.T_n$ luego la transmisión de esta película, que será un positivo, viene dada por

$$T_p = K_p\left(I_1.T_n\right)^{-\gamma_p} = K_p.K_n^{-\gamma_p}.I_1^{-\gamma_p}.I_o^{\gamma_p.\gamma_n} = K.I_o^{\gamma_p.\gamma_n}$$

donde K nuevamente es una constante positiva.



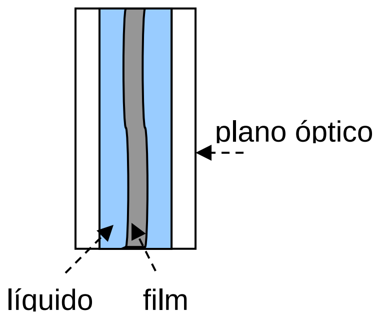
En la figura anterior se esquematiza el proceso. Ahora vemos que mediante dos pasos es posible obtener una fotografía con un factor de transmisión lineal con respecto a la intensidad I_o que se utilizó para exponerla originalmente ya que se puede encontrar una relación tal que $\gamma_n.\gamma_p = 1$

Película en un sistema óptico coherente

Cuando una película es utilizada como un elemento de un sistema óptico coherente es más apropiado trabajar con la función transmisión en campo *t(x,y)* en lugar de la transmisión en intensidad *T(x,y).* Ahora bien, podemos estar interesados en que *t(x,y)* tenga una relación lineal con la intensidad I_o a la que se expuso ó con el campo $|E_o|$ respectivo.

Uno estaría tentado a decir que *t(x,y)* es simplemente $\sqrt{T(x,y)}$, sin embargo no toma en cuenta las variaciones de fase introducidas por la película. Estas se deben principalmente a dos motivos. Uno es que el sustrato donde se deposita la gelatina no es plano desde el punto de vista óptico, el otro es que el espesor de la emulsión no es constante ya que el mismo varía con la densidad de granos de plata. Así en general tendremos $t(x,y) = \sqrt{T(x,y)}.\exp\left[i\phi(x,y)\right]$.

Si la película es introducida en un recipiente con un líquido de contacto apropiado (*liquid gate*) entonces estas variaciones de fase pueden disminuirse. Básicamente el dispositivo consiste en un sándwich del film entre dos planos ópticos de vidrio unidos por un líquido cuyo índice de refracción es cercano al de la gelatina, al del sustrato y al del vidrio (por ejemplo glicerina)

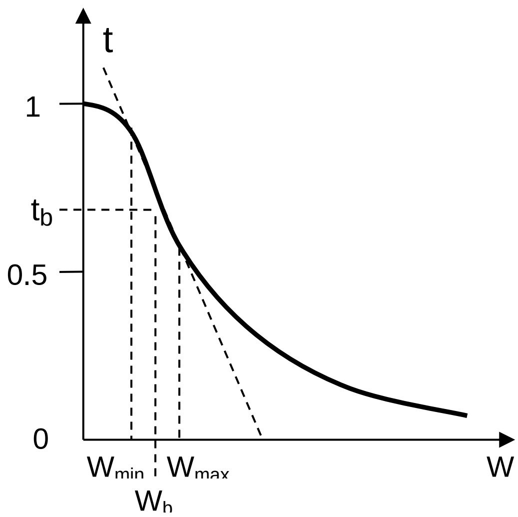


Una vez realizado este procedimiento podemos aproximar $t(x,y) = \sqrt{T(x,y)}$

Luego $$t(x,y) = K^{\frac{1}{2}}.I_o^{\frac{\gamma_n \gamma_p}{2}} = K^{\frac{1}{2}}.\left(E_o.E_o^*\right)^{\frac{\gamma_n \gamma_p}{2}} = K'.\left|E_o\right|^{\gamma_n \gamma_p}$$

Si lo que se busca es que el campo transmitido por *t(x,y)* guarde una relación lineal con el campo con el que se realizó la exposición, basta tomar $\gamma_n.\gamma_p = 1$. Si en cambio se busca que sea lineal con I_o entonces se debe tomar $\gamma_n.\gamma_p = 2$; en general se elige $\gamma_n = 1/2\,; \gamma_p = 4$

Cuando estamos interesados en que $t(x,y) = K'\left|E_o(x,y)\right|^2 = K' I_o(x,y)$ es posible obtener esta relación en una sola exposición para un pequeño rango donde $t(x,y) \propto W(x,y)$. Así, en procesado óptico coherente, muchas veces resulta útil trabajar con la curva *t* vs *W* en vez de con la curva H-D. En la figura siguiente vemos la forma típica de una de estas curvas.



Desafortunadamente esa región es muy pequeña y corresponde a una zona de la curva de H-D cerca del talón, esto es una zona subexpuesta con poco rango dinámico de la película. Si ajustamos la exposición media para trabajar en esta región de máxima linealidad entre *t* y *W* entonces, sobre cierto rango dinámico, es válido un desarrollo a primer orden que nos permite escribir

$$t(x,y) \cong t_b + \beta\left(W(x,y) - W_b\right)$$

donde W_b es la energía de *bias* y t_b la transmisión correspondiente. La pendiente de la curva en el punto de *bias* es β. (Como dato adicional y dado que esto es de interés, por ejemplo cuando usamos placas holográficas, podemos decir que la energía necesaria para trabajar con una 8E75 es aproximadamente 10 μ J/cm^2)

La ecuación anterior también puede escribirse incluyendo el tiempo de exposición $\mathcal{T}$ dentro de β, resulta así

$$t(x,y) \cong t_b + \beta'\left(I_o(x,y) - I_b\right)$$

Obviamente los materiales fotográficos, tal como los hemos estudiado hasta el momento, nos permiten sintetizar elementos (objeto de entrada, filtros, hologramas etc.) de amplitud. Esto no quiere decir que no pueda codificarse en amplitud información de fase, tal como veremos después.

La transmisión de dichos elementos será del tipo

$$\left.\begin{matrix} T(x,y) \\ t(x,y) \end{matrix}\right\} \propto |transmisión|.\exp\left[\phi(x,y)\right] \quad con\ \phi(x,y) = cte$$

Veremos a continuación que a partir de una película fotográfica es posible obtener un elemento de fase mediante un proceso llamado blanqueado.

CLASE 15

Fotografías blanqueadas

Los medios de registro cuya transmisión es del tipo $t(x,y) = \exp\left[i\phi(x,y)\right]$, aparte de ser utilizados cuando se busca un elemento que tenga esta característica, muchas veces se emplean por poseer una eficiencia de difracción mucho más alta que los de amplitud. Esto, por ejemplo, es muy deseable en hologramas, redes, etc. Más adelante, cuando estudiemos holografía, nos adentraremos un poco más en el tema de las eficiencias, por ahora sólo veamos en qué consisten estos medios y cómo es posible transformar una película fotográfica en un modulador de fase.

Existen básicamente dos tipos de blanqueo, los denominados de rehalogenación o directos y los solventes o inversos.

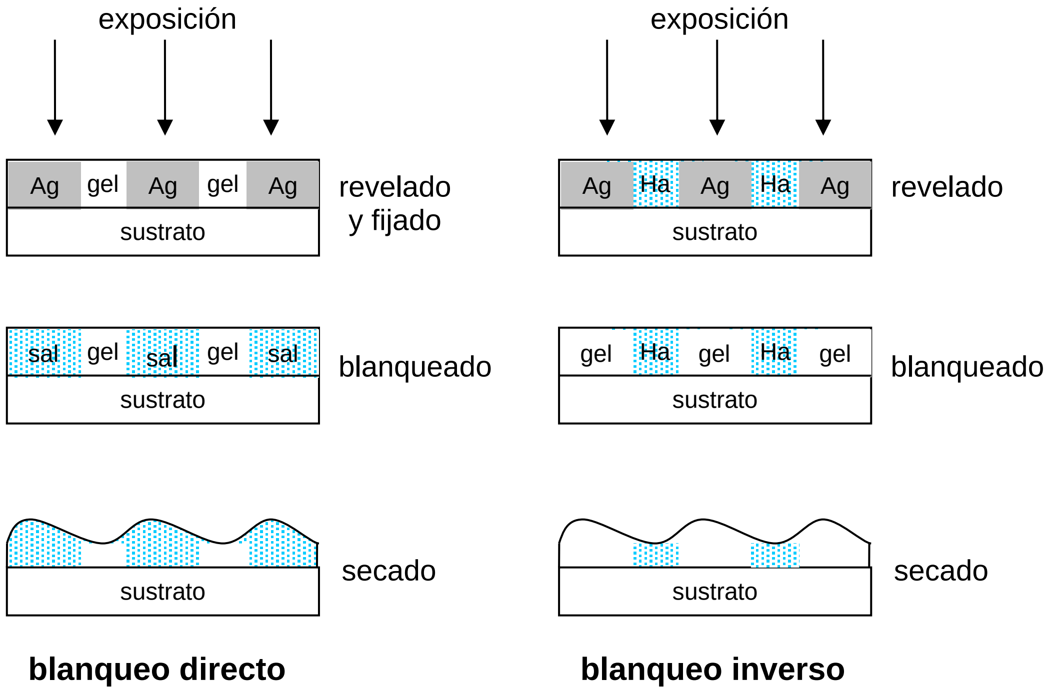
Veamos primero el blanqueo directo. En este tipo de proceso la película es revelada y fijada, como se hace habitualmente, luego se sumerge en un baño donde la plata metálica es reconvertida en una sal de plata transparente con diferente índice de refracción que la gelatina. Así pues las diferencias de intensidad registradas en la película se convierten en diferencias de camino óptico. Esto se deben a dos razones: la primera es el cambio debido a la diferencia de índices de refracción entre la gelatina y la sal de plata; la segunda es por una modulación en la superficie que se puede acentuar según el tipo de drogas empleadas. Tal modulación de superficie, en general, disminuye mucho para frecuencias espaciales mayores a las 100 l/mm. En el caso de hologramas, donde se registra la luz proveniente de objetos difusores, además es registrado el speckle provocado por la iluminación coherente. El speckle, usualmente, es un ruido no deseado con frecuencias espaciales menores a 100 l/mm, que en el caso de los blanqueos directos es amplificado por sumarse las diferencias de camino producidas por el cambio de índice de refracción y la modulación en superficie. Un baño blanqueador típico es el ferrocianuro de potasio [K_3 Fe $(CN)_6$] que se utiliza por proveer alta eficiencia y ser relativamente estable con el transcurso del tiempo. Existen otras drogas más eficientes pero las películas así tratadas luego ennegrecen con el tiempo.

Hablamos recién del efecto no deseado que produce el speckle, ya que da origen a una gran cantidad de luz dispersada (scattering) cuando se ilumina la película fotográfica una vez sometida al proceso total. Para evitar este efecto es conveniente usar el blanqueo inverso o solvente.

En este proceso la película es expuesta y luego revelada pero no fijada. Así en la gelatina hay zonas con plata metálica (las que recibieron luz) y zonas con haluro de plata. Luego del revelado se introduce la película en un baño que disuelve la plata metálica, así ahora los haluros de plata constituyen la sal transparente con distinto índice de refracción que la gelatina. En este proceso la modulación en superficie no se suma al cambio de índices por lo que el efecto que produce la luz dispersada disminuye notoriamente. Un blanqueado de este tipo es por ejemplo el Kodak R-9 que es una combinación de dicromato de potasio y ácido sulfúrico.

Para lograr una buena eficiencia con cualquiera de los procesos mencionados es necesario que las películas estén sobreexpuestas, aproximadamente 10 veces más que para un elemento de amplitud (para la placa 8E75 que mencionamos antes sería 100 μ J/cm^2). Claramente estos procesos no son lineales

En el siguiente dibujo se esquematizan las etapas de los dos tipos de blanqueo analizados.



GELATINAS DICROMATADAS - SHSG

Las gelatinas dicromatadas es otro medio sensible a la luz que una vez revelado permite obtener un elemento de fase. Poseen algunas características muy deseables para realizar hologramas de fase: alta resolución, baja absorción y dispersión de la luz, altísima eficiencia de difracción (casi el 90%), etc. Sin embargo este material no se vende comercialmente y es muy difícil manejar todos los parámetros que dan lugar a un registro de alta calidad. Su resolución es prácticamente a nivel molecular, superando las 5000 l/mm. El proceso físico-químico que durante la exposición produce el registro de la imagen aún no está completamente comprendido; es más muchos autores opinan cosas totalmente opuestas. Sin embargo un modelo más o menos aceptado es el siguiente:

En el proceso de sensibilización un sustrato cubierto con gelatina (similar al de las placas fotográficas) se sumerge en una solución de dicromato de amonio. Luego se extrae y se deja secar en una atmósfera saturada con amoníaco para que el dicromato no cristalice sobre la superficie. Cuando la placa es expuesta en el rango violeta-azul

de longitudes de onda (luego la sensibilidad baja muchísimo) iones de Cr $^{+6}$ hexavalentes pasan a ser Cr $^{+3}$ trivalentes formando encadenamientos cruzados de moléculas de gelatina. Estos encadenamientos endurecen la gelatina en forma diferencial siendo más dura en las regiones más expuestas y menos dura en las menos expuestas. Cabe destacar que dependiendo de la humedad ambiente y de la temperatura esta reacción puede también producirse en ausencia de luz y por lo tanto arruinar la placa. Una vez expuesto, el film se lava en agua corriente a temperatura controlada, esto produce un hinchamiento de la capa de gelatina. Luego se extrae y se sumerge en isopropanol, lo que produce una rápida deshidratación y revela la imagen latente. Obviamente esta placa será muy sensible a la humedad por lo que debe sellarse con una resina que la proteja.
Los primeros hologramas así realizados partían de una gelatina sin endurecer, con lo cual en el revelado las zonas no expuestas eran eliminadas dando lugar a una modulación de superficie. Posteriormente se utilizaron gelatinas pre-endurecidas de tal forma que la exposición y el revelado originaban zonas de diferente dureza y, en consecuencia con distinto índice de refracción pero ninguna zona era eliminada. Esto dio lugar a hologramas de mejor calidad.
Dijimos que este producto no se vende y se debe fabricar en el laboratorio. La forma más sofisticada consiste en preparar uno mismo la capa de gelatina sobre el soporte deseado y con el espesor adecuado. Otra más sencilla pero no tan efectiva consiste en pasar una placa fotográfica por fijador, sin revelar, de forma tal que sean eliminados todos los haluros de plata y quede solamente el soporte con la gelatina, luego esta se sensibiliza con dicromato de amonio.
Todo este proceso es extremadamente complicado y lleno de secretos que las publicaciones habitualmente no cuentan, de esta forma resulta poco factible que uno por si solo llegue a un final feliz. Por otra parte se necesita un láser que emita en el rango de longitudes de onda antes mencionado y con una potencia adecuada ya que el material es muy poco sensible, unas mil veces menos que una placa holográfica.
Cabe destacar que el producto final sólo consta de zonas de distinta dureza de gelatina.
Una técnica alternativa y mucho más sencilla es la que permite combinar la relativamente alta sensibilidad y amplia respuesta espectral de las placas holográficas, con la alta eficiencia y relación señal ruido de las gelatinas dicromatadas. En esta técnica, denominada *Silver Halide Sensitized Gelatin*, se parte de una placa holográfica, se expone, se revela y se realza un blanqueado inverso que contiene dicromato de amonio. Este blanqueador reacciona con la plata metálica y convierte, como en las gelatinas dicromatadas, iones de Cr $^{+6}$ en Cr $^{+3}$ y estos a su vez dan lugar

a zonas de endurecimiento diferencial. Luego la placa se pasa por fijador para eliminar los haluros de plata que no fueron expuestos y posteriormente se lava y se deshidrata con isopropílico. Obtenemos así nuevamente un objeto de fase donde las diferencias de camino son producidas por las diferencias de índice de refracción entre la gelatina endurecida y la menos endurecida. En este caso la resolución está determinada por la placa holográfica empleada y no a nivel molecular como en las gelatinas dicromatadas, sin embargo su resolución es alta ya que la misma puede alcanzar las 3000 l/mm.

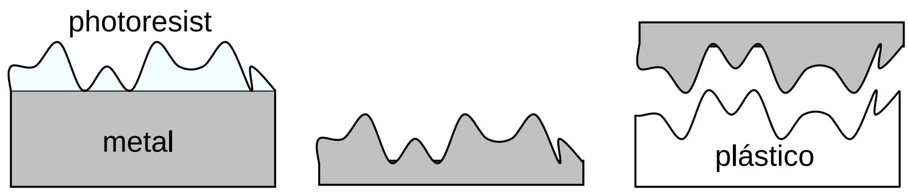
PHOTORESIST

El photoresist es una resina orgánica fotosensible que usualmente se utiliza en microelectrónica para realizar máscaras litográficas que posteriormente permitirán transferir los circuitos al soporte adecuado. En óptica en general se usa para construir redes holográficas y hologramas industriales.

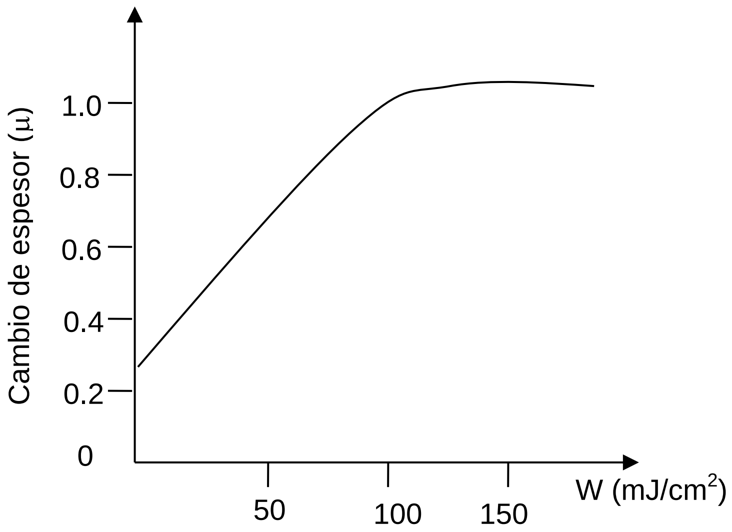
El producto usualmente se distribuye uniformemente mediante un spinner sobre un sustrato, se expone a la luz y luego se revela con un producto especial (por ejemplo el Shipley AZ-303 A) o con tricloroetileno.

Existe el photoresist positivo en el cual las áreas expuestas se disuelven en el revelado y el negativo en el cual se disuelven las zonas no expuestas. Es más utilizado el primero de ellos (por ejemplo el Shipley AZ-1350). En cualquiera de los casos lo que se obtiene es una capa transparente con modulación de superficie. Posee aproximadamente el mismo rango espectral y la misma sensibilidad que una gelatina dicromatada, su resolución es menor.

En general para las aplicaciones ópticas industriales se aplica sobre un sustrato de metal, se expone, se revela y luego mediante un ataque químico o un bombardeo iónico las diferencias de altura del photoresist se transfieren al metal. Posteriormente la pieza metálica grabada se utiliza como un sello patrón para copiar hologramas o redes sobre un plástico o resina, que son los elementos que se comercializan. El proceso se esquematiza en la siguiente figura



Cuando el photoresist AZ-1350 es tratado con revelador AZ-303 A se obtiene una profundidad de grabado aproximadamente lineal con la energía de exposición recibida.



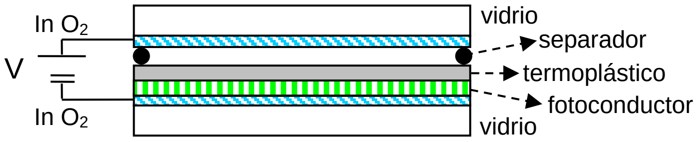
Hasta ahora los materiales analizados no son ni reutilizables ni aptos para trabajar en tiempo real. Veamos a continuación un ejemplo de material regrabable.

FOTOTERMOPLÁSTICOS

Los fototremoplásticos se encuentran dentro de la categoría de materiales de registro que permiten generar elementos de fase, donde la intensidad de entrada es convertida en deformaciones de la superficie del mismo.

Estos dispositivos si bien no trabajan en tiempo real, presentan la ventaja de poder obtener la imagen *in situ* y sin la utilización de líquidos o procesos complicados. Estas características lo hacen muy apropiados para su empleo en holografía en tiempo real con aplicaciones interferométricas (más adelante vamos a ver en qué consiste esta técnica).

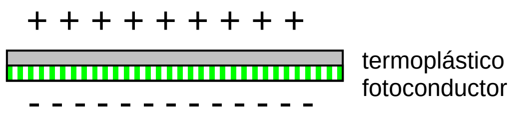
Los componentes básicos del dispositivo de grabado son los siguientes: En general se utiliza un sustrato transparente (usualmente vidrio) cubierto con una capa conductora transparente de óxido de Indio (ITO), sobre esta se deposita un material fotoconductor (el más usado es poly-n-vinil carbazole [PVK] sensibilizado con trinitrofluorenona [TNF])de aproximadamente 1 μm de espesor. Por último se deposita una capa del orden de 0.7 μm de termoplástico, siendo uno de los que ofrecen mejores resultados un copolímero de 85% de estireno y 15% de octal-decil-metacrilato. En general se cubre con otro electrodo evaporado sobre vidrio. Todas las capas, incluso el termoplástico, son transparentes.



Veamos cómo se realiza un ciclo de trabajo:

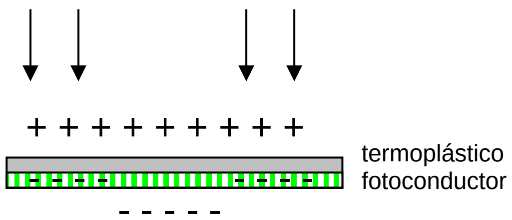
Inicialmente, a oscuras, se aplica una descarga eléctrica de modo que la superficie del termoplástico quede cargada positivamente.

1. Primera carga

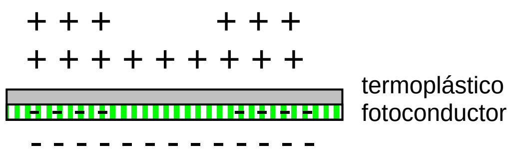


En un segundo paso, cuando el dispositivo es iluminado, el fotoconductor permite la migración de las cargas negativas donde incidió la luz. Esto no cambia la carga pero disminuye el potencial de la superficie

2. Exposición



En una tercera instancia, y para lograr una distribución no uniforme de cargas sobre la superficie del termoplástico, el dispositivo se vuelve a cargar como en el primer paso. Como resultado, una cantidad adicional de cargas se depositan en las zonas expuestas, en forma proporcional al potencial reducido y, en consecuencia, a la exposición recibida.

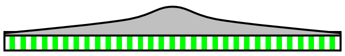


3. Segunda carga

Así sobre la superficie del termoplástico tenemos una distribución de cargas que será proporcional a la cantidad de luz recibida.

En un cuarto paso el termoplástico se calienta (aproximadamente 60 ºC), aplicándole una corriente alterna a las superficies conductoras hasta su ablandamiento. Así el termoplástico sufre una deformación superficial originada por la presencia no uniforme de cargas y se afina donde recibió luz. Luego se deja enfriar a temperatura ambiente y el registro se torna estable.

4. Revelado



El material es reutilizable ya que aplicándole una temperatura mayor que la empleada para su revelado (entre 70 y 100 ºC) el termoplástico se ablanda y la tensión superficial produce el borrado de las variaciones de espesor. Todo el ciclo puede realizarse en tiempos que oscilan entre 1 y 5 segundos.

Los dispositivos comerciales de este tipo resisten unos doscientos ciclos de grabado y borrado, luego se deteriora el material en forma irreversible.

El dispositivo es sensible a las longitudes de onda del espectro visible, alcanza una eficiencia de 30 % y permite registrar frecuencias espaciales entre las 500 y 1200 l/mm. La energía de exposición requerida va entre 10 a 100 μ J/cm^2.

CRISTALES FOTORREFRACTIVOS

Los cristales fotorrefractivos se caracterizan por sufrir una variación del índice de refracción ante la iluminación con una longitud de onda adecuada. Dependiendo de las condiciones experimentales (potencia, longitud de onda de los haces de registro y del haz de lectura), y uso que se les va a dar, se deberá elegir convenientemente el cristal fotorrefractivo a utilizar. El efecto fotorrefractivo ha sido observado en una variedad de materiales electro-ópticos ($BaTiO_3$, $KNbO_3$, $LiNbO_3$, GaAs, $Bi_{12}SiO_{20}$, $Bi_{12}GeO_{20}$, etc.).

Existen cristales, como el $LiNbO_3$, que mantienen el registro holográfico durante mucho tiempo (años). Ellos son utilizados como memorias holográficas, dado que los mismos poseen una gran capacidad para almacenar información (en un cristal de 1 cm^3 pueden almacenarse 10000 hologramas). Requieren láseres de gran potencia para ser grabados y luego de un proceso adecuado pueden leerse sin ser dañados.

Otros, como los pertenecientes a la familia de los selenites ($Bi_{12}SiO_{20}$, $Bi_{12}GeO_{20}$) son mucho más sensibles pero no son aptos para almacenar información. Se los utiliza en cambio como medio holográfico dinámico. Para ello es preciso que tanto el registro como el borrado de la información se realicen en tiempos cortos, de manera de no aumentar el tiempo de procesado de las señales.

Entre los cristales fotorrefractivos utilizados para procesado dinámico el $Bi_{12}SiO_{20}$ (BSO) es uno de los más empleados. Veamos a continuación algunas de sus características y propiedades más relevantes.

Estos cristales son sensibles en el rango verde-azul del espectro (rango en el cual suelen usarse los haces de escritura) y muy poco sensibles en el rango del rojo (color que se emplea entonces para la lectura, evitando el borrado de la información

registrada). En algunos casos (mezcla de 4 ondas degenerada) la lectura también se realiza con la misma longitud de onda y el holograma está continuamente registrándose y destruyéndose.
Con respecto a la eficiencia de difracción, si el registro se lleva a cabo con haces de escritura de intensidades del orden de los 10 mW/cm^2 en el rango verde-azul, formando entre sí un ángulo de 5º, se obtiene una eficiencia de 0.1-1%. Sin embargo para lograr estas especificaciones, es necesario aplicar un campo eléctrico externo del orden de los 3-5 KV/cm. El BSO, debido a la alta simetría de su red cristalina, es isótropo, con índice de refracción que varía desde n=2.86 para el violeta (λ=400 nm) a n=2.51 para el rojo (λ=700 nm). Sin embargo, ante la presencia de un campo eléctrico se produce una anisotropía que induce en el material, por efecto Pockels, una birrefringencia lineal proporcional a la intensidad de campo.

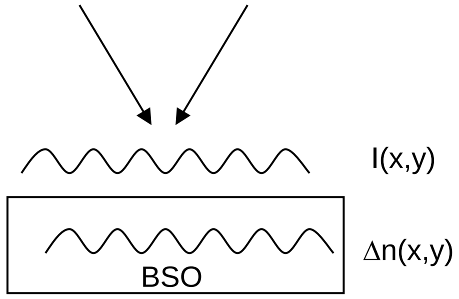
Se denomina efecto fotorrefractivo a la variación en el índice de refracción que sufren algunos materiales cuando se los ilumina con una distribución de luz no uniforme. Estos materiales deben presentar las siguientes características:

- Ser electro-ópticos, es decir, debe modificarse su índice de refracción ante la presencia de un campo eléctrico.

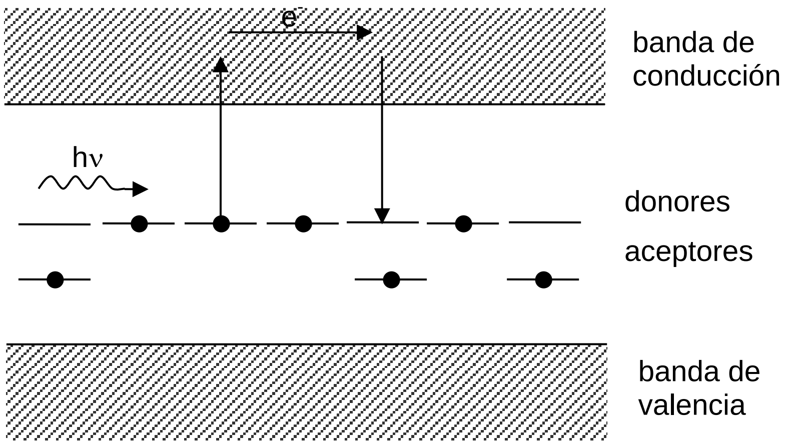
- Ser fotoconductores, esto es, ser aislantes o semi-aislantes en la oscuridad y conductores bajo la acción de una energía luminosa que permita generar portadores libres.

- Poseer trampas de cargas, esto es, centros parcialmente ocupados con cargas eléctricas que pueden fotoexcitarse. Esto permite disponer de niveles donores y aceptores de electrones y huecos dando lugar a la migración de cargas cuando estas son fotoexcitadas.

En general sobre el cristal se registrará la intensidad resultante de la interferencia de dos haces luminosos, esto es, una distribución de franjas brillantes y oscuras, lo que provocará la migración de cargas fotoexcitadas y por lo tanto una redistribución de cargas libres. Por efecto electro-óptico, esta distribución dará lugar a una modificación local del índice de refracción del material, generando entonces una red de fase en el cristal.



Si el cristal presenta donores de cargas, al excitarlo con un haz luminoso con adecuada intensidad y longitud de onda, se generan fotocorrientes. Las cargas pueden provenir de impurezas o defectos en la red cristalina. Como en la oscuridad el material es poco conductor, las cargas se encuentran “quietas”, pero al iluminar el cristal las regiones fotoexcitadas liberan los electrones que se difunden o se desplazan bajo la acción de un campo eléctrico externo y que luego son atrapados por centros aceptores presentes en las regiones oscuras. El proceso se esquematiza en la figura.

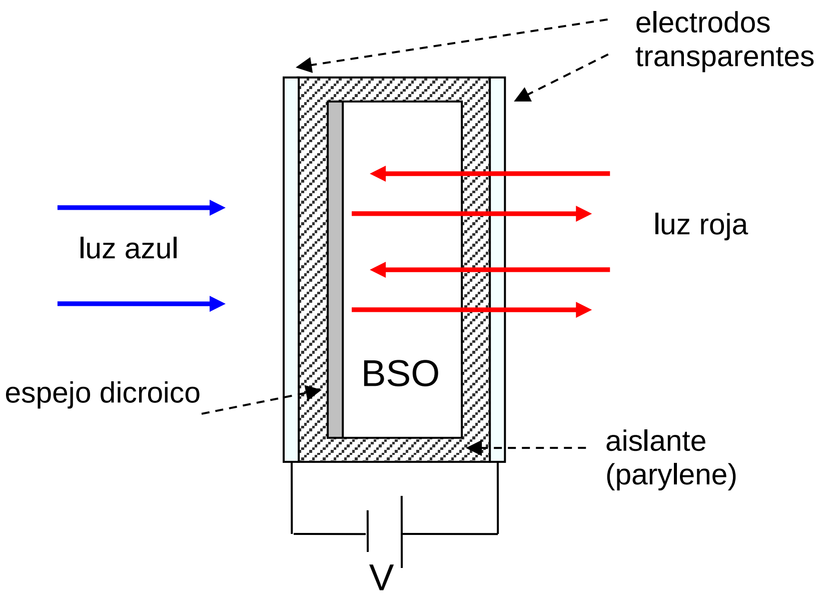


La migración de cargas fotoexcitadas puede deberse, en un material como el BSO, a dos mecanismos: difusión o arrastre. En el mecanismo de difusión, como la distribución luminosa excita a los centros donores, se crea una densidad de fotoelectrones no uniforme que entonces difunden para intentar equiparar la densidad espacial de electrones con la de los sitios donores ionizados. De esta manera se crea una distribución de cargas modulada por la distribución de luz, generando entonces un campo eléctrico que modifica localmente el índice de refracción. En el mecanismo de arrastre, los electrones fotoexcitados son arrastrados por un campo eléctrico externo hasta ser atrapados por los sitios oscuros. Es importante destacar que, en cristales como el BSO, el mecanismo de difusión es suficiente para generar la distribución de cargas si la frecuencia espacial de franjas luminosas es muy alta, situación en la cual se presenta el inconveniente de que la eficiencia de difracción disminuye notablemente. Por el contrario, con una frecuencia espacial baja, la eficiencia aumenta, pero ya no es suficiente el mecanismo de difusión para que se produzca la migración de cargas, siendo necesario entonces inducir este proceso mediante la aplicación de un campo eléctrico externo.

Dependiendo del corte del cristal (según el corte cambia su birrefringencia) este material puede utilizarse para distintos procesos. Uno de los cortes es apropiado para registrar hologramas en tiempo real mediante mezcla de cuatro ondas. Como mencionamos, si para escritura y lectura se utiliza la misma longitud de onda, el proceso se llama mezcla degenerada y si las longitudes de onda son distintas la mezcla es no degenerada. También cambiando la dirección del campo eléctrico aplicado puede utilizarse para amplificación óptica mediante la mezcla de dos ondas. En este caso un haz de alta energía pero sin información le transfiere su potencia a uno de baja energía pero que contiene la información que se desea amplificar.
En otro de los cortes posibles se utiliza para construir un dispositivo llamado PROM (Pockels read-out optical modulator) cuyo funcionamiento veremos a continuación.

PROM – Pockels read out optical modulator

El PROM es una interfase electro-óptica que consta de un cristal fotorrefractivo, típicamente BSO, el cual exhibe efecto Pockels y fotoconductividad cuando es iluminado con luz azul.
Como ya mencionamos, el efecto Pockels es una anisotropía que se hace presente en el cristal cuando es sometido a un campo eléctrico. Si el campo está ausente el cristal se comporta como un medio isótropo, de lo contrario se convierte en un cristal birrefringente.
Un esquema de este dispositivo es el siguiente



El espesor del cristal oscila entre 0.2 a 1 mm y el mismo está cubierto por una capa aislante de parylene de 3 a 10 μm. A la vez sobre la capa aislante se evaporan electrodos transparente de óxido de Indio (In O_2). Si el dispositivo se utiliza, como habitualmente, para trabajar en reflexión, se deposita antes del parylene y sobre un lado del cristal, una capa dicroica que actúe como un espejo para la luz roja.

Un ciclo típico de operación del cristal es el siguiente:

Un voltaje V_o (aproximadamente 2000V) es aplicado entre electrodos, dado que el cristal está a oscuras no es conductor y por lo tanto queda cargado a ese voltaje. Al iluminarlo con un patrón de luz azul el dispositivo se descarga en aquellas zonas iluminadas de forma tal que la distribución de voltaje a través del cristal es

$$V_C = V_o \exp\left(-KW\right)$$

donde K es una constante positiva y W es la energía de exposición.

La lectura se realiza iluminando uniformemente con luz roja, por ejemplo un láser de He-Ne, linealmente polarizado. Esta luz prácticamente no altera la conducción del cristal, esto es no genera cargas libres, ya que el mismo es aproximadamente 200 veces menos sensible al rojo que al azul.

Dijimos al comienzo que el BSO presentaba efecto Pockels, esto es que presentaba una birrefringencia proporcional al campo eléctrico aplicado. Luego en las zonas originalmente oscuras habrá máxima diferencia de índices de refracción y en aquellas previamente iluminadas con luz azul será mínima.

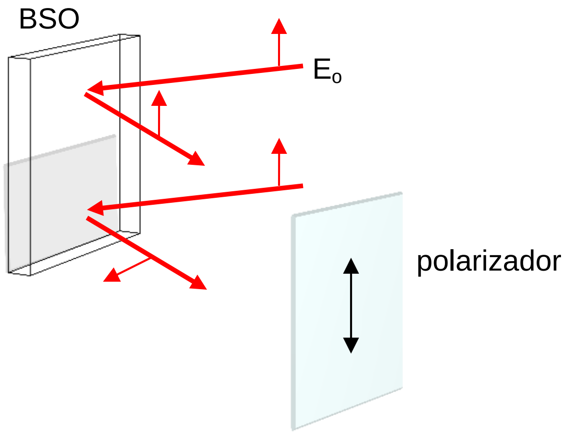
El voltaje aplicado y el espesor del cristal se eligen de forma tal que las zonas oscuras actúen como láminas de media onda, girando así 90º la polarización del haz rojo. Esto es si el haz del láser de He-Ne entró polarizado verticalmente, luego de reflejarse en el espejo dicroico y pasar nuevamente por el cristal, saldrá polarizado horizontalmente de las zonas oscuras y verticalmente de las claras. Esta codificación de niveles de iluminación en polarizaciones puede reconvertirse en cambios de intensidad si se hace pasar el haz emergente por un polaroid.

Si por ejemplo el polaroid se pone con su eje vertical entonces la amplitud del campo a la salida será

$$E = E_o \cos\left[\frac{\pi}{2}\frac{V_C}{V_o}\right]$$

donde E_o es la amplitud de la luz roja de lectura y se elige $V_o = V(\lambda/2)$

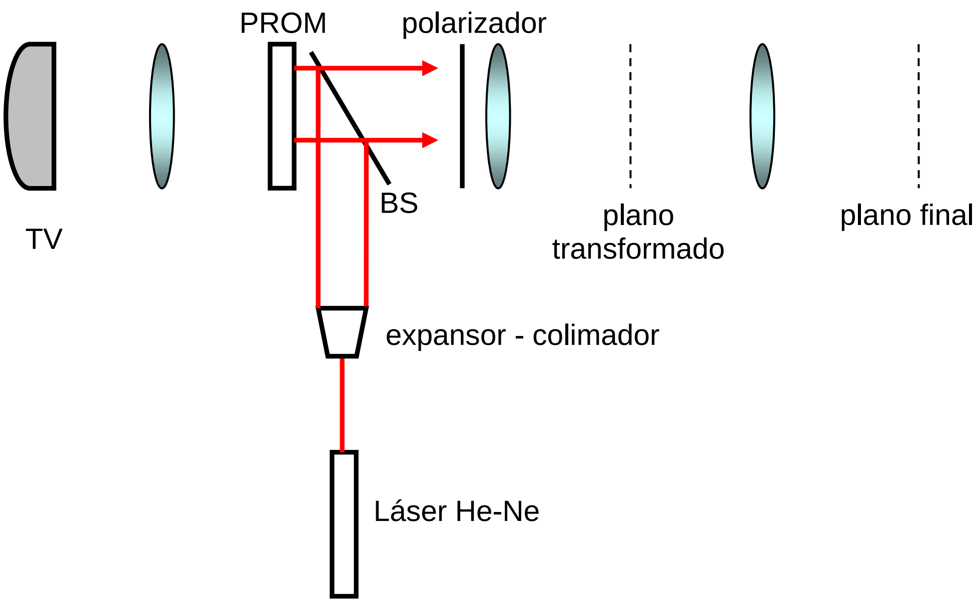
Supongamos que originalmente iluminamos con luz azul la parte superior del cristal, entonces tendremos la siguiente respuesta



Para borrar la información se le aplica el voltaje V_o y se lo ilumina uniformemente con un flash de luz azul.

Este dispositivo puede trabajar aproximadamente a 30 Hz lo que lo hace útil para emplearlo como material de registro a tiempo real. El tamaño en el que puede adquirirse es de aproximadamente 2.5x2.5 cm, su resolución alcanza las 50 l/mm y su sensibilidad para λ= 400 nm es del orden de 5 μ J/cm^2 (el voltaje cae a e^{-1})

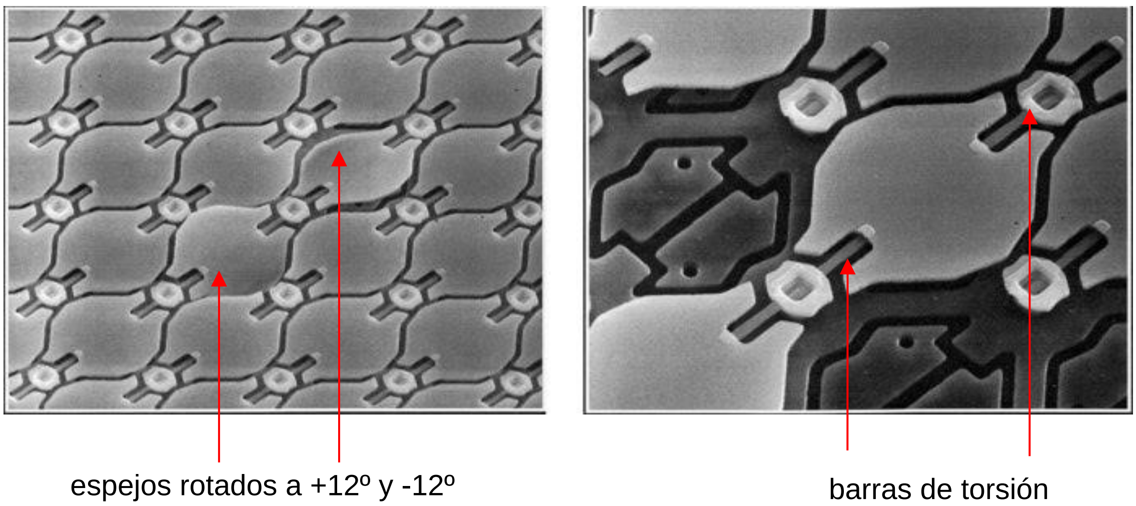
Una de las grandes utilidades del mismo es la posibilidad de convertir imágenes incoherentes en coherentes y así emplearlas en un procesador. Veamos por ejemplo el siguiente esquema

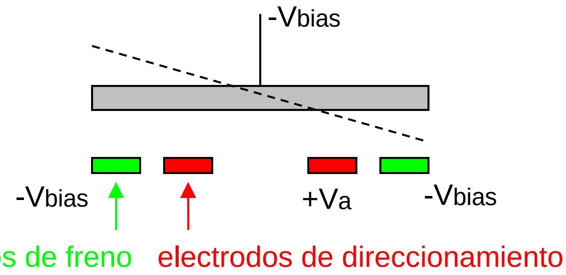


ESPEJOS DEFORMABLES

Los espejos deformables pueden dividirse en aquellos diseñados para modular la amplitud del frente de ondas y los que fueron diseñados para modular su fase. Dentro de los primeros el dispositivo más avanzado fue desarrollado por Texas Instruments (DMD, Digita Mirror Device) y está orientado a sistemas de proyección de imágenes de alta definición. Su tamaño y resolución depende del modelo pero es de aproximadamente 1 ó 2 cm^2 y contiene más de 1 millón de micro-espejos móviles, cada uno de 10 a 7.6 μm de lado. Su factor de llenado es del 90%. La eficiencia de difracción es de 86% y puede trabajar a 4KHz con señales binarias o 120 Hz a 8 bits. Los espejos pueden orientarse, mediante fuerzas electrostáticas, en ángulos que van de +12º (estado ON) a -12º (estado OFF). En el estado ON el haz de iluminación es redirigido a una lente de proyección, mientras que en el estado OFF la luz es enviada fuera de dicha lente. Los niveles de gris se logran regulando cuanto tiempo del ciclo de barrido cada espejo permanece en un estado ON u OFF.

En la figura se puede observar el dispositivo y un esquema de su funcionamiento

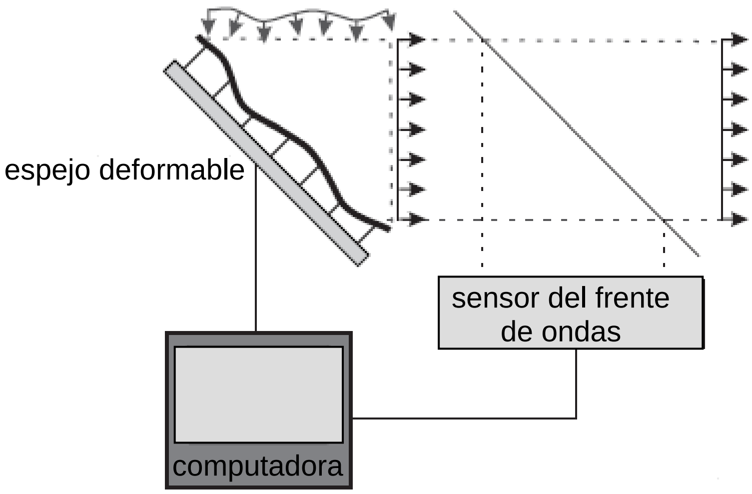




Los espejos están montados sobre una estructura que está conectada, por medio de dos barras de torsión que actúan como bisagras, a dos puntos de soporte. El espejo se halla a un potencial negativo –Vbias y debajo del mismo se hallan dos electrodos de orientación que pueden cargarse alternativamente, uno u otro, a un potencial positivo +Va de forma que se establece una fuerza atractora que permite deflectar al espejo. Asimismo un segundo par de electrodos, cargados a –Vbias actúan como freno cuando la torsión alcanza un máximo valor.

El segundo tipo de espejos deformables es aquel orientado a realizar óptica adaptativa o de corrección de aberraciones del frente de ondas (por ejemplo conformación de haces láser, astronomía, visión, etc). Estos dispositivos en general se utilizan en conjunto con un equipo que permite evaluar la forma del frente de ondas, luego el espejo se deforma de manera tal que introduce la fase conjugada de la aberración.

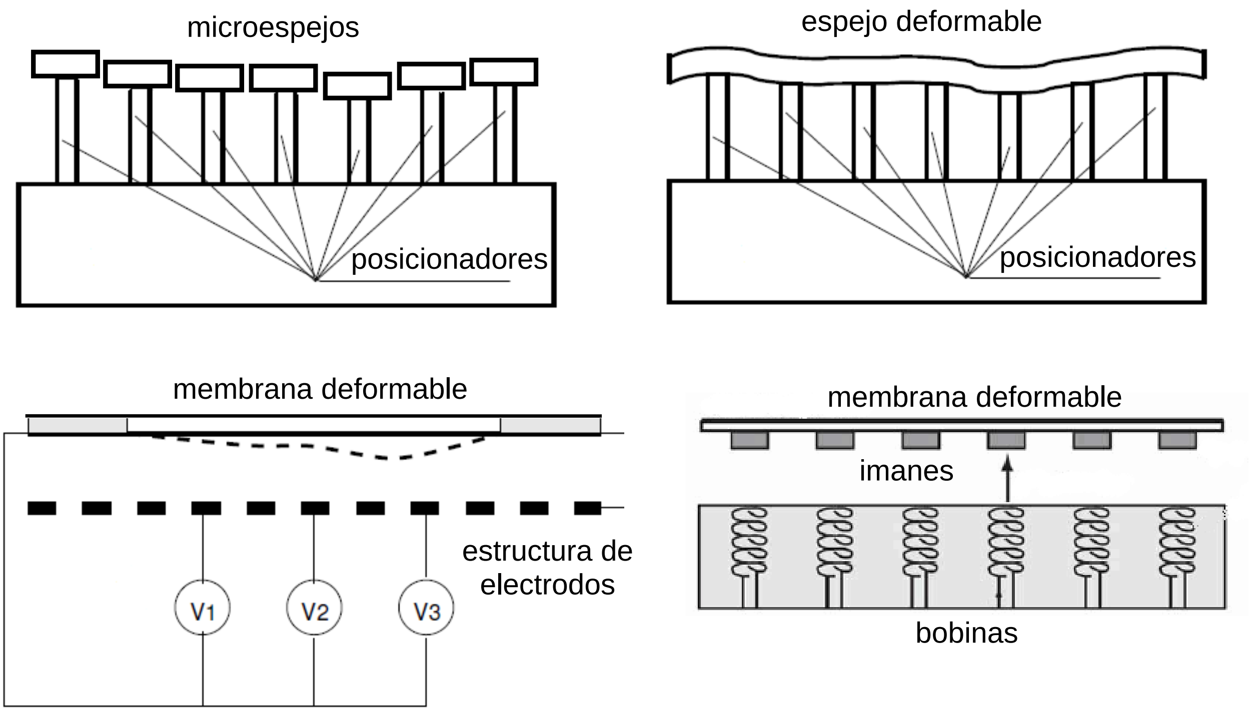
Usualmente este proceso se efectúa casi a tiempo real de manera que todo el ciclo debe ser muy rápido lo que impone que el número de elementos a modificar no sea demasiado elevado.



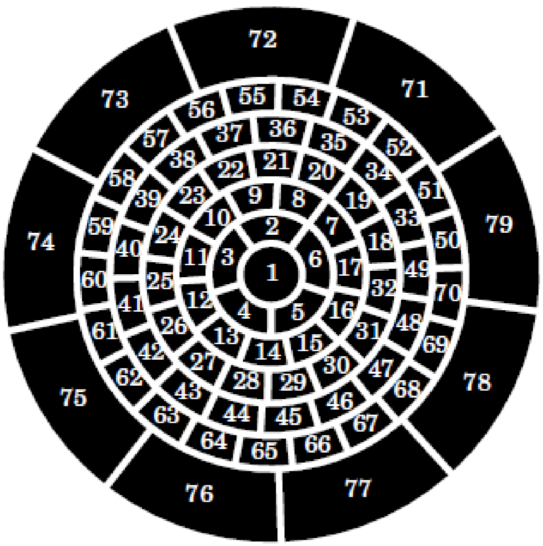
La superficie espejada puede estar compuesta por espejos independientes, conocidos como espejos segmentados, o pueden ser una superficie continua. Los segmentados presentan la desventaja de poseer pérdida de energía debido por un lado, al espacio muerto entre espejos y por el otro, a los órdenes difractados debido a la superficie pixelada. Poseen la ventaja de no presentar acople ente espejos.

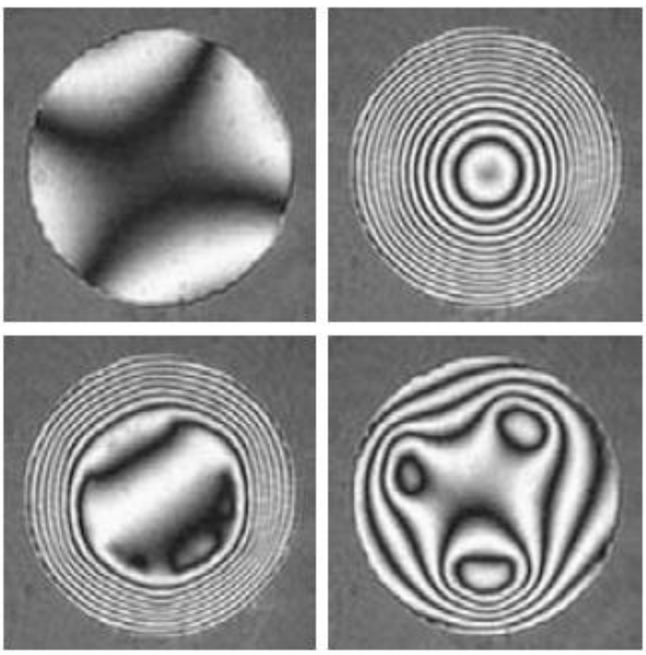
Los de superficie continua son más aptos para pequeñas estructuras, tienen alta eficiencia pero evidencian un cierto grado de acople entre elementos. A su vez estos pueden subdividirse en los que la superficie que se deforma es más o menos rígida (un fino vidrio espejado) y el control de la misma se realiza a través de piezoeléctricos o los que la superficie consiste en una membrana cuya forma se controla a través de electrodos o microbobinas. Los que están basados en el uso de piezoeléctricos pueden trabajar a muy alta frecuencia (del orden del KHz) pero presentan algo de histéresis por lo que no son aptos para determinadas aplicaciones.

En todos los casos el proceso consiste en evaluar el frente de ondas, analizar el tipo de aberración presente en términos de polinomios de Zernike y compensar las mismas deformando el modulador de forma de introducir la fase conjugada. Cuál es el grado de polinomio que se puede compensar depende de los grados de libertad (elementos de control) que posea el dispositivo y del máximo desplazamiento que pueda introducirse en cada elemento. Usualmente el número de controladores es menor a 100 y los desplazamientos típicos son de algunos micrones. En la siguiente figura se esquematizan algunos de los dispositivos anteriormente mencionados.



Un ejemplo de disposición de controladores se muestra a continuación, así como interferogramas obtenidos para distintos tipos de voltajes aplicados, en donde se evidencian las deformaciones del frente de ondas



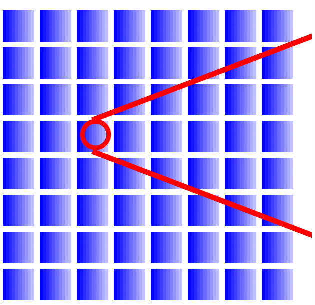


Información suplementaria sobre este tipo de moduladores del frente de ondas y óptica adaptativa puede obtenerse, por ejemplo, en
http://www.ti.com/lit/an/dlpa008a/dlpa008a.pdf
http://www.ti.com/lsds/ti/dlp/dlp-getting-started.page
http://www.okotech.com/images/pdfs/catwww3.pdf
http://www.thorlabs.com/NewGroupPage9.cfm?ObjectGroup_ID=3258

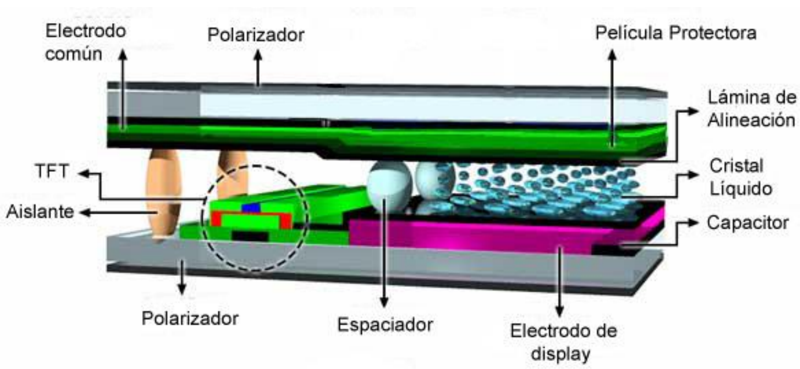
PANTALLAS DE CRISTAL LÍQUIDO CLASE 16

Los avances en el campo de la microelectrónica han permitido construir pantallas de televisión de cristal líquido con una alta resolución (LCTV). La utilización de tales dispositivos como moduladores espaciales de luz ha despertado el interés de muchos investigadores durante los últimos años ya que poseen algunas características interesantes, como ser: relativo bajo precio, trabajan a velocidad de video o sea unos 30 cuadros por segundo, brindan una resolución bastante alta y permiten obtener modulaciones puras de fase o de amplitud. Entonces vamos a comenzar describiendo como funciona una LCTV.

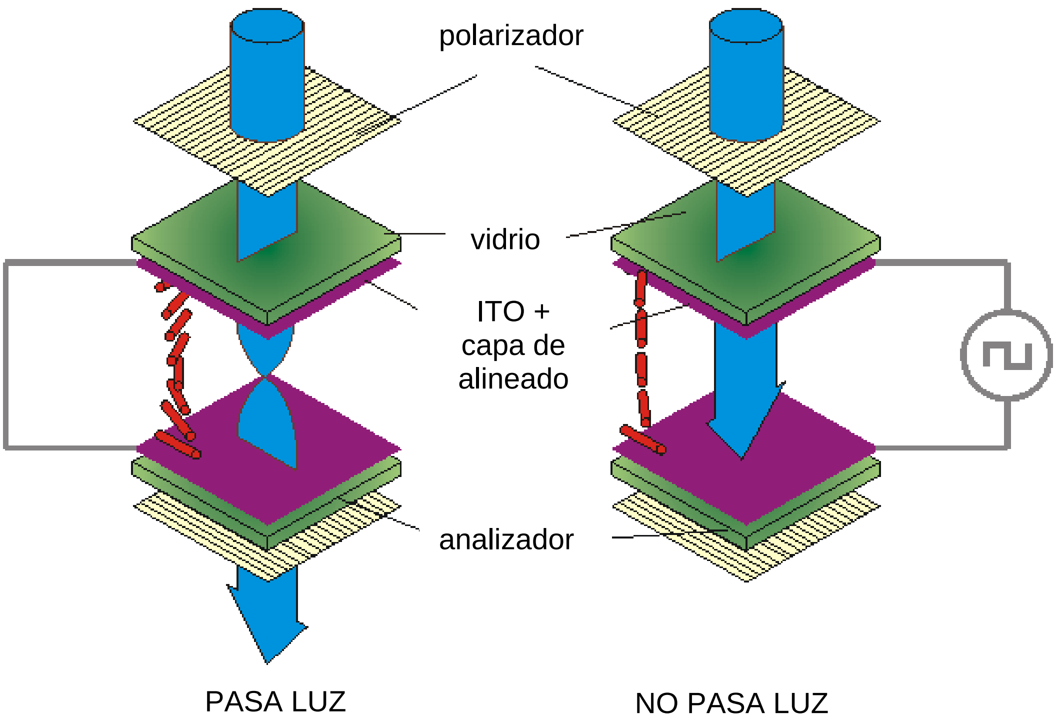
En estas cada píxel consiste en una celda de cristal líquido nemático. Las moléculas de cristal líquido tienen una estructura alargada y pueden alinearse mecánicamente mediante un rayado muy fino que se realiza sobre una lámina que se halla depositada sobre la cara interna de los vidrios que constituyen la celda. Además estos vidrios tienen evaporados electrodos transparentes que permiten aplicar un voltaje entre ambas caras de la celda. En el caso de las pantallas de matriz activa un transistor permite mantener el voltaje aplicado constante durante cada ciclo en el que se le envía señal. En las pantallas que se usan comercialmente como displays la orientación de las moléculas en contacto con la primera cara esta rotada aproximadamente 90 grados con respecto a aquellas en contacto con la segunda, de forma que a lo largo de la celda las moléculas presentan una estructura helicoidal (twisted nematic). Por último sobre ambas caras se hallan adheridos dos polarizadores que, como veremos, son los que permiten trabajar a la celda de la forma en que usualmente lo hace.



pantalla

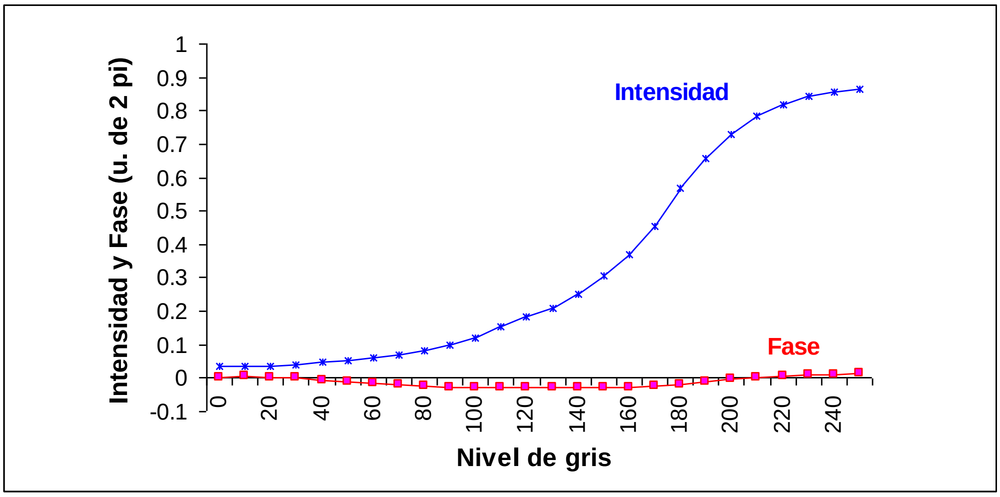


El primer polarizador se halla orientado de modo que su eje coincida con la orientación de las moléculas de la primera cara. Las moléculas de cristal son birrefringentes y van haciendo que la polarización de la luz incidente rote a medida que avanza en el cristal. Cuando no hay campo aplicado entre electrodos, al emerger la luz por la cara opuesta de la celda, rotó aproximadamente 90 grados. El eje del segundo polarizador se halla cruzado al primero y en consecuencia la intensidad transmitida es máxima. Cuando hay aplicado un voltaje máximo entre electrodos, las moléculas se orientan paralelas al campo aplicado, su birrefringencia disminuye y la luz atraviesa la celda sin rotar su polarización de modo que a la salida tenemos intensidad mínima. Así podemos lograr toda una gama de atenuaciones de la luz en función del voltaje aplicado.

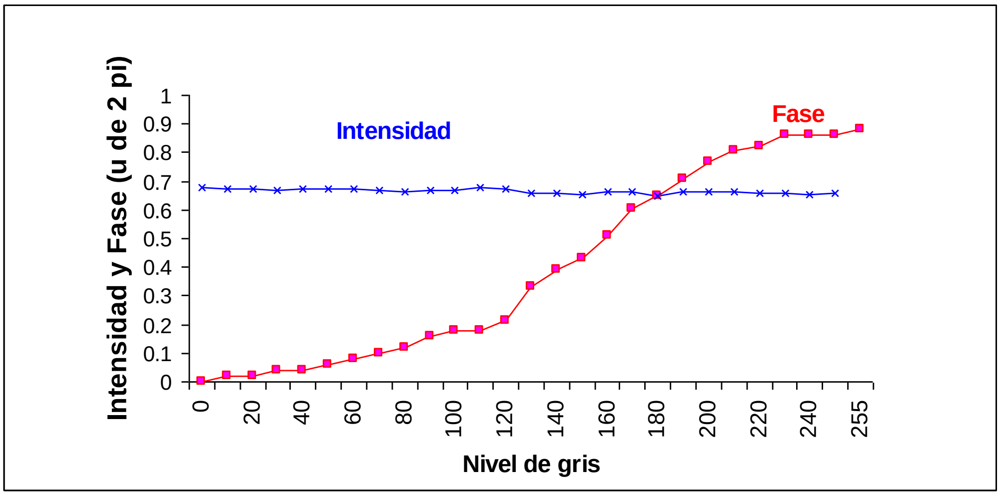


Sin embargo con esta disposición no es posible alcanzar una modulación desacoplada de amplitud y fase, aunque para los fines comerciales para los que habitualmente se usan las pantallas no tiene importancia. Para que un dispositivo de este tipo sea útil en procesado óptico se debe lograr que actúe ya sea como modulador puro de fase o puro de amplitud. Esto es posible si además de los polarizadores se utilizan entre ellos y los vidrios, láminas retardadoras que permitan generar estados de polarización elíptica adecuados a partir de los cuales se logran obtener dichas modulaciones.

Por ejemplo veamos la respuesta para LCTVs extraídas de un video-proyector marca SONY con resolución VGA (640 x 480). El tamaño de la pantalla es aproximadamente el de una diapositiva de fotografía y la distancia centro a centro entre píxeles 40 μm.
Si se emplea como fuente de iluminación la línea azul de un láser de Ar (457 nm) es posible obtener las modulaciones mostradas en las siguientes figuras. [A.Marquez, C.Iemmi, I.Moreno, J.Davis, J.Campos, M.Yzuel. *Quantitative predictions of the modulation behavior of twisted nematic liquid crystal displays based on a simple physical model.* Opt. Eng. **40**, 2558 (2001)]



Modulación de amplitud

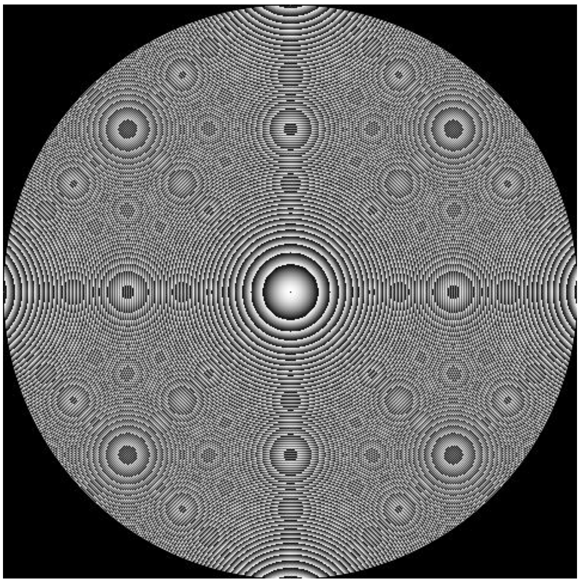


Modulación de fase

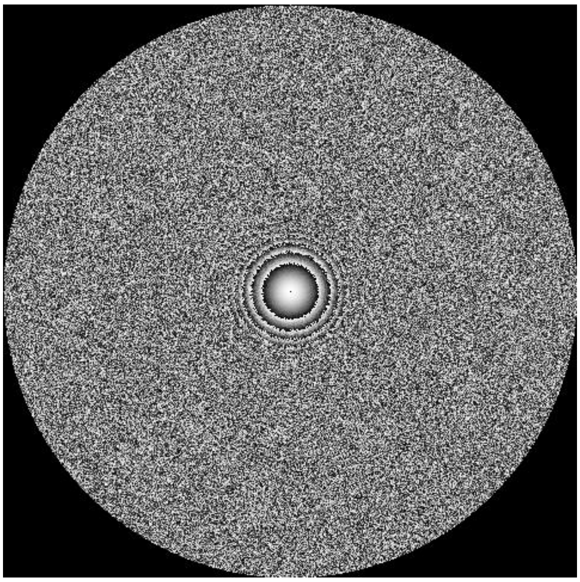
Para longitudes de onda menores no se alcanza una modulación en fase de 2 pi. En general es más interesante lograr una modulación pura de fase que de amplitud. Por ejemplo con este tipo de pantallas tenemos así una distribución de píxeles a los cuales mediante una LUT se le envía el nivel de gris necesario para obtener el desfasaje deseado. Es decir tenemos un elemento de fase programable píxel a píxel con el cual pueden obtenerse respuestas que usualmente no se consiguen con otro tipo de dispositivos y además puede modificarse a velocidad de video.

Veamos dos ejemplos en los que este dispositivo actúa como lente difractiva.

En el primero de ellos se combinaron varias lentes difractivas de diferente distancia focal, de manera de mejorar algunas de las características del sistema. Primero veamos como se obtiene una lente difractiva simple en una LCTV. No es más que la distribución de fase cuadrática que describe una lente común, calculada módulo 2 pi y enviada a través de una PC a la LCTV. En este ejemplo vamos a ver como aumentar la profundidad de enfoque (DOF) mediante una lente compuesta (esto tiene un gran interés en litografía donde es importante tener una gran DOF para obtener una imagen estable sobre el photoresist). La lente difractiva simple tiene una distancia focal de 100 cm. Diseñamos una lente compuesta conformada por 33 lentes simples, cuyas distancias focales iban de 92 a 108 cm con incrementos de 0.5 cm. Se vio que la mejor manera de obtener esta lente compuesta era eligiendo en forma random píxeles de cada una de las lentes simples [C.Iemmi, J.Campos, J.Escalera, O.López-Coronado, R.Gimeno, M.Yzuel *Depth of focus increase by multiplexing programmable diffractive lenses.* Opt. Exp. **14**, 10207 (2006)].



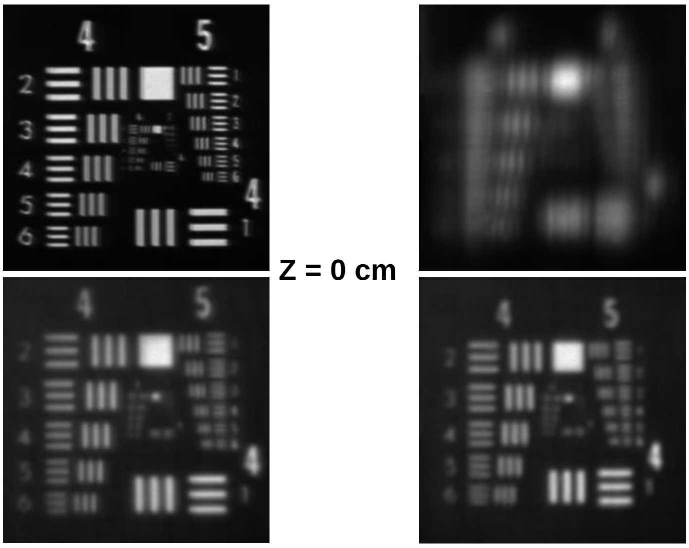
Lente simple



Lente compuesta

Lente simple

Lente compuesta

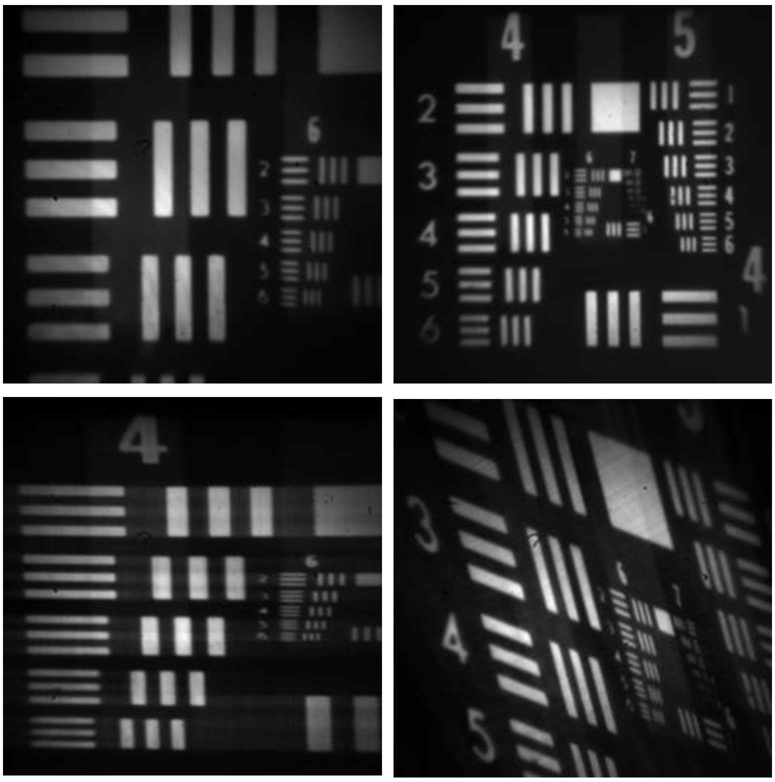


Z = 0 cm **Z = 10cm**

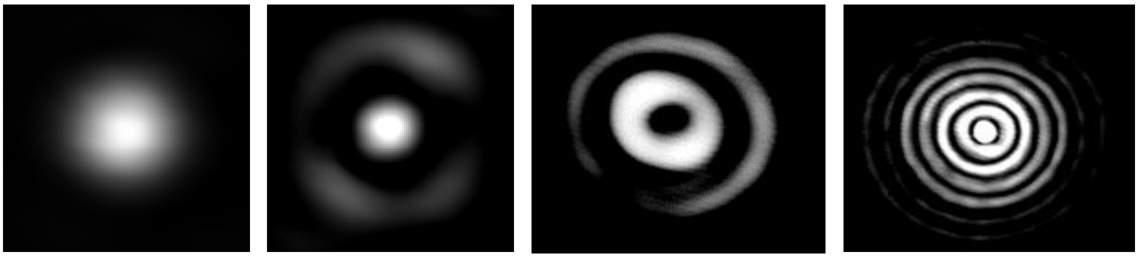
El plano de mejor enfoque para la lente simple corresponde a Z=0. A Z=+- 10cm la imagen esta muy deteriorada; vemos en cambio que para la lente compuesta la imagen mantiene su nitidez en todo el rango para la que fue programada.

Otro ejemplo es aquel en el que mediante un sistema de moduladores se puede realizar un zoom programable, anamórfico. Mediante la combinación de lentes cilíndricas con distintas orientaciones y distintas distancias focales, es posible magnificar, demagnificar y distorsionar la imagen de un objeto. A continuación se muestran algunos resultados [C.Iemmi, J.Campos, *Anamorphic zoom system based on liquid crystal displays* J. Eur. Opt. Soc. – Rapid Pub. **4**, 9029 (2009)]

.



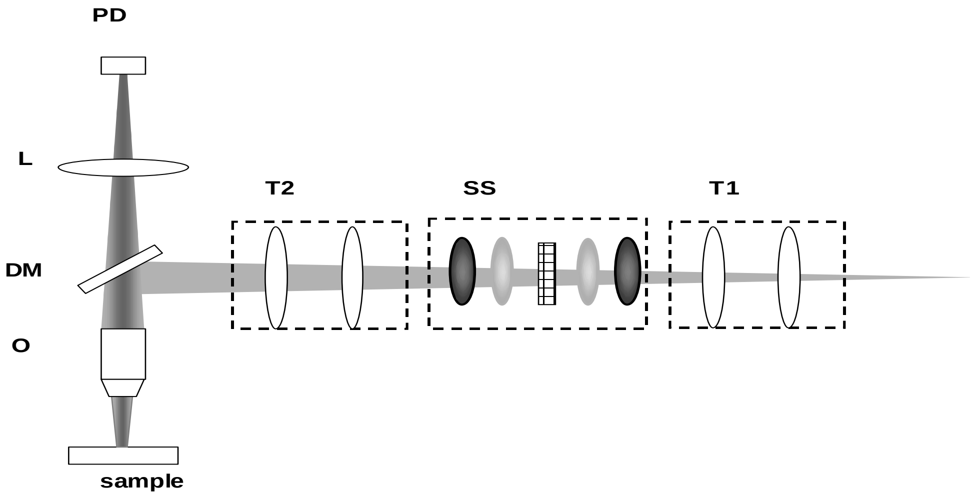
Un dispositivo como el visto también puede utilizarse para redireccionar y controlar la forma de una haz láser, siendo estas propiedades muy interesantes para aplicaciones de microscopía y para la generación de pinzas ópticas. Así es posible, por ejemplo, hacer que un haz de luz se focalice en una región transversal más ancha pero con mayor profundidad de foco en la dirección de propagación (filtro apodizante) o, por el contrario que el spot sea más concentrado (filtro hiper-resolvente). Asimismo pueden programarse distribuciones de fase que permitan obtener haces no difractivos con distribuciones de luz arbitrarias



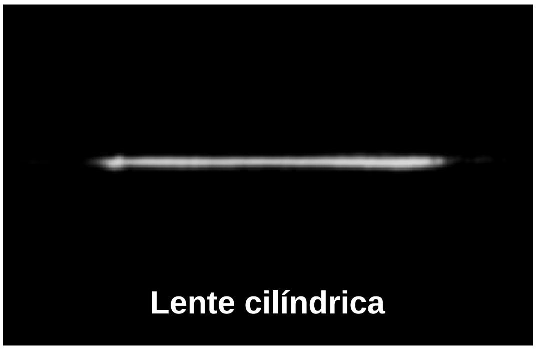
apodizante hiper-resolvente haces no difractivos

Veamos el otro ejemplo en donde la pantalla de cristal líquido es empleada para redireccionar un haz laser

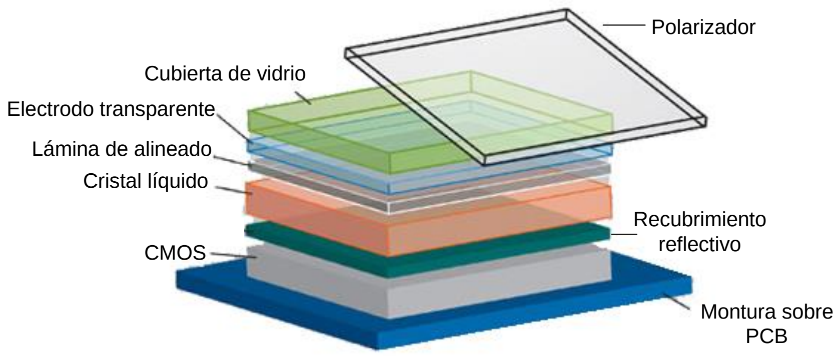
En la microscopía de barrido, por ejemplo la de fluorecencia, la imagen se va obteniendo luego de componer la señal obtenida en cada punto de scanneo de la muestra. En estos dispositivos el sistema de barrido es una pieza clave ya que debe ser rápido y preciso. Usualmente consisten en combinaciones de espejos que son rotados mediante diferentes técnicas que involucran dispositivos mecánicos. Otros consisten en un array de microlentes que son rotadas de manera de efectuar una cobertura completa del objeto a estudiar. Puede realizarse un sistema de barrido basado en una pantalla de cristal líquido que al actuar como elemento de fase permite re-direccionar el haz de barrido a voluntad. En este montaje el haz del láser utilizado para iluminar la muestra pasa por un sistema telescópico de forma de adaptar la zona que se desea iluminar de la LCTV. El conjunto LCTV, polarizadores y láminas permiten introducir los desfasajes deseados en el frente de ondas. Un segundo sistema telescópico ajusta nuevamente el área iluminada del objetivo de microscopio. Un espejo dicroico permite reflejar un haz y transmitir el correspondiente a la fluorescencia de la muestra.



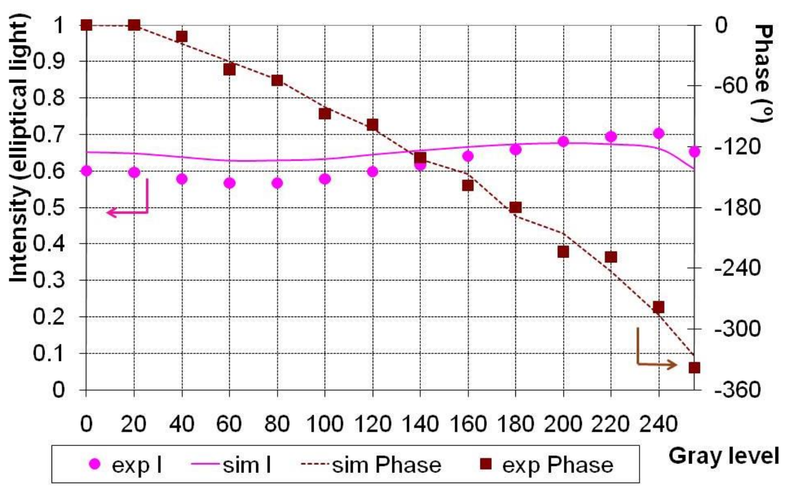
Se muestran los resultados en los que se programaron fases lineales que permiten desplazar de forma arbitraria el haz a través del plano donde esta el objeto bajo estudio. Además permite generar barridos lineales (programando una lente cilíndrica) o estudiar que sucede a distintas profundidades (programando lentes de distintas distancias focales) [M.Capeluto, C. LaMela, C.Iemmi, M.Marconi, *Scanning mechanism based on a programmable liquid crystal display*, Opt. Comm. **232**, 107, (2004)].

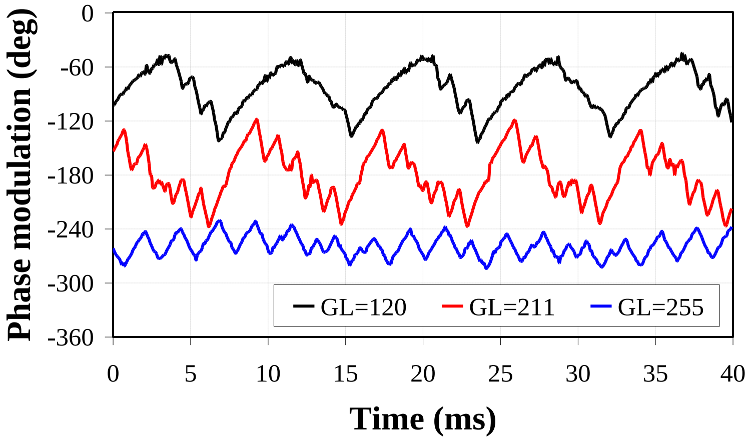


En los últimos años ha sido desarrollada una nueva tecnología de LCDs , dando lugar a un dispositivo denominado LCoS (liquid cristal on silicon). Las pantallas LCoS están basadas en el desarrollo de la tecnología CMOS, la cual permite obtener a un precio relativamente accesible un dispositivo que posee muy alta resolución, con píxeles de aproximadamente diez micrones y un factor de llenado superior al 90%. Dado que estas pantallas actúan por reflexión es posible, en principio, obtener una alta modulación de fase sobre el haz incidente ya que el mismo efectúa un recorrido de ida y vuelta dentro del cristal líquido.



A pesar de estas características tan favorables, se encontró que existe un cierto grado de depolarización de la luz y que este efecto puede llevar a una degradación en la aplicación de un LCoS como modulador espacial de luz. Un estudio detallado mostró que el problema está originado en la forma en la que se envía la señal eléctrica. En estos dispositivos los distintos niveles de gris (o fase) se logran mediante pulsos que no cambian su voltaje (como lo hacían en las LCTVs por transmisión) sino su serie temporal. Esto es, se aplican más o menos pulsos para lograra el mismo efecto que aplicando más o menos voltaje. Esto se traduce en que lo que aparentemente es un nivel de gris o fase constante, en realidad está fluctuando en el tiempo. Una muestra de este fenómeno puede verse en las siguientes figuras, donde se grafica la respuesta temporal obtenidas con una pantalla tipo LCoS marca Philips modelo X97C3A0 controlada por el dispositivo LC-R2500 provisto por Holoeye [A. Lizana, I. Moreno, A. Marquez, C. Iemmi, E. Fernandez, J. Campos, M. Yzuel, *Time fluctuations of the phase modulation in a liquid crystal on silicon display: Characterization and effects in diffractive optics,* Opt. Exp. **16**, 16711 (2008)]. Esta pantalla, de 2.46 cm de diagonal, es de cristal líquido nemático girado 45 grados, posee una resolución de 1024 X 768 pixeles con un factor de llenado del 93% y brinda 256 niveles de gris.



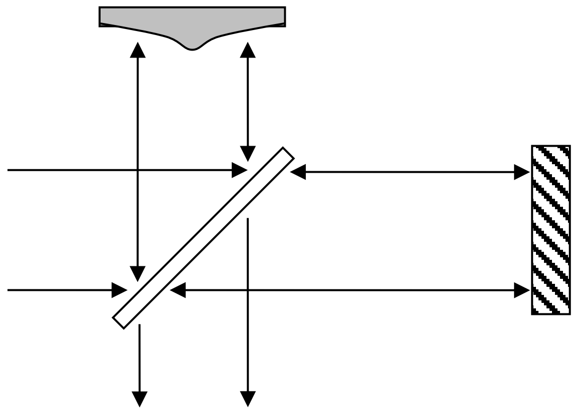


Más recientemente surgieron dispositivos de este tipo pero diseñados específicamente como moduladores espaciales de fase (esto es no son del circuito comercial perteneciente al mercado de los video-proyectores) en los que las moléculas de cristal líquido no se hallan dispuestas helicoidalmente sino en forma paralela. De esta forma se logra modulación de fase sólo con polarizadores ya que no existe modulación de amplitud acoplada. Asimismo se mejoró la forma en la que se envía el tren de pulsos, haciendo la variación temporal de la fase menos notoria.

VIII. HOLOGRAFÍA

INTRODUCCIÓN

En 1948 Dennis Gabor propuso un nuevo método formador de imágenes para el que no hacían falta lentes. El denominó al mismo reconstrucción del frente de ondas. En este proceso se deseaba registrar tanto la amplitud como la fase de las ondas luminosas provenientes del objeto. Como es sabido los medios de registro son sensible solamente a la intensidad de la onda luminosa por lo que resulta imposible registrar una fase absoluta. Lo que si puede realizarse es convertir la información de fase en variaciones de intensidad y así registrar estas últimas. Una técnica standard que realiza esta tarea es la interferometría, en donde las variaciones de fase introducidas por un objeto con respecto a un frente de ondas de referencia pueden ser registradas como variaciones de intensidad. Para lograr esto, obviamente, el objeto debe ser iluminado en forma coherente. Imaginemos por ejemplo un interferómetro de Michelson en el que uno de los espejos es reemplazado por el objeto a testear y el otro provee el haz de referencia



Gabor demostró así que la amplitud y fase de la luz proveniente de un objeto podía registrarse en una placa fotográfica (aunque esta sólo responde a cambios de intensidad) y luego reconstruirse convenientemente. Llamó a tal registro *holograma* (del griego holo:todo; grama:escritura).

Gabor publicó los primeros trabajos relacionados con el tema en los años 1949 y 1951 y estaban orientados a mejorar la resolución de los microscopios electrónicos. [D.Gabor, *Microscopy by Reconstructed Wavefronts*, Proc. Roy. Soc. **A197**, 454, (1949) ; D.Gabor, *Microscopy by Reconstructed Wavefronts II*, Proc. Phys. Soc. **B64**, 449, (1951)].

Debido a que en ese momento no se disponía de una fuente luminosa altamente coherente el tema no fue tomado con interés. Por otra parte, como veremos, resultaba muy difícil observar una imagen así obtenida.

Fue recién a comienzos de los años ´60 que, junto con el surgimiento del láser, dos investigadores de la Universidad de Michigan, Emmeth Leith y Juris Upatnieks, encontraron la forma de hacer del holograma algo realmente asombroso. La calidad y realismo de las imágenes obtenidas pronto llamó la atención de cientos de investigadores que se volcaron al tema. Surgió así una variedad enorme de aplicaciones de la técnica (interferometría holográfica, filtrado espacial, reconocimiento de caracteres, etc) que hicieron que su propiedad más efectista, esto es, la visión 3D del objeto, fuese una de las menos importantes.

Gabor recibió el premio Nobel de física en 1971 y Leith y Upatnieks una mención especial.

Analicemos algunas de las características del holograma de un objeto:

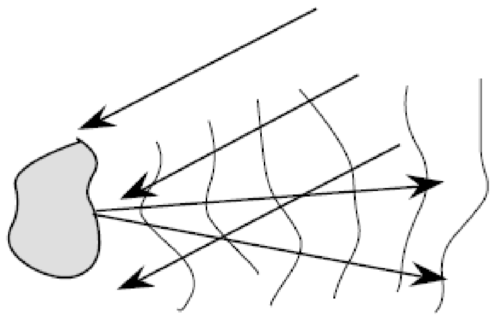
- Para su registro se utiliza, entre otros medios, material fotográfico cuya principal diferencia con el utilizado habitualmente en forma comercial es su alta resolución.

Sin embargo en una fotografía sólo se registra la intensidad proveniente del objeto mientras que en un holograma registramos intensidad y fase relativa, en realidad es la fotografía de un diagrama de interferencia.

- La holografía no necesita de lentes formadoras de imágenes por lo tanto no tiene problemas de enfoque ni profundidad de campo.
- Un holograma no requiere pasarse de positivo a negativo pues sólo sería un cambio de fase en π que el ojo no registra.
- Si registramos el holograma de un objeto difusor y luego lo dividimos, a partir de cada trozo es posible reconstruir la totalidad del objeto. Esta afirmación requiere una aclaración: debemos imaginar que es como observar el objeto por una ventana, si la misma es más chica lo seguiremos viendo pero se reduce el campo de visión, también veremos que se pierde nitidez. La posibilidad de retener toda la información en cada una de las partes se debe a que el objeto difusor emite luz en todas las direcciones, luego cada punto del mismo ilumina aproximadamente a toda la placa fotográfica, es decir sobre la misma queda registrado el diagrama de interferencia del haz de referencia con la luz proveniente del objeto.

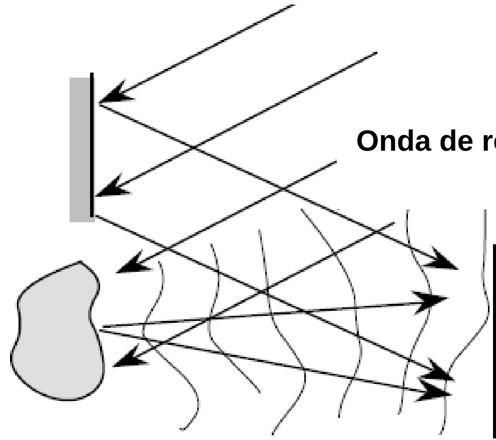
PROCESO DE SÍNTESIS Y RECONSTRUCCIÓN DE UN HOLOGRAMA

La luz proveniente del objeto la podemos describir en forma compleja como $O = O_o e^{i\varphi_o}$



Deseamos retener la información de la amplitud O_o y de la fase φ_o y para ello debemos estudiar cómo sintetizar el holograma y luego cómo reconstruirlo. Vimos que la información relativa a la fase podíamos obtenerla a partir de la interferencia de esta onda objeto con un haz de referencia al que denotaremos $r = r_o e^{i\varphi_r}$.

Así resulta necesario que O y r sean coherentes. Una forma de lograrlo es montando una especie de interferómetro de Young generalizado en el cual el papel de una de las ranuras lo juega el objeto y el de la otra ranura lo hace el haz de referencia, tal como se esquematiza en el siguiente dibujo



La distribución de intensidad sobre la placa holográfica vendrá dada por

$$I(x,y)=(O+r)(O+r)^{*}=|O|^{2}+|r|^{2}+2O_o\,r_o\cos(\varphi_o-\varphi_r)$$

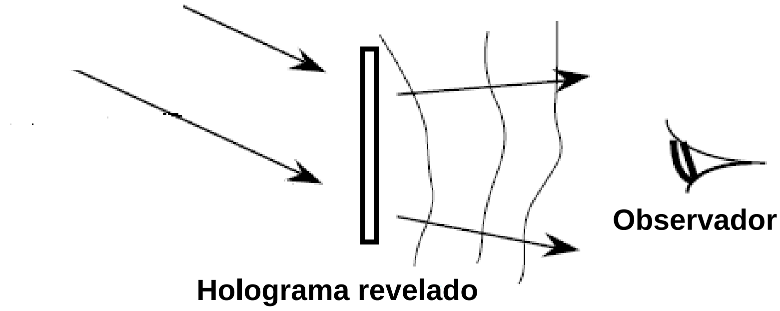
El primer término corresponde a lo que se registraría en una fotografía convencional del objeto (lente mediante), el segundo término es un fondo constante debido a la iluminación del haz de referencia, mientras que el tercer término, correspondiente a la interferencia de los dos haces, contiene toda la información de la onda objeto ya que se hallan presentes O_o y φ_o. El registro de este diagrama de interferencia constituye el holograma. Asumamos ahora que trabajamos en la zona de exposición de la película en la cual la transmisión de la misma es lineal con la intensidad recibida. Recordemos que en esa zona la transmisión en amplitud venía dada por

$$t(x,y)\cong t_b+\beta'\left(I_o(x,y)-I_b\right)$$

Tomemos, sólo para simplificar la expresión, que $I_b=|r|^{2}$, luego reemplazando la expresión hallada para la distribución de intensidad en la de la transmisión, tenemos

$$t(x,y)\cong t_b+\beta'\left(|O|^{2}+O\,r^{*}+O^{*}r\right)$$

Ahora bien, una vez que esta información está almacenada en la placa, debemos analizar cómo extraerla. Para ello utilicemos un haz de reconstrucción $R=R_o\,e^{i\,\varphi_R}$



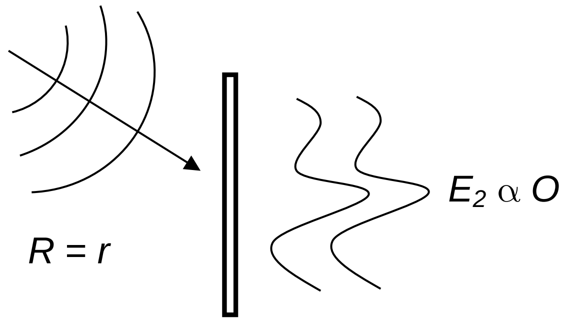
El campo emergente de la placa holográfica será

$$E(x,y) = t(x,y)\cdot R = \underbrace{\left(t_b + \beta'|O|^2\right)R}_{E_1} + \underbrace{\beta' r^* O R}_{E_2} + \underbrace{\beta' r O^* R}_{E_3}$$

El primer término (*E_1*) no es de interés ya que no aporta ninguna información sobre la fase del objeto. Analicemos que sucede con los otros dos términos

- Si tomamos como haz de reconstrucción un haz idéntico al de referencia, esto es $R = r$, entonces

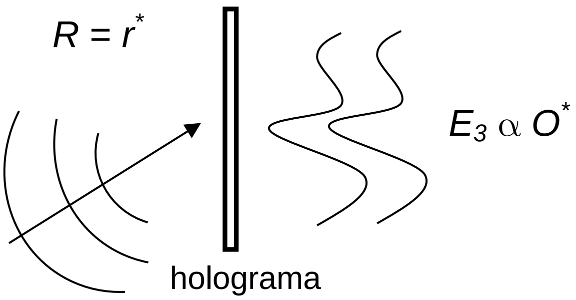
$$E_2 = \beta' r^* O R = \beta' |r|^2 O$$
$$E_3 = \beta' r O^* R = \beta' r r O^*$$



Obtenemos que *E_2* a menos de una constante provee el mismo campo que la onda objeto O; no así *E_3* donde no nos independizamos de la fase del haz de referencia r

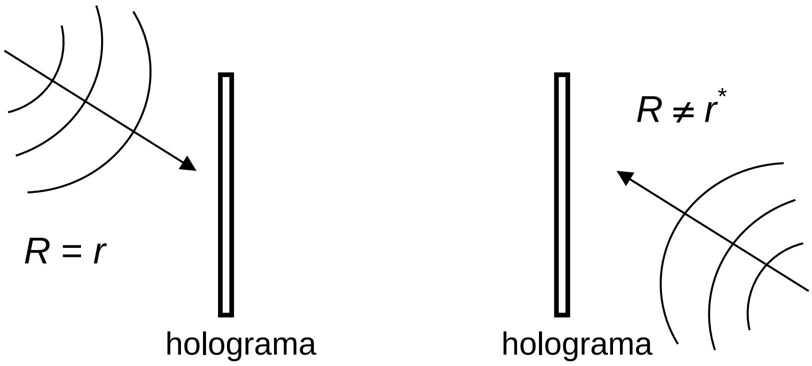
- Si tomamos como haz de reconstrucción un haz conjugado al de referencia, esto es $R = r^*$, entonces

$$E_2 = \beta' r^* O R = \beta' r^* r^* O$$
$$E_3 = \beta' r O^* R = \beta' |r|^2 O^*$$



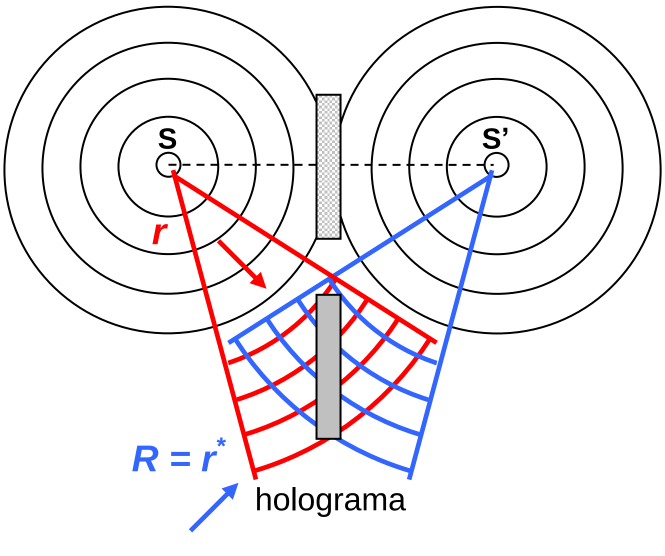
En este caso el término E_2 no es independiente de la fase de r^* y el término E_3, si bien sí lo es, provee un campo proporcional a la onda objeto conjugada O^*.

Analicemos el dibujo anterior y veamos porqué r^* se grafica de esa manera. Uno estaría tentado a pensar que el haz conjugado de r es directamente el que incide desde, o converge, hacia un punto diametralmente opuesto, es decir



Sin embargo vemos que en este caso la curvatura de ambos haces no se cancelan sobre el holograma.

Grafiquemos entonces el caso más general de dos fuentes conjugadas S y S' que emiten ondas esféricas



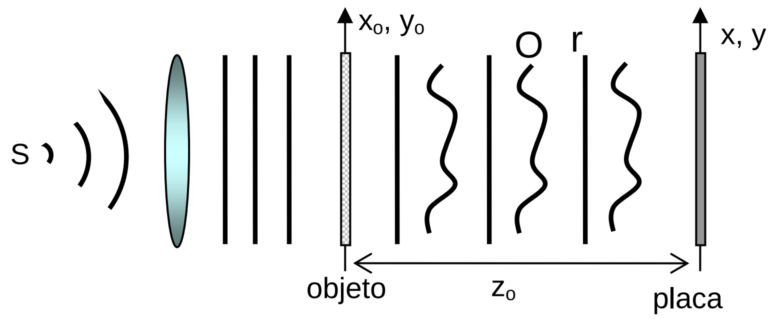
Vemos que sólo en el caso en el que la placa holográfica estuviese ubicada sobre la línea de puntos que une las dos fuentes sucede lo anteriormente descripto, de manera de cancelar sus curvaturas sobre el medio de registro. En caso contrario, si la placa está ubicada tal como hemos visto con anterioridad, entonces r^* debe converger hacia S', tal como se esquematiza en la figura.

DISPOSICIÓN DE GABOR

La disposición geométrica ideada por Gabor se muestra en la siguiente figura. Se asume que el objeto del que se desea obtener el holograma tiene un factor de transmisión bastante alto y puede escribirse como

$$t_o\left(x_o, y_o\right) = \overline{t_o} + \Delta t_o\left(x_o, y_o\right) \quad ; \quad \left|\overline{t_o}\right| >> \left|\Delta t_o\right|$$

Donde $\overline{t_o}$ es la transmitancia promedio y Δt_o son los apartamientos de dicho promedio.



Debemos recordar que en la época que Gabor diseñó esta experiencia no existían fuentes altamente coherentes como los láseres, de hecho él utilizó una lámpara de mercurio filtrada. De esta forma debió minimizar la diferencia de caminos a fin de que el haz objeto y el de referencia pudiesen interferir. Así iluminó con un haz plano un objeto cuya transmisión provocaba que una alta proporción de la luz incidente no se desviase (aquella relacionada con $\overline{t_o}$) y constituyese el haz de referencia. Otra fracción, mucho menos intensa, era difractada a causa de las variaciones de transmitancia Δt_o y constituía el haz objeto. Así la intensidad sobre la placa holográfica puede expresarse como

$$I\left(x,y\right) = \left|r + O\left(x,y\right)\right|^2 = \left|r\right|^2 + \left|O\left(x,y\right)\right|^2 + r^* O\left(x,y\right) + r\, O^*\left(x,y\right)$$

Consideremos que para el registro del holograma se trabaja en la zona lineal de la placa y que tomamos, como lo hicimos anteriormente, $I_b = |r|^2$. Además como r es una onda plana que incide normalmente entonces $r = r^*$. Luego podemos escribir la transmisión de la misma como

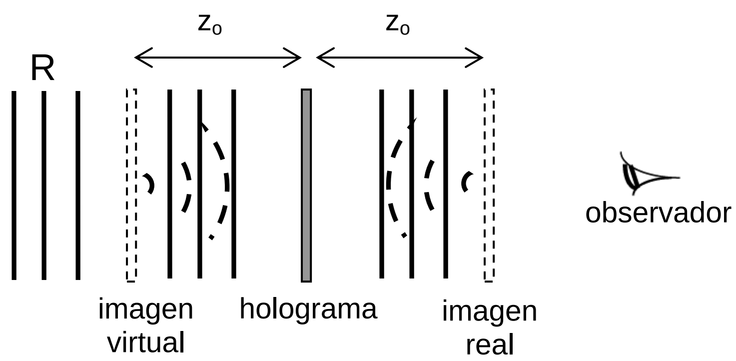
$$t(x,y) \cong t_b + \beta'\left(|O|^2 + O\,r + O^* r\right)$$

Para la reconstrucción del holograma se emplea nuevamente una onda plana con incidencia normal y amplitud *R*, con lo cual el campo transmitido viene dado por

$$Rt(x,y) \cong R\left(t_b + \beta'|O|^2\right) + \beta' R\,r\,O(x,y) + \beta' R\,r\,O^*(x,y)$$

El primer término es una onda plana que pasa por la placa casi sin dispersarse. El segundo término es proporcional a $O(x,y)$ y parece originarse en un punto virtual z_o detrás del holograma. Análogamente el tercer término es proporcional a $O^*(x,y)$ y da lugar a una imagen real a una distancia z_o delante de la placa.

La mayor limitación de este tipo de hologramas es que tanto el haz originado en el primer término como aquellos que constituyen las imágenes se hallan centrados en el eje del holograma, que es justamente la dirección de observación. Por este motivo esta disposición se conoce como *in line.* Al observar un holograma de este tipo siempre veremos dos imágenes superpuestas, una enfocada y otra fuera de foco, además como la luz de fondo es coherente y se superpone con la luz de la imágenes, estas interfieren e importa si lo registrado es un positivo o negativo ya que tendrán contraste invertido



DISPOSICIÓN DE LEITH-UPATNIEKS

CLASE 17

La mayor diferencia entre esta disposición y la de Gabor reside en que el haz de referencia no pasa a través del objeto sino que es introducido en forma desplazada y con una dada inclinación. Esta geometría es llamada *off axis*.

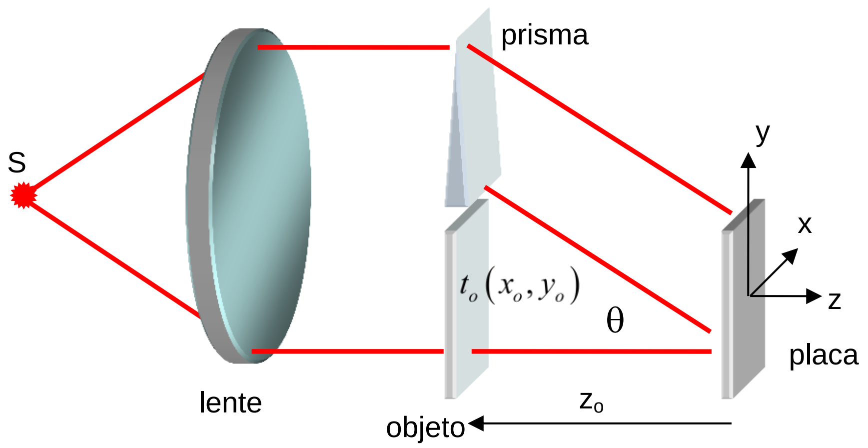
El primer trabajo donde se reporta esta disposición fue en la publicación: E.Leith, J.Upatnieks *Reconstructed Wavefronts and Communication Theory* JOSA **52**, 1123 (1962).

Otros trabajos sucesivos fueron:

E.Leith, J.Upatnieks *Wavefront Reconstruction with Continuous Tone Objects* JOSA **53**, 1377 (1963).

E.Leith, J.Upatnieks *Wavefront Reconstruction with Diffused Illumination and 3-D Objects* JOSA **54**, 1295 (1964).

La arquitectura original propuesta para el registro de este tipo de hologramas se esquematiza en la siguiente figura



La luz emergente de la fuente puntual S es colimada por la lente. Una porción del frente de ondas plano pasa a través del objeto con transmitancia $t_o(x_o, y_o)$. Otra porción del frente de ondas pasa por el prisma que lo desvía formando un ángulo θ con respecto a la normal a la placa. Así sobre el material de registro inciden dos ondas coherentes, la que pasó por el objeto y la que deflectó el prisma con dirección $\vec{k} = \frac{2\pi}{\lambda}(0, -\sin\theta, \cos\theta)$. La distribución de campo sobre la placa será

$$E(x,y) = r_o \exp\left[-i\frac{2\pi}{\lambda}\sin\theta \cdot y\right] + O(x,y)$$

donde elegimos sobre la placa el plano z = 0. La distribución de intensidad será

$$I(x,y)=|r|^2+|O(x,y)|^2+r_o\exp\left[i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O(x,y)+r_o\exp\left[-i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O^*(x,y)$$

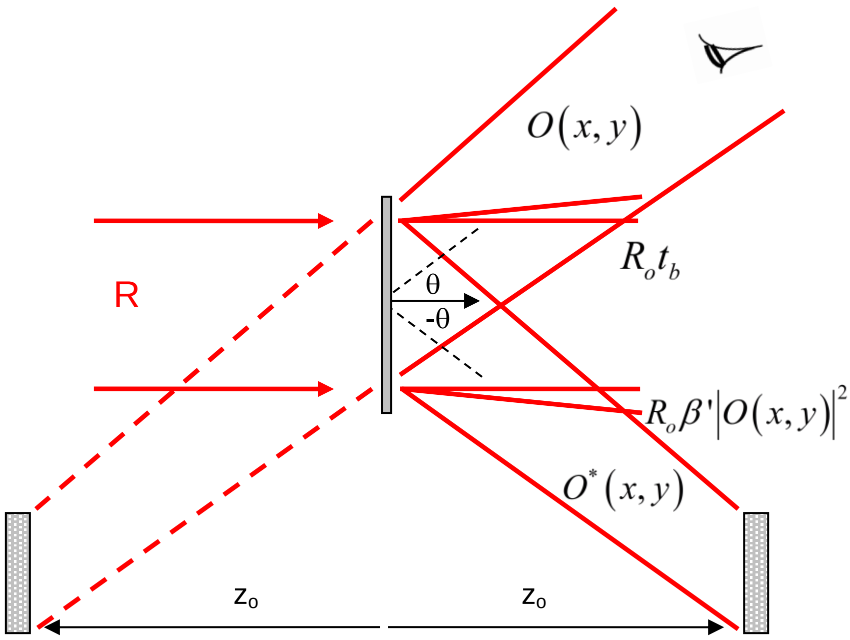
Nuevamente vamos a suponer que trabajamos en una zona del material de registro donde vale que la transmisión del campo es lineal con la intensidad de registro y tomamos $I_b=|r|^2$. Así el factor de transmisión de la placa será

$$t(x,y)=t_b+\beta'\left(|O(x,y)|^2+r_o\exp\left[i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O(x,y)+r_o\exp\left[-i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O^*(x,y)\right)$$

Luego si para reconstruir el holograma se emplea un haz plano que incide normalmente, de modo que $R=R_o$, entonces el campo emergente será

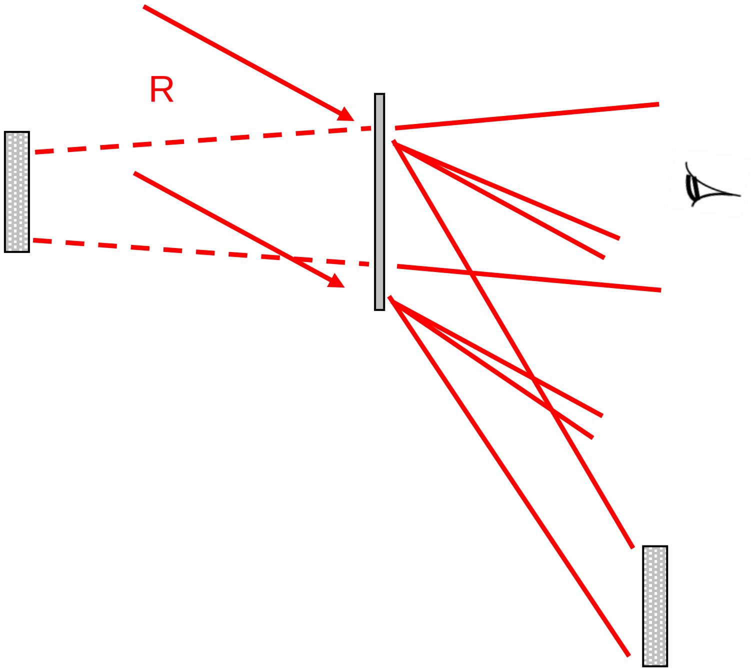
$$E(x,y)=Rt(x,y)=R_o\left(t_b+\beta'|O(x,y)|^2\right)+\beta' R_o r_o\exp\left[i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O(x,y)+$$
$$\beta' R_o r_o\exp\left[-i\frac{2\pi}{\lambda}\sin\theta\cdot y\right]O^*(x,y)$$

El primer término es directamente transmitido y no aporta ninguna información ya que $R_o t_b$ es una onda plana y $R_o\beta'|O(x,y)|^2$ representa a un haz difractado cuyo apartamiento del eje dependerá de las frecuencias espaciales de $|O(x,y)|^2$. El segundo término corresponde a una onda que lleva la información de $O(x,y)$ y que se propaga formando un ángulo θ con respecto a la normal a la placa. El tercer término se propaga con dirección $-\theta$ y contiene la información relacionada con $O^*(x,y)$



Vemos que en este caso las imágenes están separadas angularmente y por lo tanto se las puede observar en forma independiente. Por otra parte, dado que el fondo coherente no se superpone con la imagen de interés, estos haces no interfieren y en consecuencia no importa si el holograma es un positivo o un negativo. Otro detalle a tener en cuenta es que la frecuencia espacial de la portadora, la cual está determinada por el ángulo θ con que incidió el haz de referencia, debe elegirse de forma tal que sea mucho mayor que las frecuencias espaciales propias del objeto, pues así quedan bien separados los haces proporcionales a $O(x,y)$ y $O^*(x,y)$ con respecto al proporcional a $R_o\beta'|O(x,y)|^2$.

Si en vez de incidir normalmente con el haz de reconstrucción, lo hacemos de manera que sea igual al de referencia $R = r$, el objeto se reconstruirá en igual posición que la que se encontraba originalmente el objeto en el momento de registro



HOLOGRAMAS BI Y TRI DIMENSIONALES

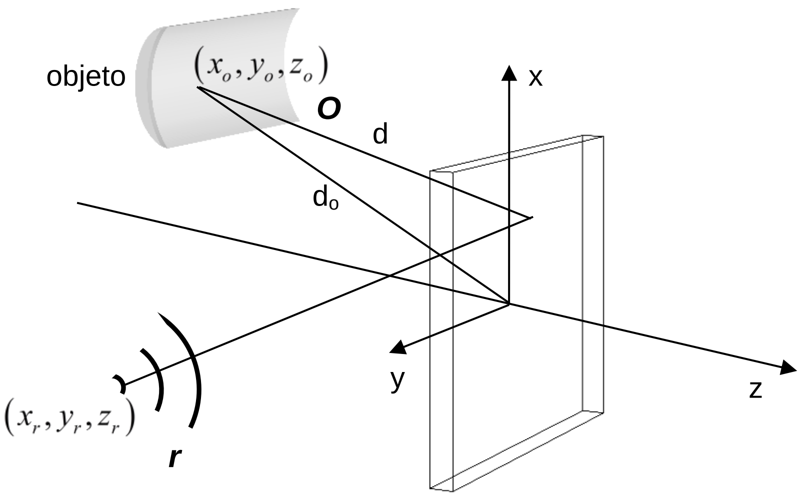
Vimos que en la disposición de Leith-Upatnieks la introducción de un haz de referencia con un cierto ángulo hace que la información registrada en el holograma tenga una frecuencia espacial media tal que permite la separación de las imágenes en el espacio.

El período de las franjas asociado con la frecuencia de la portadora es el que determina si el medio de registro puede considerarse como un medio bidimensional o

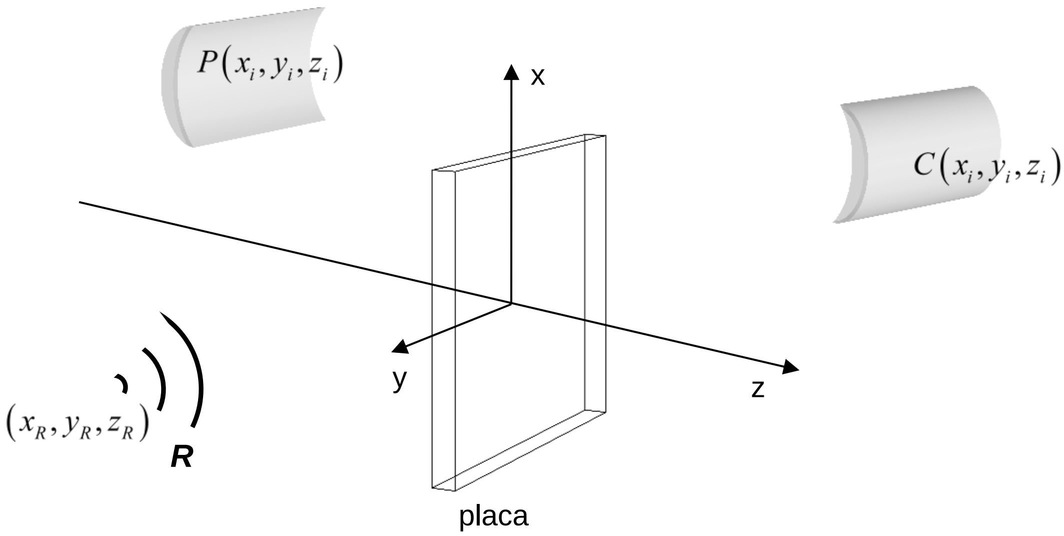
tridimensional. Esto es, la relación entre la frecuencia espacial de la portadora y el espesor del medio de registro establece si el holograma es plano o de volumen.
Supongamos que el centro del objeto está angularmente separado del haz de referencia por un ángulo θ, entonces el período de las franjas será aproximadamente $d \sim \frac{\lambda}{\sin\theta}$ donde λ es la longitud de onda de registro. Cuando d es del orden o menor que el espesor del medio de registro, el holograma comenzará a comportarse como un medio tridimensional. Por ejemplo un termoplástico o photoresist con los que obtenemos modulación de superficie se comportarán como hologramas planos, en cambio el tipo de respuesta de una placa holográfica (el espesor de la emulsión varía dependiendo de la marca entre 7 y 20 μm) dependerá del ángulo θ en particular. En general los hologramas grabados en cristales fotorrefractivos serán de volumen
Vamos a estudiar en detalle la difracción por medios bi y tri dimensionales, pero comenzaremos analizando el caso de hologramas planos que es más sencillo.

HOLOGRAMAS PLANOS

Dentro de la categoría *hologramas planos* podemos realizar una subdivisión de acuerdo a la geometría y elementos que se emplearon para su registro. Para analizar mejor estas situaciones veamos con algún detalle la matemática del proceso. Vamos a introducir entonces la siguiente notación

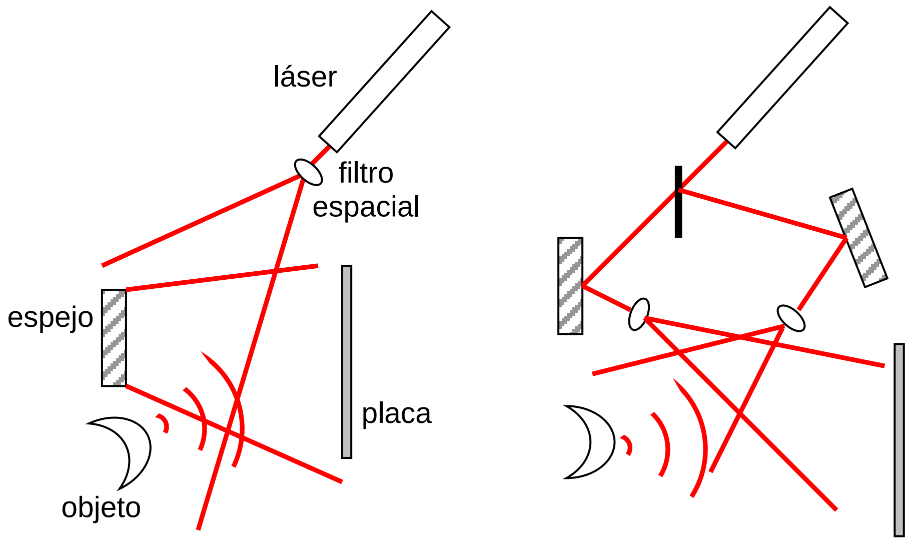


El sistema de coordenadas tiene como origen a la placa holográfica. Un punto del objeto está caracterizado por las coordenadas (x_o, y_o, z_o). El haz de referencia proviene de una fuente puntual ubicada en el punto (x_r, y_r, z_r) y en el caso de ser un haz plano se toma $z_r \to \infty$. El haz de reconstrucción *R* también proviene de una fuente puntual ubicada en $\left(x_R, y_R, z_R\right)$ y si es un haz plano nuevamente se toma $z_R \to \infty$. La iluminación del holograma con el haz de reconstrucción da origen a una imagen primaria *P* y a una conjugada *C*. Un punto imagen se caracteriza por las coordenadas $\left(x_i, y_i, z_i\right)$.



Holograma de Fresnel

Este tipo de hologramas es el que se produce cuando el objeto se halla cercano a la placa. El fenómeno de difracción involucrado es el descripto por la aproximación de Fresnel. Algunas disposiciones experimentales típicas para registrar esta clase de hologramas se esquematizan a continuación



Cada punto objeto es emisor de ondas esféricas por lo que la perturbación total que produce en el plano del holograma viene dada por

$$O(x,y,z) = \frac{-i}{\lambda_1} \int \int_{-\infty}^{\infty} O(x_o, y_o, z_o) \cos\left(\hat{d},\hat{n}\right) \frac{\exp(ik_1 d)}{d} dx_o dy_o$$

con $d = \left[(x_o - x)^2 + (y_o - y)^2 + (z_o - z)^2 \right]^{1/2}$.

El haz de referencia sobre el holograma será $r(x,y,z) = r_o \dfrac{\exp(ik_1 l)}{l}$ donde

$l = \left[(x_r - x)^2 + (y_r - y)^2 + (z_r - z)^2 \right]^{1/2}$

Para simplificar un poco los cálculos los haremos en dos dimensiones y luego extenderemos el resultado a tres.

En la aproximación de Fresnel y recordando que la placa se halla en Z = 0 tenemos que la onda objeto viene dada por

$$O(x) = \frac{\exp(-ik_1 z_o)}{\sqrt{i\lambda_1 z_o}} \exp\left(\frac{-ik_1 x^2}{2z_o}\right) \int_{-\infty}^{\infty} O(x_o) \exp\left(\frac{-ik_1 x_o^2}{2z_o}\right) \exp\left(\frac{ik_1 x\, x_o}{z_o}\right) dx_o$$

Luego en la aproximación cuadrática a una onda esférica el haz de referencia toma la forma

$$r = \frac{r_o}{\sqrt{z_r}} \exp\left[-ik_1 \left(z_r + \frac{x_r^2}{2z_r} + \frac{x^2}{2z_r} - \frac{x\, x_r}{z_r} \right) \right]$$

Análogamente para el haz de reconstrucción tenemos que

$$R = \frac{R_o}{\sqrt{z_R}} \exp\left[-ik_2 \left(z_R + \frac{x_R^2}{2z_R} + \frac{x^2}{2z_R} - \frac{x\, x_R}{z_R} \right) \right]$$

Notemos que k_1 en principio puede ser distinto de k_2 ya que la longitud de onda de registro no tiene porque coincidir con la de reconstrucción.

Ahora bien, vimos con anterioridad que los términos difractados de interés eran E_2 y E_3.

$$E(x,y) = t(x,y) \cdot R = \underbrace{\left(t_b + \beta' |O|^2\right) R}_{E_1} + \underbrace{\beta' r^* O\, R}_{E_2} + \underbrace{\beta' r\, O^* R}_{E_3}$$

Para calcularlos debemos tener en cuenta que cuando iluminamos el holograma con el haz de referencia *R*, es el mismo holograma el que actúa como una abertura difractora. Así tenemos que

$$E_2\left(x_i\right)=\beta'\frac{r_o}{\sqrt{z_r}}\exp\left(ik_1 z_r\right)\exp\left(ik_1\frac{x_r^2}{2z_r}\right)\frac{R_o}{\sqrt{z_R}}\exp\left(-ik_2 z_R\right)\exp\left(-ik_2\frac{x_R^2}{2z_R}\right)\frac{\exp\left(ik_2 z_i\right)}{\sqrt{i\lambda_2 z_i}}\cdot$$
$$.\exp\left(ik_2\frac{x_i^2}{2z_i}\right)\frac{\exp\left(-ik_1 z_o\right)}{\sqrt{i\lambda_1 z_o}}\cdot\int\int_{-\infty}^{\infty}O\left(x_o\right)\exp\left(\frac{ik_1 x^2}{2z_r}\right)\exp\left(\frac{-ik_1 x\,x_r}{z_r}\right)\exp\left(\frac{-ik_2 x^2}{2z_R}\right).$$
$$.\exp\left(\frac{ik_2 x\,x_R}{z_R}\right)\exp\left(\frac{-ik_1 x_o^2}{2z_o}\right)\exp\left(\frac{ik_1 x\,x_o}{z_o}\right)\exp\left(\frac{ik_2 x^2}{2z_i}\right)\exp\left(\frac{-ik_1 x^2}{2z_o}\right)\exp\left(\frac{-ik_2 x\,x_i}{z_i}\right)dx_o dx$$

Escribamos esta ecuación de una manera más simplificada, para ello llamamos

$$a=\frac{k_1}{2z_r}-\frac{k_2}{2z_R}+\frac{k_2}{2z_i}-\frac{k_1}{2z_o}\quad ;\quad b=\frac{x_r}{\lambda_1 z_r}-\frac{x_R}{\lambda_2 z_R}-\frac{x_o}{\lambda_1 z_o}+\frac{x_i}{\lambda_2 z_i}$$

con lo cual podemos escribir

$$E_2\left(x_i\right)=Cte\frac{\exp\left(-ik_1 z_o\right)}{\sqrt{z_o}}\int\int_{-\infty}^{\infty}O\left(x_o\right)\exp\left(\frac{-ik_1 x_o^2}{2z_o}\right)\exp\left(ia\,x^2\right)\exp\left(-i2\pi b\,x\right)dx\,dx_o$$

En realidad para considerar la difracción en el holograma deberíamos integrar la variable *x* entre los límites de la placa y no entre $\pm\infty$, pero veremos este efecto más adelante.

Recordando nuevamente cuál es la aproximación cuadrática a una onda esférica (ver las expresiones de *r* y *R*) puede decirse que en la ecuación hallada para $E_2\left(x_i\right)$, la expresión $\frac{O\left(x_o\right)}{\sqrt{z_o}}\exp\left[-ik_1\left(z_o+\frac{x_o^2}{2z_o}\right)\right]$ es una onda esférica con centro en $\left(x_o,-z_o\right)$ observada desde el origen de coordenadas. Ahora bien, escribamos la ecuación anterior de la siguiente forma

$$E_2\left(x_i\right)=Cte\frac{\exp\left(-ik_1 z_o\right)}{\sqrt{z_o}}\int_{-\infty}^{\infty}O\left(x_o\right)\exp\left(\frac{-ik_1 x_o^2}{2z_o}\right)\left\{\int_{-\infty}^{\infty}\exp\left(ia\,x^2\right)\exp\left(-i2\pi b\,x\right)dx\right\}dx_o$$

La integral sobre dx_o nos daría el campo que produce todo el objeto y la integral sobre dx es la que contempla el efecto del holograma. Hagamos una analogía con lo visto para sistemas formadores de imágenes y recordemos la expresión

$$E_i\left(x_i\right)=\int E_o\left(x_o\right)h\left(x_i,x_o\right)dx_o$$

En ese caso para que la imagen fuese lo más parecida posible al objeto, la función respuesta al impulso del sistema debía aproximarse a una función delta de dirac.
En este caso si queremos que la onda difractada por el holograma sea lo más parecida posible a la onda objeto, entonces debemos pedir que la integral entre llaves tienda a una función delta. Luego eso es cierto si se cumple que *a = 0*. En ese caso

$$\int_{-\infty}^{\infty}\exp\left(ia\,x^2\right)\exp\left(-i2\pi b\,x\right)dx=\int_{-\infty}^{\infty}\exp\left(-i2\pi b\,x\right)dx=\delta\left(b\right)$$

Pedir que *a = 0* implica que $z_i=\dfrac{1}{\dfrac{1}{z_R}+\dfrac{\lambda_2}{\lambda_1}\left(\dfrac{1}{z_o}-\dfrac{1}{z_r}\right)}$. Por otra parte tenemos que

$$E_2\left(x_i\right)=Cte\frac{\exp\left(-ik_1z_o\right)}{\sqrt{z_o}}\int_{-\infty}^{\infty}O\left(x_o\right)\exp\left(\frac{-ik_1x_o^2}{2z_o}\right)\delta\left(b\right)dx_o$$

Esta integral vale 0 siempre que $b \neq 0$. Entonces obtenemos una onda difractada sólo cuando $b = 0$ lo que implica que $x_i=\left[\dfrac{\lambda_2}{\lambda_1}\left(\dfrac{x_o}{z_o}-\dfrac{x_r}{z_r}\right)+\dfrac{x_R}{z_R}\right]z_i$

La ecuación hallada para z_i es una condición de focalización análoga a la ecuación de las lentes $\dfrac{1}{d_o}+\dfrac{1}{d_i}=\dfrac{1}{f}$.

Extendiendo los resultados hallados a tres dimensiones tenemos que

$$x_i=\left[\frac{\lambda_2}{\lambda_1}\left(\frac{x_o}{z_o}-\frac{x_r}{z_r}\right)+\frac{x_R}{z_R}\right]z_i$$

$$y_i=\left[\frac{\lambda_2}{\lambda_1}\left(\frac{y_o}{z_o}-\frac{y_r}{z_r}\right)+\frac{y_R}{z_R}\right]z_i$$

$$z_i=\left[\frac{1}{z_R}+\frac{\lambda_2}{\lambda_1}\left(\frac{1}{z_o}-\frac{1}{z_r}\right)\right]^{-1}$$

Estas serán las coordenadas de la imagen primaria $P\left(x_i,y_i,z_i\right)$.

Haciendo lo propio para el término E_3 obtenemos las coordenadas para la imagen conjugada $C\left(x_i, y_i, z_i\right)$

$$x_i = \left[\frac{\lambda_2}{\lambda_1}\left(-\frac{x_o}{z_o} + \frac{x_r}{z_r}\right) + \frac{x_R}{z_R}\right] z_i$$
$$y_i = \left[\frac{\lambda_2}{\lambda_1}\left(-\frac{y_o}{z_o} + \frac{y_r}{z_r}\right) + \frac{y_R}{z_R}\right] z_i$$
$$z_i = \left[\frac{1}{z_R} + \frac{\lambda_2}{\lambda_1}\left(-\frac{1}{z_o} + \frac{1}{z_r}\right)\right]^{-1}$$

Vemos, a partir de estas ecuaciones, que cuál imagen es real y cuál es virtual depende de z_i y en consecuencia de z_o , z_r y z_R. Si $z_i > 0$ la imagen es real, si $z_i < 0$ la imagen es virtual. Así ambas imágenes, la primaria $P\left(x_i, y_i, z_i\right) \sim O$ y la conjugada $C\left(x_i, y_i, z_i\right) \sim O^*$ puede ser reales o virtuales.

Notemos, por otra parte, que si la longitud de onda de registro λ_1 es igual a la longitud de onda de reconstrucción λ_2 y además la posición de la fuente puntual que genera el haz de referencia coincide con la posición de la usada para la reconstrucción, entonces la imagen primaria coincide en posición y tamaño con el objeto original utilizado para el registro del holograma. Esto es $x_r = x_R$, $y_r = y_R$, $z_r = z_R$ entonces $x_i = x_o$, $y_i = y_o$, $z_i = z_o$.

En otras circunstancias veamos qué pasa con el aumento. Para ello tenemos que relacionar un diferencial de objeto con un diferencial imagen.

$$\frac{\Delta x_i}{\Delta x_o} = \frac{\Delta y_i}{\Delta y_o} = \pm \frac{\lambda_2}{\lambda_1}\frac{z_i}{z_o} \equiv M$$

Reemplazando por la expresión de z_i obtenemos el aumento M como función de los parámetros de síntesis y reconstrucción

$$M = \frac{1}{1 - \frac{z_o}{z_r} \pm \frac{\lambda_1}{\lambda_2}\frac{z_o}{z_R}}$$ donde el signo + es para P y el – para C

Si $\lambda_1 = \lambda_2 \quad ; \quad z_r = z_R \quad \Rightarrow \quad M = 1$ para *P*.

Si $\lambda_1 = \lambda_2 \quad ; \quad z_r = -z_R \quad \Rightarrow \quad M = 1$ para *C*.

Si $\lambda_1 = \lambda_2 \quad ; \quad z_r = z_R = \infty \quad \Rightarrow \quad M = 1$ para *P* y *C*.

El aumento angular está dado por la relación $M_A = \frac{\alpha_i}{\alpha_o}$. En la aproximación paraxial podemos tomar $\tan\alpha_i \cong \alpha_i \cong \frac{\Delta x_i}{z_i}\,;\, \tan\alpha_o \cong \alpha_o \cong \frac{\Delta x_o}{z_o}$ con lo cual podemos escribir

$$M_A = \frac{\Delta x_i}{\Delta x_o}\frac{z_o}{z_i} = \pm\frac{\lambda_2}{\lambda_1}$$

Vemos de esta relación porqué Gabor estaba interesado en el método holográfico para aumentar la resolución de un microscopio electrónico. La relación entre la longitud de onda de un electrón y la longitud de onda del visible es aproximadamente 100000 X.

Holograma de Fraunhofer

Cuando el tamaño del objeto que se registrará en el holograma es muy pequeño a comparación de su distancia a la placa $\left(\frac{x_o^2 + y_o^2}{\lambda} \ll z_o\right)$ estamos en condiciones de difracción de Fraunhofer. Una situación típica donde esto sucede es, por ejemplo, en el holograma de un grupo de pequeñas partículas tales como las encontradas en una suspensión o las producidas por un aerosol.

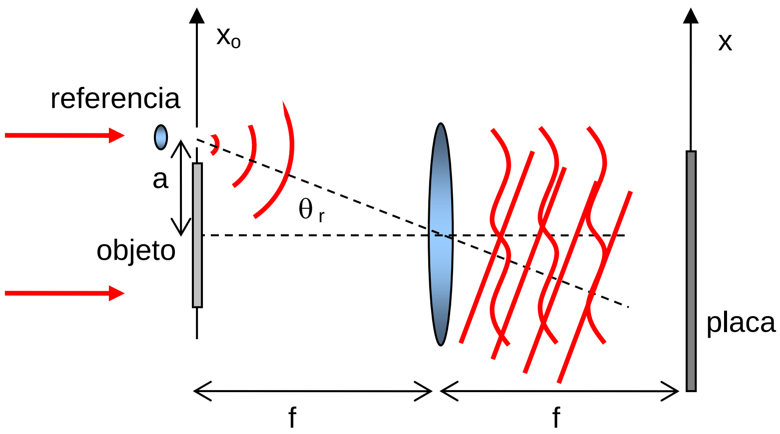
La ventaja de tener un holograma de una distribución de este tipo en vez de una fotografía es que se puede ir estudiando por distintos planos a z_o = cte.

Como sabemos la figura de difracción de Fraunhofer registrada en el holograma corresponde a la transformada de Fourier del objeto. Existe otro tipo de holograma en el que también se registra la transformada de Fourier en vez del objeto mismo y es el que veremos a continuación.

Holograma de Fourier

Se define como holograma tipo Fourier a aquel en el que se registra sobre la placa holográfica, la interferencia de dos ondas tales que corresponden a las transformadas de Fourier del objeto y de la referencia, respectivamente. Normalmente esto implica que el objeto sea bidimensional o de espesor limitado. Este tipo de hologramas, como veremos más adelante, es de suma importancia para la construcción de filtros para procesado óptico de señales.

Una disposición típica para registrar esta clase de hologramas se esquematiza en la siguiente figura.



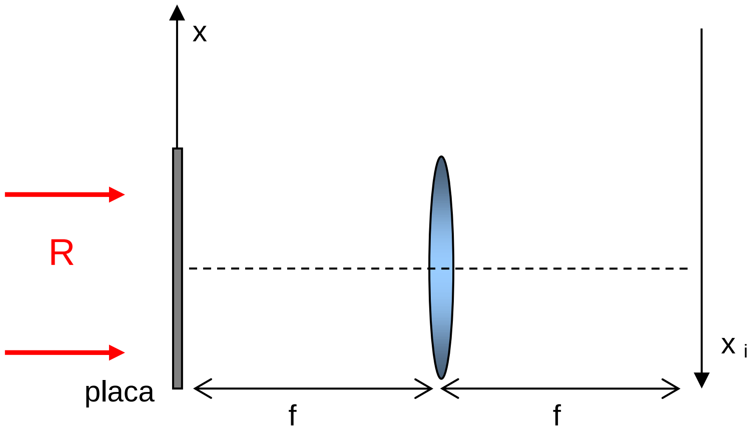
Un haz colimado monocromático ilumina a un objeto de transmisión $t_o\left(x_o, y_o\right)$ y a una lente que lo concentra sobre un pinhole que actúa como fuente puntual de referencia. Tanto el objeto como el pinhole se encuentran en el plano focal anterior de la lente transformadora. La placa holográfica se sitúa en el plano focal posterior de la misma de forma que las ondas que llegan son proporcionales a la transformada de Fourier de la transparencia y del pinhole $r\left(x_o\right)=cte\,\delta\left(x_o-a, y_o\right)$. Así sobre la placa llegarán las ondas

$$O(x,y)=\mathbb{F}_{f_x f_y}\left[t_o\left(x_o,y_o\right)\right] \quad ; \quad r(x,y)=r_o \exp\left[-i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]$$

Luego la intensidad sobre la misma será

$$\begin{aligned} I(x,y) &= \left[O(x,y)+r(x,y)\right]\left[O(x,y)+r(x,y)\right]^* = \\ &= \left|\mathbb{F}_{f_x f_y}\left[t_o\left(x_o,y_o\right)\right]\right|^2 + r_o^2 + r_o\,\mathbb{F}_{f_x f_y}\left[t_o\left(x_o,y_o\right)\right]\exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right] + \\ &+ r_o\,\mathbb{F}^*_{f_x f_y}\left[t_o\left(x_o,y_o\right)\right]\exp\left[-i\frac{2\pi}{\lambda}\sin\theta_r\, x\right] \end{aligned}$$

Como es habitual suponemos que el factor de transmisión de la placa, una vez revelada, es lineal con $I(x,y)$ y elegimos $r_o^2 = I_b$. Luego para reconstruir el holograma se utiliza el siguiente dispositivo



Iluminamos al holograma con un haz plano, monocromático que incide normalmente. Luego inmediatamente a la salida de la placa tenemos

$$E(x,y) = R_o \cdot t(x,y) = R_o \left(t_b + \beta' \left| \mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right] \right|^2 \right) +$$
$$+\beta' R_o r_o \mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right] \exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right] +$$
$$+\beta' R_o r_o \mathbb{F}^*_{f_x f_y}\left[t_o(x_o,y_o)\right] \exp\left[-i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]$$

En el plano de observación se obtendrá $\mathbb{F}^{-1}_{x_i y_i}\left[E(x,y)\right]$. Analicemos cada término separadamente. Los dos primeros términos corresponderán a

$$\mathbb{F}^{-1}_{x_i y_i}\left[R_o t_b\right] = \delta(x_i,y_i)$$
$$\mathbb{F}^{-1}_{x_i y_i}\left[\beta' R_o \mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right] . \mathbb{F}^*_{f_x f_y}\left[t_o(x_o,y_o)\right]\right] = \beta' R_o \mathbb{F}^{-1}_{x_i y_i}\left[\mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right]\right] \otimes$$
$$\otimes \mathbb{F}^{-1}_{x_i y_i}\left[\mathbb{F}^*_{f_x f_y}\left[t_o(x_o,y_o)\right]\right] = \beta' R_o t_o(x_i,y_i) \odot t_o(x_i,y_i)$$

El tercer término será

$$\mathbb{F}^{-1}_{x_i y_i}\left[\beta' R_o r_o \mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right] \exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]\right] =$$
$$= \beta' R_o r_o \mathbb{F}^{-1}_{x_i y_i}\left[\mathbb{F}_{f_x f_y}\left[t_o(x_o,y_o)\right]\right] \otimes \mathbb{F}^{-1}_{x_i y_i}\left[\exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]\right]$$

Si tenemos en cuenta que $\sin\theta_r \cong \theta_r \cong \dfrac{a}{f}$ y la definición de frecuencia espacial tenemos que

$$\exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right] = \exp\left[i\frac{2\pi}{\lambda}\frac{a}{f}x\right] = \exp\left[i2\pi\, f_x\, a\right] \Rightarrow$$

$$\Rightarrow \mathbb{F}^{-1}_{x_i y_i}\left[\exp\left[i2\pi\, f_x\, a\right]\right] = \delta\left(x_i + a, y_i\right)$$

Con lo cual

$$\mathbb{F}^{-1}_{x_i y_i}\left[\beta' R_o r_o\, \mathbb{F}_{f_x f_y}\left[t_o\left(x_o, y_o\right)\right]\exp\left[i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]\right] =$$

$$= \beta' R_o r_o\, t_o\left(x_i, y_i\right) \otimes \delta\left(x_i + a, y_i\right) = \beta' R_o r_o\, t_o\left(x_i + a, y_i\right)$$

Análogamente para el cuarto término tendremos

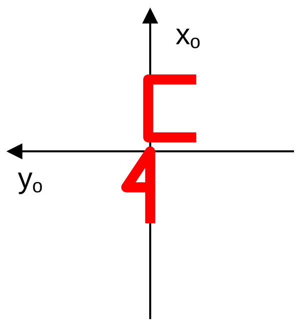
$$\mathbb{F}^{-1}_{x_i y_i}\left[\beta' R_o r_o \mathbb{F}^{*}_{f_x f_y}\left[t_o\left(x_o, y_o\right)\right]\exp\left[-i\frac{2\pi}{\lambda}\sin\theta_r\, x\right]\right] = \beta' R_o r_o\, t_o^{*}\left(-x_i + a, -y_i\right)$$

Así en el plano de salida la distribución de campo viene dada por

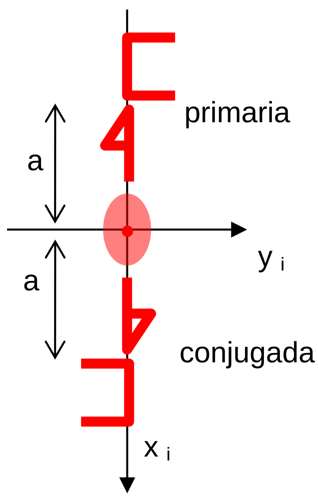
$$E\left(x_i, y_i\right) = A\delta\left(x_i, y_i\right) + B t_o\left(x_i, y_i\right) \odot t_o\left(x_i, y_i\right) + C t_o\left(x_i + a, y_i\right) + D t_o^{*}\left(-x_i + a, -y_i\right)$$

con A, B, C, D constantes.

Las imágenes de interés, como siempre, son la primaria y la conjugada representadas por el tercer y cuarto término respectivamente. Ambas imágenes son reales y están sobre el mismo plano ya que el objeto y el haz de referencia también lo están.



objeto en el
plano de entrada



imágenes en el
plano de salida

Luego veremos un dispositivo similar que permite sintetizar filtros para emplear en el procesado de señales. Cabe destacar que acá estamos viendo los distintos tipos de hologramas, qué es lo que se registra y cuál es la respuesta pero no entramos en el detalle de su síntesis. En el caso particular de la generación de filtros aplicados al procesado de señales, si bien se usan hologramas tipo Fourier, los mismos no se sintetizan ópticamente sino que son generados por computadora. Es decir los conceptos físicos involucrados son los mismos pero el proceso es numérico, luego el holograma así calculado se implementa en un medio material que permite utilizarlo en un procesador óptico. Luego veremos cómo hacerlo.

Holograma rainbow

CLASE 18

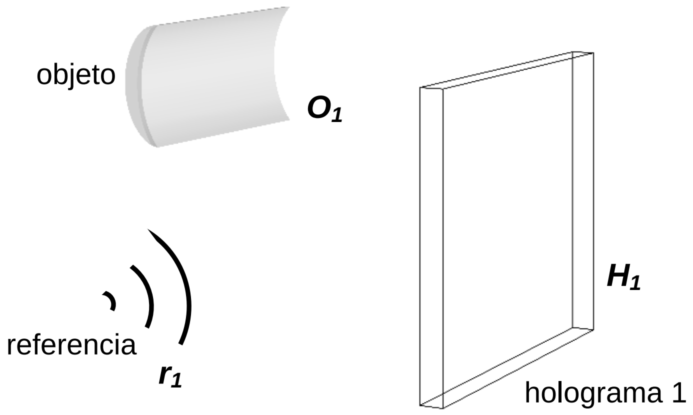
Los hologramas que hemos visto hasta ahora necesitan que tanto para su registro como para su reconstrucción se emplee una fuente coherente. En el caso del registro, esto es evidente, dado que se busca que la onda objeto interfiera con la utilizada como referencia. En el proceso de reconstrucción también debe emplearse una fuente puntual cuasimonocromática ya que el holograma, al actuar como una red de difracción, genera una imagen para cada longitud de onda con un tamaño y posición distintos.

La necesidad de utilizar iluminación coherente para su reconstrucción es una característica que muchas veces limita el empleo de hologramas para distintos propósitos. Por ejemplo, todos los hologramas comerciales no podrían existir si este requerimiento fuese siempre necesario.

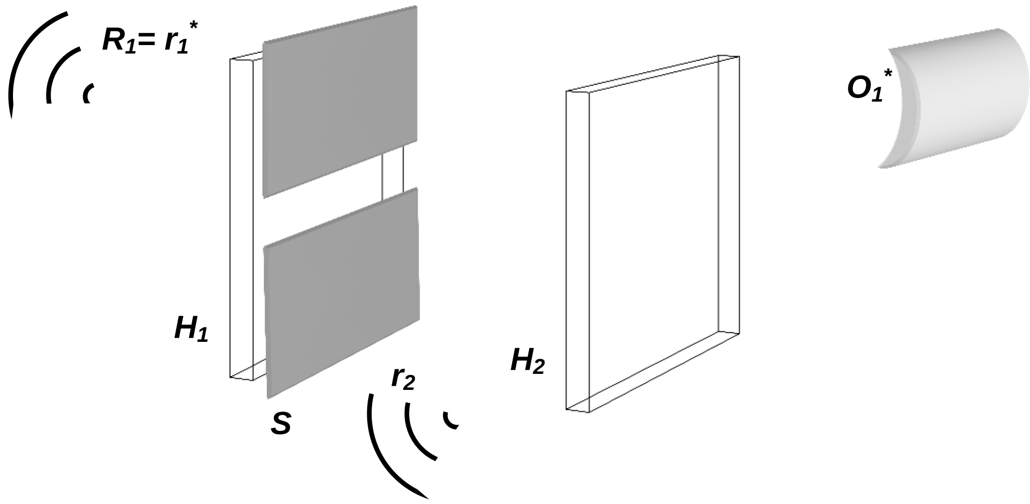
En el año 1969 Benton ideó un tipo de holograma que permitía evitar este requisito [S.A.Benton *Hologram reconstruction with extended light sources*. JOSA **59**, 1545 (1969)]. En un principio se llamaron *white light transmisión holograms* y luego fueron conocidos como *rainbow* ya que la imagen que produce aparece coloreada como el arco iris por la descomposición de la luz empleada para su iluminación.

Veremos entonces el procedimiento para realizar este tipo de hologramas.

En una primera etapa se registra un holograma tipo Fresnel, tal como los que hemos estudiado. Llamaremos a este holograma H_1, cuya síntesis se esquematiza en la siguiente figura



La reconstrucción de este holograma se realiza de una forma particular. Para ello se utiliza un haz convergente tal que $R_1 = r_1^*$ de forma tal que permita obtener una imagen real proporcional a O^*. Por otra parte, y empleando la propiedad que poseen los hologramas de objetos difusores, de poder reconstruir una imagen a partir de una porción del mismo, se pone H_1 en contacto con una ranura S tal como se muestra en la figura



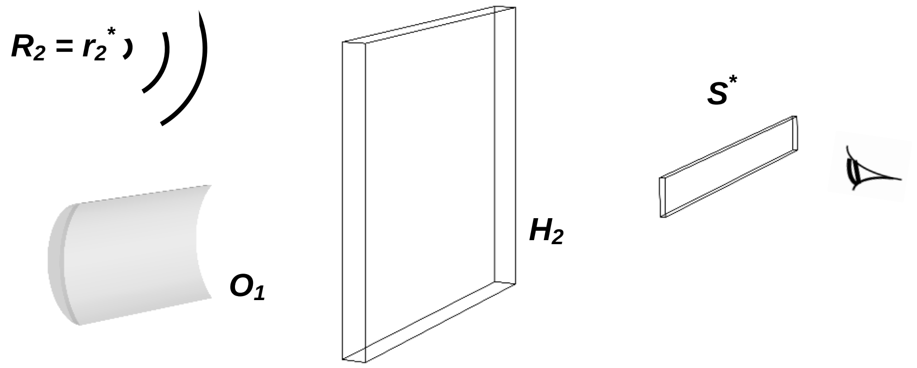
Simultáneamente a la reconstrucción de H_1 se registra un segundo holograma H_2 de forma tal que para este último el objeto compuesto a registrar sea el conjunto $O_2 = O_1^* + S$. Se utiliza como haz de referencia un haz r_2 convergente sobre la placa. El factor de transmisión de este segundo holograma, bajo las hipótesis de linealidad que siempre hacemos, será

$$t_2\left(x, y\right) = t_b + \beta'\left(\left|O_1^* + S\right|^2 + r_2^*\left(O_1^* + S\right) + r_2\left(O_1 + S^*\right)\right)$$

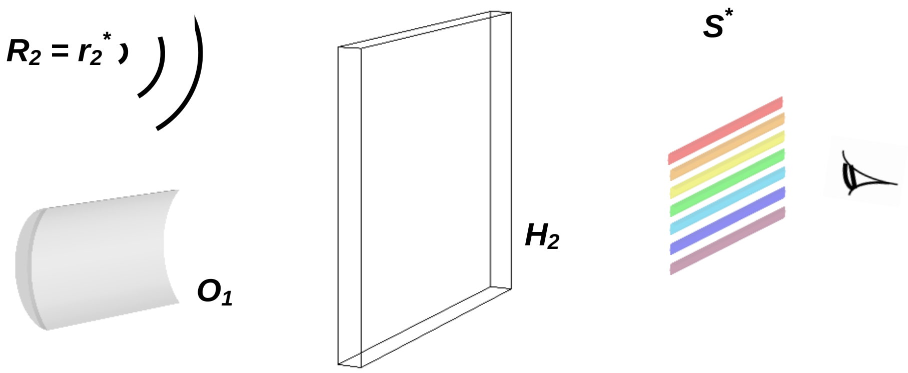
Luego si se efectúa una reconstrucción de H_2 con un haz $R_2 = r_2^*$, obtenemos

$$E \propto R_2 t_2\left(x,y\right) = r_2^*\left(t_b + \beta'\left|O_1^* + S\right|^2\right) + r_2^* r_2^* \beta'\left(O_1^* + S\right) + \beta'\left|r_2\right|^2\left(O_1 + S^*\right)$$

Vemos que en el último término aparece la onda objeto original O_1. Basta seguir la deducción de las ecuaciones para las posiciones de las imágenes anteriormente realizada, para concluir que de acuerdo a la distribución empleada O_1 será una imagen virtual y S^* una imagen real. De este modo tenemos el siguiente esquema de reconstrucción

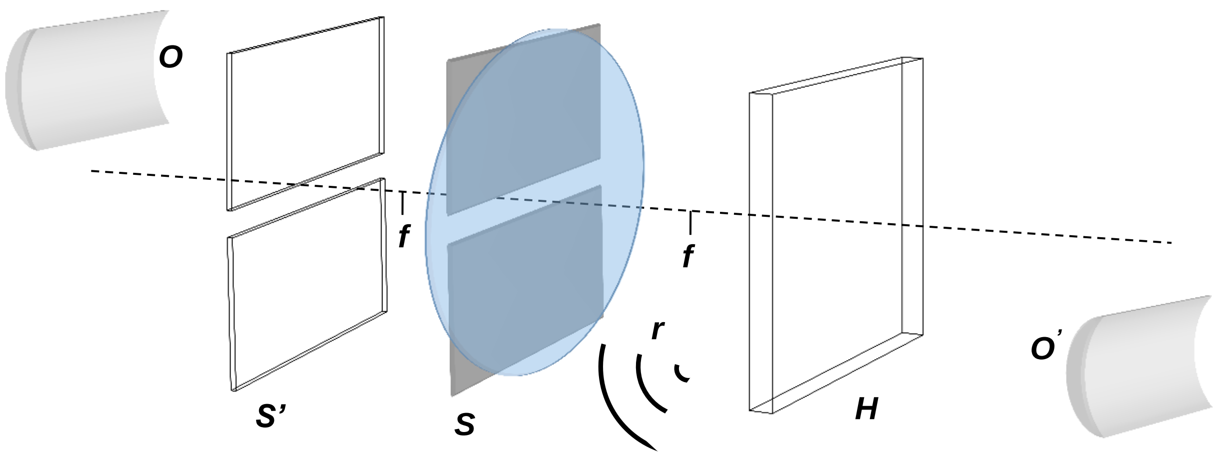


Así, si la reconstrucción se efectúa con una sola longitud de onda, el observador debe "espiar "a través de la ranura S^* para ver la imagen virtual O_1. Dado que la ranura es muy fina, al mirar a través de ella sólo veremos una porción del objeto. Si en vez de iluminar H_2 con una luz monocromática se emplea una fuente de luz blanca, se verá un continuo de ranuras, cada una con una posición caracterizada por la longitud de onda con la que se ve coloreada y de acuerdo a esto se verá una porción del objeto O_1 de ese color.



Lo que realmente se ve al observar este tipo de hologramas es el objeto coloreado con una gama de colores que van del violeta al rojo. Existe una pequeña deformación dada por los cambios de aumento, además se ha perdido el paralaje vertical, pero dado que los ojos perciben mucho más el paralaje horizontal que el vertical (ya que uno está al lado del otro y no debajo) la sensación de tridimensionalidad no se pierde.
Los hologramas comerciales son de este tipo sólo que se ven por reflexión porque están espejados del lado de atrás. Cuando uno diseña la configuración para realizar el holograma debe considerar cuál será la posición conveniente del haz de reconstrucción final.

Existen algunas variantes de esta técnica. Una de ellas es la que permite realizar el holograma rainbow en un solo paso. Conceptualmente es el mismo fenómeno el que se utiliza, sólo cambia la disposición experimental de registro.
Para ello se utiliza una lente convergente ubicada de forma tal que produzca una imagen virtual de la ranura y una real del objeto



El objeto compuesto para el holograma será *O'* (imagen real de *O*) y *S'* (imagen virtual de *S*). Ahora se está aproximadamente en la situación analizada anteriormente. Cabe destacar que en esta arquitectura la reconstrucción origina una imagen pseudoscópica del objeto (profundidad invertida)

Holograma de fase

En los cálculos que hemos realizado para los distintos tipos de hologramas siempre supusimos que la transmisión de los mismos era descripta por una función real $t(x,y) \cong t_b + \beta'\left(I_o(x,y) - I_b\right)$. Estos hologramas actúan como elementos de amplitud según lo que hemos visto en materiales de registro. Esto es, si en general el factor de transmisión de un elemento óptico puede escribirse como $t(x,y)=|t(x,y)|\exp\left(i\phi(x,y)\right)$, los hologramas tratados hasta ahora difractaban el haz de reconstrucción y retenían la información de amplitud y fase del objeto mediante variaciones de $|t(x,y)|$. Veremos luego que estos hologramas tienen muy baja eficiencia, por este motivo se busca realizar registros en materiales donde la difracción venga dada por cambios del índice de refracción o modulación de la superficie en el holograma.

Veamos entonces cómo sería un holograma de fase (por ejemplo realizado en material fotográfico blanqueado, foto-termoplásticos, photoresist, etc).

Por simplicidad supondremos $|t(x,y)|=1$. Consideremos que el haz objeto viene dado por $O(x)=O_o\exp(i\varphi_o)$ y que el haz de referencia se expresa como $r(x)=r_o\exp(i\varphi_r)$. Para su reconstrucción el holograma es iluminado con un haz dado por $R(x)$.

Llamemos $\phi(x)$ a la modulación en fase debida a la exposición $W(x)=I(x)\mathcal{T}$ que produce un cambio en el índice de refracción $n(x)$ o en la altura $h(x)$. Vamos a suponer que existe una relación lineal tal que $\phi(x)\propto I(x)$. Luego si la intensidad sobre la placa es

$$I(x)=|O(x)+r(x)|^2=O_o^2+r_o^2+2O_or_o\cos(\varphi_o-\varphi_r)$$

la función transmisión de la misma puede escribirse como

$$\begin{aligned}t(x)&=\exp\left(\propto i\,O_o^2\right)\exp\left(\propto i\,r_o^2\right)\exp\left(\propto i\,2O_or_o\cos(\varphi_o-\varphi_r)\right)=\\&=K\exp\left(ia\cos\theta\right)=K\left[\cos\left(a\cos\theta\right)+i\sin\left(a\cos\theta\right)\right]\end{aligned}$$

donde tomamos

$$k\equiv\exp\left(\propto i\,O_o^2\right)\exp\left(\propto i\,r_o^2\right);\quad a\equiv\propto 2O_or_o\,;\quad \theta\equiv\varphi_o-\varphi_r$$

Expandiendo en funciones de Bessel tenemos que

$$\cos\left(a\cos\theta\right)=J_o\left(a\right)+2\sum_{n=1}^{\infty}\left(-1\right)^n J_{2n}\left(a\right)\cos\left(2n\theta\right)$$

$$\sin\left(a\cos\theta\right)=2\sum_{n=0}^{\infty}\left(-1\right)^n J_{2n+1}\left(a\right)\cos\left(\left(2n+1\right)\theta\right)$$

donde $J_n\left(a\right)$ es la función de Bessel de primera especie y orden n. Luego podemos expresar la transmisión como

$$t\left(x\right)=K\left\{J_o\left(a\right)+2\sum_{n=1}^{\infty}\left(-1\right)^n J_{2n}\left(a\right)\cos\left(2n\theta\right)+i2\sum_{n=0}^{\infty}\left(-1\right)^n J_{2n+1}\left(a\right)\cos\left(\left(2n+1\right)\theta\right)\right\}$$

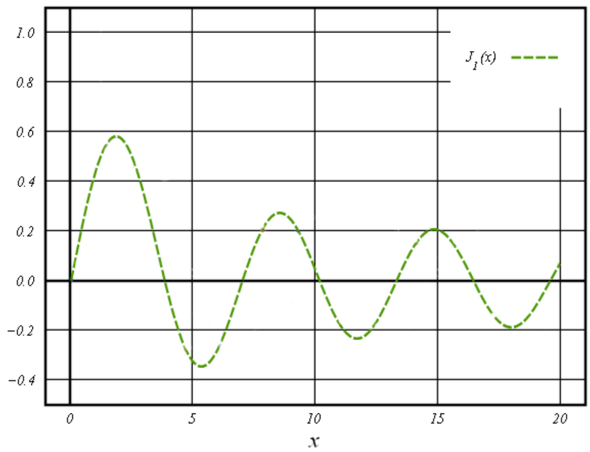
En esta expresión cada $J_n\left(a\right)$ representa a la amplitud del n-ésimo orden difractado. Como vemos no sólo se hallarán presentes los órdenes ± 1 como en los hologramas de amplitud pero esto era esperable si recordamos qué sucedía con una red sinusoidal de amplitud y con una de fase. Si analizamos la expresión para $t\left(x\right)$ veremos que el término que da lugar a las imágenes de interés es el correspondiente a *n*=0 ya que

$$K\left[i2J_1\left(a\right)\cos\theta\right]=iK\,J_1\left(a\right)\left[\exp\left[i\left(\varphi_o-\varphi_r\right)\right]+\exp\left[-i\left(\varphi_o-\varphi_r\right)\right]\right]=$$
$$=\exp\left[i\left(\propto O_o^2+\propto r_o^2+\frac{\pi}{2}\right)\right]J_1\left(\propto 2O_or_o\right)\left[\exp\left[i\left(\varphi_o-\varphi_r\right)\right]+\exp\left[-i\left(\varphi_o-\varphi_r\right)\right]\right]$$

Un término da lugar a la imagen primaria y otro a la imagen conjugada. Si iluminamos en la reconstrucción con un haz $R\left(x\right)=r\left(x\right)$ la imagen primaria será

$$P=r_o\exp\left[i\left(\propto O_o^2+\propto r_o^2+\frac{\pi}{2}\right)\right]J_1\left(\propto 2O_or_o\right)\exp\left(i\varphi_o\right)$$

Teniendo en cuenta la propiedad que establece que $J_1\left(\propto 2O_or_o\right)\cong\propto O_or_o$ si $\propto 2O_or_o$ es pequeño entonces



$$P = \propto r_o^2 \exp\left[i\left(\propto O_o^2 + \propto r_o^2 + \frac{\pi}{2}\right)\right] O_o \exp\left(i\varphi_o\right)$$

Vemos que a menos de una amplitud constante y una fase global sin importancia se reconstruye la onda objeto.

Eficiencia de hologramas planos

Calculemos a continuación la eficiencia de difracción de los hologramas estudiados, esto es, qué proporción de la intensidad de la luz incidente para reconstruir el holograma es realmente dirigida a la imagen de interés. Veamos primero el caso de un holograma de amplitud.

- Holograma de amplitud $\left(t(x,y) = \left|t(x,y)\right|\right)$

  Vamos a suponer que se registra el diagrama de interferencia más simple, esto es, una red cosenoidal. Luego la transmisión será $\left|t(x)\right| = A + B\cos\left(2\pi \frac{x}{d}\right)$ donde *A* es la transmitancia promedio, *B* es la modulación y *d* es el período. Dado que $0 \leq \left|t(x)\right| \leq 1$ y asumiendo, tal como lo hacemos siempre, que la amplitud de la onda difractada varía linealmente con la modulación, la máxima eficiencia se logrará con

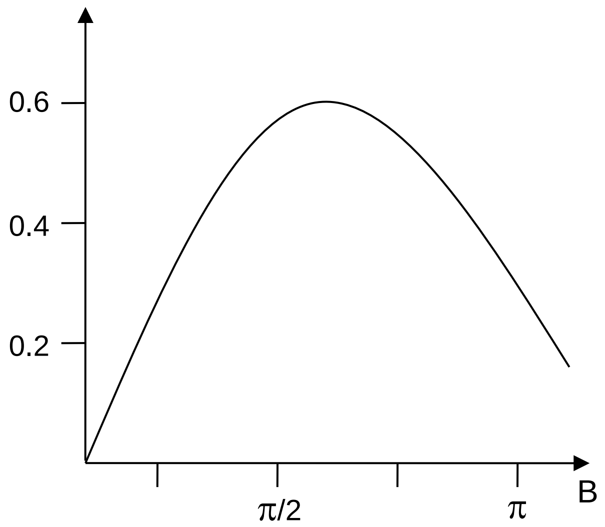
$$\left|t(x)\right| = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi \frac{x}{d}\right) = \frac{1}{2} + \frac{1}{4}\exp\left(i2\pi \frac{x}{d}\right) + \frac{1}{4}\exp\left(-i2\pi \frac{x}{d}\right)$$

  Así la máxima amplitud en un orden difractado será 0.25 de la amplitud incidente, luego la eficiencia de difracción será $(0.25)^2$ ~ 6%

- Holograma de fase $\left(t(x,y) = \exp\left[i\phi(x,y)\right]\right)$

  Nuevamente tomamos el diagrama de interferencia más simple, esto es, $\phi(x) = A + B\cos\left(2\pi \frac{x}{d}\right)$ de manera que $t(x) = \exp[iA]\exp\left[iB\cos\left(2\pi \frac{x}{d}\right)\right]$

  El factor de fase constante puede dejarse de lado. La otra exponencial, tal como hemos visto en hologramas de fase, conduce a las funciones $J_n(B)$ y la amplitud de la onda imagen de interés es la $J_1(B)$. Si graficamos la amplitud difractada como función de la modulación de fase *B* obtenemos la siguiente figura.



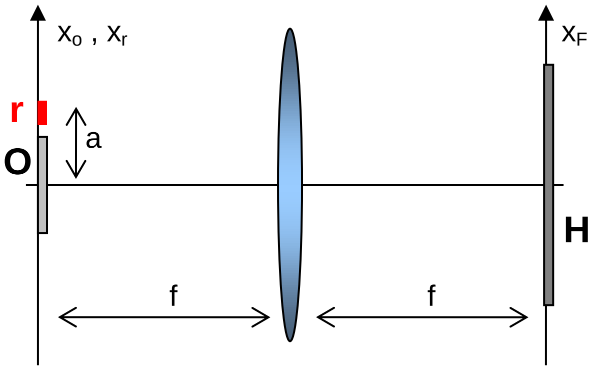
Luego la máxima eficiencia de difracción para un holograma plano de fase es $J_1^2(B) = 0.339$ o sea ronda el 34%. De este resultado vemos porqué interesan los medios de fase como materiales de registro.

FACTORES QUE AFECTAN LA RESOLUCIÓN DE UNA IMAGEN HOLOGRÁFICA

Existen varios factores que pueden limitar la resolución de una imagen holográfica, por ejemplo: el tamaño y ancho de banda de las fuentes empleadas para el registro y la reconstrucción, la resolución y tamaño del medio de registro, las aberraciones en el frente de onda introducidas por la disposición geométrica empleada, speckle, etc.
Haremos un estudio muy superficial de algunos de estos inconvenientes, analizándolos separadamente, aunque es evidente que en la práctica se pueden presentar en forma conjunta.

Efecto del tamaño finito de las fuentes y del holograma

Para estudiar este efecto es conveniente, por simplicidad, analizar el caso de un holograma tipo Fourier considerando además que trabajamos en dos dimensiones.
Hagamos un esquema del montaje empleado para el registro del mismo



Sobre el holograma llegan las ondas[1]

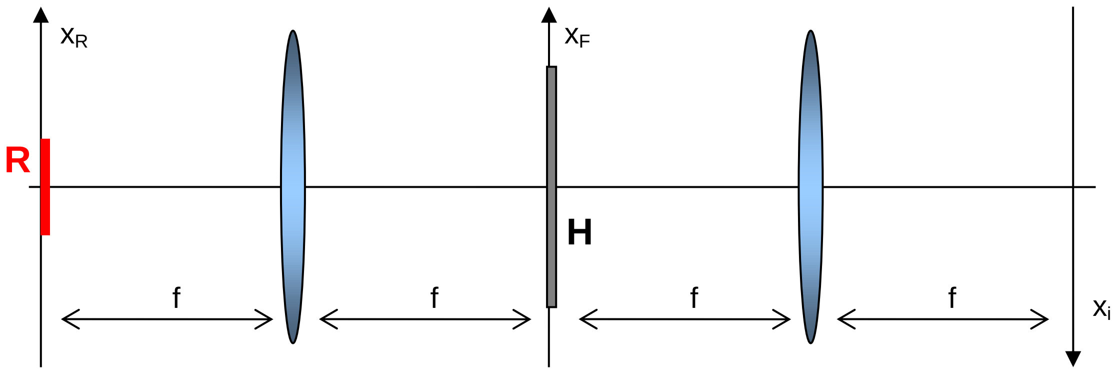
$$O(x_F)=\frac{1}{\sqrt{-i\lambda f}}\mathbb{F}_{\frac{x_F}{\lambda f}}\left[O(x_o)\right]\propto\int_{-\infty}^{\infty}O(x_o)\exp\left[\frac{-i2\pi}{\lambda f}x_o x_F\right]dx_o$$

$$r(x_F)=\frac{1}{\sqrt{-i\lambda f}}\mathbb{F}_{\frac{x_F}{\lambda f}}\left[r(x_r)\right]\propto\int_{-\infty}^{\infty}r(x_r)\exp\left[\frac{-i2\pi}{\lambda f}x_r x_F\right]dx_r$$

Luego la distribución de intensidad viene dada por

$$I(x_F)=\left|r(x_F)\right|^2+\left|O(x_F)\right|^2+r(x_F)O^*(x_F)+r^*(x_F)O(x_F)$$

Una vez registrado el holograma la reconstrucción se realiza con una fuente cuya extensión es distinta de cero.



El haz de reconstrucción viene dado por

[1] Se debe prestar atención a que hemos elegido una notación en la que la función y su transformada de Fourier se denominan con la misma letra, se diferencian una de otra a partir de la variable de la que dependen indicando en que plano se está trabajando.

$$R(x_F) = \frac{1}{\sqrt{-i\lambda f}} \mathbb{F}_{\frac{x_F}{\lambda f}}\left[R(x_R)\right] \propto \int_{-\infty}^{\infty} R(x_R) \exp\left[\frac{-i2\pi}{\lambda f} x_R x_F\right] dx_R$$

Luego la imagen primaria correspondiente a O será

$$P(x_i) \propto \frac{1}{\sqrt{-i\lambda f}} \mathbb{F}_{x_i}^{-1}\left[r^*(x_F) O(x_F) R(x_F)\right] \propto \mathbb{F}_{x_i}^{-1}\left[\mathbb{F}_{\frac{x_F}{\lambda f}}^*\left[r(x_r)\right] \mathbb{F}_{\frac{x_F}{\lambda f}}\left[O(x_o)\right] \mathbb{F}_{\frac{x_F}{\lambda f}}\left[R(x_R)\right]\right]$$

$$P(x_i) \propto \int\int\int_{-\infty}^{\infty}\int R(x_R) O(x_o) r^*(x_r) \exp\left[\frac{-i2\pi x_F}{\lambda f}(x_R + x_o - x_r - x_i)\right] dx_F\, dx_r\, dx_R\, dx_o$$

A partir de esta fórmula general analizaremos distintos casos

i) las fuentes que originan r y R son puntuales y el holograma es infinito

Este es el caso totalmente ideal en el que tenemos $r(x_r) = \delta(x_r - a)\,; R(x_R) = \delta(x_R)$ y el holograma es infinitamente extenso, con lo cual

$$P(x_i) \propto \int_{-\infty}^{\infty}\int O(x_o) \exp\left[\frac{-i2\pi x_F}{\lambda f}(x_o - a - x_i)\right] dx_F\, dx_o =$$

$$= \int_{-\infty}^{\infty} O(x_o)\, \delta\big(x_o - (a + x_i)\big) dx_o \propto O(a + x_i)$$

Como vemos se reconstruye exactamente el objeto desplazado a una distancia *a* del origen, igual a la que se encontraba la referencia del centro del objeto.

ii) las fuentes que originan r y R son puntuales y el holograma es finito

Supongamos que la extensión del holograma es $2L_H$. En ese caso para la imagen primaria se obtiene

$$P(x_i) \propto \int_{-\infty}^{\infty}\int O(x_o)\,\mathrm{rect}\left(\frac{x_F}{2L_H}\right)\exp\left[\frac{-i2\pi x_F}{\lambda f}(x_o - a - x_i)\right]dx_F\, dx_o =$$

$$= \int_{-\infty}^{\infty} O(x_o)\,\mathrm{sinc}\left[\frac{2L_H}{\lambda f}(x_o - a - x_i)\right]dx_o$$

Notemos que si L_H tiende a infinito la función sinc tiende a la función delta y estamos en el caso anterior. En caso contrario, tenemos una convolución del objeto con la transformada de Fourier de la forma del holograma [para ver que es una convolución recordemos que al ser la función sinc par podemos cambiar en el argumento $(x_o - a - x_i)$ por $(-x_o + a + x_i)$].

Así cada punto de la imagen correspondiente a un punto objeto no será exactamente un punto sino la transformada de la forma del holograma, por ejemplo en este caso un sinc^2. El ancho característico de esta función (el ancho de su máximo principal) vendrá dado por $\Delta x_i \cong \frac{\lambda f}{L_H}$, luego habrá frecuencias espaciales en el objeto que no serán reproducidas en la imagen dado que existirá una frecuencia de corte $\frac{1}{\Delta x_i} \cong \frac{L_H}{\lambda f}$ (cabe aclarar que acá estamos suponiendo que la lente no es un factor limitante y que la falta de resolución se debe solamente al tamaño finito del holograma). Recordemos que con las redes de difracción sucedía lo mismo, el tamaño de los órdenes de difracción dependía del la zona iluminada de la red.

Por ejemplo si en este caso el holograma es una placa de 5 cm de lado, *f* = 20 cm y λ = 632.8 nm, entonces la frecuencia de corte es aproximadamente 400 l/mm.

iii) Supongamos ahora que el tamaño del holograma es tal que no introduce difracción y que el objeto es puntual, pero el tamaño de la fuente empleada como referencia y el correspondiente al de reconstrucción vienen descriptos por las funciones $r(x_r) = \mathrm{rect}\left(\frac{x_r}{2L_r}\right)$ y $R(x_R) = \mathrm{rect}\left(\frac{x_R}{2L_R}\right)$ respectivamente.

Tomemos que tanto la fuente usada como referencia así como la empleada para la reconstrucción se hallan en el origen y el objeto puntual está ubicado en $x_o = A$

de modo que $O\left(x_o\right)=\delta\left(x_o-A\right)$. Así el campo correspondiente a la imagen primaria viene dado por:

$$P\left(x_i\right)\propto\int\int\int_{-\infty}^{\infty}\int \operatorname{rect}\left(\frac{x_r}{2L_r}\right)\operatorname{rect}\left(\frac{x_R}{2L_R}\right)\delta\left(x_o-A\right)\exp\left[\frac{-i2\pi x_F}{\lambda f}\left(x_R+x_o-x_r-x_i\right)\right]dx_F\, dx_r\, dx_R\, dx_o=$$

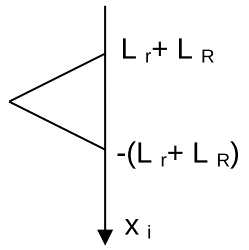
$$=\int\int_{-\infty}^{\infty}\int \operatorname{rect}\left(\frac{x_r}{2L_r}\right)\operatorname{rect}\left(\frac{x_R}{2L_R}\right)\exp\left[\frac{-i2\pi x_F}{\lambda f}\left(x_R+A-x_r-x_i\right)\right]dx_F\, dx_r\, dx_R=$$

$$=\int\int_{-\infty}^{\infty} \operatorname{rect}\left(\frac{x_r}{2L_r}\right)\operatorname{rect}\left(\frac{x_R}{2L_R}\right)\delta\left(x_r-\left(x_R+A-x_i\right)\right)dx_r\, dx_R=$$

$$=\int_{-\infty}^{\infty} \operatorname{rect}\left(\frac{x_R+A-x_i}{2L_r}\right)\operatorname{rect}\left(\frac{x_R}{2L_R}\right)dx_R$$

Dado que las dos funciones son reales, esto es una correlación, además al ser pares coincide con la convolución. Así cada punto imagen se hallará ensanchado y vendrá descripto por una función triangulo

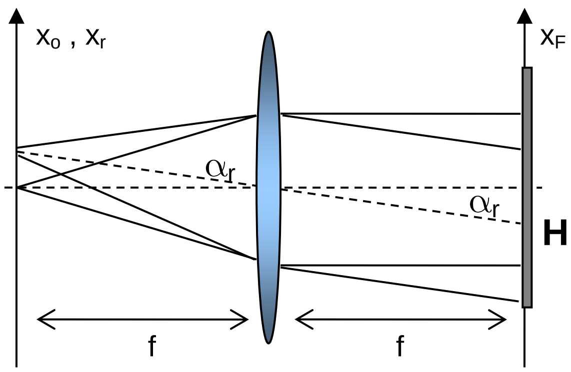
La frecuencia de corte en este caso será $\frac{1}{\Delta x_i}\cong\frac{1}{2\left(L_r+L_R\right)}$



Efecto del ancho de banda de las fuentes

Podemos estimar la relación resolución-ancho de banda considerando al holograma como una red de difracción simple.

Supongamos que dos ondas planas inciden sobre la placa formando entre sí un ángulo α_r. Por ejemplo en el caso de un holograma de Fourier el esquema sería el siguiente



Sobre la placa H tendremos una transmitancia que varía aproximadamente en forma sinusoidal con frecuencia $f_H = \dfrac{\sin\alpha_r}{\lambda_1}$. Vamos a tomar que $\bar{\lambda}_1$ es la longitud de onda media de registro y $\Delta\lambda_1$ es el ancho de banda. Con lo cual el rango de longitudes de onda empleadas está en el rango $\bar{\lambda}_1 - \dfrac{\Delta\lambda_1}{2} \leq \lambda_1 \leq \bar{\lambda}_1 + \dfrac{\Delta\lambda_1}{2}$ y la dispersión en frecuencias espaciales viene dada por $\Delta f_H = \left|\dfrac{-\sin\alpha_r}{\bar{\lambda}_1^2}\right|\Delta\lambda_1$.

Para reconstruir el holograma empleamos una onda plana incidente con un ángulo $\alpha_R = \alpha_r$ y longitud de onda λ_2 tal que $\bar{\lambda}_2 - \dfrac{\Delta\lambda_2}{2} \leq \lambda_2 \leq \bar{\lambda}_2 + \dfrac{\Delta\lambda_2}{2}$, donde $\bar{\lambda}_2$ es la longitud de onda media de reconstrucción y $\Delta\lambda_2$ es el ancho de banda.

El ángulo de difracción α_i de la imagen primaria vendrá dado por la ecuación de la red, esto es $\sin\alpha_i - \sin\alpha_R = \lambda_2 f_H$. Debido al ancho de banda de la fuente de registro y de reconstrucción habrá una dispersión de ángulos dada por

$$\Delta\left(\sin\alpha_i\right) = \left(\sin\alpha_i\right)_{\max} - \left(\sin\alpha_i\right)_{\min}$$

donde

$$\left(\sin\alpha_i\right)_{\max} = \lambda_{2\max} f_{H\max} = \left(\bar{\lambda}_2 + \frac{\Delta\lambda_2}{2}\right)\left(\frac{\sin\alpha_r}{\bar{\lambda}_1} + \frac{\sin\alpha_r}{2\bar{\lambda}_1^2}\Delta\lambda_1\right)$$

$$\left(\sin\alpha_i\right)_{\min} = \lambda_{2\min} f_{H\min} = \left(\bar{\lambda}_2 - \frac{\Delta\lambda_2}{2}\right)\left(\frac{\sin\alpha_r}{\bar{\lambda}_1} - \frac{\sin\alpha_r}{2\bar{\lambda}_1^2}\Delta\lambda_1\right)$$

Luego

$$\Delta\left(\sin\alpha_i\right) = \frac{\bar{\lambda}_2}{\bar{\lambda}_1}\left(\frac{\Delta\lambda_1}{\bar{\lambda}_1} + \frac{\Delta\lambda_2}{\bar{\lambda}_2}\right)\sin\alpha_r$$

Ahora cada punto en la imagen será un manchoncito de tamaño característico $f\,\Delta\left(\sin\alpha_i\right)$.

Efecto de las aberraciones

El tema referente a las aberraciones escapa del tratamiento de este curso ya que es muy compleja su formulación. Un estudio bastante riguroso acerca de aberraciones en hologramas se encuentra en dos trabajos tradicionales [R.Meier, JOSA **56**, 219 (1966); R.Meier, JOSA **57**, 895 (1967)].
Sólo diremos que para tener en cuenta las aberraciones deben considerarse términos superiores en el desarrollo en serie que efectuamos para las distancias, es decir, habría que trabajar con expresiones del tipo

$$r_{10} \cong z + \frac{\left(x_1^2 + y_1^2\right)}{2z} + \frac{\left(x_0^2 + y_0^2\right)}{2z} - \frac{x_1 x_0}{z} - \frac{y_1 y_0}{z} - \frac{\left(x_1 - x_0\right)^4}{8z^3} - \frac{\left(y_1 - y_0\right)^4}{8z^3} \ldots\ldots\ldots\ldots$$

Meier encontró que si se toma $z_o = z_r$ la aberración esférica es nula y además si se cumple que $z_o = \pm z_R$ no existe coma. Por otra parte estableció que en general si se emplea la misma longitud de onda y se trabaja con haces de referencia y reconstrucción planos y conjugados, entonces se anulan las aberraciones primarias: esférica, coma, astigmatismo, curvatura de campo y distorsión.

HOLOGRAMAS DE VOLUMEN

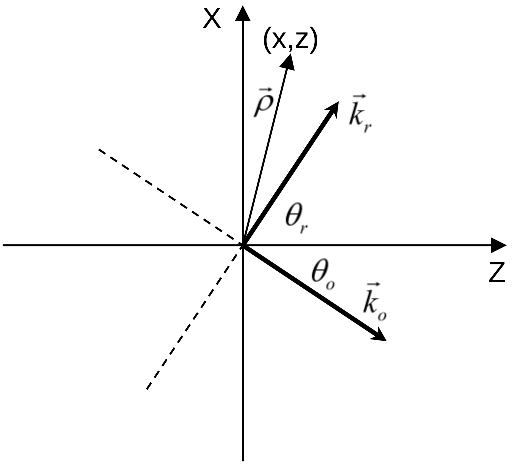
CLASE 19

Anteriormente dijimos que un holograma de volumen es aquel en el cual el espesor del medio de registro es del orden o mayor que el espaciado de las franjas de interferencia registradas. Virtualmente casi todos los hologramas realizados sobre materiales fotográficos son de volumen, esto es, en casi todas las situaciones donde se requiere un preciso análisis del fenómeno debería tenerse en cuenta el espesor del medio de registro. Los hologramas planos en realidad constituyen un caso especial y salvo en algunas situaciones particulares, considerar un holograma como plano sólo es una primera aproximación a la situación real.
Acá desarrollaremos la teoría de redes tridimensionales en forma simplificada ya que la formulación rigurosa consistiría en si misma un curso especial. Un trabajo pionero en el tema es [H.Kogelnik *Coupled wave theory for thick hologram gratings* Bell Syst. Tech. J. **48**, 2909 (1969)]

Franjas de interferencia en tres dimensiones

Consideremos la situación más elemental consistente en dos haces que interfieren, dando como resultado un conjunto de franjas de espaciado y dirección constante.

De acuerdo a la siguiente figura, asumamos que los vectores de onda $\vec{k}$ de los haces que interfieren están en el plano *x-z*.



El sistema de coordenadas está centrado en el medio de registro y todas las direcciones se refieren a las mismas dentro del medio. Las direcciones fuera del medio se relacionan con las interiores mediante la ley de Snell.

Las dos ondas planas que interfieren son la objeto *O* y la de referencia *r* que se propagan formando ángulos θ_o y θ_r , respectivamente, con el eje *z*. El tratamiento está hecho en forma general y no siguiendo estrictamente la figura de modo que los ángulos θ_o y θ_r en cada caso particular deberán reemplazarse teniendo en cuenta la convención de signos siguiente: los ángulos son positivos en sentido antihorario y negativos en sentido horario.

Si $\vec{\rho}$ es el vector posición de un punto (x,z) en el medio y tomamos $\left|\vec{k}_o\right| = \left|\vec{k}_r\right| = k_n = \dfrac{2\pi n}{\lambda}$ donde λ es la longitud de onda en el vacío y n el índice de refracción del medio, entonces en dicho punto los campos serán:

$$O = O_o \exp\left(i\,\vec{k}_o.\vec{\rho}\right) = O_o \exp\left(i\,\varphi_o\right) \quad ; \quad r = r_o \exp\left(i\,\vec{k}_r.\vec{\rho}\right) = r_o \exp\left(i\,\varphi_r\right)$$

con

$$\vec{\rho} = x\,\hat{x} + z\,\hat{z}\,; \quad \vec{k}_o = k_n\left(\sin\theta_o\,\hat{x} + \cos\theta_o\,\hat{z}\right); \quad \vec{k}_r = k_n\left(\sin\theta_r\,\hat{x} + \cos\theta_r\,\hat{z}\right)$$

luego

$$\varphi_o(x,z) = k_n x \sin\theta_o + k_n z \cos\theta_o \; ; \quad \varphi_r(x,z) = k_n x \sin\theta_r + k_n z \cos\theta_r$$

Así la diferencia de fase entre las dos ondas viene dada por

$$\varphi_r(x,z) - \varphi_o(x,z) = k_n \left[x(\sin\theta_r - \sin\theta_o) + z(\cos\theta_r - \cos\theta_o) \right]$$

Tendremos máximos cuando esa diferencia de fase sea igual a $2m\pi$. Por lo tanto resolviendo para x obtenemos una ecuación que define la posición de las franjas, la cual viene dada por

$$x = \left[\frac{\cos\theta_o - \cos\theta_r}{\sin\theta_r - \sin\theta_o} \right] z + \frac{m\lambda_n}{\sin\theta_r - \sin\theta_o}$$

Vemos que consisten en franjas rectas cuya pendiente es el término entre corchetes y $\frac{m\lambda_n}{\sin\theta_r - \sin\theta_o}$ su ordenada al origen. En realidad para obtener exactamente la posición de los máximos deberíamos incluir el desfasaje de las dos ondas en el origen pero, como veremos, una translación de las franjas no afecta los cálculos subsiguientes.

El espaciado de las franjas en la dirección *x* se puede deducir analizando cuál es la variación en *x* cuando *m* cambia en una unidad. Así el período vendrá dado por

$$\Delta x = \frac{\lambda_n}{\sin\theta_r - \sin\theta_o} = d$$

y la frecuencia espacial será

$$f = \frac{1}{d} = \frac{\sin\theta_r - \sin\theta_o}{\lambda_n}$$

Llamemos $\Delta\theta$ al ángulo entre los dos haces, esto es, $\Delta\theta = \theta_r - \theta_o$. Consideremos además que el ángulo α que forman las franjas con el eje *z* puede deducirse de la pendiente de la recta de forma tal que

$$\tan\alpha = \frac{\cos\theta_o - \cos\theta_r}{\sin\theta_r - \sin\theta_o} = \frac{2\sin\left(\frac{\theta_o + \theta_r}{2}\right)\sin\left(\frac{\theta_r - \theta_o}{2}\right)}{2\cos\left(\frac{\theta_o + \theta_r}{2}\right)\sin\left(\frac{\theta_r - \theta_o}{2}\right)} = \tan\left(\frac{\theta_o + \theta_r}{2}\right) \Rightarrow \quad \alpha = \frac{\theta_o + \theta_r}{2}$$

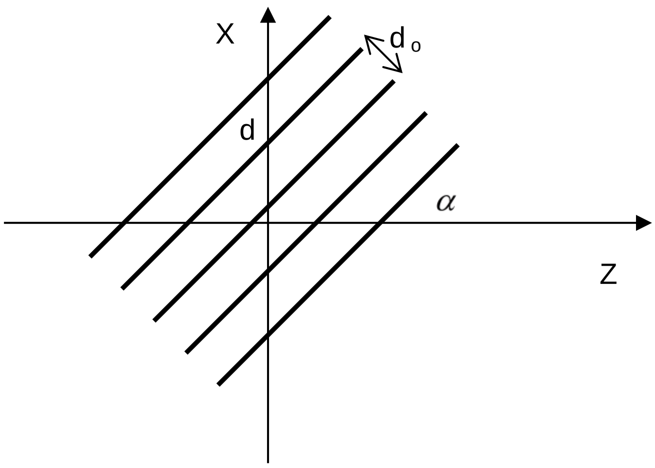
Podemos escribir además la frecuencia de la siguiente manera

$$f = \frac{1}{d} = \frac{\sin\theta_r - \sin\theta_o}{\lambda_n} = \frac{2}{\lambda_n}\cos\left(\frac{\theta_o + \theta_r}{2}\right)\sin\left(\frac{\theta_r - \theta_o}{2}\right) = \frac{2}{\lambda_n}\sin\left(\frac{\Delta\theta}{2}\right)\cos\alpha$$

Cabe destacar que *d* y *f* es el período y la frecuencia en la dirección *x* y por lo tanto son los que se obtendrían en un holograma plano. Sin embargo el período y la frecuencia en la dirección perpendicular a las franjas vienen dados por

$$d_o = d\cos\alpha \quad ; \quad f_o = \frac{2}{\lambda_n}\sin\left(\frac{\Delta\theta}{2}\right)$$

Veamos estas magnitudes en el siguiente gráfico



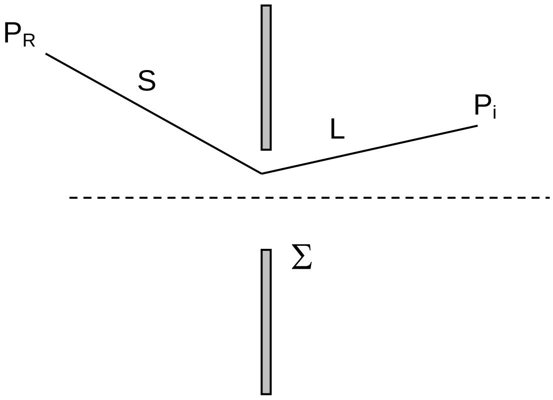
Difracción por una red tridimensional

Ahora que conocemos la orientación de las franjas de interferencia en el medio de registro se busca saber cómo se difracta la luz incidente sobre esta red tridimensional. Para ello consideremos que existe simetría de translación a lo largo del eje *y*. Analicemos primero cuál es la expresión, en la aproximación de Fraunhofer, del campo difractado en el punto (x_i, y_i) por una pantalla Σ ubicada a una distancia *z* del origen.

Hasta ahora cuando calculábamos la difracción de Fraunhofer producida por una abertura, considerábamos que la misma estaba ubicada en el origen de coordenadas. Ahora debemos encontrar una expresión más general donde se tenga en cuenta que la pantalla difractora es encuentre a una distancia *z* del origen.

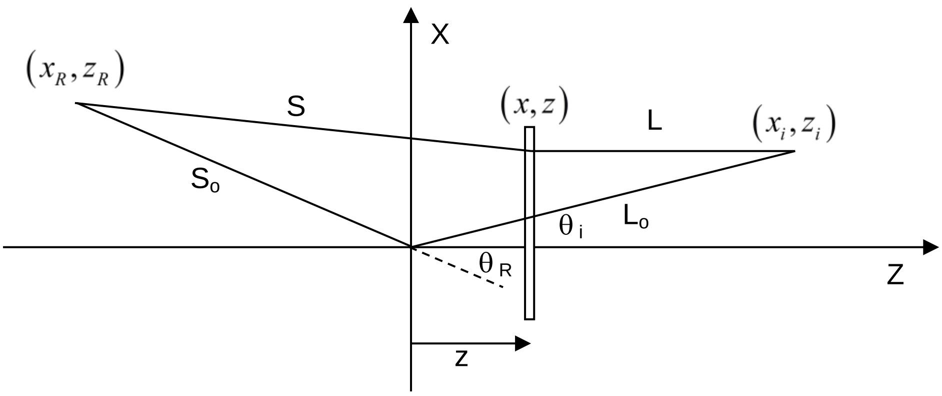
Habíamos visto que, según la teoría de Rayleigh-Sommerfeld, una fuente puntual ubicada a una distancia *S* de la pantalla producía un campo *E* a una distancia *L* de la misma cuya expresión, en una dimensión, venía dada por

$$E\left(P_i\right)=\frac{A}{\sqrt{i\lambda}}\int_{\Sigma}\frac{\exp\left[ik\left(S+L\right)\right]}{\sqrt{SL}}\,q\,d\Sigma$$



donde q es el factor de oblicuidad

Analicemos ahora esta disposición



S es la distancia a la que se halla la fuente con respecto a un punto de la pantalla, *L* es la distancia desde dicho punto hasta el de observación, en tanto S_o y L_o son las distancias al origen, de la fuente y el punto de observación, respectivamente. Luego

$$S^2=\left(x_R-x\right)^2+\left(z_R-z\right)^2\,;\quad S_o^2=x_R^2+z_R^2$$
$$L^2=\left(x_i-x\right)^2+\left(z_i-z\right)^2\,;\quad L_o^2=x_i^2+z_i^2$$

con lo cual tenemos que

$$S^2=S_o^2+x^2+z^2-2x_Rx-2z_Rz\,;\quad L^2=L_o^2+x^2+z^2-2x_ix-2z_iz$$

Dado que las distancias S_o y L_o son grandes en comparación con *x,z* de la pantalla difractora, podemos escribir en forma aproximada

$$S\cong S_o+\frac{x^2}{2S_o}+\frac{z^2}{2S_o}-\frac{x_Rx}{S_o}-\frac{z_Rz}{S_o}.......;\quad L\cong L_o+\frac{x^2}{2L_o}+\frac{z^2}{2L_o}-\frac{x_ix}{L_o}-\frac{z_iz}{L_o}.......$$

Vamos a reemplazar estas expresiones en la integral de Rayleigh-Sommerfeld tomando la aproximación a primer orden en la amplitud y a segundo orden en la fase. Así en la aproximación de Fresnel tenemos

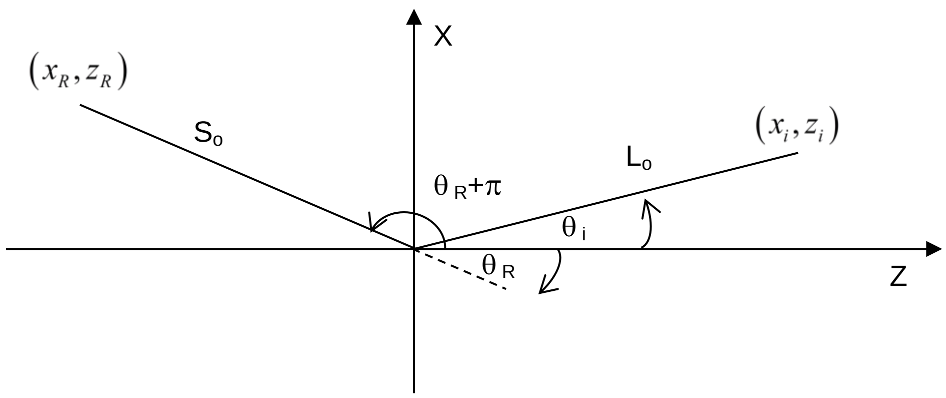
$$E\left(x_i, z_i\right) = \frac{A}{\sqrt{i\lambda S_o L_o}} \exp\left[ik\left(S_o + L_o\right)\right].$$

$$.\int_{\Sigma} \exp\left[-ik\left(\frac{x_R x}{S_o} + \frac{z_R z}{S_o} + \frac{x_i x}{L_o} + \frac{z_i z}{L_o} - \frac{x^2+z^2}{2S_o} - \frac{x^2+z^2}{2L_o}\right)\right] dx$$

La aproximación de Fraunhofer corresponde a tomar $kx^2 << S_o, L_o$; $kz^2 << S_o, L_o$ con lo cual la expresión anterior toma la forma

$$E\left(x_i, z_i\right) = Cte.\int_{\Sigma} \exp\left[-ik\left(\frac{x_R x}{S_o} + \frac{z_R z}{S_o} + \frac{x_i x}{L_o} + \frac{z_i z}{L_o}\right)\right] dx$$

Además, en esta aproximación, puede suponerse que el ángulo que forma S_o con el eje z es aproximadamente igual al que forma S con dicho eje. Análogamente el ángulo que forma L_o con el eje z es aproximadamente igual al que forma L. Luego, de acuerdo a la convención de signos anteriormente tomada y teniendo en cuenta la siguiente figura, podemos escribir



$$\frac{z_R}{S_o} = \cos\left(\theta_R + \pi\right) = \cos\theta_R \cos\pi = -\cos\theta_R \ ; \quad \frac{z_i}{L_o} = \cos\theta_i$$

$$\frac{x_R}{S_o} = \sin\left(\theta_R + \pi\right) = \sin\theta_R \cos\pi = -\sin\theta_R \ ; \quad \frac{x_i}{L_o} = \sin\theta_i$$

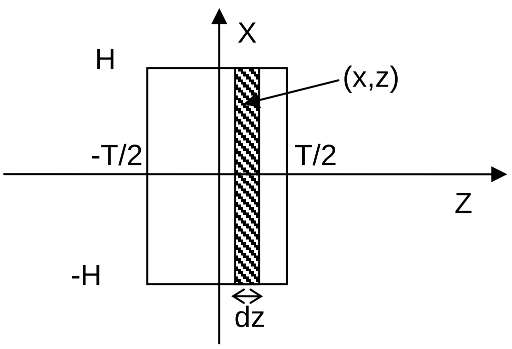
Así el campo difractado en una dirección $\left(\sin\theta_i, \cos\theta_i\right)$, debido a una pantalla difractora plana (de espesor dz) ubicada a una distancia z del origen, viene dado por

$$d\left[E\left(\sin\theta_i,\cos\theta_i\right)\right]=\left[Cte.\int_{\Sigma}\exp\left[-ik\left[\left(\sin\theta_i-\sin\theta_R\right)x+\left(\cos\theta_i-\cos\theta_R\right)z\right]\right]dx\right]dz$$

Para una red tridimensional, esto es una abertura de espesor no despreciable, supondremos que podemos sumar las contribuciones de cada elemento de espesor *dz*. Esto requiere asumir la hipótesis de que el campo difractado por cada elemento es débil, así como que la onda incidente sobre cada abertura elemental es la misma. Supongamos que el medio de registro tiene un espesor *T* y un ancho *2H*, entonces el campo difractado total será

$$E\left(\sin\theta_i,\cos\theta_i\right)=Cte.\int_{-H}^{H}\int_{-T/2}^{T/2}t\left(x,z\right)\exp\left[-ik_{nR}\left(\sin\theta_i-\sin\theta_R\right)x\right]$$
$$.\exp\left[-ik_{nR}\left(\cos\theta_i-\cos\theta_R\right)z\right]dx\,dz$$

donde $t\left(x,z\right)$ es la función transmisión de la red y $k_{nR}=\dfrac{2\pi}{\lambda_{nR}}$, con λ_{nR} la longitud de onda del haz de reconstrucción en el medio de registro.



Veamos ahora qué expresión encontramos para la función $t\left(x,z\right)$. Para ello analicemos primero cuál es la distribución de intensidad que llega a la placa en el registro de la misma. Supongamos que dicha distribución de intensidad se logra a partir de la interferencia de dos ondas planas. Luego podemos escribir la ecuación

$$I\left(x,z\right)=O_o^2+r_o^2+2O_or_o\cos\left(\varphi_r-\varphi_o\right)$$

en función de los resultados hallados anteriormente. Por un lado habíamos visto que

$$\varphi_r\left(x,z\right)-\varphi_o\left(x,z\right)=k_n\left[x\left(\sin\theta_r-\sin\theta_o\right)+z\left(\cos\theta_r-\cos\theta_o\right)\right]$$

Además

$$\tan\alpha=\frac{\cos\theta_o-\cos\theta_r}{\sin\theta_r-\sin\theta_o}\quad\Rightarrow\quad\cos\theta_r-\cos\theta_o=-\left(\sin\theta_r-\sin\theta_o\right)\frac{\sin\alpha}{\cos\alpha}$$

con lo cual es posible escribir

$$\varphi_r\left(x,z\right)-\varphi_o\left(x,z\right)=k_n\left[x\left(\sin\theta_r-\sin\theta_o\right)-z\left(\sin\theta_r-\sin\theta_o\right)\frac{\sin\alpha}{\cos\alpha}\right]=$$
$$=\frac{k_n\left(\sin\theta_r-\sin\theta_o\right)}{\cos\alpha}\left[x\cos\alpha-z\sin\alpha\right]$$

Por otra parte teníamos que

$$f=\frac{\sin\theta_r-\sin\theta_o}{\lambda_n}=\frac{2}{\lambda_n}\sin\left(\frac{\Delta\theta}{2}\right)\cos\alpha=f_o\cos\alpha$$

resultando así

$$\varphi_r\left(x,z\right)-\varphi_o\left(x,z\right)=2\pi\,f_o\left[x\cos\alpha-z\sin\alpha\right]$$

Con lo cual la expresión para la distribución de intensidad toma la forma

$$I\left(x,z\right)=O_o^2+r_o^2+2O_or_o\cos\left(2\pi\,f_o\left[x\cos\alpha-z\sin\alpha\right]\right)$$

Tomemos ahora, como siempre lo hacemos, que la transmisión en amplitud de la placa es proporcional a la intensidad de registro y que $I_b=r_o^2$. Entonces podemos escribir

$$t\left(x,z\right)=t_b+\beta'O_o^2+\beta'O_or_o\exp\left[i2\pi f_o\left(\cos\alpha\,x-\sin\alpha\,z\right)\right]+$$
$$+\beta'O_or_o\exp\left[-i2\pi f_o\left(\cos\alpha\,x-\sin\alpha\,z\right)\right]$$

Ahora que tenemos esta ecuación veamos cuales son las *condiciones de difracción de Bragg*. Para ello reemplacemos la expresión hallada para la función transmisión en la ecuación que expresa el campo difractado.

$$E\left(\sin\theta_i,\cos\theta_i\right)=Cte.\int_{-H}^{H}\int_{-T/2}^{T/2}\left\{\left[t_b+\beta' O_o^2\right]\exp\left[-ik_{nR}\left(\sin\theta_i-\sin\theta_R\right)x\right].\right.$$
$$.\exp\left[-ik_{nR}\left(\cos\theta_i-\cos\theta_R\right)z\right]+\beta' O_o r_o\exp\left[i2\pi f_o\left(\cos\alpha\, x-\sin\alpha\, z\right)\right].$$
$$.\exp\left[-ik_{nR}\left(\sin\theta_i-\sin\theta_R\right)x\right]\exp\left[-ik_{nR}\left(\cos\theta_i-\cos\theta_R\right)z\right]+$$
$$+\beta' O_o r_o\exp\left[-i2\pi f_o\left(\cos\alpha\, x-\sin\alpha\, z\right)\right]\exp\left[-ik_{nR}\left(\sin\theta_i-\sin\theta_R\right)x\right].$$
$$\left..\exp\left[-ik_{nR}\left(\cos\theta_i-\cos\theta_R\right)z\right]\right\}dx\,dz$$

Teniendo en cuenta que $\int_{-A}^{A}\exp\left(-i\,Bx\right)dx=2A\dfrac{\sin AB}{AB}=2A\operatorname{sinc}AB$ y resolviendo obtenemos

$$E\left(\sin\theta_i,\cos\theta_i\right)=Cte.\left[t_b+\beta' O_o^2\right]T\,2H\operatorname{sinc}\left[Hk_{nR}\left(\sin\theta_i-\sin\theta_R\right)\right].$$
$$.\operatorname{sinc}\left[\frac{T}{2}k_{nR}\left(\cos\theta_i-\cos\theta_R\right)\right]+$$
$$+Cte.\beta' O_o r_o\,T\,2H\operatorname{sinc}\left[H\left(k_{nR}\sin\theta_i-k_{nR}\sin\theta_R-2\pi f_o\cos\alpha\right)\right].$$
$$.\operatorname{sinc}\left[\frac{T}{2}\left(k_{nR}\cos\theta_i-k_{nR}\cos\theta_R+2\pi f_o\sin\alpha\right)\right]+$$
$$+Cte.\beta' O_o r_o\,T\,2H\operatorname{sinc}\left[H\left(k_{nR}\sin\theta_i-k_{nR}\sin\theta_R+2\pi f_o\cos\alpha\right)\right].$$
$$.\operatorname{sinc}\left[\frac{T}{2}\left(k_{nR}\cos\theta_i-k_{nR}\cos\theta_R-2\pi f_o\sin\alpha\right)\right]$$

Esta es la solución a primer orden para la difracción por una red tridimensional. Analicemos esta ecuación:

El primer término tiene un máximo cuando

$$\left.\begin{array}{l}\sin\theta_i=\sin\theta_R\\ \cos\theta_i=\cos\theta_R\end{array}\right\}\Rightarrow\theta_R=\theta_i$$

Esto corresponde al orden 0, es decir el haz incidente de reconstrucción es transmitido por el holograma directamente sin desviarse

El segundo término es máximo cuando el argumento de ambas funciones sinc es cero, esto es

$$\sin\theta_i=\sin\theta_R+\frac{2\pi f_o}{k_{nR}}\cos\alpha$$

$$\cos\theta_i=\cos\theta_R-\frac{2\pi f_o}{k_{nR}}\sin\alpha$$

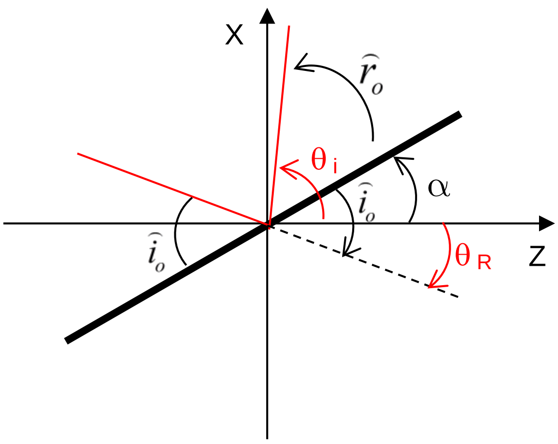
Para analizar una de estas dos condiciones escribamos la ecuación de difracción de la red que debería cumplirse, por ejemplo, para el orden 1 difractado

$$d\left(\sin\theta_i - \sin\theta_R\right) = \lambda_{nR}$$

Luego

$$\sin\theta_i = \lambda_{nR} f + \sin\theta_R = \frac{2\pi f_o}{k_{nR}}\cos\alpha + \sin\theta_R$$

Vemos que reobtenemos la primera de las ecuaciones, vale decir que una de las condiciones para que el segundo término del campo difractado por la red tridimensional sea máximo, es justamente que se cumpla la ecuación de la red.
Para analizar la otra condición estudiemos cómo se reflejaría la onda incidente sobre un plano coincidente con una de las franjas de interferencia. Sea $\hat{i}_o$ el ángulo de incidencia y $\hat{r}_o$ el de reflexión



$$\hat{i}_o = \theta_R - \alpha \Rightarrow \cos\hat{i}_o = \cos\theta_R\cos\alpha + \sin\theta_R\sin\alpha$$

$$\hat{r}_o = \theta_i - \alpha \Rightarrow \cos\hat{r}_o = \cos\theta_i\cos\alpha + \sin\theta_i\sin\alpha$$

Dado que al haber una reflexión $\hat{i}_o = \hat{r}_o$

$$\cos\hat{i}_o = \cos\hat{r}_o \Rightarrow \cos\theta_R\cos\alpha + \sin\theta_R\sin\alpha = \cos\theta_i\cos\alpha + \sin\theta_i\sin\alpha$$

$$\cos\theta_R - \cos\theta_i = \left(\sin\theta_i - \sin\theta_R\right)\frac{\sin\alpha}{\cos\alpha}$$

pero habíamos visto que

$$\sin\theta_i - \sin\theta_R = \frac{2\pi f_o}{k_{nR}}\cos\alpha \quad \Rightarrow \quad \cos\theta_R - \cos\theta_i = \frac{2\pi f_o}{k_{nR}}\sin\alpha$$

que es justamente la otra condición de máximo.

En consecuencia llegamos a que existe un máximo para el primer orden difractado, en la dirección $\left(\sin\theta_i, \cos\theta_i\right)$ cuando se cumplen las condiciones de Bragg, esto es, cuando se cumplen *simultáneamente* la ecuación de la red y la de reflexión en un plano que, al igual que las franjas de interferencia, forma un ángulo α con el eje Z.

Un razonamiento análogo se puede realizar para el tercer término del campo difractado que corresponde al orden -1.

Notemos que es imposible obtener simultáneamente un máximo en los órdenes 1 y -1 ya que el primero es máximo cuando

$$\sin\theta_i - \sin\theta_R = \frac{2\pi f_o}{k_{nR}}\cos\alpha \quad ; \quad -\left(\cos\theta_i - \cos\theta_R\right) = \frac{2\pi f_o}{k_{nR}}\sin\alpha$$

y el segundo cuando

$$-\left(\sin\theta_i - \sin\theta_R\right) = \frac{2\pi f_o}{k_{nR}}\cos\alpha \quad ; \quad \cos\theta_i - \cos\theta_R = \frac{2\pi f_o}{k_{nR}}\sin\alpha$$

Veamos cómo podemos poner estas expresiones en función de los haces de registro. Comencemos por el orden 1

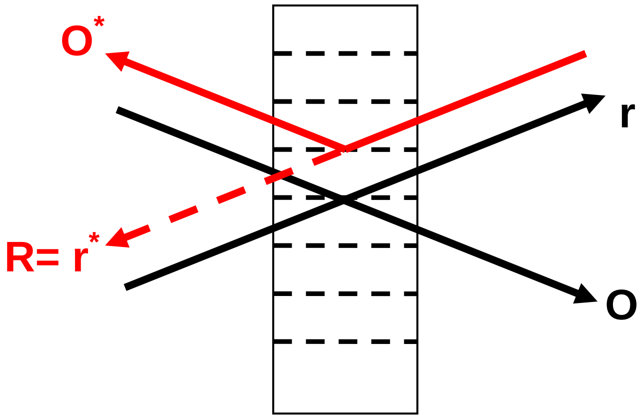
$$\sin\theta_i - \sin\theta_R = \lambda_{nR} f_o \cos\alpha = \lambda_{nR} f = \frac{\lambda_{nR}}{\lambda_n}\left(\sin\theta_r - \sin\theta_o\right)$$

además

$$\cos\theta_i - \cos\theta_R = -\lambda_{nR} f_o \sin\alpha = -\lambda_{nR} f \tan\alpha = \frac{\lambda_{nR}}{\lambda_n}\left(\cos\theta_r - \cos\theta_o\right)$$

dónde utilizamos los resultados hallados en la página 249.

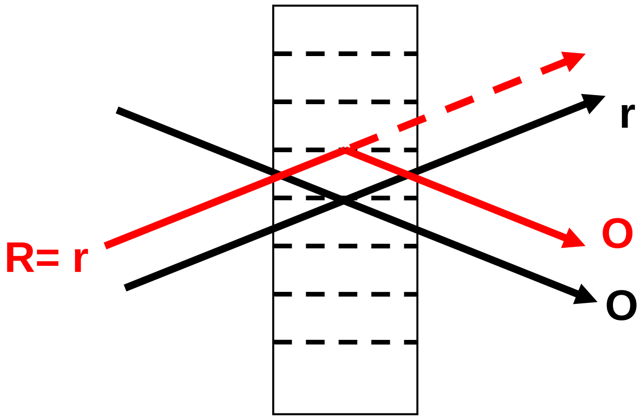
De estas ecuaciones vemos que si la longitud de onda de registro es igual a la de reconstrucción y el haz de reconstrucción se toma con igual dirección pero sentido inverso al de referencia, esto es $\cos\theta_r = -\cos\theta_R$; $\sin\theta_r = -\sin\theta_R$ tenemos que $\sin\theta_i = -\sin\theta_o$; $\cos\theta_i = -\cos\theta_o$. Luego la imagen se difracta en sentido reverso al del haz objeto pero a través del mismo paso, esto es, se obtiene el objeto conjugado. Veamos esto en el siguiente esquema



Análogamente analizando el orden -1 podemos escribir las ecuaciones

$$\left(\sin\theta_i - \sin\theta_R\right) = -\frac{\lambda_{nR}}{\lambda_n}\left(\sin\theta_r - \sin\theta_o\right); \quad \left(\cos\theta_i - \cos\theta_R\right) = -\frac{\lambda_{nR}}{\lambda_n}\left(\cos\theta_r - \cos\theta_o\right)$$

Ahora si la longitud de onda del haz de registro es igual a la del de reconstrucción y ambos haces tienen igual dirección, esto es, $\cos\theta_r = \cos\theta_R\,;\sin\theta_r = \sin\theta_R$, entonces $\sin\theta_i = \sin\theta_o\,;\cos\theta_i = \cos\theta_o$. Luego el haz imagen coincide con el haz objeto. Veamos esta situación en el siguiente dibujo



Sensibilidad a la orientación y a la longitud de onda de una red tridimensional

En realidad las consideraciones que hemos hecho hasta ahora se aplican a hologramas muy gruesos, donde el espaciado de las franjas es mucho menor que el espesor. Para la mayoría de los hologramas esto no es estrictamente cierto y existirá un rango de direcciones en las que puede incidir el haz de reconstrucción y también habrá una cierta tolerancia a un cambio en la longitud de onda. Vamos a estudiar entonces estos fenómenos

Sensibilidad a la orientación

Para determinar el rango angular dentro del cual el holograma puede reconstruirse analicemos el tercer término de la expresión de $E\left(\sin\theta_i,\cos\theta_i\right)$ que da lugar a la imagen primaria

$$P\left(\sin\theta_i,\cos\theta_i\right)=Cte.\beta' O_o r_o\, T\, 2H \,\mathrm{sinc}\left[H\left(k_{nR}\sin\theta_i-k_{nR}\sin\theta_R+2\pi f_o\cos\alpha\right)\right].$$
$$.\mathrm{sinc}\left[\frac{T}{2}\left(k_{nR}\cos\theta_i-k_{nR}\cos\theta_R-2\pi f_o\sin\alpha\right)\right]$$

Recordemos que *2H* era la dimensión del holograma y *T* su espesor. La primera función sinc es la que está relacionada con la ecuación de la red y, como sabemos, esta se cumple siempre, independientemente del grosor del holograma. Esto se confirma por el hecho de que el argumento de dicha función no depende de *T*. Llamemos $I_o=\left[Cte.\beta' O_o r_o\, T\, 2H\right]^2$ a la intensidad difractada cuando se cumplen las condiciones de Bragg, esto es, cuando las dos funciones sinc valen 1.
Sea entonces

$$I=I_o\,\mathrm{sinc}^2\left[\frac{T}{2}\left(k_{nR}\cos\theta_i-k_{nR}\cos\theta_R-2\pi f_o\sin\alpha\right)\right]$$

la intensidad difractada cuando se cumple la ecuación de la red pero no la de reflexión en los planos de Bragg. La intensidad de la onda difractada valdrá cero cuando el argumento de la función sinc sea igual a π y esto nos indicará cuál es la rotación angular tolerada por el holograma para poder reconstruir el frente de ondas. Vamos a escribir dicho argumento en función de los haces de registro, para ello tengamos en cuenta que

$$f_o=\frac{2}{\lambda_n}\sin\left(\frac{\Delta\theta}{2}\right);\quad \cos\theta_o-\cos\theta_r=2\sin\alpha\sin\left(\frac{\Delta\theta}{2}\right)$$

luego $2\pi f_o\sin\alpha=\left(\cos\theta_o-\cos\theta_r\right)\dfrac{2\pi}{\lambda_n}$ con lo cual el argumento toma la forma

$$\frac{T}{2}\left[\frac{2\pi}{\lambda_{nR}}\left(\cos\theta_i-\cos\theta_R\right)+\frac{2\pi}{\lambda_n}\left(\cos\theta_r-\cos\theta_o\right)\right]$$

Vamos a tomar que la longitud de onda de registro es igual a la de reconstrucción, así la expresión anterior queda

$$\frac{T\pi}{\lambda_n}\left[\cos\theta_i-\cos\theta_R+\cos\theta_r-\cos\theta_o\right]$$

Ahora bien, un cambio en la dirección del haz de reconstrucción $\Delta\left(\cos\theta_R\right)$ dará lugar a un cambio en la dirección de difracción $\Delta\left(\cos\theta_i\right)$ luego la función sinc^2 se anulará cuando

$$\frac{T\pi}{\lambda_n}\left[\cos\theta_i+\Delta\left(\cos\theta_i\right)-\cos\theta_R-\Delta\left(\cos\theta_R\right)+\cos\theta_r-\cos\theta_o\right]=\pi$$

Si asumimos que partimos con un haz de reconstrucción igual al de referencia y en consecuencia con una imagen igual al haz objeto, la ecuación anterior se reduce en el caso de extinción a

$$\Delta\left(\cos\theta_i\right)-\Delta\left(\cos\theta_R\right)=\frac{\lambda_n}{T}$$

Con respecto a la otra función sinc y dado que la ecuación de la red se cumple, se llega a que

$$\Delta\left(\sin\theta_i\right)-\Delta\left(\sin\theta_R\right)=0$$

Ahora bien, tenemos que

$$\begin{aligned}\Delta\left(\cos\theta_i\right)&=-\sin\theta_i\Delta\left(\theta_i\right); \quad \Delta\left(\cos\theta_R\right)=-\sin\theta_R\Delta\left(\theta_R\right)\\ \Delta\left(\sin\theta_i\right)&=\cos\theta_i\Delta\left(\theta_i\right); \quad \Delta\left(\sin\theta_R\right)=\cos\theta_R\Delta\left(\theta_R\right)\end{aligned}$$

Luego reemplazando en las ecuaciones anteriores obtenemos

$$\cos\theta_i\Delta\left(\theta_i\right)-\cos\theta_R\Delta\left(\theta_R\right)=0\Rightarrow \quad \Delta\left(\theta_i\right)=\frac{\cos\theta_R}{\cos\theta_i}\Delta\left(\theta_R\right)$$

$$-\sin\theta_i\Delta\left(\theta_i\right)+\sin\theta_R\Delta\left(\theta_R\right)=\frac{\lambda_n}{T}=-\sin\theta_i\frac{\cos\theta_R}{\cos\theta_i}\Delta\left(\theta_R\right)+\sin\theta_R\Delta\left(\theta_R\right)$$

Con lo cual

$$\Delta(\theta_R) = \frac{\lambda_n}{T}\left(\sin\theta_R - \sin\theta_i \frac{\cos\theta_R}{\cos\theta_i}\right)^{-1}$$

Pero de acuerdo a lo postulado inicialmente $(\sin\theta_i, \cos\theta_i) = (\sin\theta_o, \cos\theta_o)$ ya que tomamos $(\sin\theta_R, \cos\theta_R) = (\sin\theta_r, \cos\theta_r)$ por lo tanto podemos escribir

$$\Delta(\theta_R) = \frac{\lambda_n}{T}\left(\frac{\cos\theta_o}{\sin\theta_r\cos\theta_o - \sin\theta_o\cos\theta_r}\right) = \frac{\lambda_n}{T}\frac{\cos\theta_o}{\sin\Delta\theta}$$

donde $\Delta\theta$ es el ángulo entre el haz objeto y el de referencia.

Esta ecuación nos da la desviación tolerada en el haz de reconstrucción hasta extinguir la imagen, para un holograma de espesor *T* registrado con una longitud de onda λ_n en el medio y con ángulos de incidencia θ_o y θ_r para el haz objeto y el de referencia, respectivamente. Vemos que para $\Delta\theta$ chicos el holograma no es muy sensible a los cambios de orientación, sin embargo esta sensibilidad aumenta para ángulos y espesores grandes lo cual es coincidente con lo dicho anteriormente, esto es, que el holograma comienza a comportarse como una red tridimensional cuando la interfranja es pequeña comparada con el espesor del medio de registro.

Sensibilidad al cambio de longitud de onda

CLASE 20

Consideremos ahora el efecto que causa iluminar al holograma con una longitud de onda diversa a la usada para el registro. Nuevamente estudiemos el término correspondiente a la imagen primaria.

$$P(\sin\theta_i, \cos\theta_i) = Cte.\beta' O_o r_o\, T\, 2H \,\mathrm{sinc}\left[H\left(k_{nR}\sin\theta_i - k_{nR}\sin\theta_R + 2\pi f_o\cos\alpha\right)\right].$$
$$.\mathrm{sinc}\left[\frac{T}{2}\left(k_{nR}\cos\theta_i - k_{nR}\cos\theta_R - 2\pi f_o\sin\alpha\right)\right]$$

Habíamos visto con anterioridad que

$$f_o = \frac{2}{\lambda_n}\sin\left(\frac{\Delta\theta}{2}\right); \quad \cos\theta_o - \cos\theta_r = 2\sin\alpha\sin\left(\frac{\Delta\theta}{2}\right); \quad 2\cos\alpha\sin\left(\frac{\Delta\theta}{2}\right) = \sin\theta_r - \sin\theta_o$$

Luego los argumentos de las funciones sinc se pueden escribir de la forma

$$H\left[k_{nR}\left(\sin\theta_i - \sin\theta_R\right) + k_n\left(\sin\theta_r - \sin\theta_o\right)\right]$$
$$\frac{T}{2}\left[k_{nR}\left(\cos\theta_i - \cos\theta_R\right) + k_n\left(\cos\theta_r - \cos\theta_o\right)\right]$$

Habíamos visto que cuando empleamos una haz de reconstrucción igual al de referencia, esto es, $\left(\sin\theta_R, \cos\theta_R\right) = \left(\sin\theta_r, \cos\theta_r\right)$ y $\lambda_{nR} = \lambda_n$, teníamos un máximo para $\left(\sin\theta_i, \cos\theta_i\right) = \left(\sin\theta_o, \cos\theta_o\right)$. Queremos ver ahora qué cambio en la longitud de onda hace que este haz difractado se anule.

Sabemos que un cambio en la longitud de onda de reconstrucción $\left(k_n = k_{nR} \to k_n + \Delta\left(k_n\right)\right)$ producirá un cambio en la dirección de la onda difractada tal que $\left(\cos\theta_i \to \cos\theta_i + \Delta\left(\cos\theta_i\right)\right); \left(\sin\theta_i \to \sin\theta_i + \Delta\left(\sin\theta_i\right)\right)$. Si suponemos, al igual que antes, que se cumple la ecuación de la red, la onda primaria se anulará cuando el argumento de la segunda función sinc sea igual a π. Luego

$$H\left[\left(k_n + \Delta\left(k_n\right)\right)\left(\sin\theta_i + \Delta\left(\sin\theta_i\right) - \sin\theta_R\right) + k_n\left(\sin\theta_r - \sin\theta_o\right)\right] = 0$$
$$\frac{T}{2}\left[\left(k_n + \Delta\left(k_n\right)\right)\left(\cos\theta_i + \Delta\left(\cos\theta_i\right) - \cos\theta_R\right) + k_n\left(\cos\theta_r - \cos\theta_o\right)\right] = \pi$$

Si incluimos en estas ecuaciones las condiciones de las que partimos estas se reducen a

$$\Delta\left(k_n\right)\left(\sin\theta_o - \sin\theta_r\right) + \left(k_n + \Delta\left(k_n\right)\right)\Delta\left(\sin\theta_i\right) = 0$$
$$\Delta\left(k_n\right)\left(\cos\theta_o - \cos\theta_r\right) + \left(k_n + \Delta\left(k_n\right)\right)\Delta\left(\cos\theta_i\right) = \frac{2\pi}{T}$$

Ahora bien, por un lado tenemos que $k_n = \frac{2\pi}{\lambda_n} \Rightarrow \Delta\left(k_n\right) = -\frac{2\pi}{\lambda_n^2}\Delta\left(\lambda_n\right)$ y si consideramos las expresiones anteriormente halladas para $\Delta\left(\cos\theta_i\right)$, $\Delta\left(\sin\theta_i\right)$ llegamos a que el haz que da origen a la imagen primaria se extingue cuando

$$\left|\Delta\left(\lambda_n\right)\right| = \frac{\lambda_n^2}{T}\left(\frac{\cos\theta_o}{1-\cos\theta_r\cos\theta_o-\sin\theta_o\sin\theta_r}\right) = \frac{\lambda_n^2}{T}\frac{\cos\theta_o}{1-\cos\Delta\theta}$$

Recordando que $\Delta\theta$ es el ángulo formado entre el haz de referencia y el haz objeto, vemos que cuando este ángulo es chico el rango de longitudes de onda con los que se puede reconstruir el holograma es muy grande. En el caso contrario, cuando el haz objeto incide de un lado de la placa holográfica y el de referencia lo hace desde el otro, $\Delta\theta = \pi$ y $\Delta\left(\lambda_n\right)$ tiene el mínimo valor. De hecho este caso es el correspondiente a los hologramas de reflexión donde los planos de Bragg son casi paralelos a la superficie de la placa. Este tipo de hologramas pueden verse iluminados con luz blanca ya que el mismo actúa como un filtro interferencial reconstruyendo la imagen para una sola longitud de onda.

Hologramas de volumen de transmisión y reflexión

Tal como hemos dicho con anterioridad la mayoría de los hologramas deberían considerarse con un espesor no despreciable. Sin embargo el hecho de que una sola imagen se forme para una dada dirección de reconstrucción, se presenta solamente en los hologramas verdaderamente gruesos. Así la mayoría se encuentra en una situación intermedia.

Un holograma de volumen, ya sea de transmisión o reflexión, alcanza una eficiencia de difracción teórica del 100% y experimentalmente (por ejemplo en gelatinas dicromatadas) del 90%.

Hemos visto que hologramas de volumen de transmisión poseen una alta sensibilidad a la orientación y al cambio de longitud de onda

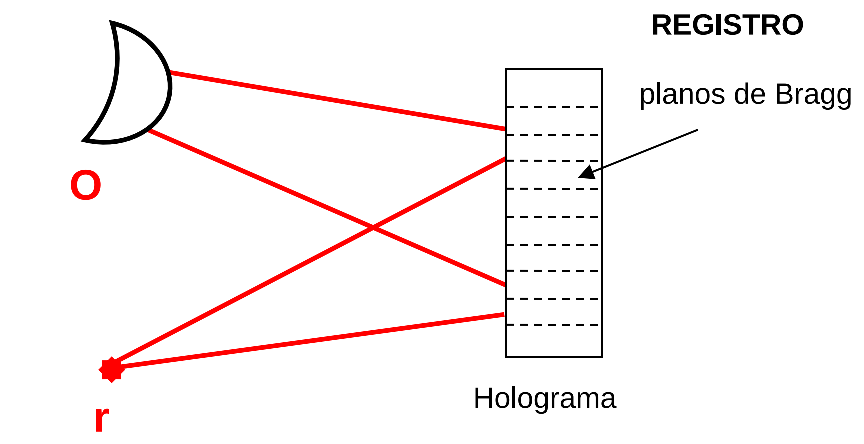
$$\Delta\left(\theta_R\right) = \frac{\lambda_n}{T}\frac{\cos\theta_o}{\sin\Delta\theta} \qquad\qquad \Delta\left(\lambda_n\right) = \frac{\lambda_n^2}{T}\frac{\cos\theta_o}{1-\cos\Delta\theta}$$

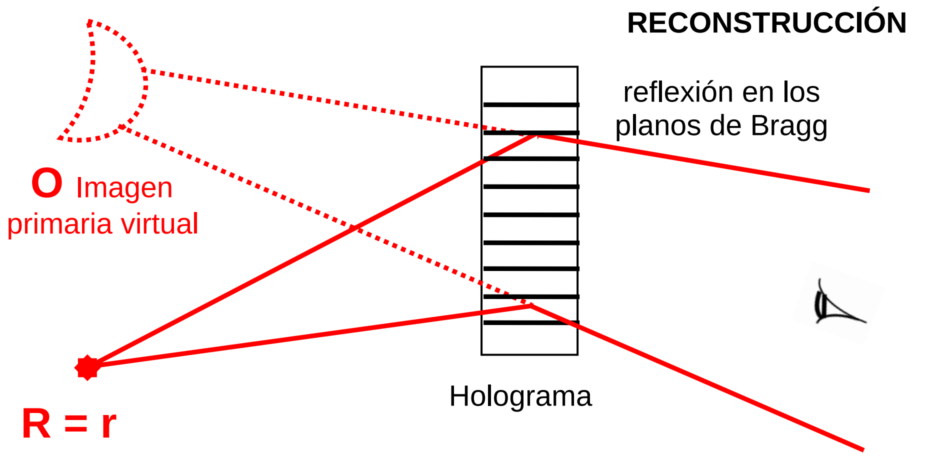
Sin embargo notemos que si bien la selectividad en la longitud de onda es máxima para hologramas de reflexión, los mismos resultan relativamente insensibles a cambios de orientación. Esto hace que estos hologramas sean muy útiles ya que pueden reconstruirse con luz blanca y desde casi cualquier dirección.

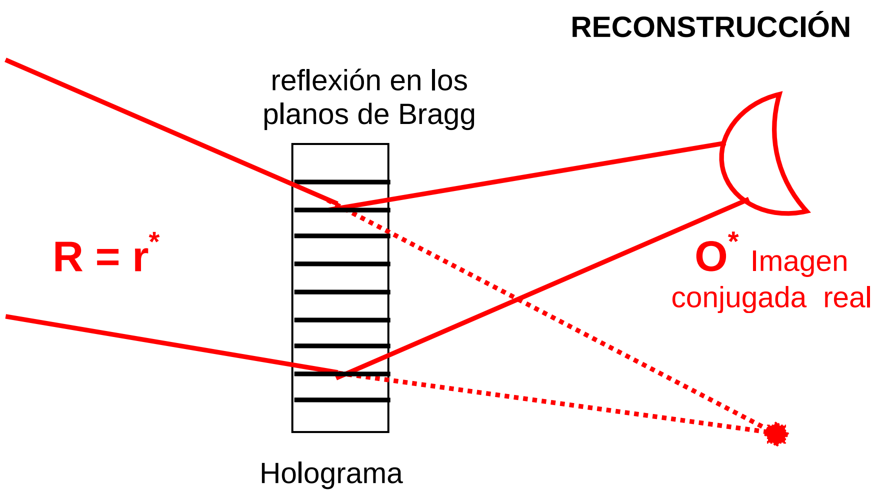
Los hologramas de reflexión fueron inventados por el ruso Denisyuk y sus primeras investigaciones se encuentran en el trabajo [Y.Denisyuk, Soviet Phys. Doklady, **7**, 543, (1962)].

De acuerdo con los análisis que realizamos previamente veamos cómo son las disposiciones geométricas para registrar y reconstruir hologramas de volumen.

Estudiemos primero el caso de los de transmisión

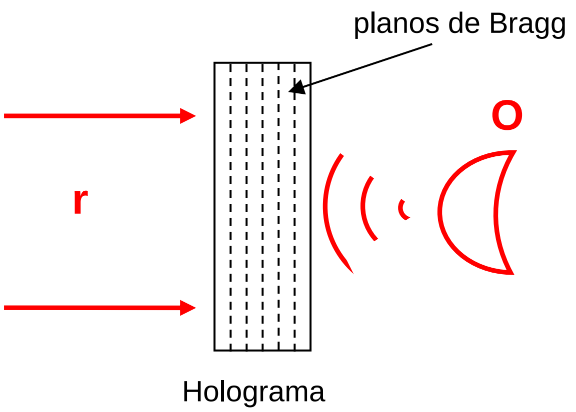




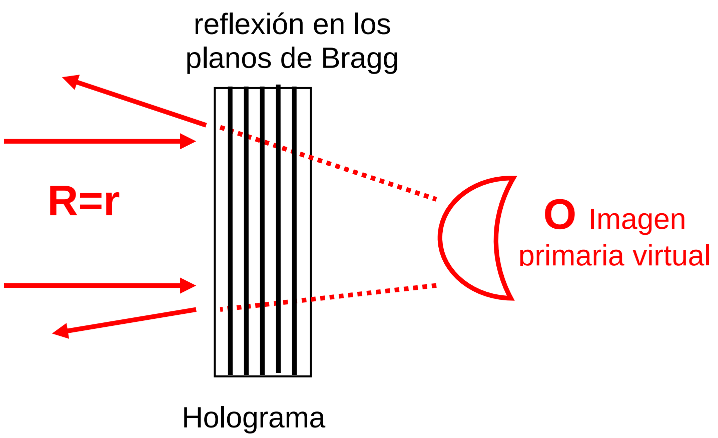


Veamos ahora el caso de reflexión

REGISTRO



RECONSTRUCCIÓN



Dado que este último tipo de hologramas es poco sensible a la orientación es posible ver la imagen de un color para un ángulo de incidencia y de otro color (cuya longitud de onda cumpla la nueva condición de Bragg) para otra orientación. Obviamente suponemos que iluminamos con luz blanca. Cabe destacar que en el proceso fotográfico pueden producirse dilataciones o contracciones que hacen que cambie el espaciado de los planos de Bragg lo que se manifiesta en un cambio de la longitud de onda reflejada con respecto a la de registro.

IX. APLICACIONES DE LA HOLOGRAFÍA

Desde la invención de la holografía han surgido numerosas aplicaciones: microscopía holográfica, construcción de diversos elementos ópticos holográficos (redes, espejos, lentes, memorias, etc.), interferometría, procesado de señales, holografía digital, etc. Nosotros, a modo de ejemplo, sólo veremos algunas de ellas

INTERFEROMETRÍA HOLOGRÁFICA

La interferometría constituye uno de los principales métodos para realizar mediciones de alta precisión. El hecho que la longitud de onda de la luz sea del orden de 0.5 μm y que la interferometría permita detectar cambios de fracciones de la longitud de onda permite tener una idea de la precisión que puede alcanzarse en estas mediciones.
La interferometría convencional puede utilizarse solamente para medir superficies altamente pulidas y de formas relativamente simples, lo que limita en cierta medida su aplicación. La interferometría holográfica extiende este rango permitiendo realizar mediciones de superficies tridimensionales de forma y acabado arbitrarios. Así resulta posible medir con altísima precisión perturbaciones existentes en medios tan diversos como piezas mecánicas sometidas a esfuerzos, flujos de aire en túneles de viento, plasmas, transferencia de calor, etc. Un amplio estudio de diversas técnicas de interferometría holográfica puede encontrarse en dos libros clásicos [R.K.Erf, *Holographic non destructive testing* New York, Academic Press] [C.M.Vest, *Holographic Interferometry*, New York, John Wiley].
Lo que se observa con esta técnica no es un holograma del objeto, sino la figura de interferencia entre la imagen en el estado inicial del objeto y la del estado final (deformado). Estudiando las franjas de interferencia puede evaluarse que tipo de perturbaciones sufrió.
Básicamente los distintos tipos de interferometría holográfica los podemos dividir en: interferometría en tiempo real o de una sola exposición, doble exposición y promedio temporal. Comencemos analizando el caso de tiempo real.

Interferometría holográfica en una sola exposición o en tiempo real

Vamos a analizar el proceso planteando las ecuaciones en una dimensión para que sean más sencillas. En una primera etapa se registra un holograma fuera de eje del objeto a estudiar. La onda proveniente del objeto será $O(x) = O_o(x)e^{i\varphi_o(x)}$ y la de referencia $r(x)=r_o\, e^{i\varphi_r}=r_o\, e^{i\,kx\sin\alpha_r}$, luego la transmisión del holograma vendrá dada como siempre por

$$
\begin{aligned}
t(x) &= t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)+\\
&\quad+\beta' r_o O_o(x)\Big(\exp\big[i\left(\varphi_o(x)-kx\sin\alpha_r\right)\big]+\exp\big[-i\left(\varphi_o(x)-kx\sin\alpha_r\right)\big]\Big)
\end{aligned}
$$

El holograma una vez revelado, se repone exactamente en la misma posición que se registró, ó se revela in situ y posteriormente se ilumina con la onda de reconstrucción $R(x)=O'(x)+r(x)$, o sea con casi la misma iluminación con la que se registró, sólo que ahora $O'(x)$ es la onda correspondiente al objeto en su nuevo estado. Esto es $O'(x)=O_o(x)e^{i\,\varphi'_o(x)}$, donde la amplitud es básicamente la misma que la anterior, ya que la perturbación es mínima, y sólo se considera la variación en la fase. Luego el campo transmitido por el holograma será $R(x)\cdot t(x)$

$$
\begin{aligned}
R(x).t(x) &= \left[t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)\right]O_o(x)\exp\left(i\varphi_o'(x)\right)+\\
&\quad+\left[t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)\right]r_o\exp\left(ikx\sin\alpha_r\right)+\\
&\quad+\beta' O_o^2(x)r_o\exp\left[i\left(\varphi_o(x)+\varphi_o'(x)-kx\sin\alpha_r\right)\right]+\beta' r_o^2 O_o(x)\exp\left(i\varphi_o(x)\right)+\\
&\quad+\beta' O_o^2(x)r_o\exp\left[i\left(\varphi_o'(x)-\varphi_o(x)+kx\sin\alpha_r\right)\right]+\\
&\quad+\beta' r_o^2 O_o(x)\exp\left[i\left(-\varphi_o(x)+2kx\sin\alpha_r\right)\right]
\end{aligned}
$$

De todos estos términos sólo nos interesan los que no contienen la información de la fase del haz de referencia. Tales términos describen a ondas difractadas en aproximadamente la misma dirección. Ellos son

$$
\left[t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)\right]O_o(x)\exp\left(i\varphi_o'(x)\right)+\beta' r_o^2 O_o(x)\exp\left(i\varphi_o(x)\right)
$$

La intensidad registrada por un detector orientado hacia estas ondas será

$$I(x)=\left[t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)\right]^2 O_o^2(x)+\left[\beta' r_o^2 O_o(x)\right]^2+$$
$$+2\left[t_b+\beta'\left(O_o^2(x)+r_o^{\,2}\right)\right]\beta' r_o^2 O_o^2(x)\cos\left(\varphi_o^{'}(x)-\varphi_o(x)\right)$$

donde vemos que tanto los dos primeros términos como el coeficiente de la función coseno del tercer término son magnitudes que no depende de la distribución de fase y la única dependencia en la variable *x* es a través de potencias de $O_o(x)$. Luego lo que se observa es aproximadamente el objeto, pero esta imagen aparece modulada por franjas dadas por $\cos\left(\varphi_o^{'}(x)-\varphi_o(x)\right)$ que sí dependen de la distribución inicial y final de fase.
Esta técnica tiene la ventaja de permitir ver en tiempo real una deformación ya que se compara instante a instante la onda objeto inicial, registrada en el holograma, con la onda objeto deformada que lo ilumina. Los inconvenientes residen en la reubicación correcta de la placa y en el cuidado que se debe tener para que la emulsión fotográfica no se deforme en el revelado y secado. Para esta técnica resultan ideales los materiales de registro fototermoplásticos ya que, como vimos, se revelan in situ.

Interferometría holográfica de doble exposición

La técnica de doble exposición, si bien no permite realizar un estudio a tiempo real de las deformaciones, no presenta el inconveniente del reposicionado y estabilidad del medio de registro ya que los dos hologramas son registrados sucesivamente en la misma placa. La diferencia fundamental con la técnica de una sola exposición reside en el hecho que el cambio sufrido por el objeto entre los dos exposiciones queda registrado permanentemente. La reconstrucción del mismo se realiza como en cualquier holograma con un haz igual al utilizado como referencia.
Analicemos cómo es esta técnica. En cada exposición se registra un holograma fuera de eje, uno antes de la deformación y otro después. La intensidad que llega a la placa en cada caso será

$$I_j(x)=O_{oj}^2(x)+r_o^2+r(x)O_{oj}^*(x)+r^*(x)O_{oj}(x) \qquad j=1,2$$

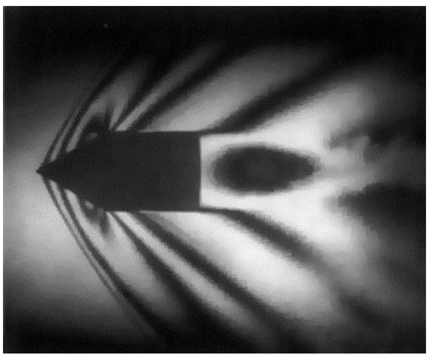
Luego, como siempre, trabajamos en la zona en la que la transmisión del holograma es lineal con la intensidad. En este caso la intensidad total será $I(x)=I_1(x)+I_2(x)$
Luego

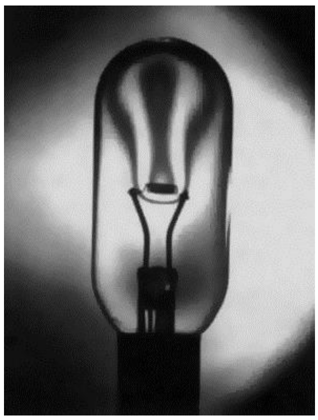
$$t(x)=t_b+\beta'\left(I_1(x)+I_2(x)-I_b\right)=t_b+\beta'\left(O_{o1}^2(x)+O_{o2}^2(x)+r_o^2\right)+$$
$$+\beta'\left[\left(O_1(x)+O_2(x)\right)r^*(x)+\left(O_1(x)+O_2(x)\right)^* r(x)\right]$$

Si se reconstruye el holograma con un haz igual al de referencia la imagen primaria queda $P(x)=\beta' r_o^2(x)\left[O_1(x)+O_2(x)\right]$. Luego la intensidad será

$$I(x)=\beta'^2 r_o^4(x)\left[O_{o1}^2(x)+O_{o2}^2(x)+O_{o1}(x)O_{o2}(x)\cos\left(\varphi_1(x)-\varphi_2(x)\right)\right]$$

En este caso nuevamente se obtiene una imagen del objeto modulada por franjas de interferencia que denotan las deformaciones sufridas por el mismo.
Otra forma de interpretar esta técnica es como un batido de las franjas registradas. Cada holograma consiste en un registro de franjas de interferencia que varían muy poco de uno a otro. El registro total consistirá en un diagrama de franjas que en algunas zonas se refuerzan y en otras se borronean formando batidos o figuras de moiré. Cuando reconstruimos el holograma, las zonas borroneadas no contribuirán con luz difractada mientras que sí lo harán aquellas zonas donde las franjas se refuerzan. La técnica de doble exposición es muy apropiada para el registro de fenómenos que varían rápidamente en el tiempo (ondas de choque, flujo de fluidos, etc), se utiliza para ello una fuente luminosa tal como un láser pulsado. Un ejemplo de este tipo de interferogramas se muestra a continuación.

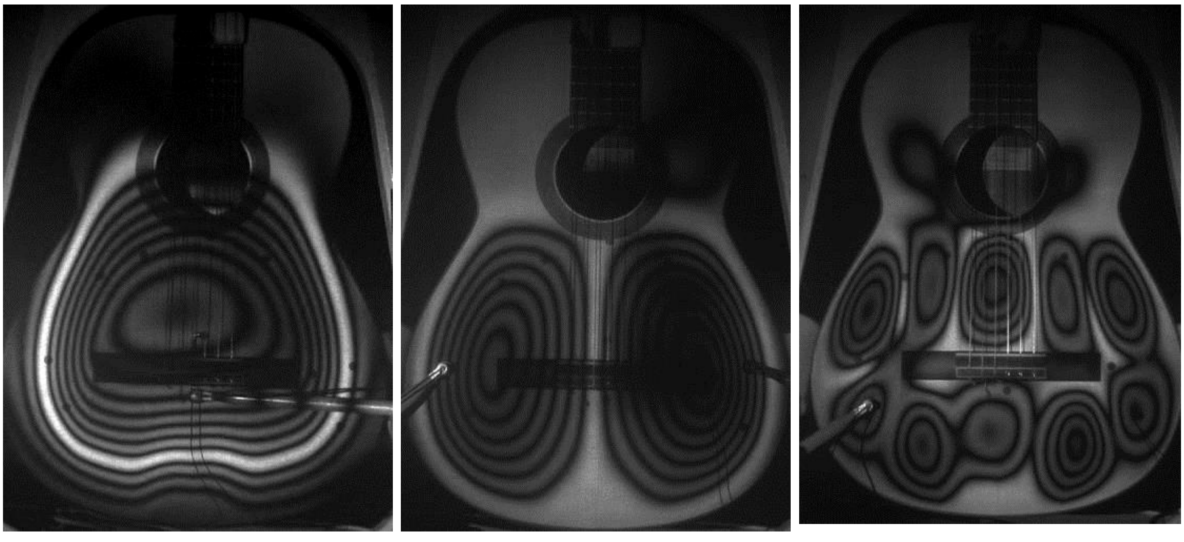




Estos hologramas fueron realizados por Brooks et.al. En el primero de ellos se ve un proyectil cuya onda de choque afecta al aire que lo rodea, en este caso la primera exposición se realiza con el aire en reposo y la segunda mientras el proyectil atravieza la zona de exposición. El segundo holograma es similar, la primer toma se realiza con la bombita apagada y la segunda haciendo pasar cierta corriente por el filamento de manera de calentar el aire que lo rodea.

Interferometría holográfica de promedio temporal

La idea de exposiciones múltiples puede extenderse al límite de casi un continuo, dando como resultado lo que se llama interferometría holográfica de promedio temporal. Esta técnica es aplicada ampliamente en el análisis de vibraciones. La idea básica del método es que dado que el holograma en sí surge de un proceso interferométrico, toda inestabilidad creará un borroneo de las franjas. Así el holograma de un objeto vibrante es un registro de la distribución de intensidad sobre la placa promediada durante el tiempo que duró la exposición.
Dado que la cantidad de luz difractada por cada región del holograma depende del contraste de las franjas, cualquier movimiento del objeto causará un desplazamiento de las franjas durante la exposición y consecuentemente una pérdida de contraste, dando lugar a una menor cantidad de luz difractada. El proceso podría pensarse como el registro de una infinidad de hologramas, uno para cada pequeño desplazamiento del objeto.
Si el objeto está vibrando en un modo normal, habrá ondas estacionarias de vibración de modo tal que encontraremos zonas (nodos) donde el movimiento es prácticamente inapreciable y otras (antinodos) donde el desplazamiento es máximo.
El holograma final dará como resultado una imagen brillante de las zonas quietas mientras que no difractará luz en las zonas donde el movimiento del objeto borroneó las franjas de interferencia.
Un ejemplo de modos de vibración de una guitarra se muestra en estos interferogramas realizados por Richardson

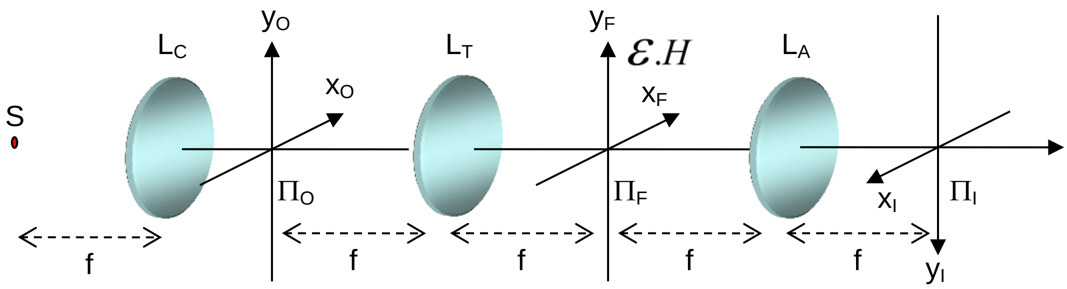


FILTROS HOLOGRÁFICOS APLICADOS AL PROCESADO DE SEÑALES

Cuando comenzamos a estudiar filtrado espacial vimos algunos ejemplos de filtros de amplitud, otros de fase y los llamados filtros compuestos con lo que la amplitud y fase de la señal de entrada era modificada simultáneamente. Habíamos dicho en esa oportunidad que si bien en algunos casos sencillos era posible realizar un filtro para la amplitud y otro para la fase, por ejemplo mediante evaporaciones de un material transparente sobre un sustrato de manera de modificar su espesor, en general esta técnica era muy difícil de implementar. Filtros compuestos más complejos requerían de un registro holográfico. Ahora que hemos visto algunos temas de holografía podemos retomar el filtrado de señales estudiando este tipo de filtros. Para ello comenzaremos con los más sencillos, esto es redes de difracción simples o compuestas utilizadas como filtros espaciales.

Redes de difracción utilizadas como filtros espaciales

Analicemos el comportamiento de una red sinusoidal utilizada como filtro espacial en un procesador coherente



Recordemos que en estos casos la función de entrada sobre el plano objeto era $E_o\left(x_o, y_o\right)$ y al plano transformado llegaba una distribución de campo dada por $\mathbb{F}\left[E_o\right] = C_1 \mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$. Sobre dicho plano se ubicaba un filtro cuya transmisión era $t_F\left(x_F, y_F\right) = C_2 H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$ donde C_1 y C_2 son constantes complejas. Así la distribución de campo detrás del filtro era proporcional a $\mathcal{E}.H$. Finalmente la lente L_A producía sobre el plano imagen del procesador la distribución $\mathbb{F}\left[\mathcal{E}.H\right] = E \otimes h$.

Tomemos ahora el caso en el que el filtro es una red sinusoidal cuya función transmisión es

$$H\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)=\frac{1}{2}\left\{1+\frac{1}{2}\exp\left[i\left(2\pi\frac{b}{\lambda f}y_F+\phi\right)\right]+\frac{1}{2}\exp\left[-i\left(2\pi\frac{b}{\lambda f}y_F+\phi\right)\right]\right\}$$

La frecuencia espacial de dicha red es $\frac{b}{\lambda f}$ y ϕ tiene en cuenta la posición de la red con respecto al eje *z*. Esto es, si no incluimos ϕ estamos asumiendo que para $y_F = 0$ tenemos un máximo de transmisión, pero nosotros deseamos tener la libertad de desplazar la red en el plano en el que se halla de manera que debemos incluir esa variable.

Con este tipo de filtro un objeto ubicado en el plano de entrada será difractado en tres imágenes sobre el plano de salida. Así dos objetos que no se superponen en el plano Π_o serán difractados en seis imágenes sobre Π_i, algunas de las cuales pueden superponerse.

Supongamos entonces que sobre Π_o tenemos una distribución de campo dada por:

$$E_o\left(x_o,y_o\right)=O_1\left(x_o,y_o-b'\right)+O_2\left(x_o,y_o+b'\right)$$

Luego

$$\mathcal{E}\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)=\mathcal{O}_1\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[-i\,2\pi\frac{b'\,y_F}{\lambda f}\right]+\mathcal{O}_2\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[i\,2\pi\frac{b'\,y_F}{\lambda f}\right]$$

donde las exponenciales aparecen debido al desplazamiento del objeto fuera del eje óptico en $\pm$ b'. Si elegimos b'= b, después del filtro el campo es:

$$\begin{aligned}\mathcal{E}.H=&\frac{1}{2}\mathcal{O}_1\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[-i\,2\pi\frac{b\,y_F}{\lambda f}\right]+\frac{1}{2}\mathcal{O}_2\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[i\,2\pi\frac{b\,y_F}{\lambda f}\right]+\\&+\frac{1}{4}\mathcal{O}_1\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[i\phi\right]+\frac{1}{4}\mathcal{O}_2\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[i\left(4\pi\frac{b}{\lambda f}y_F+\phi\right)\right]+\\&+\frac{1}{4}\mathcal{O}_1\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[-i\left(4\pi\frac{b}{\lambda f}y_F+\phi\right)\right]+\frac{1}{4}\mathcal{O}_2\left(\frac{x_F}{\lambda f},\frac{y_F}{\lambda f}\right)\exp\left[-i\phi\right]\end{aligned}$$

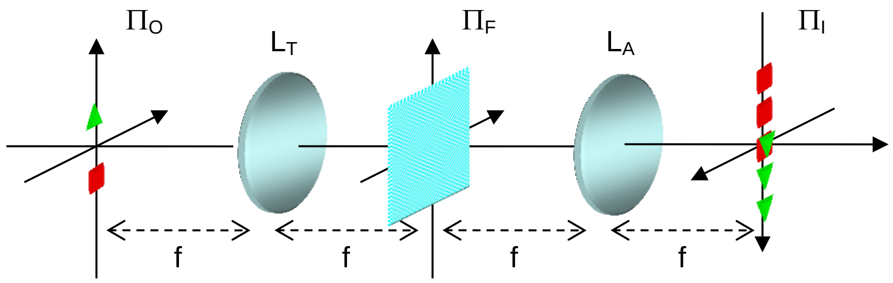
La antitransformada de esta expresión produce sobre el plano de salida una distribución dada por:

$$E_i(x_i,y_i)=\frac{1}{2}O_1(x_i,y_i-b)+\frac{1}{2}O_2(x_i,y_i+b)+\frac{1}{4}O_1(x_i,y_i)\exp[i\phi]+$$
$$+\frac{1}{4}O_2(x_i,y_i+2b)\exp[i\phi]+\frac{1}{4}O_1(x_i,y_i-2b)\exp[-i\phi]+\frac{1}{4}O_2(x_i,y_i)\exp[-i\phi]$$

De todos estos términos sólo nos interesan los que se superponen en el origen

$$\frac{1}{4}O_1(x_i,y_i)\exp[i\phi]+\frac{1}{4}O_2(x_i,y_i)\exp[-i\phi]=\frac{1}{4}\left[O_1(x_i,y_i)+O_2(x_i,y_i)\exp[-i2\phi]\right]\exp[i\phi]$$

Luego, cuando la máxima transmisión de la red coincide con el eje óptico ($\phi=0$), obtenemos la suma de los objetos. Si se desplaza la red de forma tal que $\phi=\frac{\pi}{2}$ se obtiene la resta



Pasemos a analizar otra operación matemática que vimos previamente. Con anterioridad habíamos estudiado la operación derivada a partir de un filtro de amplitud que variaba linealmente con la frecuencia combinado con uno de fase que introducía un desfasaje de $\frac{\pi}{2}$ para las frecuencias positivas y uno de $\frac{3\pi}{2}$ para las negativas. Veamos ahora como hacer esta operación de una forma mucho más sencilla.

Supongamos que sobre la misma placa holográfica registramos dos redes sinusoidales superpuestas, una con una frecuencia espacial ligeramente distinta de la otra. Esto es, sus funciones transmisión vienen dadas por

$$H_1 = \frac{1}{2}\left\{1 + \frac{1}{2}\exp\left[i\left(2\pi\frac{b}{\lambda f}y_F + \phi\right)\right] + \frac{1}{2}\exp\left[-i\left(2\pi\frac{b}{\lambda f}y_F + \phi\right)\right]\right\}$$

$$H_2 = \frac{1}{2}\left\{1 + \frac{1}{2}\exp\left[i\left(2\pi\frac{b+\varepsilon}{\lambda f}y_F\right)\right] + \frac{1}{2}\exp\left[-i\left(2\pi\frac{b+\varepsilon}{\lambda f}y_F\right)\right]\right\}$$

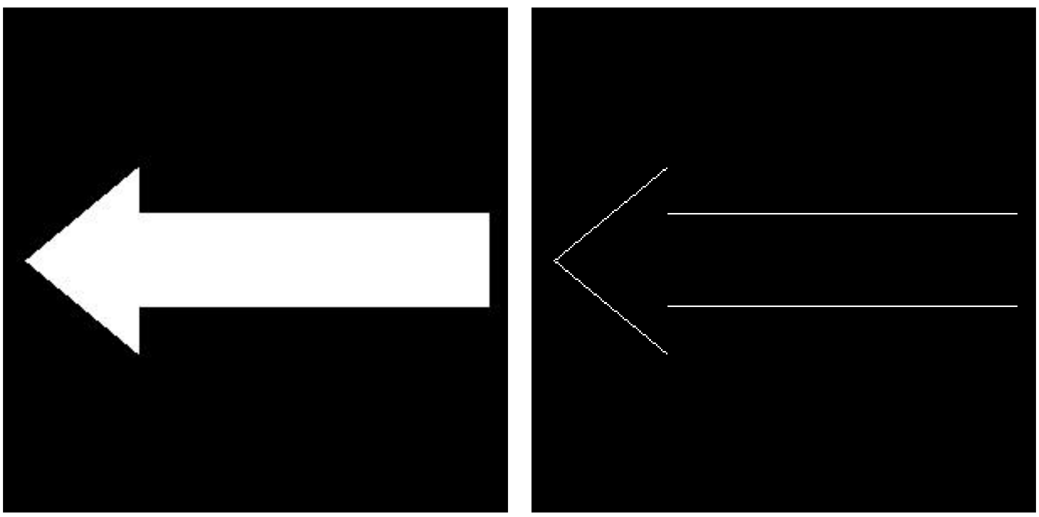
De modo tal que $H = H_1 + H_2$. Tomamos como entrada el objeto que queremos derivar $O(x_o, y_o)$. Luego cada red sobre el plano final producirá tres imágenes de este objeto. Las correspondiente al orden 0 de difracción no sufren cambios, pero analicemos por ejemplo que sucede en un orden difractado. Ahí tendremos una imagen del objeto y una imagen ligeramente corrida en la dirección y_i ya que la diferencia de frecuencias ε de las redes es muy pequeña. Si además la fase ϕ es tal que las imágenes se restan, tenemos justamente la definición de derivada. Dicho orden difractado será

$$E_i(x_i, y_i) = \ldots\ldots\ldots\frac{1}{4}O(x_i, y_i + b)\exp[i\phi] + \frac{1}{4}O(x_i, y_i + b + \varepsilon)\ldots\ldots$$

Luego si hacemos $\phi = \pi$ y $\varepsilon \to 0$ obtenemos en ese orden

$$\lim_{\varepsilon\to 0}\left[O(x_i, y_i + b + \varepsilon) - O(x_i, y_i + b)\right] \propto \frac{\partial O(x_i, y_i)}{\partial y_i}$$

Un ejemplo de esta operación se muestra en la siguiente figura

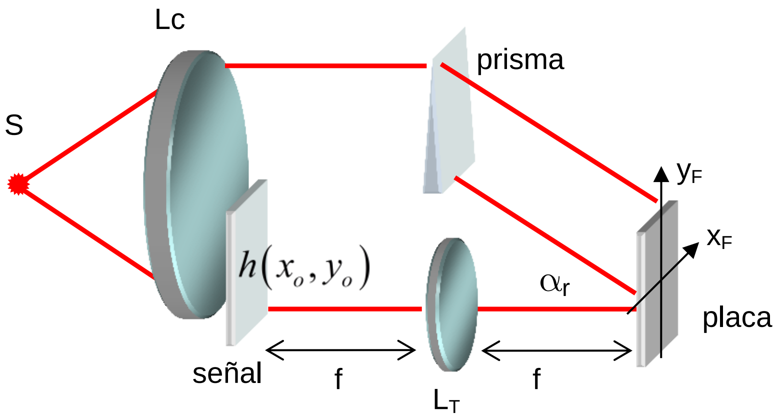


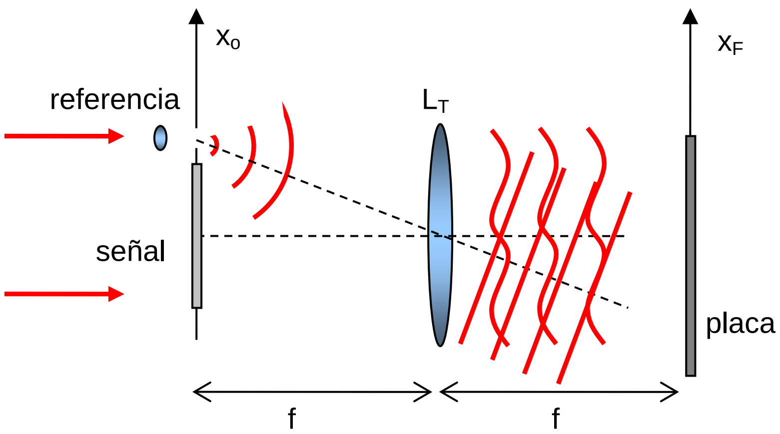
Mediante la superposición de redes sinusoidales pueden realizarse filtros para llevar acabo operaciones más complicadas tales como Laplaciano, detección de bordes, etc.

Filtro de Vander Lugt CLASE 21

La utilización de hologramas como filtros espaciales fue introducida por Vander Lugt, quien propuso una nueva técnica para la síntesis de filtros complejos [A.Vander Lugt *Signal Detection by Complex Spatial Filtering*, IEEE Trans. Inform. Theory, **IT-10**, 2 (1964)].

La principal característica de los filtros así sintetizados es que permiten registrar la amplitud y fase mediante métodos interferométricos. Originalmente para la síntesis de este tipo de filtros se utilizaban sistemas como los esquematizados en la figura. Notemos la analogía con un holograma fuera de eje tipo Fourier.





En este caso se toma como objeto la respuesta al impulso h deseada para el filtro. La lente L_T proporciona sobre el plano de registro una distribución de campo

$\frac{1}{\lambda f} H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$ proporcional a la transformada de Fourier de la señal de interés, además sobre dicho plano incide el haz de referencia $r = r_o \exp\left[-i2\pi \frac{\sin\alpha_r}{\lambda} y_F\right]$. Luego la distribución de intensidad sobre el film será

$$I(x_F, y_F) = r_o^2 + \frac{1}{\lambda^2 f^2}\left|H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right|^2 + \frac{r_o}{\lambda f} H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[i2\pi \frac{\sin\alpha_r}{\lambda} y_F\right] +$$
$$+ \frac{r_o}{\lambda f} H^*\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[-i2\pi \frac{\sin\alpha_r}{\lambda} y_F\right]$$

Vemos que la información sobre la amplitud y fase de la función $H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) = H_o\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[i\,\varphi_H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right]$ queda registrada en el tercer término. Como supusimos siempre, se realiza una exposición tal que se trabaja en el rango lineal del film de modo que la transmisión del filtro resulta $t\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) \propto I\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$.

Veamos ahora como procesa una señal. El filtro holográfico de Vander Lugt se ubica, como todo filtro en un procesador óptico coherente, en el plano transformado.

Si recordamos lo hecho para hologramas tipo Fourier, la placa se ubicaba en el foco de una lente antitransformadora y se la iluminaba con un haz de reconstrucción plano. Ahora en vez de un haz plano, sobre el filtro llega la transformada de Fourier de la señal de entrada que se desea filtrar. Esto es, si la entrada es $E_o(x_o, y_o)$, al filtro llega la distribución $\mathbb{F}\left[E_o(x_o, y_o)\right] = \frac{1}{\lambda f}\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$. Luego inmediatamente a la salida del filtro tendremos

$$t\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right).\frac{1}{\lambda f}\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) \propto \frac{r_o^2}{\lambda f}\mathcal{E} + \frac{1}{\lambda^3 f^3} H_o^2 \mathcal{E} + \frac{r_o}{\lambda^2 f^2} H\mathcal{E}\exp\left[i2\pi \frac{\sin\alpha_r}{\lambda} y_F\right] +$$
$$+ \frac{r_o}{\lambda^2 f^2} H^*\mathcal{E}\exp\left[-i2\pi \frac{\sin\alpha_r}{\lambda} y_F\right]$$

La lente antitransformadora del procesador producirá sobre el plano de salida una distribución de amplitud dada por

$$E_i(x_i,y_i) \propto r_o^2 E_o(x_i,y_i) + \frac{1}{\lambda^2 f^2}\left[h(x_i,y_i) \otimes h^*(-x_i,-y_i) \otimes E_o(x_i,y_i)\right] +$$
$$+\frac{r_o}{\lambda f}\left[h(x_i,y_i) \otimes E_o(x_i,y_i) \otimes \delta(x_i, y_i + f\sin\alpha_r)\right] +$$
$$+\frac{r_o}{\lambda f}\left[h^*(-x_i,-y_i) \otimes E_o(x_i,y_i) \otimes \delta(x_i, y_i - f\sin\alpha_r)\right]$$

El primero y segundo término no son de interés y dan una distribución de luz centrada en el origen de coordenadas. Analicemos el tercer y cuarto término.

El tercero es la convolución de h con E_o centrada en las coordenadas $(0, -f\sin\alpha_R)$ del plano de salida

$$h(x_i,y_i) \otimes E_o(x_i,y_i) \otimes \delta(x_i, y_i + f\sin\alpha_r) =$$
$$= \int\int_{-\infty}^{\infty} h(x_i - x_o, y_i + f\sin\alpha_r - y_o) E_o(x_o,y_o)\, dx_o dy_o$$

Justamente $h \otimes E$ es lo que deseamos obtener como señal filtrada.

En tanto el cuarto término da origen a la correlación cruzada de h con E_o centrada en las coordenadas $(0, f\sin\alpha_R)$ del plano de salida.

$$h^*(-x_i,-y_i) \otimes E_o(x_i,y_i) \otimes \delta(x_i, y_i - f\sin\alpha_r) =$$
$$= \int\int_{-\infty}^{\infty} E_o(x_o,y_o) h^*(x_o - x_i, y_o - y_i + f\sin\alpha_r)\, dx_o dy_o$$

Veamos cómo es esta distribución de intensidad sobre el plano final. Sea W_h el máximo ancho de la función h en la dirección y, mientras que W_E es el máximo ancho de E_o sobre dicha coordenada. Luego los tamaños característicos de cada término sobre este plano serán:

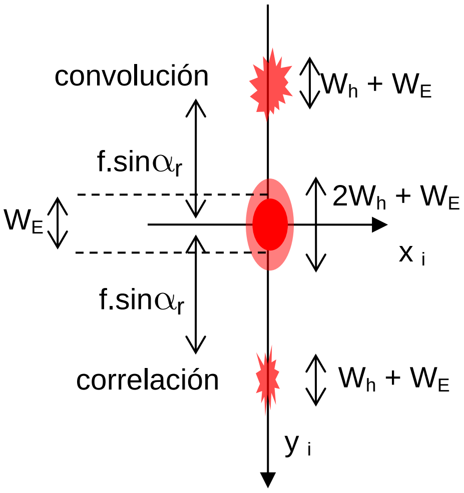
$$1- \qquad r_o^2 E_o(x_i, y_i) \to W_E$$

$$2- \qquad \frac{1}{\lambda^2 f^2}\left[h(x_i, y_i) \otimes h^*(-x_i, -y_i) \otimes E_o(x_i, y_i)\right] \to 2W_h + W_E$$

$$3- \qquad \frac{r_o}{\lambda f}\left[h(x_i, y_i) \otimes E_o(x_i, y_i) \otimes \delta(x_i, y_i + f \sin\alpha_r)\right] \to W_h + W_E$$

$$4- \qquad \frac{r_o}{\lambda f}\left[h^*(-x_i, -y_i) \otimes E_o(x_i, y_i) \otimes \delta(x_i, y_i - f \sin\alpha_r)\right] \to W_h + W_E$$

Luego esto nos da una idea de cuál es el ángulo que debemos elegir para el haz de referencia α_r de modo tal que los términos no se superpongan.



imágenes en el plano de salida

Filtro adaptado – reconocimiento de formas

Se dice que un filtro está adaptado (matched filter) a una señal particular $s(x_o, y_o)$ si su respuesta al impulso $h(x_o, y_o)$ es tal que $h(x_o, y_o) = s^*(-x_o, -y_o)$. Así la función transferencia resulta $H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) = \mathcal{S}^*\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$. Analicemos la acción de este filtro en un correlador tipo Vander Lugt, para ello centrémonos en el tercer término mencionado en el punto anterior. El mismo da

$$s^*(-x_i, -y_i) \otimes E_o(x_i, y_i) \otimes \delta(x_i, y_i + f \sin\alpha_r) =$$

$$= \int\int_{-\infty}^{\infty} E_o(x_o, y_o)\, s^*(x_o - x_i, y_o - y_i + f \sin\alpha_r)\, dx_o dy_o$$

O sea que obtenemos la correlación cruzada entre $E(x_o, y_o)$ y $s(x_o, y_o)$ que en el caso particular en que $E(x_o, y_o) = s(x_o, y_o)$ se transforma en una autocorrelación. Veamos cómo esta operación nos permite detectar una señal de interés. Para ello supongamos que el campo a la entrada del procesador consista solamente en la señal $E(x_o, y_o) = s(x_o, y_o)$ luego la distribución de campo que llega al filtro será $\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) = \mathcal{S}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)$. En consecuencia el campo emergente del filtro viene dado por

$$t\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right).\mathcal{E}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) \propto \frac{r_o^2}{\lambda f}\mathcal{S} + \frac{1}{\lambda^3 f^3}\mathcal{S}_o^2\mathcal{S} + \frac{r_o}{\lambda^2 f^2}\mathcal{S}_o^2 \exp\left[i2\pi\frac{\sin\alpha_r}{\lambda}y_F\right] + \\ + \frac{r_o}{\lambda^2 f^2}\mathcal{S}\mathcal{S}\exp\left[-i2\pi\frac{\sin\alpha_r}{\lambda}y_F\right]$$

Dado que tanto r_o como $\mathcal{S}_o^2$ son reales, el tercer término representa una onda plana, pues la curvatura de $\mathcal{S}$ fue cancelada por la de $\mathcal{S}^*$. Luego tenemos que esta onda plana, emergente del filtro, es antitransformada por la segunda lente del procesador dando origen a un punto brillante en el plano focal de la misma. Si en cambio la señal de entrada no coincide con la que fue sintetizado el filtro, las curvaturas no se compensarán y dicho término dará origen a una distribución de luz tanto menos intensa cuanto menor sea la correlación entre el objeto de entrada y la señal a detectar. Otra forma de ver esto es la siguiente: el filtro fue sintetizado como un holograma a partir de la interferencia del haz correspondiente a la señal y un haz plano de referencia. En general se utiliza un haz de reconstrucción igual al de referencia con el propósito de reobtener la onda objeto, sin embargo si iluminamos el holograma con un haz de reconstrucción igual al haz objeto lo que reobtendremos es el haz de referencia, esto es una onda plana.

Supongamos ahora que el objeto de entrada $E(x_o, y_o)$ está constituido por un conjunto de señales dentro de las cuales se encuentra la que se quiere detectar, por ejemplo la entrada es la página de un libro y se desea detectar las letras C. Vamos a llamar a esa señal particular $s_3(x_o, y_o)$. Así el objeto de entrada estará conformado por 26 señales distintas (todas las letras) entre la que se encuentra la de interés, esto es $E(x_o, y_o) = s_1(x_o, y_o) + s_2(x_o, y_o) + s_3(x_o, y_o) + \ldots\ldots + s_{26}(x_o, y_o)$. Ahora bien,

queremos demostrar que la máxima señal de correlación se obtiene para la señal de entrada a la cual está adaptado el filtro o sea que el pico de autocorrelación es mayor que el de la correlación cruzada. Dicho de otra forma, queremos ver cuando un detector en el plano de salida registra la máxima intensidad.
Para analizar esto debemos considerar primero que el pico de intensidad se producirá donde está centrada la correlación, o sea, debemos estudiar la integral

$$\int\int_{-\infty}^{\infty} s_j\left(x_o, y_o\right) s_3^*\left(x_o, y_o\right) dx_o dy_o \qquad j = 1...26$$

Además el detector registra la intensidad, luego interesa el módulo al cuadrado de esta integral. Por último cada señal $s_j\left(x_o, y_o\right)$ aportará distinta cantidad de luz así que para independizarnos de esto debemos normalizar la expresión por $\int\int_{-\infty}^{\infty} \left|s_j\left(x_o, y_o\right)\right|^2 dx_o dy_o$. Luego para el caso de la autocorrelación tenemos

$$\frac{\left|\int\int_{-\infty}^{\infty} \left|s_3\left(x_o, y_o\right)\right|^2 dx_o dy_o\right|^2}{\int\int_{-\infty}^{\infty} \left|s_3\left(x_o, y_o\right)\right|^2 dx_o dy_o} = \int\int_{-\infty}^{\infty} \left|s_3\left(x_o, y_o\right)\right|^2 dx_o dy_o$$

Mientras que para las correlaciones cruzadas obtenemos

$$\frac{\left|\int\int_{-\infty}^{\infty} s_j\left(x_o, y_o\right) s_3^*\left(x_o, y_o\right) dx_o dy_o\right|^2}{\int\int_{-\infty}^{\infty} \left|s_j\left(x_o, y_o\right)\right|^2 dx_o dy_o} \qquad \forall\ j \neq 3$$

Por la desigualdad de Schwarz vale que

$$\left|\int\int_{-\infty}^{\infty} s_j\left(x_o, y_o\right) s_3^*\left(x_o, y_o\right) dx_o dy_o\right|^2 \leq \int\int_{-\infty}^{\infty} \left|s_j\left(x_o, y_o\right)\right|^2 dx_o dy_o . \int\int_{-\infty}^{\infty} \left|s_3\left(x_o, y_o\right)\right|^2 dx_o dy_o$$

Luego

$$\frac{\left|\int\int_{-\infty}^{\infty} s_j\left(x_o,y_o\right)s_3^*\left(x_o,y_o\right)dx_o dy_o\right|^2}{\int\int_{-\infty}^{\infty}\left|s_j\left(x_o,y_o\right)\right|^2 dx_o dy_o} \leq \int\int_{-\infty}^{\infty}\left|s_3\left(x_o,y_o\right)\right|^2 dx_o dy_o$$

Con lo cual queda demostrado que siempre la correlación cruzada es menor a la autocorrelación. Ahora bien, en estos filtros si la función objeto es trasladada a una nueva posición $s_j\left(x_o-a,y_o-b\right)$ también se correrá el pico de correlación teniendo su máximo donde está centrada la señal de interés. Así estos filtros resultan invariantes frente a desplazamientos de la señal. Claramente esto no sucede para un cambio de tamaño o rotaciones. Por supuesto que habrá señales objeto más sensibles a rotaciones que otras. Por ejemplo la letra *L* será más sensible que la *O*. También en algunos casos habrá problemas de discriminación, esto es, si tenemos un filtro adaptado a la letra *C* la correlación cruzada con la letra *L* será muy baja pero tendrá problemas con una *O* ya que estas dos letras tienen mucha área de superposición.

Para solucionar estos inconvenientes se trata de desarrollar variantes de filtros que optimicen la operación de reconocimiento, pero usualmente estos no son realizables mediante métodos ópticos por lo que resulta necesario elaborar los hologramas digitalmente. Luego veremos cómo pueden sintetizarse estos hologramas.

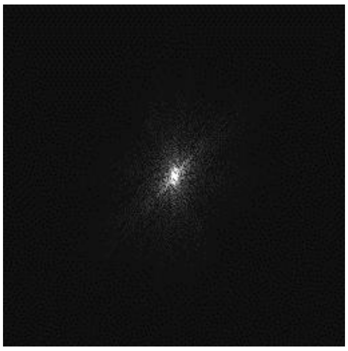
Un ejemplo de estos filtros, utilizado para evitar el problema de las rotaciones, se obtiene a partir de la descomposición de la señal de interés en armónicos circulares y luego se realiza un filtro adaptado a estas funciones que son invariantes a las rotaciones. Obviamente esto debe realizarse en forma digital.

Otros tipos de filtro están diseñados para tener un poder de discriminación mucho mayor que el filtro adaptado clásico, estos son los llamados filtros solo de fase (phase only filter POF). Para analizar como funcionan supongamos que la función transferencia del filtro es $H=H_o\exp\left[i\,\varphi_H\right]$. J.Horner y P.Gianino [Appl. Opt. **23**, 812 (1984)] demostraron que la información retenida en la fase es mucho más importante que la contenida en la amplitud, así propusieron un filtro dado por $H=\exp\left[i\,\varphi_H\right]$.

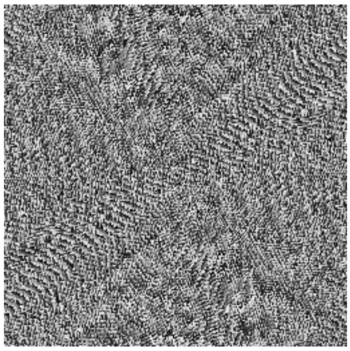
El aumento de discriminación de estos filtros puede entenderse de la siguiente manera: es sabido que en general la distribución de energía en el espectro de Fourier de un objeto complejo (no una red de difracción o algo con frecuencias muy marcadas) sigue una ley de decrecimiento del tipo *1/F*, donde *F* es la frecuencia espacial. Así la mayor parte de luz irá a las bajas frecuencias, que corresponden a los detalles más burdos, en tanto que una cantidad considerablemente menor irá a las frecuencias más altas que son las que contienen la información de los detalles más finos del objeto. Si se construye un filtro en el cual eliminamos la información de la amplitud, estamos dándole el mismo peso a todas las frecuencias de forma que logramos un aumento de los detalles más finos. Obviamente este filtro también debe realizarse de forma digital y luego transferirse a un medio que permita utilizarlo en el correlador.

A continuación se muestra una escena, su transformada de Fourier en módulo y fase y la obtención de la antitransformada a partir de la información sólo de amplitud y sólo de fase

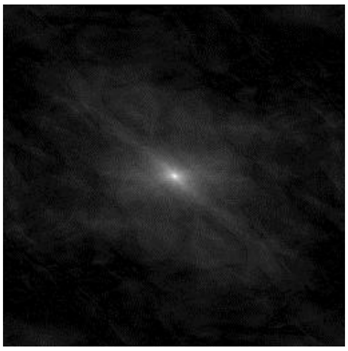
Escena



Módulo TF



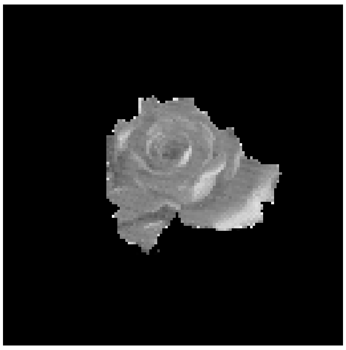
Fase TF



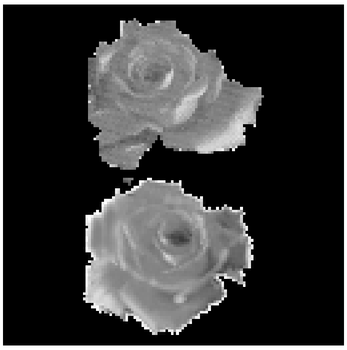
TF^{-1} Módulo TF

TF^{-1} Fase TF

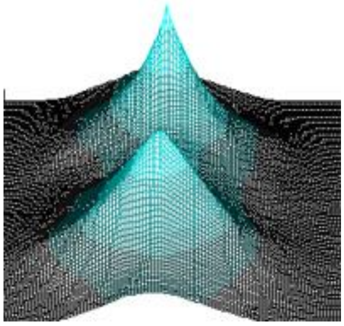
En la siguiente figura se muestra la señal a detectar, la escena de entrada y las señales de correlación obtenidas para un filtro adaptado clásico y un POF. Puede observarse un aumento importante del poder de discriminación.



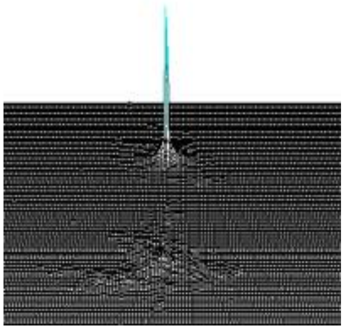
Señal a detectar



Escena de entrada



Filtro clásico



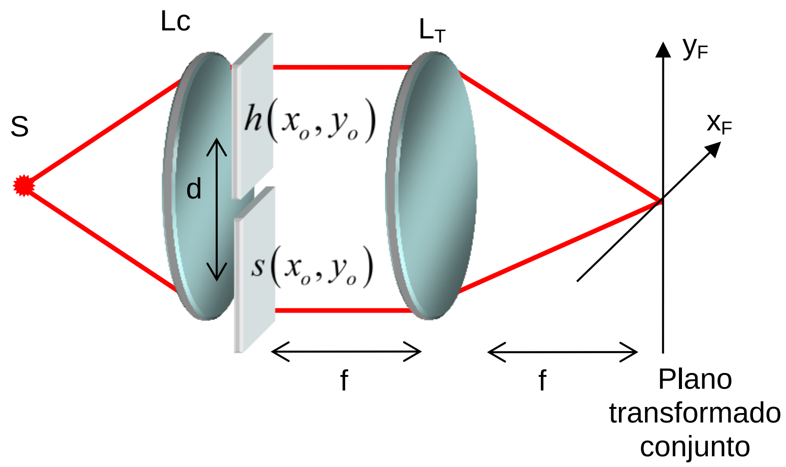
Filtro solo de fase

Hasta ahora hemos analizado el correlador de Vander Lugt y la generación de los filtros necesarios para su implementación. A continuación veremos otro tipo de correlador en el cual no hace falta el empleo de filtros en el plano de Fourier.

Correlador por transformada conjunta

El correlador por transformada conjunta (JTC por Joint Transform Correlator), ideado por Weaver y Goodman [C.Weaver, J.Goodman, *A technique for optically concolving two functions*, Appl. Opt. **5**, 1248 (1966)], ha sido diseñado originalmente para llevar a cabo la operación de convolución entre dos funciones y por ello resulta de interés para emplearlo en operaciones de filtrado.

La ventaja fundamental que presenta este dispositivo respecto del correlador 4f, se debe a que su operación está basada en un proceso de registro holográfico, y luego en un proceso de lectura empleando un haz plano, evitando así los inconvenientes de alineación. La diferencia fundamental entre ambos dispositivos, radica en que tanto la señal que se desea procesar, como la respuesta al impulso deseada, están presentes simultáneamente durante el proceso de registro, por lo que basta emplear un haz de lectura a fin de obtener la señal se salida deseada. En la figura se presenta un esquema del proceso de registro en este dispositivo. Un haz plano de longitud de onda λ ilumina el plano de entrada del correlador, donde se representan, separadas por una distancia *d* en el eje y, la escena $s(x_o, y_o)$ (señal a procesar) y la referencia $h(x_o, y_o)$ (que dará cuenta de los efectos de filtrado a efectuar sobre la escena).



La señal de entrada está representada entonces por

$$E_o(x_o, y_o) = s\left(x_o, y_o + \frac{d}{2}\right) + h\left(x_o, y_o - \frac{d}{2}\right)$$

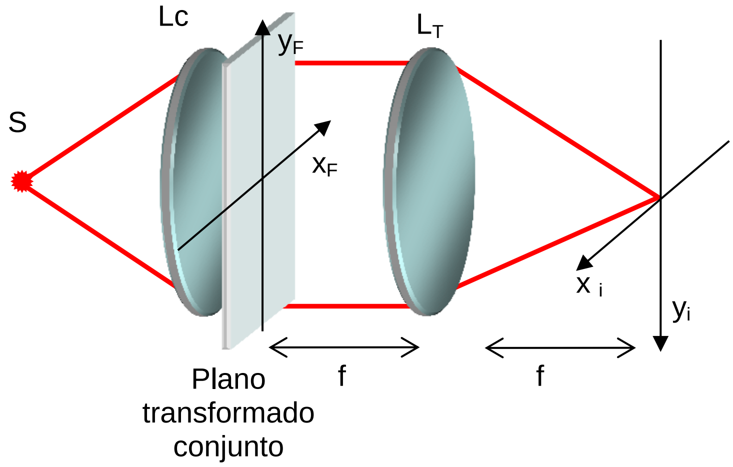
Una lente convergente L_T de distancia focal *f* es empleada para obtener en el *plano transformado conjunto JTP* (*Joint Transform Plane*), ubicado a distancia focal de la misma, la interferencia de las transformadas de Fourier de ambas señales,. Dicha transformada viene representada por

$$\mathcal{E}(x_F, y_F) = \frac{1}{\lambda f}\mathcal{S}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[i\pi\, y_F \frac{d}{\lambda f}\right] + \frac{1}{\lambda f} H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[-i\pi\, y_F \frac{d}{\lambda f}\right]$$

con lo cual la distribución de intensidad en ese plano está dada por

$$\begin{aligned} I(x_F, y_F) = \frac{1}{\lambda^2 f^2}\Bigg[& \left|H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right|^2 + \left|\mathcal{S}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\right|^2 + \\ & + \mathcal{S}^*\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right) H\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[-i\pi 2\, y_F \frac{d}{\lambda f}\right] + \\ & + H^*\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\mathcal{S}\left(\frac{x_F}{\lambda f}, \frac{y_F}{\lambda f}\right)\exp\left[i\pi 2\, y_F \frac{d}{\lambda f}\right]\Bigg] \end{aligned}$$

Esta distribución debe ser registrada en algún medio sensible a la luz situado en el plano transformado conjunto del sistema. Dicho registro constituye lo que denominaremos *transformada conjunta* del correlador. Suponemos que se efectúa un registro que posee una transmitancia lineal con la intensidad que llegó al mismo. El proceso de lectura de la transformada conjunta se esquematiza en la figura



De modo que si se la ilumina con un haz colimado de amplitud unitaria, tal como se esquematiza en la figura, sobre el plano final tendremos

$$
\begin{aligned}
E_i(x_i,y_i) = \frac{1}{\lambda f}\Big[& h(x_i,y_i)\otimes h^*(-x_i,-y_i) + s(x_i,y_i)\otimes s^*(-x_i,-y_i) + \\
& + h(x_i,y_i)\otimes s^*(-x_i,-y_i)\otimes \delta(x_i,y_i-d) + \\
& + h^*(-x_i,-y_i)\otimes s(x_i,y_i)\otimes \delta(x_i,y_i+d)\Big]
\end{aligned}
$$

Por lo tanto, la señal de salida está constituida en primer lugar, por los dos primeros términos que dan respectivamente la autocorrelación de la escena y la referencia, ambas distribuciones centradas en el origen y que no resultan de interés en este análisis. Obtenemos además las correlaciones cruzadas

$$
\begin{aligned}
& h(x_i,y_i)\otimes s^*(-x_i,-y_i)\otimes \delta(x_i,y_i-d) = \\
& = \int\int_{-\infty}^{\infty} h(x_o,y_o)\, s^*(x_o-x_i,y_o-y_i+d)\, dx_o dy_o
\end{aligned}
$$

$$h^*\left(-x_i,-y_i\right)\otimes s\left(x_i,y_i\right)\otimes\delta\left(x_i,y_i+d\right)=$$
$$=\int\int_{-\infty}^{\infty} s\left(x_o,y_o\right)h^*\left(x_o-x_i,y_o-y_i-d\right)dx_o dy_o$$

Una ubicada en las coordenadas (0, -d) y la otra en (0, d). Así, una vez definida la función transferencia asociada a un proceso particular, sólo es necesario calcular la respuesta al impulso correspondiente e invertirla y conjugarla para emplearla como señal de referencia en el correlador. En particular, para aplicaciones de reconocimiento, donde se requiere una señal de salida dada por la correlación entre la escena y la referencia, basta con representar ambas señales en el plano de entrada del correlador. Notemos que en esa situación, las dos señales de correlación entre la escena y la referencia que se obtienen, son una la reflexión especular de la otra.

Este correlador puede interpretarse como una experiencia de Young generalizada en la cual en vez de simples aberturas iguales tenemos objetos. Si estos objetos (alguno de la escena y el de referencia) son iguales entonces en el plano transformado los espectros se superpondrán e interferirán dando lugar a franjas. Luego al iluminar estas franjas, las mismas se comportarán como una red de difracción y darán lugar a órdenes (picos de correlación). Si los objetos no son semejantes los espectros no coinciden y por lo tanto al no interferir no se registrarán franjas.

Estos correladores poseen la ventaja de no utilizar filtros, cuyo posicionado siempre es crítico, y además son aptos para procesado en tiempo real ya que la entrada puede representarse en una pantalla de cristal líquido, el espectro se puede captar con una CCD y luego representarse en una pantalla de cristal líquido para finalmente captar la correlación con una CCD.

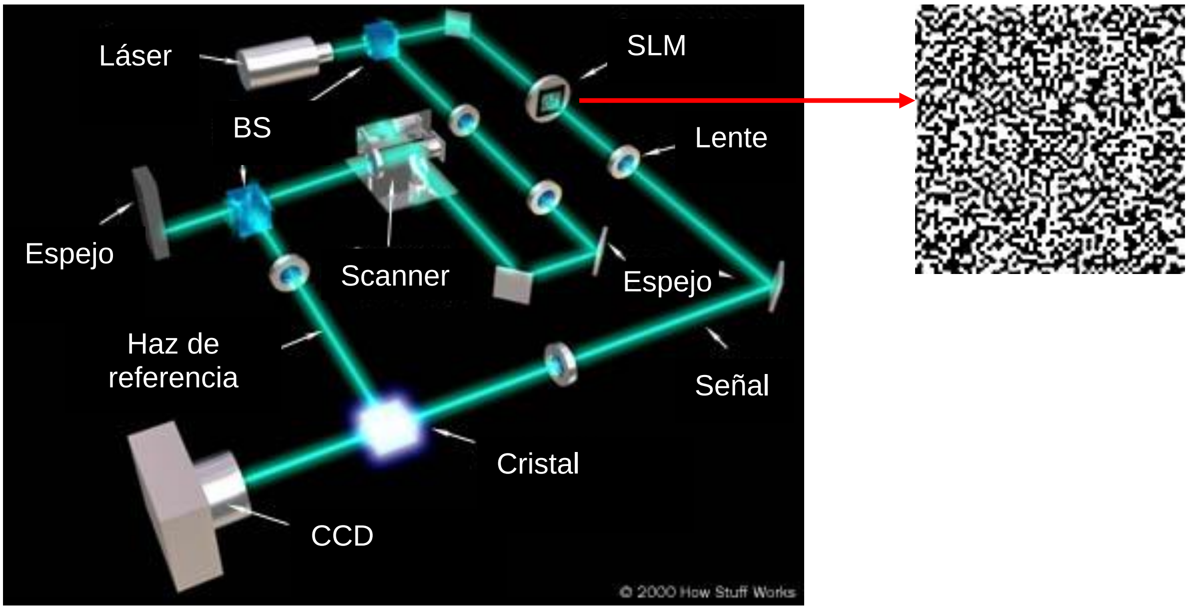
Cabe destacar que cuando este tipo de correlador es utilizado para el reconocimiento de formas su poder de discriminación es equivalente al del filtro adaptado. Sin embargo existen técnicas de procesado en la señal de entrada y en el plano de registro que permiten llevar a cabo un filtrado equivalente al que se realiza con un filtro POF.

MEMORIAS HOLOGRÁFICAS

El almacenado óptico de datos ha sido un tema de gran interés durante las últimas décadas. Desde la aparición comercial de CDs y DVDs su uso creció exponencialmente así como la necesidad de aumentar su capacidad de almacenamiento y su velocidad de acceso. Los primeros CDs fueron evolucionando a elementos multicapas y a medios donde el grabado y lectura se hacen con longitudes de onda cada vez más cortas con el propósito de disminuir el tamaño del spot del láser utilizado para esas tareas. Sin embargo en todos estos dispositivos el proceso de grabado y lectura es serial, esto es, la información es almacenada y leída bit a bit.

Un desarrollo alternativo es el de las memorias holográficas basadas en el efecto de difracción por redes de volumen. Van Heerden [*Theory of optical information storage in solids*, Appl. Opt. **2**, 393 (1963)] propuso por primera vez que múltiples hologramas podían ser superpuestos o *multiplexados* en un medio de volumen. Luego a cada holograma podía accederse selectivamente a través de las condiciones de difracción de Bragg.

Un sistema típico para el registro y reconstrucción de memorias holográficas es el siguiente

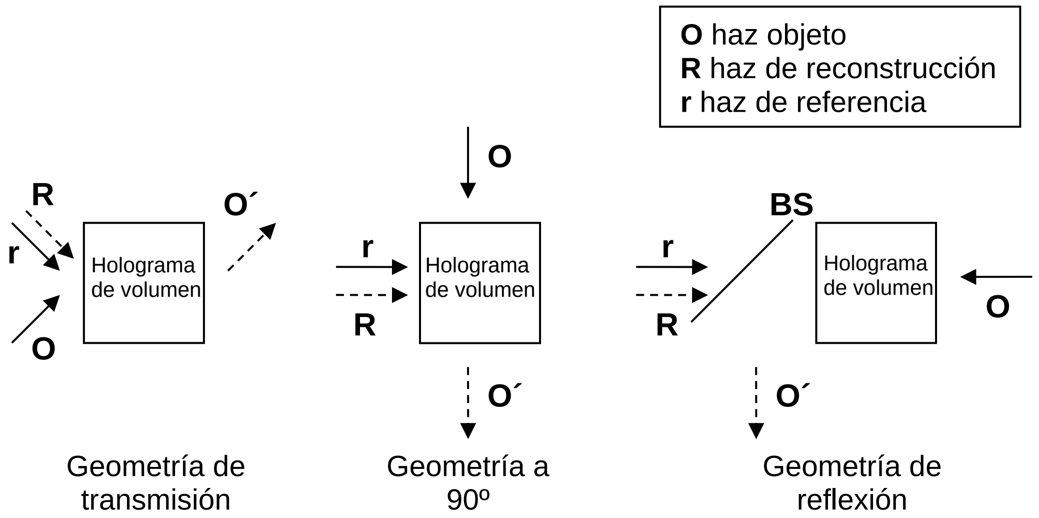


Acá el objeto consiste en un modulador espacial de luz en el que se representa en forma binaria una *página* de información. Esta entrada particular puede pensarse como una distribución de unos y ceros mediante los cuales se codifican los datos que se desean almacenar. Un sistema de lentes, dependiendo del método utilizado para grabar las memorias, forma o bien la imagen del SLM sobre el cristal o bien su transformada de Fourier. Esta página se va cambiando sucesivamente, constituyendo cada una de ellas uno de los hologramas multiplexados. A su vez el haz de referencia también cambia ligeramente de un registro a otro de manera que nuevas condiciones de Bragg se establecen para cada uno de ellos. La reconstrucción de la página deseada sólo se llevará a cabo cuando el haz empleado para ello cumpla con las condiciones de Bragg apropiadas. Esta imagen reconstruida es registrada por un detector tal como una CCD o un CMOS.

Queda claro acá cómo el proceso de escritura y lectura de la información (bits) se realiza en paralelo y no en forma secuencial como sucede en un CD o DVD por lo que resulta mucho más rápido.

El método empleado para realizar el multiplexado varía y puede ser por cambios de longitudes de onda o, como se muestra en el dibujo anterior, por medio de un sistema que cambie ligeramente el ángulo de incidencia del haz de referencia.

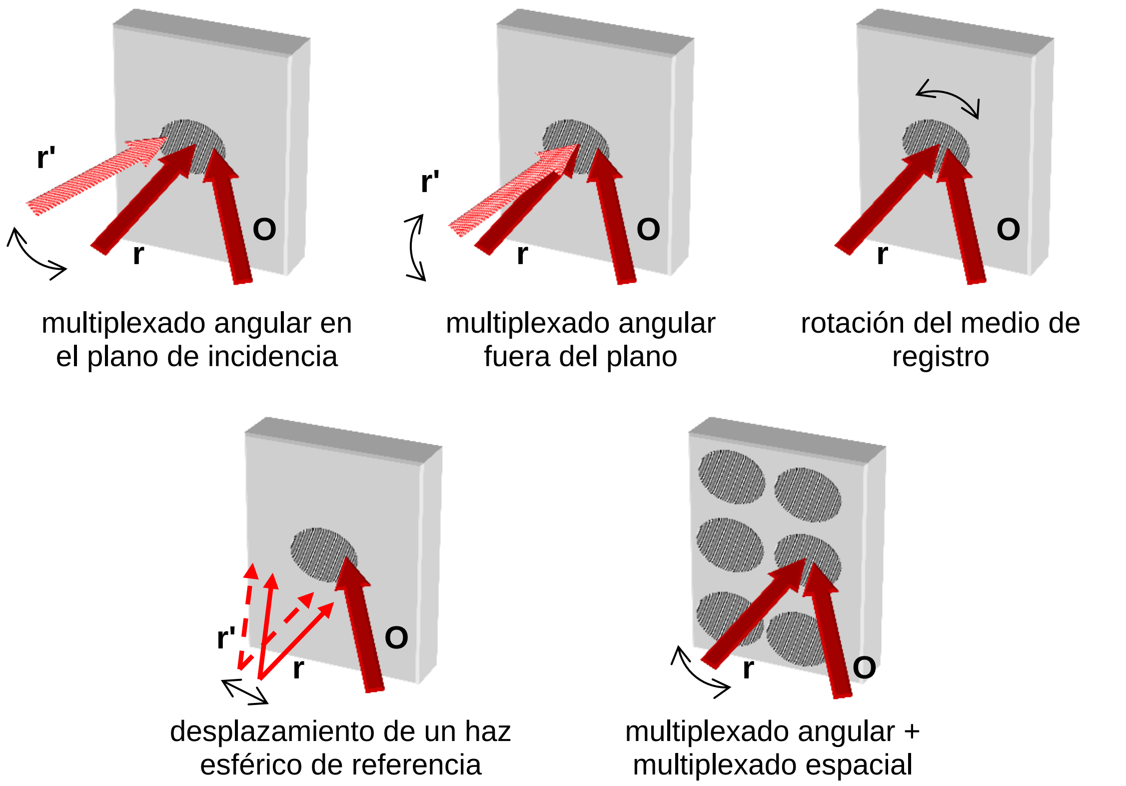
A continuación se muestran algunas geometrías de registro y reconstrucción



En general la geometría de reflexión se utiliza para el multiplexado en longitudes de onda, ya que si recuerdan de cuando estudiamos hologramas 3D, esa arquitectura brinda elementos que se comportan casi como un filtro interferencial.

Las dos primeras disposiciones se utilizan principalmente para multiplexado angular.

En la siguiente figura se esquematizan además algunas variantes de multiplexado en geometría de transmisión pero muchos de ellos también son aplicables a otras arquitecturas. En el primer caso el haz de referencia es plano y su ángulo de incidencia cambia de registro a registro en forma coplanar con el haz objeto. En el segundo caso el haz de referencia también cambia angularmente pero ahora fuera del plano de incidencia. En el tercer caso los ángulos de incidencia del haz objeto y de referencia quedan inalterados pero se gira el medio de registro. Estos dos últimos métodos se llaman fractálicos ya que en ellos no se produce una alteración significativa de las condiciones de Bragg. Estas si se modifican en el caso del primer dibujo y del cuarto, en donde un haz de referencia esférico se desplaza, en el plano de incidencia, entre exposición y exposición. En el quinto caso se muestra una combinación de uno de los métodos anteriores con un desplazamiento del medio de registro. De esta forma tenemos diversas zonas expuestas del medio de registro y en cada una de ellas varios hologramas multiplexados.



Los medios utilizados para la implementación de memorias holográficas van desde cristales de niobato de litio dopado con hierro de algunos milímetros de espesor hasta fotopolímeros de centenas de micrones de espesor. En el caso de los cristales, al ser tan gruesos, permiten grabar un gran número de hologramas ya que las condiciones de Bragg que se imponen son muy restrictivas. Esto es, para cada pequeña variación del ángulo de referencia es posible almacenar y reconstruir una página de datos distinta. Cuál es el límite? Entre otros factores viene determinado, además de por el tamaño de la página, por la eficiencia de difracción ya que cuantos más hologramas se hallen grabados cada uno de ellos difractará menos luz. En un cristal cúbico de 1 cm de lado llegaron a grabarse 10000 páginas de 320X220 bits [G.Burr, F.Mok, D.Psaltis *Storage of 10000 holograms in LiNbO$_3$*, CLEO (1994)]. Esta cantidad de información si bien es asombrosa, aún es menor que la almacenada en un CD. Así el proceso de acceso a los datos, al ser en forma paralela y poder realizarse por ejemplo variando el

ángulo de incidencia con moduladores acusto-ópticos, es mucho más rápido que el empleado en la tecnología convencional de CDs pero su poder de almacenamiento es menor. Alternativamente comenzaron a desarrollarse los llamados *discos holográficos 3D* que, en general, usan un fotopolímero cuyo espesor es mucho menor que el de un cristal y en consecuencia almacenan menos cantidad de información por área, pero que fácilmente pueden grabarse en distintas zonas tal como se esquematizó en el último dibujo de la figura anterior. Por otra parte la eficiencia de difracción es mayor. Si bien aún son dispositivos no comerciales, se demostró en los últimos años que en un dispositivo del tamaño de un CD se pueden almacenar del orden de 6 Terabits. Podemos comparar esta cantidad con la capacidad de un Blu-Ray (25 Gigabits por capa @ 405 nm), un DVD (4.7 Gigabits por capa @ 650 nm) y un CD (700 Megabits @ 780 nm). Cabe destacar que la velocidad de grabación y lectura de un HD-DVD o Blu-Ray es de 36.5 Mb/s y que la de un disco holográfico (HVD) es de 1Gb/s.
Un análisis detallado de este tema puede encontrarse en [*Holographic data storage* H.Coufal, D.Psaltis, G.Sincerbox , Springer (2000)]

X. DIGITALIZACIÓN DE IMÁGENES CLASE 22

Como hemos visto hasta el momento en los procesadores en donde se utilizan elementos generados enteramente mediante procesos ópticos, las funciones matemáticas que representan la transmisión de las escenas y los filtros son de naturaleza continua. Vimos que un filtro como el de Vander Lugt puede ser generado holográficamente pero si se desea implementar alguno de los otros filtros mencionados posteriormente se debe recurrir a técnicas computacionales ya que ninguno de ellos puede ser sintetizado ópticamente. Lo mismo ocurre si introducimos no linealidades en un JTC o utilizamos un dispositivo optoelectrónico que permita trabajar a tiempo real.
Ahora bien, las computadoras actúan sobre señales del tipo numéricas, por ello antes de procesar la imagen, la misma debe ser digitalizada, esto es representada mediante un arreglo de números. Analicemos primero entonces, como una imagen pasa de ser continua a discreta y que condiciones deben cumplirse para que en la etapa de digitalización no se pierda información. Como paso siguiente, en general, la imagen es procesada en el plano de frecuencias (por ejemplo para la generación del holograma),

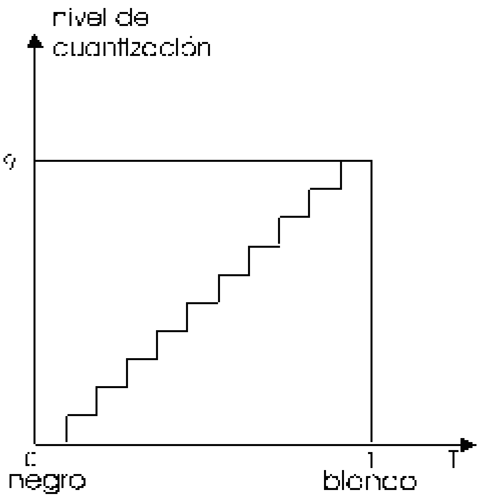
por lo que se hace necesario también analizar cómo se transforma Fourier una señal digital.

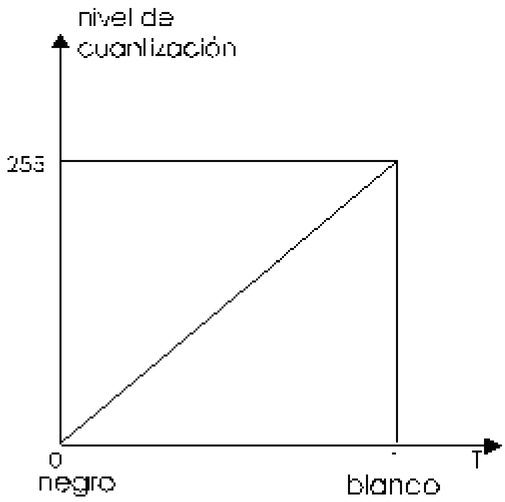
DIGITALIZACIÓN

Para que una imagen del tipo analógico pueda pasar de su estado continuo a una representación numérica en la computadora deben realizarse ciertos pasos intermedios de transformación.

En primer lugar la imagen debe ser registrada por algún dispositivo electrónico o generada numéricamente y luego debe almacenarse en la memoria de una computadora. La resolución espacial de esta imagen dependerá del número de píxeles que contenga, es decir de cómo se realizó el muestreo de la señal. En la siguiente figura se muestra un ejemplo de una imagen digitalizada usando distintos tamaños de píxel.

Otro parámetro a tener en cuenta al analizar la fidelidad de la imagen es el nivel de cuantización, es decir, cuales serán los distintos valores numéricos que se le asignarán a cada píxel de la imagen. Para esto se debe seguir alguna convención, en la que se determine una cantidad finita de niveles equidistantes. Para ejemplificar esto último se puede pensar en una imagen con distintos tonos de gris, si la transmisión T es 0 para negro y 1 para blanco, una convención sería cuantizar con 10 niveles, donde el 0 corresponda a T=0 y el 9 a T=1. Otra convención sería cuantizar en 256 niveles, donde el nivel 0 sigue correspondiendo a T=0 pero ahora es el nivel 255 el que corresponde a T=1. En la siguiente figura se esquematiza la cuantización de ambos casos. En el primero la computadora puede representar la imagen con 10 niveles de gris, en cambio en el segundo cuenta con 256.





El nivel de cuantización dependerá del número de bits con el que trabaje el conversor analógico digital. En cuanto al número de píxeles que deben escogerse para que la pérdida de información se reduzca al mínimo o sea inexistente, existe un teorema que contempla este punto en base al estudio del proceso de discretización de una función continua. Recibe el nombre de teorema del muestreo o de Shannon y es el tema a tratar a continuación.

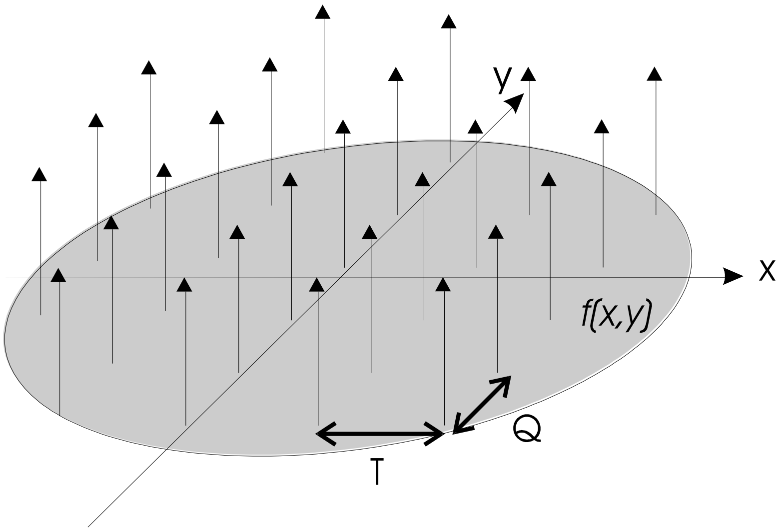
TEOREMA DEL MUESTREO

Sea *f*(*x,y*) una función continua cuya transformada de Fourier es finita. El objetivo es representar esta función a través de un arreglo de puntos.

Muestrearla de forma regular a un intervalo determinado involucra el hecho de aceptar de la función solo aquellos valores que coinciden con múltiplos de este intervalo. Matemáticamente, esta operación equivale a multiplicarla por la función peine de Dirac:

$$p(x,y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta(x-nT)\,\delta(y-mQ)$$

donde *T* es el intervalo de muestreo en la dirección *x* y *Q* es el intervalo en la dirección *y*.



Así la función *f*(*x*,*y*) muestreada se expresará de la forma:

$$f_M(x,y)=\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}f(x,y)\delta\left(x-nT\right)\delta\left(y-mQ\right)$$

Tomando transformada de Fourier a ambos miembros se obtiene

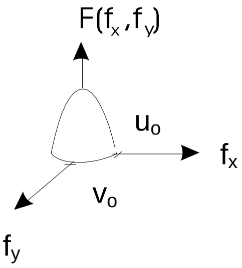
$$F_M\left(f_x,f_y\right)=F\left(f_x,f_y\right)\otimes P\left(f_x,f_y\right)$$

donde $P(f_x,f_y)=\dfrac{1}{TQ}\displaystyle\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\delta\left(f_x-\frac{n}{T}\right)\delta\left(f_y-\frac{m}{Q}\right)$ corresponde a la transformada de la función peine de Dirac. Reemplazando en la expresión anterior queda:

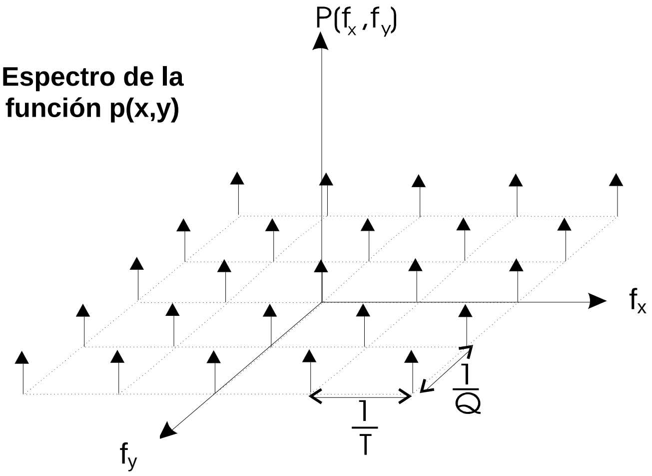
$$\begin{aligned}F_M(f_x,f_y)&=\frac{1}{TQ}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}F(f_x,f_y)\otimes\delta\left(f_x-\frac{n}{T}\right)\delta\left(f_y-\frac{m}{Q}\right)\\&=\frac{1}{TQ}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}F\left(f_x-\frac{n}{T},f_y-\frac{m}{Q}\right)\end{aligned}$$

Esto indica que el espectro de la función muestreada se obtiene repitiendo el espectro de la función sin muestrear en los puntos (*n/T*, *m/Q*) en el espacio de frecuencias. En la siguiente figura se ejemplifica lo obtenido hasta ahora.

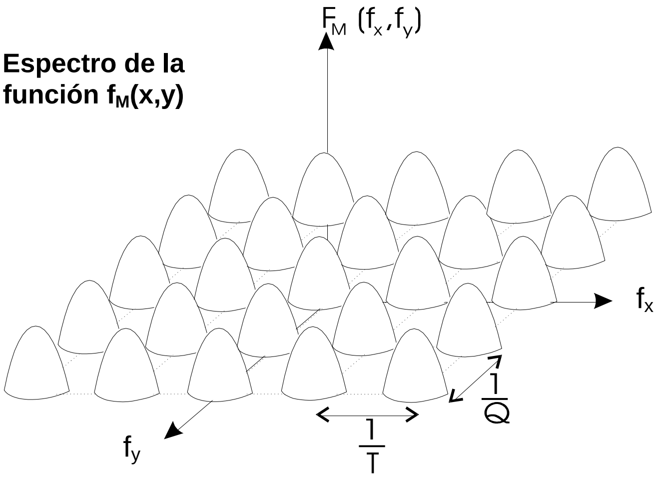
Espectro de la función f(x,y)



Espectro de la función p(x,y)



Espectro de la función f_M(x,y)

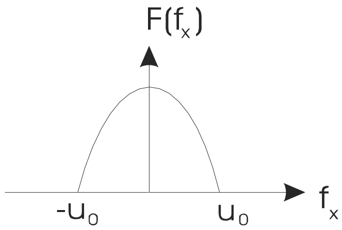


Analizando el espectro de la función muestreada se puede ver que debe existir alguna relación entre el tamaño de $F(f_x, f_y)$ y las distancias $1/T$ y $1/Q$ entre los distintos espectros como para que estos no se superpongan. Si u_o y v_o son las máximas frecuencias de $F(f_x, f_y)$ en las direcciones f_x y f_y respectivamente, las situaciones planteadas son las siguientes:

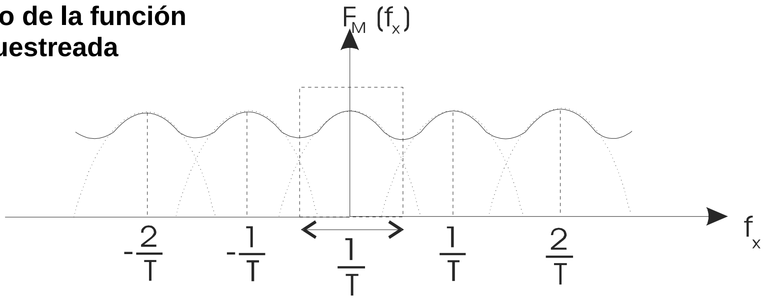
- Si los intervalos de muestreo son tal que $1/T \geq 2u_o$, $1/Q \geq 2v_o$ entonces los espectros no se superpondrán, dando lo que se muestra en la figura anterior.

- Si por el contrario, los intervalos son tal que $1/T < 2u_o$, $1/Q < 2v_o$ entonces los espectros tendrán porciones superpuestas y se formará lo que se llama *aliasing*[1], haciendo imposible la reconstrucción de la señal. Esta situación está representada en forma unidimensional en la siguiente figura

Espectro de la función



Espectro de la función muestreada



Para que la superposición no ocurra y se pierda información es necesario entonces, pedir que se cumpla la condición $\frac{1}{T} \geq 2u_o \qquad \frac{1}{Q} \geq 2v_o$

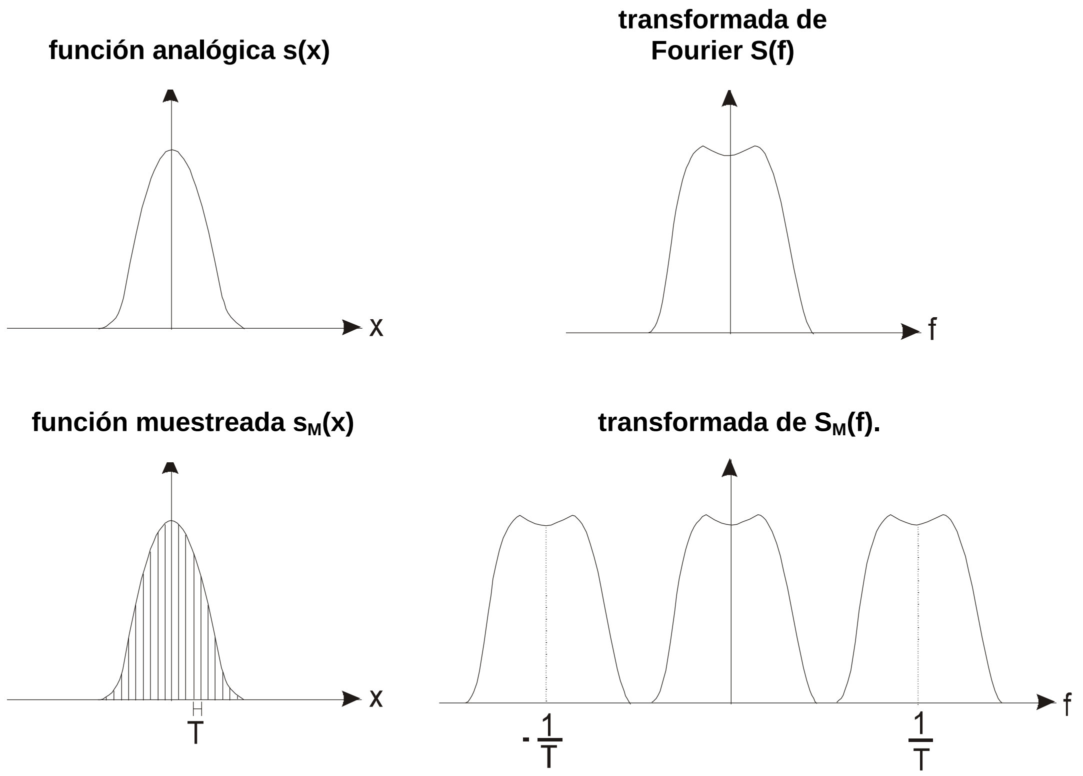
Esto se conoce como condición de Shannon y se expresa diciendo que para muestrear una función sin perder información, la mínima frecuencia de muestreo debe ser dos veces la frecuencia más alta contenida en la señal.

TRANSFORMADA DE FOURIER DISCRETA DE SEÑALES DIGITALES

Para simplificar tomemos una función unidimensional $s(x)$ y con transformada de Fourier $S(f)$. Si esta función es muestreada a intervalo T, m veces, se obtendrá otra señal, sea $s_M(x)$, cuya transformada de Fourier será $S_M(f\,)$ que corresponde a $S(f\,)$

[1] El término *aliasing* se refiere a un falso traslado. Es decir, si se considera uno solo de los espectros, la porción del mismo que queda afuera del rango de frecuencias determinado por los intervalos de muestreo se dice que es “falsamente trasladada” dentro de ese rango. La expresión “falsamente” viene a cuenta de que en realidad estas frecuencias espurias presentes en el espectro provienen de los componentes vecinos.

repetida con período 1/*T* y multiplicada por 1/*T* como indica el teorema del muestreo. En la siguiente figura se grafican las cuatro funciones.



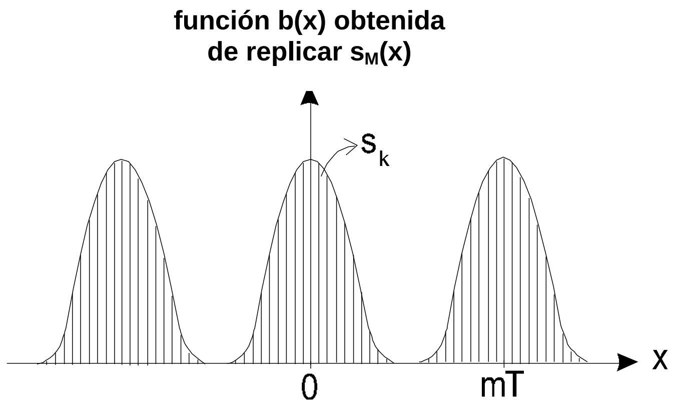
Si ahora a la señal $s_M(x)$ se la repite con período *mT* se obtendrá una función *b*(*x*) dada por

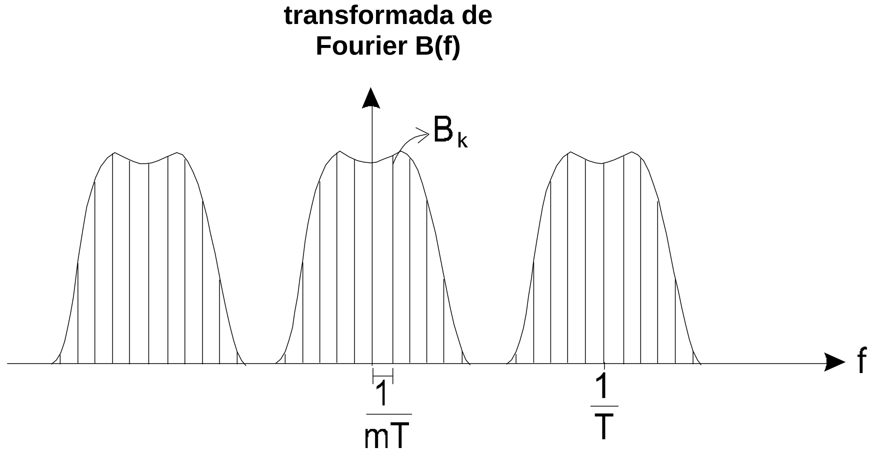
$$b(x) = \sum_{k=-\infty}^{\infty} s_M(x) \otimes \delta(x - kmT)$$

cuya transformada de Fourier *B*(*f*), corresponde a $S_M(f\ )$ multiplicando por 1/*mT* y muestreando a un intervalo 1/*mT*.

$$B(f) = \frac{1}{mT} \sum_k S_M(f)\,\delta\left(f - \frac{k}{mT}\right)$$

Ambas funciones se grafican en la siguiente figura.





Esta transformada se la puede expresar también a través de los valores discretos B_k

$$B(f) = \sum_{k=-\infty}^{\infty} B_k \delta\left(f - \frac{k}{mT} \right)$$

Estos valores B_k corresponden justamente a la discretización de $B(f)$. Para hallar la expresión de B_k, entonces habrá que calcular $S_M(f)$ a partir de la señal digitalizada $s_M(x)$

$$s_M(x) = \sum_n s(nT)\delta(x - nT)$$

Transformando Fourier y sabiendo que $\mathbb{F}\left(\delta\left(x - nT\right)\right) = \exp\left(-2\pi i n T f\right)$ se tiene que

$$S_M(f) = \sum_n s(nT)\exp(-2\pi i n T f)$$

que corresponde a la transformada de Fourier de la señal digital $s_M(x)$ (debemos recordar que *s(nT)* son simplemente números). Reemplazando esta expresión en

$$B(f) = \frac{1}{mT}\sum_k S_M(f)\delta\left(f - \frac{k}{mT}\right)$$

se obtiene

$$\begin{aligned}B(f) &= \frac{1}{mT}\sum_k\left[\sum_n s(nT)\exp(-i2\pi nTf)\right]\delta\left(f - \frac{k}{mT}\right)\\ &= \frac{1}{mT}\sum_k\sum_n s(nT)\exp\left(-i2\pi k\frac{n}{m}\right)\delta\left(f - \frac{k}{mT}\right)\end{aligned}$$

Comparando esta expresión con la anteriormente hallada para *B(f)* en término de los B_k se obtiene que

$$B_k = \frac{1}{mT}\sum_n s(nT)\exp\left(-i2\pi k\frac{n}{m}\right)$$

Esta ecuación representa la transformada discreta de Fourier y relaciona los *m* valores muestreados $s(nT)$ con los *m* valores discretos B_k.

Como B(f) es periódica sólo se necesita calcular la componente central y de esta forma, se obtienen los *m* B_k. La expresión discreta de la antitransformada está dada por la ecuación:

$$s(nT) = T\sum_{k\,(en\;periodo\,1/T)} B_k\exp\left(i2\pi k\frac{n}{m}\right)$$

Estas dos últimas ecuaciones expresan el hecho de que $s(nT)$ (s_k en la figura) y B_k son transformada y antitransformada discreta respectivamente.

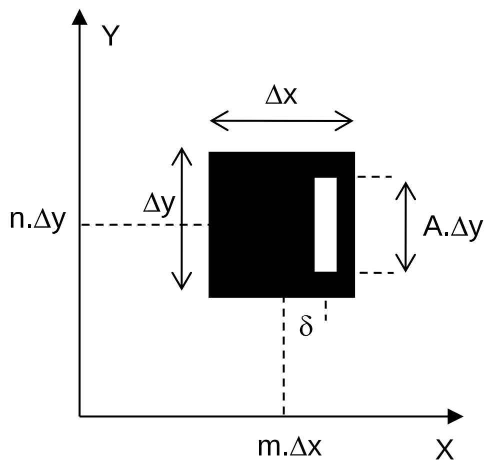
Una vez establecida la forma en que una imagen es expresada como un arreglo numérico y luego llevada al plano de Fourier de manera también digital, el paso siguiente sería generar el holograma computacionalmente. Existen varias técnicas para esto último, algunas de las cuales veremos a continuación.

HOLOGRAMAS GENERADOS POR COMPUTADORA (CGH)

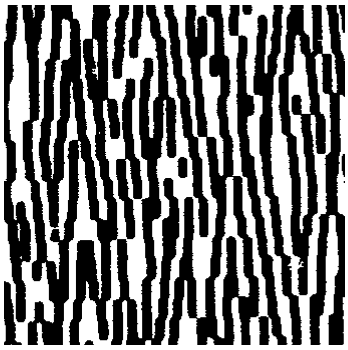
La generación de un holograma por métodos computacionales requiere básicamente de tres pasos, cálculo, representación gráfica y transferencia a un medio físico.
La obtención del valor de amplitud y fase en cada punto (píxel) del holograma lo hemos visto en el punto anterior. Debemos recordar que en general se trabaja con hologramas tipo Fourier, por lo que una vez obtenida la transformada de Fourier discreta debemos analizar como codificar los valores hallados.
Desde el punto de vista histórico, y considerando los sistemas de transferencia existentes, resultaba muy difícil representar fielmente un holograma digital en un medio de amplitud utilizando una impresión que no fuese binaria. Así surgieron básicamente tres métodos para codificar la amplitud y fase correspondientes a cada punto del holograma:

Método de fase de desvío propuesto por Lohmann

Los trabajos en los que se describe este método son [B.Brown, A.Lohmann *Complex spatial filtering with binary mask* Appl. Opt. **5**, 967 (1966)], [A.Lohmann, D.Paris, *Binary Fraunhofer Holograms, generated by computer*, Appl. Opt. **6**, 1739 (1967)]
En este método tenemos una malla de M x N píxeles (número de puntos en los que fue muestreado el objeto) de los cuales debe codificarse la amplitud y la fase. Dicha codificación se realiza mediante aberturas rectangulares cuyo tamaño, dentro del píxel, determina la amplitud y su posición la fase.



El píxel (m, n) tiene, como todos, tamaño Δx, Δy y su posición es ($m\Delta x$, $n\Delta y$); la cantidad de luz que deja pasar depende del alto de la abertura $A\Delta y$ (el ancho se deja fijo en este método) y la fase está determinada por el desplazamiento δ de la abertura respecto al centro del píxel. En la siguiente figura se muestra una porción ampliada de un holograma de este tipo



Representación aditiva de números complejos

Este método se halla detallado en las publicaciones [W.Lee *Sampled Fourier transform hologram generated by computer*, Appl. Opt. **9**, 639 (1970)], [C.Burckhardt, *A simplification of Lee´s method of generating holograms by computer*, Appl. Opt. **9**, 1949 (1970)].
El mismo se basa en que un número complejo puede descomponerse como

$$\mathbb{C} = A\exp(i\varphi) = A_{-1}\exp\left(-i\frac{2\pi}{3}\right) + A_o + A_1\exp\left(i\frac{2\pi}{3}\right)$$

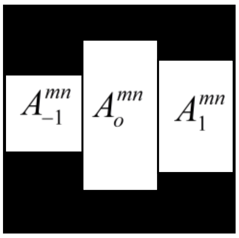
con

$$A_{-1} = \frac{\mathbb{C}\exp\left(i\frac{2\pi}{3}\right) + \mathbb{C}^*\exp\left(-i\frac{2\pi}{3}\right)}{3}$$

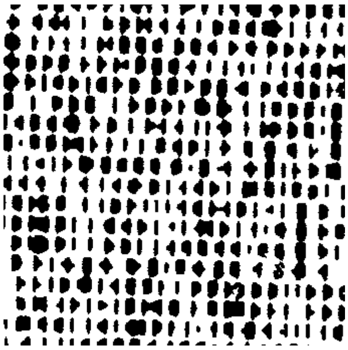
$$A_o = \frac{\mathbb{C} + \mathbb{C}^*}{3}$$

$$A_1 = \frac{\mathbb{C}\exp\left(-i\frac{2\pi}{3}\right) + \mathbb{C}^*\exp\left(i\frac{2\pi}{3}\right)}{3}$$

En esta técnica también se divide el holograma en M x N celdas subdivididas en tres aberturas con posición y ancho fijo, tal como se ilustra en la figura



A continuación se muestra una porción ampliada de un holograma sintetizado por este método

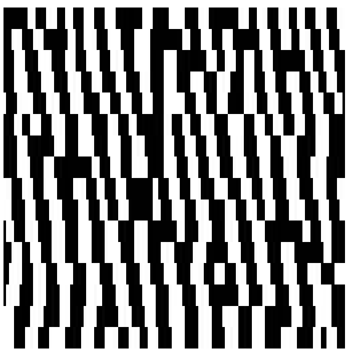


Codificación de hologramas binarizados

Este método se describe en la publicación [W.Lee, *Binary Computer Generated Holograms* Appl. Opt. **18**, 3661 (1979)]
En este caso se introduce en los cálculos una onda de referencia al igual que en un holograma verdadero, así se obtiene una función del tipo

$$t(x,y) \propto r^2 + A^2(x,y) + 2r.A(x,y)\cos\left[2\pi\alpha x - \varphi(x,y)\right]$$

donde $A(x,y)\exp\left[i\varphi(x,y)\right]$ representa la onda objeto y $r\exp\left[i2\pi\alpha x\right]$ la onda de referencia. Luego se binariza este interferograma digital mediante un algoritmo que encuentra las posiciones de las franjas.

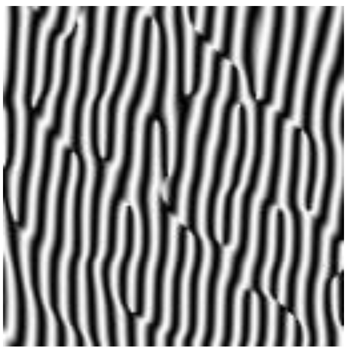


Una vez generados estos hologramas deben transferirse a un medio físico que permita utilizar el filtro en un procesador óptico. Hay medios más o menos sofisticados que permiten llevar a cabo este propósito. Una de las formas más elementales en la que esto puede hacerse es realizar una impresión del holograma en papel y luego realizar una fotorreducción. Otro, un poco más elaborado, es realizar directamente la impresión utilizando una impresora de artes gráficas de alta resolución (aproximadamente 3000 dpi) que proporciona el filtro sobre película fotográfica.

Como dato curioso, existe una fotografía en donde se ve a Lohmann y sus colaboradores, pintando con pincel un holograma sobre una pared para después reducirlo fotográficamente. Recordemos que cuando Lohmann propuso su método para sintetizar CGHs no existían impresoras y prácticamente no había computadoras. Constituye todo un desafío estudiar y proponer la utilización de un elemento para el que aún no están desarrollados los métodos de implementación.

Independientemente de cuál sea el método de transferencia estos hologramas binarios presentan algunos inconvenientes ya que si se desea superponer hologramas para hacer un filtro compuesto cada vez hay más zonas oscuras que bloquean las claras. Por otra parte al ser las franjas representadas por rectángulos, es muy difícil elegir arbitrariamente la frecuencia y dirección de la portadora.

El advenimiento de nuevos métodos de transferencia permitió la representación de CGHs directamente en niveles de gris [C. Iemmi, S. Ledesma, J. Campos, M. Villarreal, *Gray-level computer generated hologram filters for multiple-object correlation*, Appl. Opt. **39**, 1233 (2000)]. Esto hace que sean verdaderos interferogramas que pueden sumarse y que permiten elegir arbitrariamente, a través de la portadora, donde irá a parar la correlación. Esta técnica permite implementar el procesado multicanal y es aplicable ya sea utilizando material fotográfico (mediante máquina de copiado AGFA Alto) o directamente en pantallas de cristal líquido. A continuación se ve una porción ampliada de un holograma generado en niveles de gris



HOLOGRAFÍA DIGITAL

Con la aparición de elementos optoelectrónicos de alta resolución se comenzó a estudiar el reemplazo de las técnicas ópticas convencionales por procesos híbridos. Así surgió la holografía digital, a partir de métodos interferométricos y el uso de una cámara CCD, que permite almacenar digitalmente la información correspondiente a la amplitud y fase de una onda difractada por un objeto y luego realizar una reconstrucción numérica en una computadora [U. Schnars y W. Jüptner *Direct Recording of Holograms by a CCD target and numerical reconstruction*, Appl. Opt. **33**, 179, (1994), Yamaguchi, Zhang, *Phase-shifting digital holography*, Opt. Lett. **22**, 1268, (1997)] o simplemente generar la imagen a través de una pantalla LCD [L.Gonçalves Neto, D.Roberge, Y.Sheng, *Programmable optical phase-mostly holograms with coupled-mode modulation liquid-crystal television*. . Appl. Opt. **34**, 1944, (1995)].
Tal como hemos visto los procesos convencionales de registro de hologramas se llevan a cabo, en general, sobre materiales fotosensibles que requieren de algún tipo de revelado. Si bien existen materiales que permiten un proceso en tiempo real, tales como los cristales fotorrefractivos, todos ellos presentan el inconveniente de requerir del medio en donde se realizo el registro para efectuar a posteriori su reconstrucción.
Una idea atrayente, para evitar la presencia física del medio de registro en el momento de la reconstrucción, es la adquisición del holograma mediante una cámara CCD (Charged Coupled Device) y su posterior digitalización para poder reconstruirlo computacionalmente mediante integrales de difracción. Esta matriz de números también podría representarse ópticamente si se dispusiese de un medio capaz de representar los valores de dicha matriz píxel a píxel.
De acuerdo a lo que hemos visto al estudiar holografía el hecho de utilizar una onda portadora fuera de eje, si bien permite separar el orden cero de la imagen primaria y la conjugada, impone serias restricciones sobre la resolución espacial del medio a utilizar
La resolución actual de las cámaras CCD, si bien está muy por debajo de la de una placa holográfica, permite registrar hologramas en los cuales el haz de referencia se encuentra prácticamente alineado con el haz objeto. Este hecho, en principio, traería los inconvenientes propios de los hologramas en línea. Veamos cómo pueden superarse.
Como vimos, al no contar con medios que registren la fase de un haz luminoso sino sólo su intensidad, es necesario recurrir a la interferencia del haz objeto con un haz de referencia para que en este diagrama quede codificada la fase. Sin embargo si mediante una técnica apropiada fuese posible conocer la distribución de fase del haz objeto y no se necesitase de esta codificación, entonces sería posible reconstruir sólo

dicho haz sin la presencia del orden cero de difracción, ni del haz conjugado. Esto es, el holograma sería lo que se conoce como un kinoform.

Existe una técnica de interferometría, llamada por corrimiento de fase [K. Creath *Phase-measurement Interferometry Techniques*, E. Wolf, Progress in Optics XXVI Cap. 5, (1988)], que permite obtener la distribución de amplitud y fase del haz de interés. En este método se utiliza un interferómetro por división de amplitud, tipo Michelson, en uno de cuyos brazos hay un espejo que proporciona el haz de referencia y en el otro se halla el objeto que proporciona la onda a la cuál se le quiere medir la distribución de fase. La imagen de interferencia es adquirida por una cámara CCD. A diferencia de la técnicas interferométricas usuales en este método la fase del haz de referencia es desplazada, en forma controlada, con respecto a la fase del haz a testear, y la intensidad resultante es registrada para diferentes valores de corrimiento relativo. Algunas técnicas cambian la fase escalonadamente entre mediciones y otras integran la intensidad a medida que la fase es corrida. La cantidad de corrimientos también varía.

Se necesita un mínimo de tres mediciones, ya que son tres las incógnitas a develar, surgidas de la ecuación para la intensidad en el patrón de interferencia:

$$I = I_0[1+\gamma\cos(\varphi)]$$

La intensidad I_0 máxima, la modulación de las franjas de interferencia γ y la fase ϕ del frente de onda.

Uno de los métodos más utilizados es el que se realiza a partir de 4 variaciones de la fase controlada $\alpha = 0, 1/2\pi, \pi$ y $3/2\pi$. Así la distribución de intensidad, registrada por la CCD, para cada uno de los interferogramas se pueden escribir de la siguiente forma:

$$I_1(x,y) = I_0(x,y)\left\{1+\gamma\cos\left[\phi(x,y)\right]\right\}$$

$$I_2(x,y) = I_0(x,y)\left\{1+\gamma\cos\left[\phi(x,y)+\frac{1}{2}\pi\right]\right\} = I_0(x,y)\left\{1-\gamma\sin\left[\phi(x,y)\right]\right\}$$

$$I_3(x,y) = I_0(x,y)\left\{1+\gamma\cos\left[\phi(x,y)+\pi\right]\right\} = I_0(x,y)\left\{1-\gamma\cos\left[\phi(x,y)\right]\right\}$$

$$I_4(x,y) = I_0(x,y)\left\{1+\gamma_0\cos\left[\phi(x,y)+\frac{3}{2}\pi\right]\right\} = I_0(x,y)\left\{1+\gamma_0\sin\left[\phi(x,y)\right]\right\}$$

La modulación y fase en cada punto está dada por:

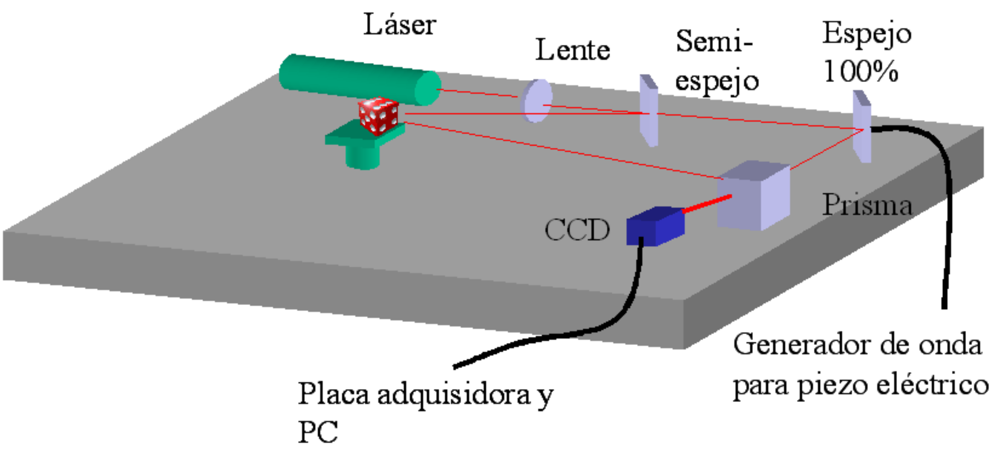
$$\gamma(x,y)=\frac{\sqrt{\left[I_4(x,y)-I_2\left(x,y\right)\right]^2+\left[I_1(x,y)-I_3(x,y)\right]^2}}{2I_0}$$

$$\varphi(x,y)=\arctan\left(\frac{I_4(x,y)-I_2\left(x,y\right)}{I_1(x,y)-I_3(x,y)}\right)$$

Como en cualquier otra ecuación en donde se calcula el arcotangente, subyace una ambigüedad en la fase, que sólo se puede determinar al comparar los signos del numerador (*sen* ϕ) y del denominador (*cos* ϕ).

Una vez obtenida la distribución de fase sobre el plano de registro de la CCD es posible reconstruir el haz objeto ya sea numéricamente o empleando un medio capaz de modular la fase de un haz incidente sobre el. Cabe destacar que al tratarse del holograma de un objeto difusor, guardar la información de la amplitud y fase del haz objeto o sólo la correspondiente a la fase es prácticamente equivalente a la hora de reconstruir el mismo.

En la siguiente figura se esquematiza un posible montaje para los registros interferográficos.



Analicemos primero cómo se reconstruye el holograma numéricamente. Para ello supongamos que conocemos la distribución compleja *U(x,y)* en un plano, luego la distribución en otro plano se puede encontrar por medio de la integral de Fresnel-Kirchhoff. Supongamos que la distribución de fase sobre el plano de la CCD, es decir el plano del holograma, viene dada por $U(x,y)=\exp\left[i\phi(x,y)\right]$, luego es posible reconstruir el objeto, numéricamente, aplicando dicha integral de propagación hasta el plano objeto. La misma resulta:

$$U(\xi,\eta)=\frac{iA}{\lambda d}\exp\left[-i\frac{\pi}{\lambda d}\left(\xi^2+\eta^2\right)\right]\times\iint_{(x,y)}U(x,y)\exp\left[-\frac{i2\pi}{\lambda d}\left((\xi-x)^2+(\eta-y)^2\right)\right]dx\,dy$$

donde *A* es la amplitud de la onda incidente, *d* es la distancia del plano de registro al plano objeto y λ es la longitud de onda utilizada para el registro. En principio y dado que esta ecuación representa la aproximación de Fresnel, esta sería válida si se cumple la condición $d^3>>\frac{\pi}{4\lambda}\left[(\xi-x)^2+(\eta-y)^2\right]^2$. Para una dimensión aproximada de holograma de 1.2 *cm* x 1.2 *cm* (tamaño del CCD en la cámara), *d* debería ser mayor que 47 *cm*. Sin embargo vimos al comienzo del curso que esta condición era una sobre-exigencia ya que por el método de la fase estacionaria se obtenía difracción de Fresnel a distancias mucho menores. En este caso particular existe otra limitación que impone un límite a la proximidad y tamaño del objeto. Tal como anteriormente hemos comentado, la resolución de una CCD es muy inferior a la de una placa holográfica, por ello la configuración debe permitir que el haz objeto y el de referencia se hallen prácticamente alineados. Además debe tenerse en cuenta que el patrón de interferencia que la CCD registra consiste en realidad en un diagrama de speckle modulado por franjas de interferencia. Así resulta necesario que el tamaño de los granos de speckle sean relativamente grandes lo que implica que el objeto debe encontrarse lo suficientemente alejado de la cámara. Por otra parte si el objeto se halla cerca de la CCD los haces emergentes del mismo, y que lleguen a la cámara, formarán ángulos mayores y en consecuencia aumentará la frecuencia espacial de las franjas, siendo imposible su resolución.

La intensidad en el plano objeto se puede calcular tomando $I(\xi,\eta)=\left|U(\xi,\eta)\right|^2$

En realidad la función $U(\xi,\eta)$ es discreta y se calcula a partir de una matriz de *N x N* valores que representan a la distribución $U=(x,y)$. Sea Δx y Δy el período de muestreo en cada eje de coordenadas sobre el plano de la CCD, luego las variables continuas ξ y η deben reemplazarse por $r\Delta\xi$ y $s\Delta\eta$, donde *r* y *s* son números enteros.

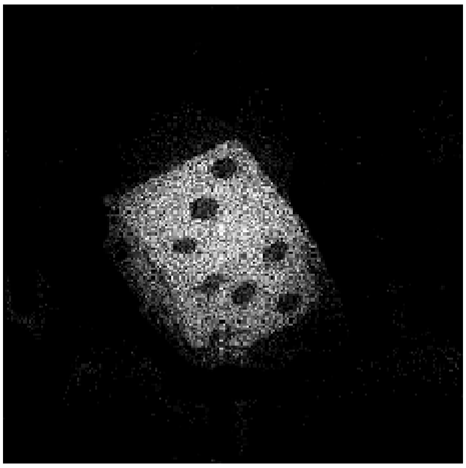
Como vimos, en este caso, la representación discreta de la ecuación de Fresnel está dada por la siguiente ecuación:

$$U\left(r,s\right)=\exp\left[-i\frac{\pi}{\lambda d}\left(r^2\Delta\xi^2+s^2\Delta\eta^2\right)\right].$$
$$.\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}U(k,l)\exp\left[-i\frac{\pi}{\lambda d}\left(k^2\Delta x^2+l^2\Delta y^2\right)\right]\times\exp\left[i2\pi\left(\frac{kr}{N}+\frac{ls}{N}\right)\right]$$

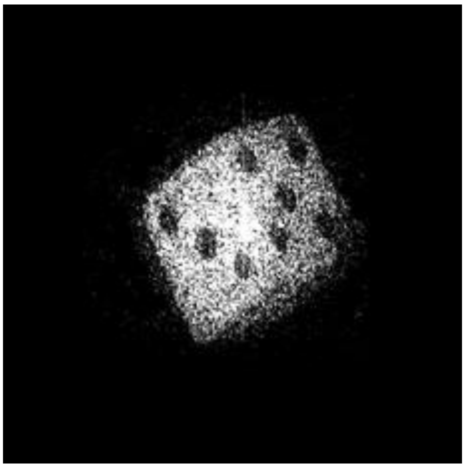
donde $U\left(r,s\right)$ también es una matriz de N x N puntos que describe la distribución de fase y amplitud de la reconstrucción del objeto.

Desde el punto de vista numérico, la ecuación anterior es una representación de la aproximación de Fresnel en términos de una transformada discreta de Fourier. Esto es relevante, ya que entonces se pueden aplicar algoritmos conocidos para realizar la transformada rápida de Fourier (FFT – Fast Fourier Transform). Debe tenerse en cuenta, aunque no necesariamente con los últimos algoritmos publicados, que en general para aplicar la FFT se requiere que el número de puntos N sea una potencia de 2.

Alternativamente a la reconstrucción numérica, si se posee un medio físico en el que directamente se puedan representar valores de fase, tal como una pantalla de cristal líquido, entonces a través de la tarjeta gráfica de una PC se envían los valores de fase calculados y estos son representados en la LCTV. Al iluminar este dispositivo con un láser se reconstruye el haz objeto del holograma con la ausencia del orden 0 y la imagen conjugada. En la siguiente figura se muestra el objeto (un dado) reconstruido numéricamente y por medios ópticos



Reconstrucción numérica



Reconstrucción óptica